Sociological Research Online

◆ Abstract

In this paper we are presenting a simple simulation of the Internet World-Wide Web, where one observes the appearance of web pages belonging to different web sites, covering a number of different thematic topics and possessing links to other web pages. The goal of our simulation is to reproduce the form of the observed World-Wide Web and of its growth, using a small number of simple assumptions. In our simulation, existing web pages may generate new ones as follows: First, each web page is equipped with a topic concerning its contents. Second, links between web pages are established according to common topics. Next, new web pages may be randomly generated and subsequently they might be equipped with a topic and be assigned to web sites. By repeated iterations of these rules, our simulation appears to exhibit the observed structure of the World-Wide Web and, in particular, a power law type of growth. In order to visualise the network of web pages, we have followed N. Gilbert's (1997) methodology of scientometric simulation, assuming that web pages can be represented by points in the plane. Furthermore, the simulated graph is found to possess the property of small worlds, as it is the case with a large number of other complex networks.

Keywords:

◆ Introduction

1.1

The aim of this paper is to present a simple simulation of the structure of the World-Wide Web (from hereon referred as the 'web') of the Internet. The web is composed of linked hyper-documents, i.e., web pages possessing links to

other web pages and constituting a huge directed graph. Each web page belongs to a web site and covers a number of different thematic topics. Internet statistics in general report that the distribution of web pages follows Lotka type power laws.

1.2

The goal of our simulation is to reproduce the structure of the observed web and the dynamics of its growth. We are attempting to do this by using a small number of simple assumptions. This is a short description of how our simulation proceeds: First, we start with a random number of web pages and we equip each one of them with a topic concerning its contents. Second, links between web pages are established according to whether web pages share or do not share common topics. Next, new web pages are randomly generated and subsequently they are equipped with a topic and they are assigned to web sites.

1.3

What we obtain after repeated applications of the previous rules in our simulation appears to preserve the observed structure and characteristics of the Internet World-Wide Web (WWW). In particular, the simulated web obeys a power law type of growth as the real web does too. We are visualising the simulated network of web pages through N. Gilbert's (1997) methodology of a scientometric simulation, in which we are assuming that web pages are represented by points in the plane. Since our simulation is based on certain simplified assumptions and rules, our aim is to reproduce the existing structure of the Web just qualitatively rather than quantitatively (even in a scaled form).[1] We are doing this by showing that our simulation (i) exhibits a scale-free power-law distribution and (ii) it possesses the *small-world* property (Watts & Strogatz, 1998; Watts, 1999), as both of these characteristics have been shown to be met within a large number of complex networks including the real Web (Albert & Barabαsi, 2002).

◆ Discussion of the Simulation

2.1

The reason that a simple simulation as ours, which is based on a set of very basic rules about the articulation of links connecting information units, succeeds to reproduce the complex structure of the World-Wide Web is a consequence of the fact that the WWW has essentially developed in un-planned and rather un-regulated ways in particular with respect to what concerns the addition of new web pages and new links between them.[2] However, the inadvertent development of the Web does not imply that the involved actors at the micro-level (users, surfers, consumers, citizens, activists, companies, organisations, public services etc., who add their web pages and navigate through them) are driven without their own intentions or motivations. It is exactly such a micro-level polyphonic composition of the Web (which recent regulatory plans on the semantic level[3] attempt to harness so that it might not become a cacophony) that explains the complex nature of the Web at the macro-level. Another way to say the same thing is by remarking that complex systems as the Web display 'emergent behaviour', which is not inherent in or derived from a knowledge of their constituent parts (Holland, 1998).

2.2

Thus, the macro-effects in a social simulation as that of the Web that is based on simple rules (although the emergent outcome is a complex network) is often a good example of unintended consequences[4] in social actions, as Hegselmann & Flache (1998) have already argued in a plea for the paradigmatic value of cellular automata based modelling. As Michael Macy (1998) puts it: "Computational models of emergent cooperation can restore the explanatory power of the unintended consequences of symbolic collective action, in which egoism is a variable instead of a constant. A dynamical theory of microsocial interaction, formalized using artificial agents, appears to be a viable and promising new direction for theoretical research on prudent cooperation, as well as solidarity mobilized around emergent norms and identities." It is exactly in this direction that our simulation targets: by providing a flexible instrument to test and experiment on a variety of hypotheses ruling microsocial interaction mediated through the Web, we aim to contribute to an understanding of the emergent structures of this complex network.

2.3

In fact, the underlying basic assumption of our simulation is just that the World-Wide Web of the Internet is a huge database consisting of millions of interlinked web pages. The physical location of web pages is inside certain computer hosts (all interconnected through the Internet), which are called web servers, and each of which may include a large number of web pages. Without loss of generality, each web server can be considered as part of a web site (Bray, 1996). In our simulation we are going to model the generation of web pages hosted in web sites.

2.4

The way we would like to see the relationship between web pages and web sites is by an analogy to the couple of (scientific) papers and authors. Of course, papers and web pages are of the same modality (physical or electronic documents) but authors and web sites appear to be heterogeneous, the former being human and the latter non-human (artefacts). However, conceiving socio-technical change as a complex process involving a whole network of human, social and technical resources makes appealing the idea that agency can be performed by both humans and machines.[5] Furthermore, the relation between a web page and the human who has possibly constructed it is quite weaker than the relationship between a (scientific) paper and its author. It might be the case that a web page is constructed by somebody who has not written its contents (copy-right issues are very perplexing on cyberspace) and not all web pages necessarily contain information about their authors (or constructors). Not to mention that there are web pages automatically produced by some software programs (scripts, autonomous intelligent agents etc.). Hence, releasing a web page from its 'authors' and assigning it to its virtual location, the web site where it 'lives,' fixes a more stable association to it and makes perfect sense in cyberspace.

2.5

Furthermore, independently of where they are physically located, web pages are distinguished by their content, their format and their associations (links) to/from other web pages. In fact, the geographic allocation of web pages is not practically and experientially sensed by Internet users ('web surfers') because

by a simple click they may virtually travel across the web universe. As long as the network bandwidth is sufficiently high, instantaneous retrieving (downloading) a web page withholds the geographical origin of its physical storage. Everything on the web is directly reachable, possibly a few clicks away.[6]

2.6

As for the content of web pages, for the purposes of this simulation, we are going to isolate two of its attributes, topics and links.

