‘Beauty’ of the World Wide Web — Cause, Goal, or Principle

Sándor Dominich, Júlia Góth, Mária Horváth, Tamás Kiezer

University of Veszprém, Department of Computer Science
8200 Veszprém, Egyetem u. 10, Hungary
Email: {dominich,goth,kiezer}@dcs.vein.hu

Abstract. It is known that the degree distribution in the World Wide Web (WWW) obeys a power law whose degree exponent exhibits a fairly robust behaviour. The usual method, linear regression, used to construct the power law is not based on any, probably existing, intrinsic property of the WWW which it is assumed to reflect. In the present paper, statistical evidence is given to conjecture that at the heart of this robustness property lies the Golden Section. Applications of this conjecture are also presented and discussed.

1 Introduction

The experimental discovery by Faloutsos et al. [7] that the degree distribution for Web pages and Internet nodes follows a power law with a fairly robust degree exponent value was a basic milestone towards the emergence of a new science of the Web. The formulation of the principle of preferential attachment [5] triggered research into and stimulated ideas towards trying to explain, using generative models [13], why the Web link topology evolves according to a power law. Pennock et al. [15] as well as Adamic and Huberman [3] showed that this principle is not necessarily valid in the real Web; modified principles were proposed to better explain the development of a power law for degree distribution in the real Web.

Kahng et al [11] investigated the question of why the degree exponent exhibits a fairly robust behaviour, just above 2. Using a directed network model in which the number of vertices grows geometrically with time, and the number of edges evolves according to a multiplicative process, they established the distribution of in- and out-degrees in such networks. They arrived at the result that if the degree of vertex grows at a higher pace than the edges then the in-degree distribution is independent of the ‘details’ of the network.

The usual method, that of linear regression, used to construct the Power Law is not based on any ‘internal’ property of the Web network—it is a mere reflection of some deeper structure. The generative models proposed thus far are incomplete. They only model growth, and fail to take into account that nodes and links are also destroyed (not just added). It is not known how the processes of growth and extinction go on in the Web, how they relate to each other to give birth to what we observe as a power law.

In the present paper, based on the robustness property of the degree exponent, a different approach is proposed: it is conjectured that, at the present scale of the Web, at the heart of this robustness property lies the Golden Section. The Golden Section is one of the most ancient and overdone yet evergreen topics in mathematics. It is also far-reaching in several other fields, e.g., art, architecture, biology, music, physics. There is a common agreement that it always relates — subjectively — to a notion of 'beauty' of the field. For example, it is believed that rectangles whose width-to-height ratio is the Golden Section are the most pleasing to the human eye, or that the timing of musical pieces is considered to be most pleasing to human ears when in Golden Section. In this paper, it is beleived that the evolution of the Web link topology may have an intrinsic property that is reflected in the Golden Section, and this is the expression of an inner beauty of the Web.

After a brief overview of several degree exponent values obtained experimentally, statistical evidence is given to conjecture that the degree exponent value varies around a Golden Section-based value. Using number theoretic results, this conjecture is then used, on the one hand, to propose a method, referred to as F-L method, for the construction of the Power Law for the real Web portion under focus, and, on the other hand, to give a theoretical underpinning for the application of high degree walks in crawling and searching in peer-to-peer networks. Also, formal relationships between the Golden Section and the LCD as well as Bollobás models are shown.

2 Power Law

If the probability P that a discrete random variable V assumes values equal to or greater than some value v is given by

$$P(V \geq v) = \left(\frac{m}{v}\right)^k, \quad m > 0, k > 0, v \geq m, \tag{1}$$

we say that V follows Pareto's Law [1, 10]. It follows from (1) that:

$$P(V < v) = 1 - \left(\frac{m}{v}\right)^k, \tag{2}$$

which is the distribution function $F(v)$ of V; it is differentiable with respect to v, the derivative is continuous (absolutely continuous). V has density function $f(v) = F'(v) = m^k \cdot v^{-(k+1)}$. The function $f(v)$ is referred to as a Power Law [18], and it is usually written in the following general form:

$$f(v) = C \cdot v^{-\alpha}, \tag{3}$$

where C is a — problem-dependent — constant, α is referred to as the degree exponent. For visualisation purposes, the Power Law is represented in a log-log plot as a straight line obtained by taking the logarithm of (3):

$$\log f(v) = \log C - \alpha \times \log v\ . \tag{4}$$

log v is represented on the abscissa, log $f(x)$ on the ordinata, $-\alpha$ is the slope, log C is the intercept. Given two sequences of values $X = (x_1,..., x_n)$ and $Y = (y_1, ..., y_n)$. If the correlation coefficient $r(X, Y)$ suggests a fairly strong correlation — i.e., it is close to $|1|$ — between X and Y at a log scale, then a regression line can be drawn to exhibit a relationship between X and Y; using the slope and the intercept of the regression line the corresponding Power Law can be written.

