

39: *Webometry: Measuring the Complexity of the World Wide Web*

RALPH H. ABRAHAM

1 INTRODUCTION

The World Wide Web (WWW) has grown explosively in five years from a novel idea of Tim Berners-Lee to the nervous system of a new planetary society. One wonders what to make of this, and perhaps the various opinions correspond to the historical paradigms. Here are four of them .

A. In the paradigm of ancient Greece and the Middle Ages, humans stood helplessly in an autonomous harmony of forces, celestial and terrestrial. Occasional divine disharmonies wrought havoc. In this view, the WWW is seen as a new and suspicious god. Whether like Zeus or Eros, only time may tell.

B. In the paradigm of the Renaissance, humans were seen as potential partners of the gods, able to harness divine forces to human will by magical means. From this platform, the WWW is a new partner for advancing our most ambitious or foolish whims. By black magic as it were, or white, only time will tell.

C. In the religion of the Enlightenment and its derivative, modern science, humans create and control all. In this view, the WWW is just another machine, like the world economy. It exists because we thought it might be useful to business.

D. In the postmodern worldview, of the General Evolution Research Group, or of Rupert Sheldrake for example, the terrestrial, human, and celestial spheres are all in a process of concomitant coevolution, as in the embryogenesis of a new planetary society. In this habit of thought, the WWW may be regarded as the neurogenesis of the global brain, intrinsic to, and essential for, the overall coevolution of the all and everything.

This paper belongs to this last paradigm. It is our view that the WWW is essential to our further evolution, but that in order for this further evolution to have a favorable outcome, we must participate in the emerging consciousness of the global brain, and thus, we must visualize, observe, and interact with, the explosion of the

WWW. It is because of this belief that we have developed the tools of webometry which are described in this paper: the tools of Web Watch. Morphogenesis requires self-reference.

The works of Eric Chaisson, Peter Russell, Ervin Laszlo, and Rupert Sheldrake (listed in the bibliography) may be consulted for more details on this new paradigm.

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, connectionism views the intelligence of the network as primarily derived from its connections, as opposed to its nodes. While the number and sophistication of nodes may increase during neurogenesis, a maximum population is eventually attained. Meanwhile, the network of connections develops during embryogenesis, but then continues indefinitely. This is the physiological basis of learning, for example.

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, $g(i,j)$, denoting the strength of the connection. All of this data, the $g(i,j)$, may be set out in a single tableau, which is a square matrix of size N , the total number of nodes. After maturity is attained by the evolving neural net, this number may be regarded as fixed, although perhaps enormously large. The further evolution, such as learning, is then manifest by changes in this large matrix of real numbers.

And it is this matrix which we wish to observe, in Operation Web Watch, and to present to the web-literate public, the cybercitizens of the future planetary society, in order to empower self-reflection on this morphogenetic process, in which we may consciously participate in the creation of the future.

3 VISUALIZATION OF MASSIVE NEURAL NETS

Suppose given a massive neural net, that is, for which the size, N , may be on the order of tens or hundreds of thousands. How to observe its instantaneous state, or a sequence of states, to understand its evolution? In this paper we present only one of many possible strategies, already inherent in the neural net approach: the view of the matrix of connection strengths as a two-dimensional image. This may be done in shades of gray, or through translation by a color look-up table. There are two serious problems with this approach. Nevertheless, we advocate it here, and plan to pursue it in further work.

The first problem is in the massive size of the image. As computer screens and printed pages are generally limited to a size of one thousand or so, the literal image of a matrix of size N as conceived here must cover many computer screens, or many pages of print. The obvious solution to this problem of massive size is an intentional reduction of resolution, by pixel averaging for example.

The second problem is in the fictitious representation of the nodes in linear order, that is, as a one-dimensional geographic space, when in fact, the ordering given by the index (i) is arbitrary, or logical, or anything but geographical. In case there is a geometric or geographical map for the nodes of the neural net, its dimension is usually greater than one, and so the representation within a one-dimensional space is forced and artificial.

Note: Complex dynamical systems with geometric reference spaces have been discussed in the literature. For example, with a two-dimensional reference space, the connection matrix may be embedded in four dimensions, giving rise to a four-dimensional image.

Worse yet, these two problems aggravate each other. For averaging neighboring pixels, when the proximity of nodes has no natural significance, may destroy all significance in the image, providing a very foggy (that is, fractal) visualization of the net.

Nevertheless, we feel this approach has a certain promise, as fractal geometry provides tools for studying foggy (fractal) images. And here we propose just one of these tools: the pointwise fractal dimension. By computing the fractal dimension of the large matrix at each point, we obtain another matrix of the same size. This

derived matrix may be viewed as a topography of complexity, a parameter of considerable significance in the context of morphogenesis, even of foggy images. And furthermore, the derived image of the complexity of the original image may be expected to behave well under pixel averaging, or other resolution reducing transformations. For this invariance under scaling is a characteristic of fractals.

