ELSEVIER

BioSystems 46 (1998) 209–212

BioSystems

Webometry: measuring the synergy of the world-wide web

Ralph H. Abraham *

Visual Math Institute, POB 7920, Santa Cruz, CA 95061-7920, USA

Abstract

This is the second progress report on the webometry project: acquisition of data regarding the density of links on the world-wide web (WWW). We illustrate the primary visualization strategy, the synergy matrix, in the case of a model subnet of nine nodes. © 1998 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: World-wide web; Connectivity; Synergy; Webometry

1. Introduction

In an earlier paper (Abraham, 1997) we viewed the explosive growth of the world-wide web (WWW) as the neurogenesis phase in the embryogenesis of a new planetary civilization. To empower this emergent phenomenon with self-reflection, we proposed strategies for the visualization of the complexity of the WWW, seen in the context of connectionism, that is, as a neural net. The pointwise fractal dimension of a massive matrix was the basis of these strategies. Our fundamental idea in this paper, also based on connectionism, is that the intelligence and morphogenesis of the WWW has to do with the density and growth of links between different nodes. To explore this, we create a subweb, by selecting certain groups of domains as the nodes of a web. We then use the AltaVista index of the WWW as a source of data on the number of pages at each node, and number of links between one node and another. The synergy, a relative density defined below, is then computed from this data, for all pairs of nodes, and recorded in a matrix of nonnegative real numbers, the synergy matrix. We view the WWW as the nervous system of the emerging planetary society, and propose to image the neurogenesis of this process, as an ongoing project of Operation Webwatch.

2. Connectionism

The mathematics of morphogenesis, complex dynamical systems theory, is the basis of our strategies for visualizing the web. Thus we view the web as a neural net, that is, a massive web of neurons or nodes. While neurons are not dumb,

* E-mail: abraham@vismath.org; http://www.vismath.org

PII S0303-2647(97)00100-7

connectionism views the intelligence of the network as primarily derived from its connections, as opposed to its nodes.

In the simple models for neural nets provided by the mathematics of complex dynamical systems, the connections are represented by real numbers. Given two nodes, $n(i)$ and $n(j)$, the connection from the first to the second is represented by a single real number, $s(i,j)$, denoting the strength of the connection. All of this data, the $s(i,j)$, for i, j from 1 to N, may be set out in a single tableau, which is a square matrix of size N, the total number of nodes. In this paper we are going to use a small subweb, with $N=9$, as an example.

3. An exemplary subweb

We wish to compute an exemplary connection matrix in detail, so we must choose a subweb. We now consider a set of domains (campus computer networks) comprising the University of California (UC) system, for which $N=9$. We might as well have chosen the 45 nations of Europe, or the five continents, and most probably the synergy matrices and images for these systems will be posted in due time on the WWW at http:/hNww.vismath.org/webwatch. We go on now to illustrate, using this small subweb, the creation of the synergy matrix, and its associated image.

The domains of the UC system, and their sizes, are shown in Table 1. The size is the number of HTML documents at the site, as reported by Alta Vista (http://www.altavista.digital.com) in response to the advanced query, host: ucsc.edu, on 2 December, 1996.

Table 1

Campus	Domain name	Size, S
Davis	ucdavis.edu	42 607
Berkeley	berkeley.edu	93 779
San Francisco	ucsf.edu	19 502
Santa Cruz	ucsc.edu	11 807
Santa Barbara	ucsb.edu	21 155
Los Angeles	ucla.edu	44 886
Riverside	ucr.edu	8070
Irvine	uci.edu	28 226
San Diego	ucsd.edu	46 995

4. The connectivity matrix defined

Our next step is to obtain the raw data from Alta Vista for the connectivity matrix, as described in our first report, Webometry #1. In this N by N matrix, we record the number $C(i,j)$ of pages (that is, HTML documents) of the ith node which contain one or more links to any document of the jth node. This is the definition of the connectivity number. Assuming that the ith node is uniquely identified by a single URL fragment $U(i)$ for each i from 1 to N, the connectivity number is returned by Alta Vista to the advanced (Report Count Only) query: host:$U(i)$ AND link:$U(j)$.

5. An exemplary connectivity matrix

In the case of our model subnet, the UC system, we have $N=9$, and the URL fragments listed in Table 1. Performing the 81 queries and recording the results in an 9×9 matrix, which could be done automatically by a PERL script for example, we obtain the exemplary connectivity matrix shown in Table 2. The entry in the ith row and jth column is the connectivity number, $C(i,j)$.

6. The synergy matrix defined

We now normalize connectivity numbers according to the scheme suggested in our report Webometry #1. To recall this, imagine a matrix in which all the rows correspond to the pages of the ith node, and all the columns correspond to the pages of the jth node. The number of rows is thus the size (number of pages) of the ith node, $S(i)$, but call it H for height. Similarly the number of columns of our matrix is the size of the jth node, $S(j)$, but call it W, for width. Then the area of our matrix, an array of small cells of a unit

Table 2

	D	B	SF	SC	SB	LA	R	I	SD
D	17081	400	50	58	63	103	52	123	157
B	200	61725	118	122	155	300	62	400	300
SF	22	122	5802	26	10	30	7	60	25
SC	40	185	28	9908	33	43	23	65	85
SB	107	200	22	89	11407	130	72	89	186
LA	86	300	31	51	90	26758	56	200	153
R	21	58	3	12	17	28	7027	52	47
I	105	300	41	68	74	139	74	19394	179
SD	132	500	26	68	97	192	76	200	24479

area, is the product of the height times the width, $H*W$. Each small cell in this matrix corresponds to a single chosen page of the ith node (say A.html) and a single chosen page of the jth node, say B.html.