2.7

The topic of a web page is an issue, about which the web page includes information. For simplicity we assume that each web page refers to a single topic. For instance, a web page topic could be scientific, educational, financial, political, social, cultural, religious, philosophical, artistic, recreational, athletic or related to computers, science and technology, health, business, entertainment, news and media, sports etc. (or any finer thematic subcategory of the previous).

2.8

Because of the architecture of the World-Wide Web, web pages are hyper-linked documents containing not only a variety of media (text, image, audio, video, animation etc.) but also links to other web pages. Usually there is a relation between topics and existing links in a web page. The tendency is that web pages on a certain topic to be linked with other pages in the same (or related) topic(s). Also, sometimes by tracing the existing links in a web page, one could clarify the topic of the page.

2.9

Here we should say that although the concept of 'topics' might appear to be a rather simplistic assumption, we argue that, as signifiers of content, topics together with links might serve as the starting thick indicators of the information content of web pages in a simulation like ours which aims to uncover emergent structures based on very simple rules. Of course, subsequently a more detailed analysis would be needed taking into account more thin indicators such as various discursive or communicative 'web genres' or 'cyber-genres' (Crowston & Williams, 1999; Shepherd & Watters, 1999). Needless to say these more detailed assumptions would further complicate the simulation and widen its scope.

2.10

As for web links, an interesting way of thinking about web pages and their content in information or knowledge is suggested by Deborah Heath and her co-workers, who discuss the contribution of the Web in an 'interactive knowledge production', arguing that web links are at the same time "vehicles for travel within and between Internet domains and online documents, and part of the content of such areas" (Heath *et al.*, 1996, p. 456). By mapping out the part of the Web (mainly medical web databases), in which these scholars were interested, they managed to obtain important 'clues' for understanding the interactions and blockages between users or visitors of the Web but also to reveal certain 'tricks' that skilled users were utilising in order to 'decode' the content of the Web in their effort to find answers to their searches and queries.

2.11

Having discussed all the main parameters intervening in our simulation, it remains that we explain what kind of mappings of the Web we are employing.

As it is known, the World-Wide Web is evaluated to consist today of several hundred millions of web pages with over a billion web links among them and it appears to grow exponentially with time ([Bharat & Broder, 1998](#)). This suggests that the pages and links of the Web may be viewed as nodes and connections in a huge graph. But besides this graph-theoretic mapping, we are also going to use another geometrical way to map the Web as an information space. In fact, in order to visualise a geometrical positioning of web pages, we follow a technique applied by Nigel Gilbert ([Gilbert, 1997](#)) in a simulation of scientometric structures. By analogy to the idea that each (scientific) paper captures some quantum of knowledge, a 'kene,' to use Gilbert's neologism, we might say each web page contains a quantum of information.[7] Similarly with kenes, in order to be able to display it graphically on a two-dimensional plane, we require that each quantum of information contained in a web page is composed of two sub-sequences of equal length, and we treat each sub-sequence as a representation of a coordinate on a plane. In this way, every web page can be assigned a position on the plane. In our simulation, each web page is composed of two coordinates, each 16 bits in length, giving a total 'web universe' of 4,294,967,296 potential web pages, which is an essentially infinite number compared with the number of web pages generated during one run of the simulation.

2.12

However, a quantum of information differs from a quantum of knowledge with respect to what these quanta display. On the one side, each web page possesses a certain content of information and so it corresponds to a certain quantum of information, the way each (scientific) paper corresponds to a certain quantum of knowledge (a kene). On the other side, although kenes uniquely determine papers, containing the knowledge represented by kenes, the situation with the information of web pages might be different. It is possible that two web pages (located in two different web sites) contain similar information, which is represented by the same position in the plane where the corresponding quantum of information is mapped. This is the case with mirrored web pages and other strong redundancies often met on the Internet. In other words, cyberspace seems to disregard the restriction on publication in science that no two papers may be published when they contain the same or similar knowledge.

◆ Description of the Simulation

3.1

In the first step of our simulation we select the initial population of web pages together with their corresponding topics, links and web sites in the following way:

- The process starts with a random selection of an initial population of web pages, which is going to generate new web pages in the subsequent steps.
- We randomly assign a topic from a finite list of topics to each web page.
- To equip web pages with links we proceed as follows. For each web page, we consider the set of other web pages with which it shares the

same topic and the set of other web pages having different topics. The original web page can be potentially linked to each page in the former set, although it is possible to have links also towards pages in the latter set. In our simulation, linking is done randomly, according to a probability p_1 for pages of the same topic and probability p_2 for pages of different topics, assuming that p_1 is much greater than p_2; apparently, this is a sort of a 'preferential attachment' linking.

- Furthermore, each web page is assigned to a web site in the same way as Gilbert (1997) was assigning authors to papers: Select a random number from a uniform distribution from 0 to 1. If this number is less than a (which is fixed as a parameter of the simulation), assign the web page to a new web site. Otherwise, select randomly a web site among the already assigned.

- To visualise the initial population of web pages by positioning them on the plane, we randomly select the same number of two-dimensional points, i.e., the quanta of contained information, representing the web pages of the initial population. Note that in this selection not necessarily all points representing the quanta of information are distinct, because it is possible to select two or more web pages corresponding to the same information

.

3.2

In the second step we are going to generate more web pages from the ones already selected in the initial step. We assume that each initial web page can potentially generate per se a new web page independently of the web site to which it (the initial web page) belongs. This means that the new generated web pages can be assigned to different web sites from the ones they were spawned. To do this we proceed as follows:

- First, a 'generator' web page is selected at random among the population of the initial web pages. Automatically, this page is going to generate a new web page.
- To associate a topic to the new generated web page, select a random number from a uniform distribution from 0 to 1. If this number is less than i, associate the web page to the topic of its generator. Otherwise, select randomly a topic from the list of topics excluding the topic of the generator. Apparently, the smaller this parameter i is, the more possible is to change topics.
- To equip the generated web page with links we proceed as in the first step. We randomly choose links from the list of web pages with which the generated one has the same topic with probability p_1 and from all the other web pages with probability p_2.
- Again as in the first step, we assign the generated web page to a web site by selecting a random number from a uniform distribution from 0 to 1 and, if this number is less than a, assigning the generated web page to a new web site; otherwise, we select randomly a web site among the ones already assigned.
- Finally, at this step, we have to decide how many new web pages are going to be generated in the following step. For this purpose, we select

randomly a number from a uniform distribution from 0 to 1 and, if this number is less than y (which is fixed as a parameter of the simulation), then in the next step the generated new web pages should be increased by one (i.e., in the third step, they become 2). Otherwise, we keep the same number of the generated new web pages in the next step (i.e., in the third step, they are still 1).