In summary, here is our proposal for viewing the morphogenetic process of a massive neural net:

- * given a large connection matrix, C
- * compute the pointwise dimension at each point, thus another large matrix, D
- * reduce the dimension as needed for viewing, to a smaller matrix, E .

Given a time series of connection matrices, compute the derivatives D and E for each, and view the time series of matrices, E , as a time-lapse movie of the morphogenesis of the net.

4 MEASURING THE WWW

Our strategy for viewing the morphogenetic process of a massive neural net may be applied to the WWW. That is indeed the main point of this paper. But how to represent the Web as a Net? There are clearly two necessary steps: to define the nodes, and to measure the connection strengths. For each of these steps there are many possibilities. Here we describe only one approach to each.

Nodes

The WWW is a tree consisting of domains, servers, and pages. There are now tens of thousands of domains, several servers in each domain, and many pages in each server. Each domain has a unique name (for example, *vismath.org*), each server has a unique name (e.g., *www.vismath.org*) and IP address (e.g., 162.227.70.1), and each page has a unique URL (e.g., *http://www.vismath.org/index.html*).

These are the main choices for nodes of the WWW. For reasons of size, mainly, let us regard domain names as the nodes of the Web. We may further reduce the size of the network to be visualized by considering only the suffices *edu* or *org*. Besides reducing to a smaller number of nodes, we might anticipate that the domains in the *com* class are relatively sparsely connected, and thus less interesting from the mathematical point of view.

Connections

The interconnections of the WWW, as a hypertext and hypermedia system, are links. Links connect pages, but pages are secondary to domains according to our choice above. Thus, given two domains, that is, nodes, we must determine all links from any page of the first domain, to any page of the second domain. Then this simple count should be normalized. That is, regarding the number of all pages of all servers of the first domain as a width, and all pages of all servers of the second domain as a height, we obtain a rectangle, the area of which (the product of the two page counts) may be regarded as contributing to the probability of a link. Thus, the connection strength we are proposing here is the ratio of the number of links to the product of the width and the height. A more precise measure might take into account the byte size of pages, or equivalently, the total storage served by each domain. However, this data is much more expensive to obtain.

In any case, the data to construct the massive connection matrix for the entire WWW is to be collected by a Web crawler or robot, not just once, but repeatedly, according to our larger plan. And fortunately for this program, a number of Web crawlers are already at work collecting links for indices of the WWW. This is to be the basis for further work in this project.

5 CONCLUSION

We have described a complete, step-by-step, procedure for the visualization of the complexity and morphogenesis of the World

Wide Web. The implementation of this procedure, our next goal, aims at the installation of a website in which, like a weather report, the current web image, and movies of earlier web images, are available for browsing. The stages of this implementation, in review, are:

- obtain connection matrix data for domains *.org, *.edu from a web crawler;
- transform to a matrix of pointwise fractal dimension;
- reduce by pixel averaging;
- post as GIF images on the web.

We see this as a relatively simple program, the first step being the most difficult. For this first step we see two options: one is to write our own web crawler, the other is to enter into partnership with one of the existing WWW-index services, such as: Alta Vista, Yahoo, Excite, etc.

Acknowledgments

Thanks to my class, Webology, at the University of California at Santa Cruz, Spring 1996, for the opportunity of testing these ideas on an unsympathetic audience, and to Don Foresta of the University of Paris for suggesting this idea in the first place. In a joint research project currently under way, we hope to actually carry out the fractal dimension strategy, presenting our results on the WWW at <http://www.vismath.org/webometry>. Many thanks to the London School of Economics and the University of Paris for grants making this research possible, and to the Istituto di Scienze Economiche of the University of Urbino for hospitality during the writing of this paper.

References

- Abraham, Fred D., Dynamical modeling and research of collective cognition, *J. World Futures*, to appear.
- Abraham, Ralph H., *Complex dynamics*, Santa Cruz, CA: Aerial Press, 1991.
- Abraham, Ralph H., Frank Jas, and Willard Russell, *The Web Empowerment Book*, New York: Springer-Verlag, 1995.
- Chaisson, Eric, *The Life Era*, New York: Atlantic Monthly Press, 1987.

Farmer, J. Doyne, E. Ott, and J. Yorke, Fractal dimension, *Physica D*, 7 (1983), p. 153.

Grossberg, Stephen and Michael Kuperstein, *Neural Dynamics of Adaptive Sensory-motor Control*, New York: Pergamon Press, 1989.

Laszlo, Ervin, *Evolution: the Grand Synthesis*, Boston: New Science Library, 1987.

Mandelbrot, Benoit, *The Fractal Geometry of Nature*, New York: W. H. Freeman, 1877/1982.

Russell, Peter, *The Global Brain*, Los Angeles: J. P. Tarcher, 1983.

Sheldrake, Rupert, *A New Science of Life*, London: Blond and Briggs, 1981.