Now, let this small cell be colored black if there is one or more links from the page A.html of the ith node to the page B.html of the jth node, and otherwise suppose it is white. Do likewise for each of the $H*W$ small cells. Then the total number of black cells is the connectivity number, $C(i,j)$, defined above. Also, the greyness (relative blackness) of the entire matrix of $H*W$ cells is this number, $C(i,j)$, divided by the total number of cells, $H*W$, which is none other than the product, $S(i)*S(j)$. This quotient,

$$s(i,j) = C(i,j)/(S(i)*S(j)),$$

is the overall greyness of our imaginary matrix of area $H*W$, and this is the number we call the synergy from the ith node to the jth node. The synergy matrix is the $N \times N$ square matrix of these relative greyness numbers.

7. An exemplary synergy matrix

We now carry out the process defined above, in the context of the exemplary subweb with $N=9$ displayed in Table 1. Using the connectivity matrix $[C(i,j)]$ for this subweb given in Table 2 and the sizes $S(i)$ given in Table 1, we obtain the 9×9 synergy matrix for the UC system, shown in Table 3.

8. An exemplary synergy image

Of course this data is difficult to understand, and at the Visual Math Institute we are naturally predisposed to turn it into a visual display. Ignoring the very large self-synergy numbers found on the descending diagonal of this matrix, corresponding to the narcissism of university (and perhaps all) domains, we adopt a simple grey scale to represent the synergy data, from white (no synergy, or zero) to black (high synergy, >45). The resulting grey scale image is shown in Fig. 1, with the scale to the right. In this scale we have used only ten shades of grey for pedagogical simplicity. In the application to larger subwebs, such as the nations of Europe, we will use a more refined color code.

Note that the image is only slightly more understandable than the numerical display of Table 3. And for a larger subnet, such as the 250 nations for which Alta Vista has data, the image would look more like a photograph of clouds. Even so, many interesting questions might arise from contemplation of the image. These questions are cybersociological, and we hope that cybersocial scientists will consider them. Our own approach, as presented in Webometry # 1, is to admit that the synergy image is fractal, and resort to methods of fractal geometry to convert the image into another complimentary fractal, which represents directly the complexity of the web. However, some questions familiar from the sociology of science could be addressed directly from the synergy image. Recall that according to the findings

Table 3

	D	B	SF	SC	SB	LA	R	I	SD
D	940.9	10.0	6.0	11.5	7.0	5.4	15.1	10.2	7.9
B	5.0	701.0	6.5	11.0	7.8	7.1	8.2	15.1	6.8
SF	2.7	6.7	1525.3	11.3	2.4	3.3	4.5	10.9	2.7
SC	8.0	16.7	12.2	7107.4	13.2	8.1	24.1	19.5	15.3
SB	11.9	10.1	5.3	35.6	2548.9	13.7	42.2	14.9	18.7
LA	4.5	7.1	3.6	9.6	9.5	1329.5	15.5	15.8	7.3
R	6.1	7.7	1.9	12.6	10.0	7.7	10790.0	22.8	12.4
I	8.7	11.3	7.5	20.4	12.4	11.0	32.5	2434.3	13.5
SD	6.6	11.4	2.8	12.3	9.8	9.1	20.0	15.1	1108.4

of Ruth Benedict, synergy correlates with the prosperity and happiness of a society, see Abraham (1988, 1989). Therefore, we may expect paranoia to emerge in a cybersociety of low synergy. On the other hand, we might discover cybersocieties from the synergy image by locating subnets of high synergy.

9. Conclusion

We have described a complete, step-by-step procedure for the measurement of the connectivity of a subweb of the WWW. In a previous paper, Webometry #1, we described the application of this procedure to all 30 million Alta Vista indexed domains of the WWW, as a gedanken experiment. The practical implementation of this procedure naturally requires the restriction to a subweb of manageable size. At present, we consider $N = 1000$ manageable. We hope in future to apply these methods to societies of nations, such as the Council of Europe, the United Nations, and so on. As a kind of weather map for the climate of international health, a distant early warning system for outbreaks of international paranoia and other disorders might be created.

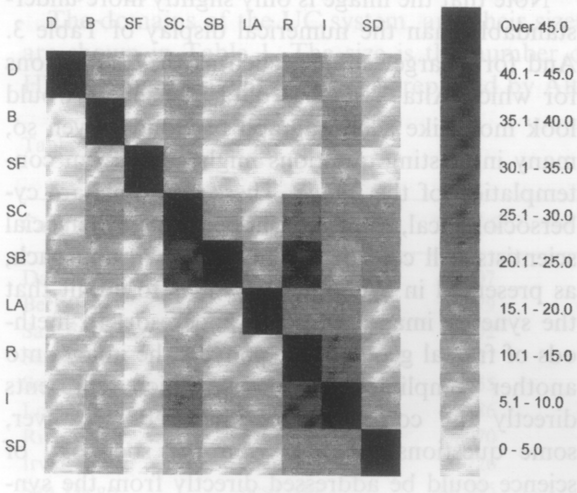

Fig. 1. Synergy of the UC web, December 1996.

Acknowledgements

Thanks to Don Foresta of the Electronic Language Laboratory (Paris) for suggesting this idea. In a joint research project currently under way, we hope to actually carry out the strategy described here, presenting our results on the WWW at http://www. vismath.org/webwatch.

References

Abraham, R., 1988. Social and international synergy: a mathematical model. International Synergy Journal 3, 18–26.

Abraham, R., 1989. Social synergy and cognitive maps. ICIS Forum 19, 45–51.

Abraham, R., 1997. Measuring the complexity of the world-wide web. World Futures 50, 785–791.