- To position the new generated web pages on the plane, we proceed as follows (Gilbert, 1997). Suppose that the generator web page is represented by the point X on the plane (i.e., the quantum of information of the generator) and X_{lnk} are the points of the web pages with which X has links. Then the generated web page is the point X' produced as:

$$X' = X + (X_{lnk} - X)(1 - m)/2,$$

where m is a value between zero and one, which increases randomly but monotonically for each successive link.

3.3

The next step repeats the above rules acting on the original population of web pages increased by one and the purpose is to generate one or two new web pages. Continuing in this way, at the step k (k (2), the initial web pages might increase up to $k(k - 1)/2$.

◆ Results

4.1

To simulate the expansion of the World Wide Web we have written a program in Matlab.[8] In our simulation we have used the values of parameters as shown in Table 1:

**Table 1: Parameters for the simulation run**

Parameter	Description	Value
a	Probability threshold to introduce new web sites	0.2
i	Probability threshold to introduce new topics	0.5
y	Probability threshold to increase the number of new web pages	0.0025
T	Total number of topics	200
p_1	Probability threshold to introduce a link with a page of the same topic	0.5
p_2	Probability threshold to introduce a link with a page of a different topic	0.0005

4.2

A sensitivity analysis[9] has shown that small variations in the values of these parameters make little difference to the form of the output.

4.3

An animation of a two-dimensional display of our simulation is shown in Animation 1. The square is the surface on which the web pages are located when they are represented by their corresponding two-dimensional quanta of information. The simulation has been run for 1000 time steps. Each dot represents a web page and the position of a dot is given by the x and y coordinates of the two-dimensional quantum of information that corresponds to the web page. Blue dots denote the randomly selected 500 web pages at the initial step. Red dots denote the generated new web pages in the subsequent steps. The total number of simulated web pages created in the simulation was 2844 and the number of the web sites that hosted these web pages was 541. Furthermore, we should say that in this animation, each frame corresponds to 25 steps of the simulation.
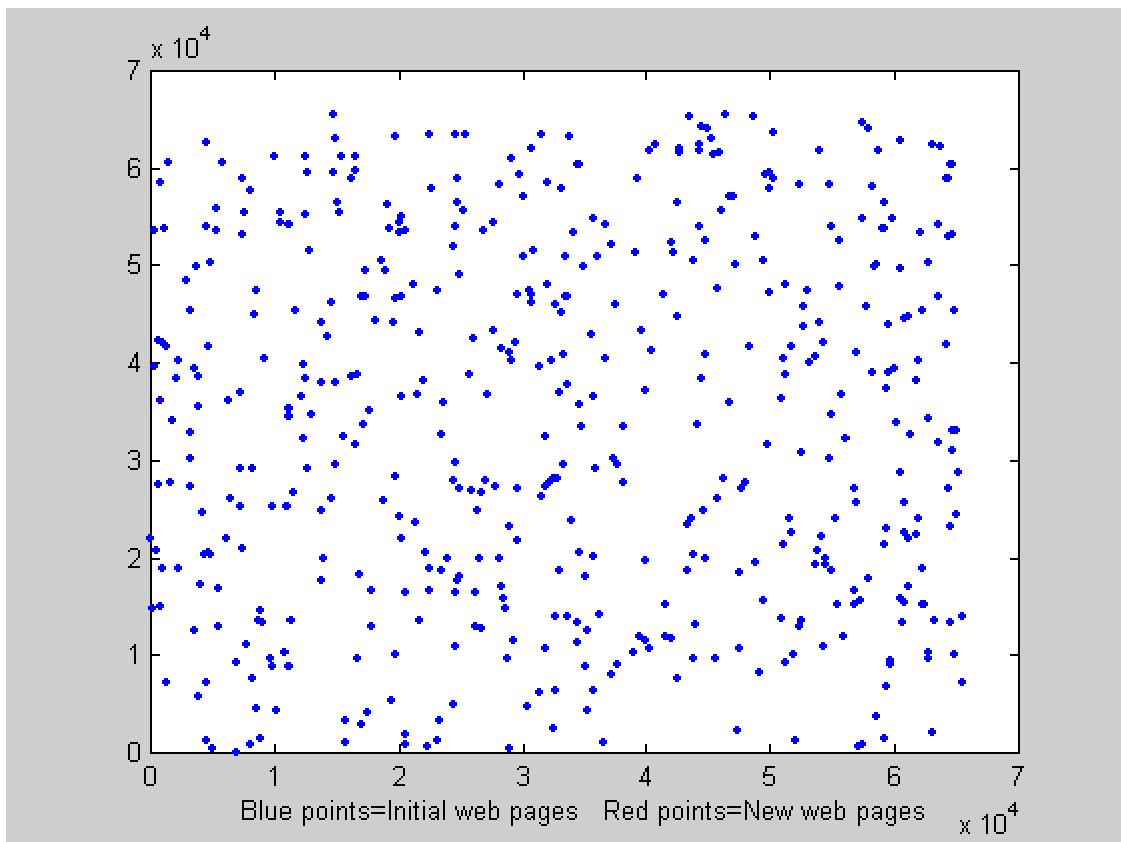
Animation 1. Web pages generation every 25 time units