3 Degree Exponent Values for the World Wide Web

Faloutsos et al. [7] arrived at the result that, using data provided by the National Laboratory for Applied Networks Research between the end of 1997 and end of 1998, the tail of the frequency distribution of an out-degree — i.e., the number of Internet nodes and Web pages with a given out-degree — is proportional to a Power Law. Their observation was that the values of the exponent seemed to be almost constant: 2.15; 2.16; 2.2; 2.48.

Barabási et al. [5] — using 325,729 HTML pages involving 1,469,680 links from the *nd.edu* domain — confirmed the earlier results obtained for the values of the degree exponent. They obtained the value 2.45 for out-degree, and 2.1 for in-degree.

In [6], two experiments are described using two web crawls, one in May and another one in October 1999, provided by Altavista, involving 200 million pages and 1.5 billion links. The results arrived at were the same in both experiments: the values of the degree exponent were estimated to be 2.1, 2.54, 2.09, 2.67, 2.72 for out-links distribution.

The values obtained earlier for the degree exponent were also confirmed by Pennock et al. [15], who found — using 100,000 Web pages selected at random from one billion URLs of Inktomi Corporation Webmap; they binned the frequencies using histograms — that the exponent for out-degree was 2.72, whereas 2.1 for in-degrees. Similar exponent values were obtained for the in-degree distribution for category specific homepages: 2.05 for companies and newspapers, 2.63 for universities, 2.66 for scientists, and 2.05 for newspapers.

Shiode and Batty [16] assessed the Power Law for Web country domain names in- and out-link distribution as of 1999. Their results for the Power Law exponent were the following values: 2.91, 1.6, 2.98, 1.46, 2.18, 2.

Adamic and Huberman [3] report on an experiment involving 260,000 sites, each representing a separate domain name. The degree exponent was estimated to be 1.94.

In [12], it is reported that a copy of the 1997 Web from Alexa (a company that archives the state of the Web) was used to estimate the degree exponent of the Power Law. The data consisted of about 1 Terabyte of data representing the content of over 200 million web pages. It was found that the degree exponent was 2.38.

In [4], it is reported that the value of 2.3 was found for the degree exponent, in [9] the values 2.1 and 2.38 are reported, while in [14] 2.1 and 2.7.

Friedman et al. [8], using a crawl on the *.hu* domain, assessed the power law for 11,359,640 pages and 95,713,140 links, and found the following values for exponent: 2.29 for in-degree, and 2.78 for out-degree.

Experiment 1. Using the "Barabási-data"[1], we repeated the fitting of a Power Law curve to out-degree distribution. Fig. 1 shows our results. (Computational details are given in the Appendix.)

Experiment 2. We generated the in-links frequency distribution for country domain names[2] as of January, 2004 (Fig. 2). The domain names *.gov, .org, .net, .edu, .us, .com, .mil, .um, .vi* were all considered as representing the USA, the domain names *.ac*, *.uk*, *.gb* as representing the UK, and *.fr*, *.fx* for France. The number of inlinks for every country domain name was identified using Altavista search engine's Webmasters option during 19-22 January, 2004. For example, the UK got a total of 30,701,157 in-links, the USA got 271,019,148; Albania got 2,041,573, Belgium got 3,386,900 in-links. The in-links were binned into 1,000 equally spaced intervals, the correlation coefficient was found to be –0.99 (at a log scale). The value for the Power Law exponent was found to be equal to α = 1.18 using Mathcad's *linfit* linear regression command, the approximation error was equal to 14,509.