4.4

The positioning of all the web pages on the plane at the final step of the simulation is given in the following figure:
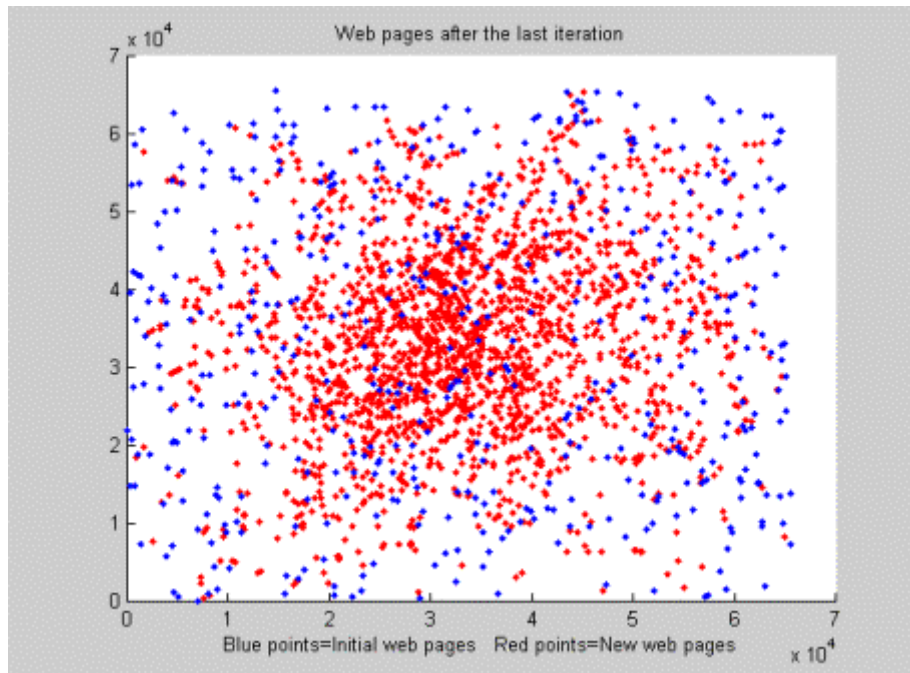
Figure 1: The simulation output

4.5

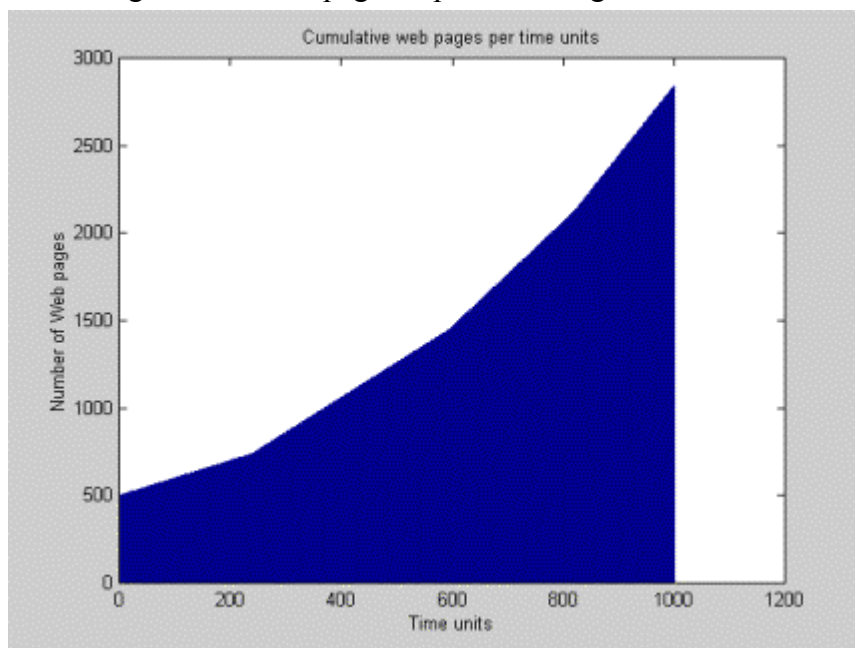The rate of growth of web pages is plotted in Figure 2:

Figure 2: Cumulative web pages per time unit

4.6

Figures 3 and 4 display the distribution of the number of out-going links per web page and in-coming links per web page, respectively. What we see is a highly skewed distribution that appears to follow some inverse square law.
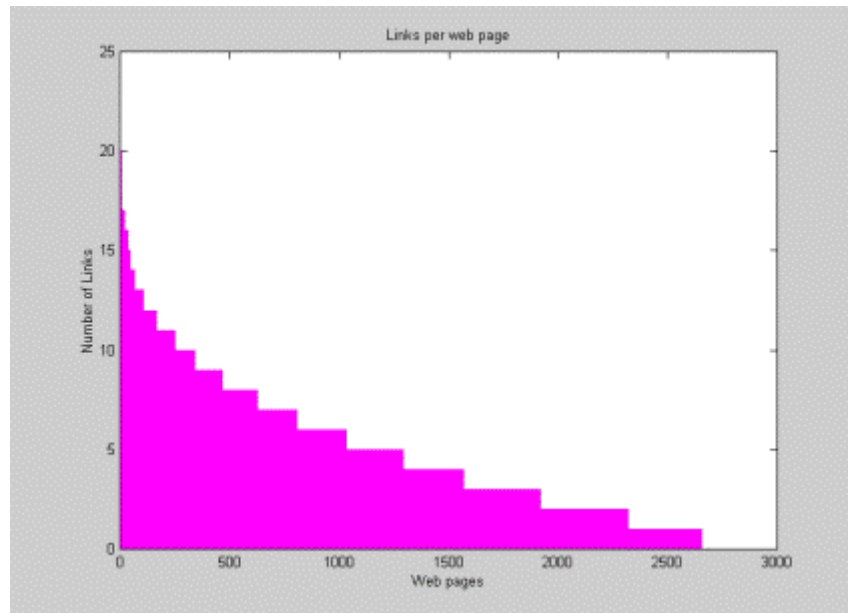
Figure 3: Out-going links per web page
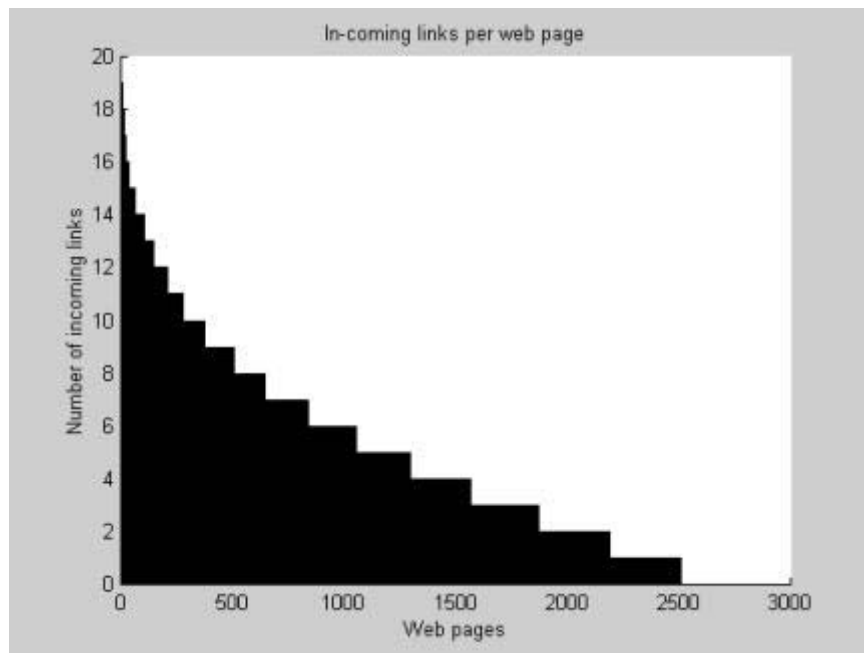

Figure 4: In-coming links per web page

4.7

Figure 5 shows the distribution of web pages per links. This distribution has a mean of 4.8256, indicating that an average page includes about 5 links to other web pages.
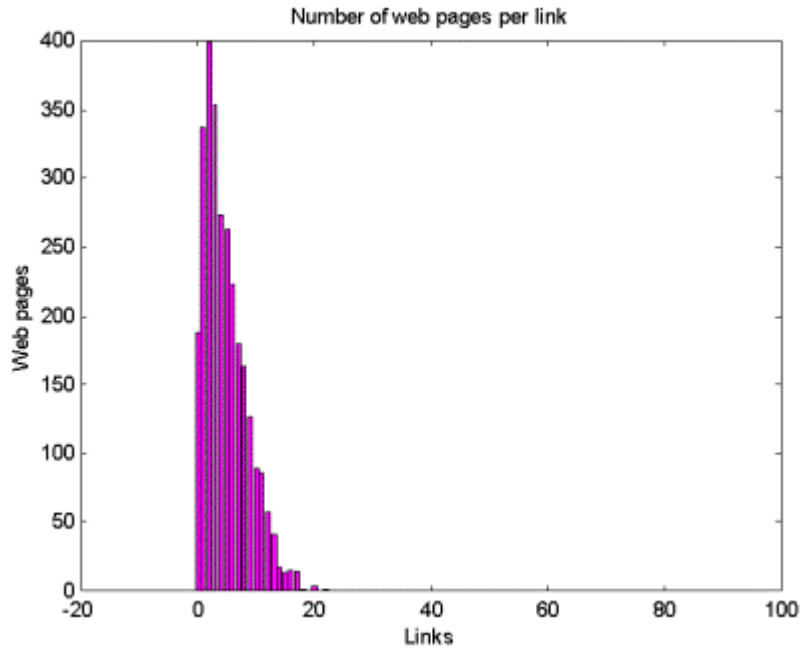
Figure 5: Links per web page

4.8

Figure 6 displays the number of web pages hosted in each of the 541 simulated web sites. We expect this distribution to follow Lotka's Law and Table 2 compares the distribution obtained from the simulation and plotted in Figure 7 with the distribution predicted from the exponentially truncated power law.

Figure 6 Web pages per web site

Table 2: Web pages per web site from the simulation compared with the exponentially truncated power law distribution

Number of Web Pages	Simulation	Distribution

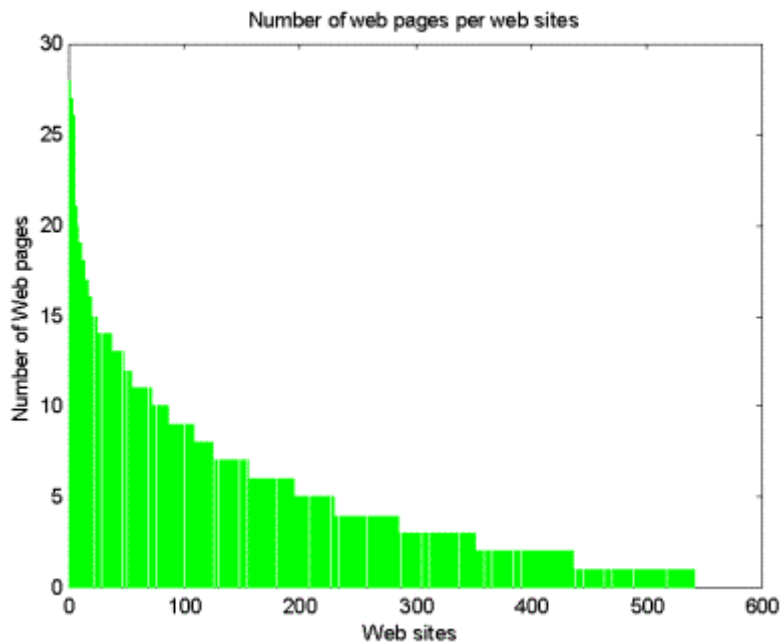|		
1	105	106
2	85	83
3	66	66
4	56	53
5	34	43
6	40	35
7	31	29
8	16	23
9	23	19
10	13	16
11	18	13
12	7	16
13	10	13
14	13	10
15	5	9
16	3	7
17	3	6
18 or more	13	19

4.9

Lotka's Law is one of the most powerful laws in scientometrics. In general, this is a power law, in which the number of authors of n papers follows an $1/n^2$ distribution. However, the data of our simulation are better fitted by an exponentially truncated power law[10] of the form:

$$p_n = an^{-k}e^{-n/g},$$

where *a, k, g* are the parameters. In our simulation, we have used a least square optimisation in order to evaluate these parameters and we have found them to take the following values:
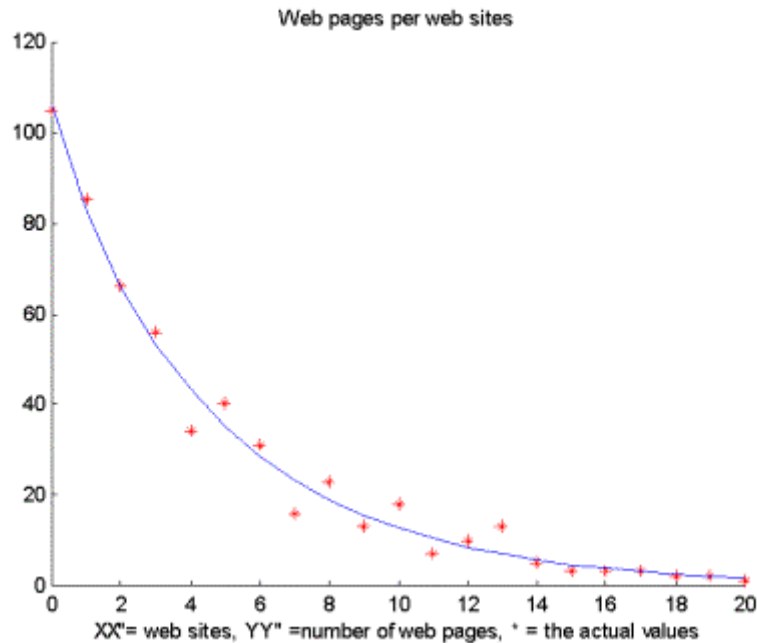
$a = 128.4421$
$k = 0.0783$
$g = 5.1872$