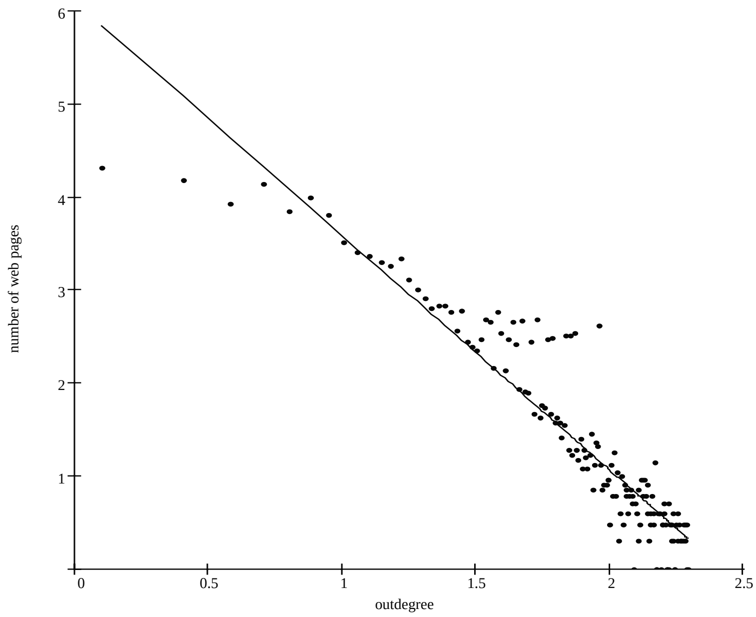

Fig. 1. World Wide Web Power Law. The frequency (i.e., number of Web pages) of the outdegrees of Web pages plotted at a log-log scale. The points represent real values, the straight line represents the regression line fitted to the real values. The correlation coefficient is equal to r = –0.94, the Power Law exponent is equal to α = 2.5

---

1 Provided at http://www.nd.edu/~networks/database/index.html; downloaded January 2, 2004

2 Taken from http://www.webopedia.com/quick_ref/topleveldomains.

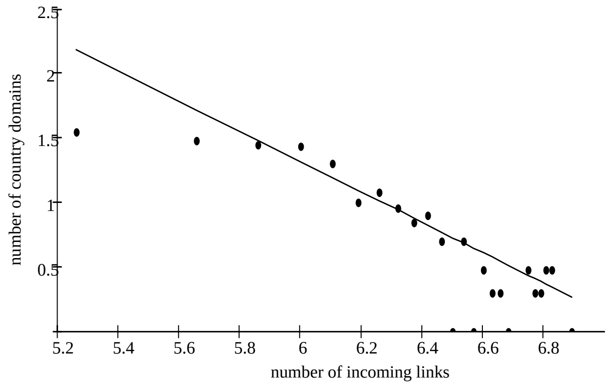

Fig. 2. Log-log plot of the Power Law for the in-links of country domain names as of January, 2004. The correlation between the number of in-links and the corresponding number of country domain names was found to be –0.99, whereas the value of the power law exponent was 1.18

The estimated values obtained experimentally thus far for the exponent of the Power Law for degree distribution in then Web Power Law are summarised in Table 1.

Table 1. Estimated values obtained experimentally thus far for the exponent of the Power Law for degree distribution in the World Wide Web

Source (experiment)	**Degree exponent value**
Faloutsos et al. (1999)	2.15; 2.16; 2.2; 2.48
Barabási et al. (2000)	2.1; 2.45
Broder et al (2001)	2.1; 2.72; 2.09; 2.67; 2.54
Pennock et al. (2002)	2.1; 2.72, 2.05; 2.05; 2.63; 2.66
Kumar et al. (1998)	2.38
Adamic, Huberman (2000)	1.94
Shiode, Batty (2000)	2.91; 1.6; 2.98; 1.46; 2.18; 2
Albert (2000)	2.3
Gil et al. ()	2.1; 2.38
Pandarungan (2002)	2.1; 2.7
Friedman et al. (2003)	2.29; 2.78
Experiment 1 (see text)	2.5
Experiment 2 (see text)	1.18

4 Statistics of the Experimentally Obtained Degree Exponent Values

Let us conceive the different degree exponent values obtained experimentally (Table 1) as being a sample drawn from a population [17] consisting of degree exponent values (the population may consist, for example, of the degree exponent values obtained using the data of all Web crawlers, all domain names, etc.). Our sample has size $N = 34$. The mean M of the sample is equal to

$$M = \frac{1}{N}\sum_{i=1}^{N}\alpha_i = 2.284 \quad (5)$$

The standard deviation s of the sample is equal to

$$s = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\alpha_i - M)^2} = 0.392 \quad (6)$$