Figure 7: Web pages per web site

4.10

In the above figure 7, blue line represents the distribution p_n that has been derived through the least square optimisation and red dots represent the actual values which have been obtained in the simulation.

4.11

As the web is a directed graph, each web page can be characterised by the number of in-coming, k_{in}, and out-going, k_{out}, links. Barabαsi & Albert (1999) have investigated the probability distributions, $P(k)$, that a randomly selected web page has exactly k_{in} or k_{out} links, respectively, and they found that $P(k)$ decayed via a power law of the form $P(k) \approx k^{-y}$ at large k. Indeed, Barabαsi & Albert found that $y = 2.45$ for out-going links and $y = 2.1$ for in-coming links, a result that was confirmed in a parallel study by Ravi Kumar and co-workers (Kumar 1999). In the present simulation, we have computed that $y = 2.3637$ for out-going links and $y = 2.3641$ for in-coming links, which verifies that our simulation produces a scale-free network such as the real World-Wide Web.[11]
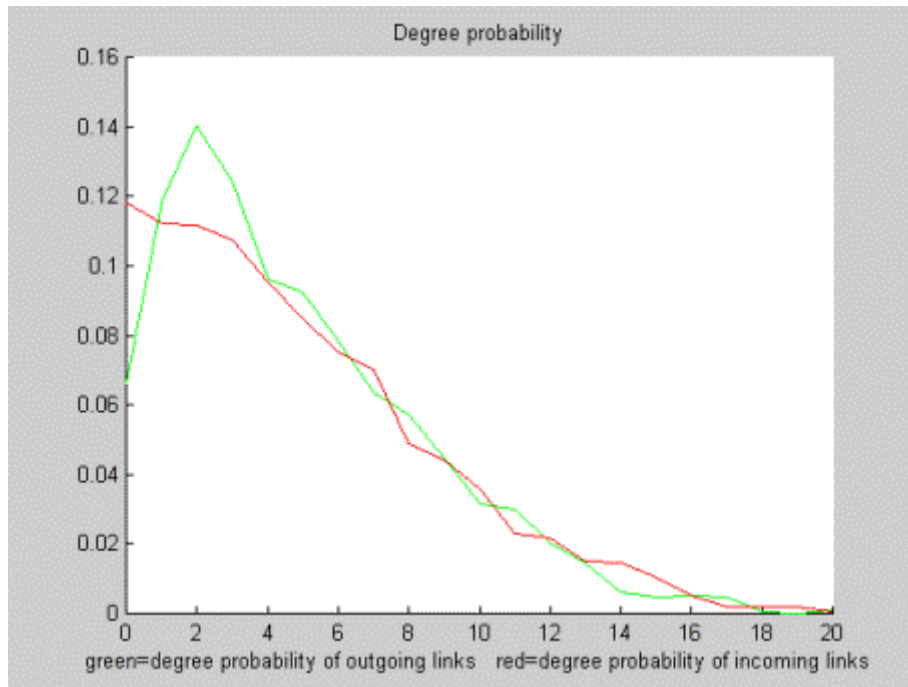
Figure 8: Probability distribution of links

4.12

During the last few years, it has been recognised that the Internet and the World-Wide Web are best described as complex systems together with a very wide range of other natural, biological, social, scientometric etc. systems (Albert & Barabαsi, 2001). A common feature that is shared by all these complex systems is the so-called *small world* property (Watts & Strogatz, 1998; Watts, 1999). Patterns of small worlds can be detected from two basic statistical quantities, the *clustering coefficient C* and the *average path length d.*

4.13

To give the definitions of the above quantities, we consider a directed *graph G = (V,L)*, where *V* is the set of *N* nodes and *L* is the set of links among the nodes. Two nodes are called *adjacent* if they are linked. Then, the *clustering coefficient C* of the graph *G* is defined as follows: Fix a node *u* and consider the set A_u of all the adjacent nodes to *u*; let C_u be the ratio of the number of all the existing links among the nodes of A_u divided by the number of all possible links among these nodes; then *C* is the average over all nodes, i.e.,

$$C = N^{-1}\Sigma_{u \in V} C_u.$$

4.14

The second quantity is defined as follows: Fix two nodes *u* and *w* and consider the *shortest path length d(u,w)* between these nodes, i.e., the minimum number of links that must be traversed in order to reach node *w* starting from node *u*; the average path length of the node *u* is defined as $d(u) = N^{-1}\Sigma_{w \in V} d(u,w)$ and the *average path length d* of the graph *G* is defined as $d = N^{-1}\Sigma_{u \in V} d(u)$.

4.15

Let us now consider the case that the graph *G* is a *random graph*. In their classic definition, Erdφs and Rınyi define a random graph as consisting of *N* nodes, each of which has a probability *p* of being connected to another link (Erdφs & Rınyi, 1959). In general, in a graph *G = (V,L)* of *N* nodes, the *degree*

k_u of a node u is the number of all the links from u to all its adjacent nodes and the average degree of the graph G is $<k> = N^{-1}\sum_{u \in V} k_u$. Now, for a random graph, one easily computes that the clustering coefficient is $C^{rand} \approx <k>/N$ and the average path length is $d^{rand} \approx lnN/ln<k>$.

4.16

According to Watts and Strogatz, a graph has the small world property if (i) d is close to d^{rand} but (ii) C is much larger than C^{rand} (Watts & Strogatz, 1998). In the above simulation of the web that we have performed, for $N = 2844$ and $= 4.8256$, we have computed that (i) $C = 0.1998$, while, $C^{rand} = 0.0017$, and (ii) $d = 5.4892$, while $d^{rand} = 5.0529$. Therefore, our simulated web is indeed a small world as it happens with the real World-Wide Web. As a matter of fact, studying the web at the level of web pages and as a directed graph, Albert, Jeong & Barabαsi (1999) found[12] for a sample of $N = 325,729$ that $d = 11.2$ and predicted, using finite size scaling, that for the full web of 800 million nodes d would be around 19. Adamic has computed for the web at the site level and as an undirected graph that $N = 153,127$, $= 35.21$, $C^{rand} = 0.00023$, $C = 0.1078$, $d^{rand} = 3.35$ and $d = 3.1$ (Adamic, 1999).

◆ Conclusions

5.1

The aim of our simulation was to reproduce the emergent structure of the World-Wide Web as a complex scale-free growing network, which possesses the small-world property, by a small number of underlying assumptions and rules. Future research along the lines of this simulation should focus on adopting less simplistic and more realistic assumptions and rules. For example, we could include a multi-topic composition of web pages, detach the selection of links from topics and include a life cycle in the duration of web pages. Moreover, we could run such a more general simulation over a longer number of iterations in order to form a population of web pages, which is comparable with their actual number (either of the whole Web or of its restriction over a particular sector, domain, country etc.). Such a more general simulation might be possible to reproduce what Broder *et al*. (2000) have described as the 'bow tie'-looking model of the Web. Using a web crawl consisting of roughly 200 million web pages and 1.5 billion links, Broder and co-workers have built a very detailed database model of the Web. Then, they have showed that the Web consists of five distinct regions characterised by whether nodes have just out-going links, just in-coming links, both types of links or no links at all and whether nodes are connected to the 'bow tie knot' (consisting of nodes with both out-going and in-coming links) or not. In this way, Broder *et al*. have concluded that the macroscopic structure of the web is considerably more intricate than suggested by earlier experiments on a smaller scale.

5.2

Furthermore, we should stress the importance of this type of web simulations for policy issues of the Information Society. Usually such policies refer to the regulation of the Internet and deal with such issues as copy-right, pricing, privacy, universal access, 'cybersquatting', security, hackers' attacks, spread of viruses, etc. (Drucker & Gumpert, 1999). As e-commerce and e-business

transactions are more extensively used, many interested stakeholders pressure for more regulation despite a libertarian or communitarian concern for privacy and public freedom. Therefore, in what concerns various policy scenarios of future developments on the Web, this type of simulation might be very useful in providing early warnings about new opportunities, challenges in risks as far as these simulations could also model the agencies of web-actors besides the content and linkage of web pages. In this way, detailed simulations of the Web (based on realistic assumptions about the behaviour of web-actors) might become a flexible tool for policy-makers to test and experiment on future scenarios and possible choices of socially accountable decisions and best practices of informational polity, which would guarantee a socially robust and sustainable Information Society.

◆ Notes

[1] Taking into account that the characteristics of complex networks (a) depend on the population of networked agents and (b) they imply the small-world property in a relative way (that is by being compared with the corresponding random graph), a strict quantitative comparison does not make any sense here. For instance, in the article of Albert & Barabαsi (2002), one can see a large number of different networks which are considered to be complex networks only in terms of the occurrence of the above mentioned characteristics of complex networks, without any other quantitative comparison on them.

[2] The authors wish to thank the referees of the paper for providing useful suggestions on the sociological discussion of the Web.

[3] This is the initiative of the 'semantic web' currently launched by the World-Wide Web Consortium (W3C). According to Berners-Lee, Handler & Lassila (2001): "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation".

[4] Unintended or unanticipated consequences (Merton, 1936) are of tremendous interest to sociologists in the tradition of Merton's (1963) discussions of manifest versus latent functions of purposive action. They also play a central role in more recent sociological developments such as Giddens's (1984) structuration theory.

[5] This is the dogma of the 'Actor Network Theory,' in which the development and stabilisation of scientific and technological objects (facts and artefacts) results from the construction of heterogeneous networks as concrete alignments between human actors, natural phenomena, and social or technical intervening aspects (Callon, 1986; Latour, 1987). However, not everybody would subscribe to these ideas; for some this is a controversial theory and sometimes serious objections are risen against it (Collins, 1994).

[6] Albert et al. (1999) have estimated the degree of connectivity of the web concluding that any two web pages are separated by an average of '19 clicks.'

[7]We are tempted to call 'infene' this quantum of information captured by web pages in analogy with Gilbert's neologism.

[8] The Matlab program of this simulation is available online at: <http://nicomedia.math.upatras.gr/simulations/ssweb.m>

[9] The results of the sensitivity analysis to the variation of the simulation parameters are given in the following table:

a	i	y	T	p_1	p_2	Average number of links	Clustering coefficient	Number of pages	Number of sites
0.18	0.5	0.0025	200	0.50	0.0005	6.7681	0.1953	4019	728
0.20	0.5	0.0025	200	0.50	0.0005	4.8256	0.1998	2844	541
0.22	0.5	0.0025	200	0.50	0.0005	4,5022	0,2029	2678	609
0.20	**0.4**	0.0025	200	0.50	0.0005	2.6180	0.2283	1500	298
0.20	**0.5**	0.0025	200	0.50	0.0005	4.8256	0.1998	2844	541
0.20	**0.6**	0.0025	200	0.50	0.0005	4.4691	0.2059	2528	475
0.20	0.5	**0.0022**	200	0.50	0.0005	4.8256	0.1998	2844	541
0.20	0.5	**0.0025**	200	0.50	0.0005	4.8256	0.1998	2844	541
0.20	0.5	**0.0028**	200	0.50	0.0005	4.8256	0.1998	2844	541
0.20	0.5	0.0025	**180**	0.50	0.0005	6.0627	0.2010	3364	694
0.20	0.5	0.0025	**200**	0.50	0.0005	4.8256	0.1998	2844	541
0.20	0.5	0.0025	**250**	0.50	0.0005	5.4559	0.1761	4018	822
0.20	0.5	0.0025	200	**0.47**	0.0005	2.9335	0.2025	1653	325
0.20	0.5	0.0025	200	**0.50**	0.0005	4.8256	0.1998	2844	541
0.20	0.5	0.0025	200	**0.55**	0.0005	4.3863	0.2313	2358	449
0.20	0.5	0.0025	200	0.50	**0.0003**	2.9620	0.2417	1791	371
0.20	0.5	0.0025	200	0.50	**0.0005**	4.8256	0.1998	2844	541
0.20	0.5	0.0025	200	0.50	**0.0010**	5.8110	0.1525	3011	596

[10] Exponentially truncated powers laws have been used by Newman (2000) in a study of scientific co-authorship networks.