Because all the degree exponent values α_i lie in the open interval (2; 3), the mean μ, whether sample or population ('true') mean, should also lie in this same interval. We may ask ourselves the question of whether there exists positive integer numbers p such that the hypothesis: "$\mu = \sqrt{p}$" be supported. Candidate values for p are 4, 5, 6, 7, 8. Using the

$$z\text{-score}(\mu) = \left|\frac{M - \mu}{s / \sqrt{N - 1}}\right| \quad (7)$$

the following values are obtained:

$$z\text{-score}(\sqrt{4}) = 4.163, \quad z\text{-score}(\sqrt{5}) = 0.7,$$
$$z\text{-score}(\sqrt{6}) = 2.43, \quad z\text{-score}(\sqrt{7}) = 5.309, \quad z\text{-score}(\sqrt{8}) = 7.988.$$

The 95% confidence interval for the score is $-2.035 < z\text{-score}(\mu) < 2.035$ ($t_{0.975} = 2.035$, $N - 1 = 33$ degrees of freedom), only z-score($\sqrt{5}$) lies within this interval. Thus, the 95% confidence interval for the mean $\mu = \sqrt{5} = 2.236$ is as follows:

$$2.145 < \mu < 2.422 \quad (8)$$

We may hence say that there is statistical support to assume that the sample comes from a population with mean $\mu = \sqrt{5}$. Thus, the Power Law for the degree distribution in the World Wide Web may be written in the following form:

$$f(x) \approx C \cdot x^{-\sqrt{5}} \quad (9)$$

i.e., the number $f(x)$ of Web nodes having degree x is proportional to $x^{-\sqrt{5}}$.

4 Golden Section, Fibonacci and Lucas Numbers

In this part, those properties of the Golden Section, the Fibonacci and the Lucas numbers are recalled which are of interest to us in connecting them with the Web Power Law.

4.1 Golden Section

The Golden Section (*aka* Golden Ratio, Golden Mean, Divine Proportion) is denoted by φ, and defined as the smallest root of the equation:

$$x^2 - x - 1 = 0; \quad \varphi = (\sqrt{5} - 1)/2 \approx 0.61803398875$$

The other root is $\Phi = (\sqrt{5} + 1)/2 \approx 1.61803398875$. The following relationships hold:

$$\sqrt{5} = 2\varphi + 1, \quad \varphi \cdot \Phi = 1$$

A straightforward connection between the degree exponent as defined in (9) and the Golden Section is as follows:

$$\mu = \sqrt{5} = 2\varphi + 1$$

4.2 Fibonacci Numbers

The *Fibonacci numbers* are defined as $F_0 = 0$, $F_1 = 1$, $F_n = F_{n-1} + F_{n-2}$, $n \geq 2$, i.e., 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ... The ratio of the consecutive numbers (i.e., 5/8 = 0.625; 8/13 = 0.615; 13/21 = 0.619;…) has limit equal to the Golden Section, namely:

$$\lim_{n \to \infty} \frac{F_n}{F_{n+1}} = \varphi \qquad (10)$$

The Golden Section and the Fibonacci numbers are related by Binet's formula:

$$F_n = \frac{1}{\sqrt{5}} \left(\Phi^n - (-\varphi)^n \right) \qquad (11)$$

from which it follows that:

$$(-1)^n \cdot \varphi^{2n} + F_n \cdot (2\varphi + 1) \cdot \varphi^n = 1, \quad n = 0, 1, 2, ... \qquad (12)$$

4.3 Lucas Numbers

If the recurrence relation $L_n = L_{n-1} + L_{n-2}$, $n \geq 2$, is initialised with the numbers $L_0 = 2$, $L_1 = 1$, then one obtains the *Lucas numbers* L_n: 2, 1, 3, 4, 7, 11, 18, 29, 47, 76,… . It can be shown, e.g., using induction on n, that the Fibonacci and Lucas numbers are bound by the following relationship:

$$L_n = F_{n-1} + F_{n+1}, \quad n \geq 1 \tag{13}$$

5 Constructing the Web Power Law using Fibonacci and Lucas numbers

In this part, we propose a method based on the Golden Section, Fibonacci and Lucas numbers to construct the Web Power Law for a web portion under focus. Also, experimental evidence will be given to demonstrate the application of the method in practice.