[11]Although it appears as a disparity in our simulation's computation (y = 2.3637 for out-going links and y = 2.3641 for in-coming links) in contrast with what Barabαsi & Albert (1999) have found (? = 2.45 for out-going links and y = 2.1 for in-coming links), we must stress the fact that our simulation is based on a rather simplistic model of the real Web, while Barabαsi & Albert have collected information about a large sample of the real Web composed of over 800 million web pages. What is important is that the results of a simple simulation as ours are exhibiting the same structure with that of the real

World-Wide Web: the connectivities of the simulated web pages follow too a scale-free power-law distribution.

[12] Again the reason that our simulation has resulted a different value of average path length d than what was computed by Albert, Jeong & Barabasi (1999) is due to the fact that our simulation was based on a rather simplistic model of the World-Wide Web, while Albert *et al*. had focused on a very large sample of the real Web. However, it is in both cases the comparison with the corresponding random graph that implies the property of the small-world.

 References

ADAMIC. L. (1999). 'The Small World Web' in *Proceedings of ECDL'99*, Lecture Notes in Computer Science, Vol. 1696, pp. 443-452. Berlin: Springer Verlag. <http://www.parc.xerox.com/istl/groups/iea/papers/smallworld/smallworld.pdf>.

ALBERT, R., and BARABASI, A.-L. (2001). 'Statistical Mechanics of Complex Networks', LANL arXiv:cond-mat/0106096, <http://xxx.lanl.gov/abs/cond-mat/0106096>.

ALBERT, R., JEONG, H., and BARABASI, A.-L. (1999). 'Diameter of the World-Wide Web', *Nature*, Vol. 401, pp. 130-131.

BARABASI, A.-L., and ALBERT, R. (1999). 'Emergence of scaling in random networks', *Science,* Vol. 286, pp. 509-511.

BERNERS-LEE, T., HENDLER, J., & LASSILA, O. (2001). 'The Semantic Web', *Scientific American*, Issue 501, <http://www.sciam.com/2001/0501issue/0501berners-lee.html>.

BHARAT, K, & BRODER, A. (1998). 'A technique for measuring the relative size and overlap of public Web search engines', *WWW7/Computer Networks,* Vol. 30, No. 1-7, pp. 379-388.

BRAY, T. (1996). 'Measuring the Web', *Fifth International World Wide Web Conference,* <http://www5conf.inria.fr/fich_html/papers/P9/Overview.html>.

BRODER, A., KUMAR, R., MAGHOULL, F., RAGHAVAN, P., RAJAGOPALAN, S., STATA R., TOMKINS, A. & WIENER, J. (2000). 'Graph structure in the web', *Computer Networks*, Vol. 33, No. 1-6, pp. 309-320, <http://www.almaden.ibm.com/cs/k53/www9.final>.

CALLON, M. (1986). 'The sociology of an Actor Network' in M. Callon, J. Law & A. Rip (editors) *Mapping the Dynamics of Science and Technology*. London: Macmillan.

COLLINS, H.M. (1994) 'Bruno Latour: *We Have Never Been Modern*', *Isis*, Vol. 85, No. 4, 672-674.

CROWSTON, K., & WILLIAMS, M. (1999). 'The effects of linking on genres of web documents', in *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences; Genre in Digital Documents.* Los Alamitos CA: IEEE Computer Society, <http://crowston.syr.edu/papers/ddgen04.pdf>.

DRUCKER, S.J., & GUMPERT (eds.) (1999). *Real Law at Virtual Space: Communication Regulation in Cyberspace*. Cresskill, NJ: Hampton Press.

ERDOS, P., and RENYI, A. (1959). 'On random graphs. I', *Publicationes Mathematicae (Debrecen)*, Vol. 6, pp. 290-297.

GIDDENS, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley: University of California Press.

GILBERT, N. (1997). 'A Simulation of the Structure of Academic Science', *Sociological Research Online,* Vol. 2, No. 2, <http://www.socresonline.org.uk/socresonline/2/2/3.html>.

HEATH, D., KOCH, E., LEY, B., & MONTOYA, M. (1999). 'Nodes and queries: Linking locations in networked fields of inquiry', *American Behavioral Scientist,* Vol. 43, pp. 450-463.

HEGSELMANN, R., & FLACHE, A. (1998). 'Understanding complex social dynamics: A plea for cellular automata based modelling', *Journal of Artificial Societies and Social Simulation,* Vol. 1, No. 3, <http://jasss.soc.surrey.ac.uk/1/3/1.html>.

HOLLAND, J.H. (1998). *Emergence: From Chaos to Order.* Redwood City, CA: Addison-Wesley.

KUMAR, R., RAGHAVAN, P., RAJAGOPALAN, S., and TOMKINS, A. (1999). 'Trawling the Web for emerging cyber-communities', *Computer Networks,* Vol. 31, pp. 1481-1493.

LATOUR, B. (1987). *Science in Action*. Cambridge, MA: Harvard University Press.

MACEY, M. (1998). 'Social order in artificial worlds', *Journal of Artificial Societies and Social Simulation,* Vol. 1, No. 1, <http://jasss.soc.surrey.ac.uk/1/1/4.html>

MERTON, R.K. (1936). 'The unanticipated consequences of purposive social action', *American Sociological Review*, Vol. 1. pp. 894-920.

MERTON, R.K. (1963). *Social Theory and Social Structure*. Glencoe, IL: Free Press.

NEWMAN, M.E.J. (2000). 'Who is the best connected scientist? A study of scientific coauthorship networks', Santa Fe Working Paper 00-12-064, <http://www.santafe.edu/sfi/publications/Abstracts/00-12-064.ps>.

SHEPHERD & WATTERS (1999). 'The functionality attribute of cybergenres', in *Proceedings of the 32nd Annual Hawaii International Conference of System Sciences; Genre in Digital Documents.* Los Alamitos CA: IEEE Computer Society.

WATTS, D.J. (1999). *Small-worlds*. Princeton, NJ: Princeton University Press.

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