5.1 Fibonacci-Lucas (F-L) method

Taking into account the relationship (10), for sufficiently large values of *n*, we can write:

$$\sqrt{5} = 2\varphi+1 \approx 2\cdot\frac{F(n-1)}{F(n)}+1=\frac{F(n-1)+F(n-1)+F(n)}{F(n)}. \tag{14}$$

which — given the recursive definition of the Fibonacci numbers, and taking into account the relationship between The Fibonacci and Lucas numbers — becomes

$$\frac{F(n-2)+F(n-3)+F(n-1)+F(n)}{F(n)} = \frac{L(n-1)+L(n-2)}{F(n)} = \frac{L(n)}{F(n)} \tag{15}$$

Thus, the Web Power Law (9) re-writes in a form in which the exponent is expressed using both Fibonacci and Lucas numbers as follows:

$$f(x) \approx C\cdot x^{-\frac{L(n)}{F(n)}}. \tag{16}$$

Taking the logarithm of the relationship (16), one can write the following:

$$\log f(x) \approx \log C - \frac{L(n)}{F(n)}\log x \qquad \log f(x) + \frac{L(n)}{F(n)}\log x \approx \log C$$

$$F(n)\cdot\log f(x) + L(n)\cdot\log x \approx F(n)\cdot\log C \qquad f(x)^{F(n)}\cdot x^{L(n)} \approx C^{F(n)} \tag{17}$$

For real Web data, $f(x)$ is not a computed value but the actual frequency, while the Power Law exponent is slightly different from $L(n)/F(n)$. Let X_k denote the actual page degrees and Y_k denote the corresponding actual frequency (k = 1, 2, …, M). Then, the relationship (17) becomes:

$$F(n)\cdot\log Y_k + L(n)\cdot\log X_k \approx F(n)\cdot\log C \quad \textbf{(18)}$$

Because the relationship (18) should hold for every k = 1, 2, …, M, the mean of the left-hand side taken over all k should equal $F(n)$·log C (of course, with an inherent approximation error):

$$\frac{1}{M}\sum_{k=1}^{M} \left(F(n)\cdot\log Y_k + L(n)\cdot\log X_k \right) \approx \frac{1}{M}\sum_{k=1}^{M} F(n)\cdot\log C = \quad \textbf{(19)}$$

$$= \frac{1}{M}\cdot M\cdot F(n)\cdot\log C = F(n)\cdot\log C$$

This property makes it possible to propose the following method for constructing a specific Power Law for given real Web data.

Fibonacci-Lucas (F-L) Method for constructing the Web Power Law for degree distribution

Step 1. Establish the number of degrees (e.g., out-links) X_k (in ascending order) and their corresponding frequencies Y_k, , k = 1, 2, …, M, for the Web or Internet nodes under focus.

Step 2. Choose some n, e.g., n = 8, 9 or 19, and compute the corresponding Fibonacci number $F(n)$ and Lucas number $L(n)$ using, for example, Binet's formula (11) and formula (12) respectively (or other formulas available).

Step 3. Compute the left-hand side of the relationship (18), i.e.,

$$S_k = F(n)\cdot\log Y_k + L(n)\cdot\log X_k\,, \quad k = 1, 2, \ldots, M \quad \textbf{(20)}$$

Step 4. Compute the mean μ of S_k over all k, i.e.,

$$\mu = \frac{1}{M}\sum_{k=1}^{M} S_k \quad \textbf{(21)}$$

Step 5. Apply a correction equal to the standard deviation of S_k to compensate for the approximation errors:

$$\mu' = \mu + \mathrm{stdev}(\mu) \quad \textbf{(22)}$$

Step 6. Compute the approximate value for the constant C as follows (using the relationship (17)):

$$C = 10^{\frac{\mu'}{F(n)}} \quad \textbf{(23)}$$

Step 7. Write the specific Power Law for the real Web or Internet portion under focus as follows:

$$f(x) \approx C\cdot x^{-\frac{L(n)}{F(n)}} \quad \textbf{(24)}$$

where, as already seen, x denotes degree and $f(x)$ denotes frequency.

5.2 Experimental evidence in support of the F-L method

We give now experimental evidence to support the applicability of the F-L method proposed above in practice.

Using the data of the Appendix, we applied the F-L method using the following pairs of $F(n)$ and $L(n)$:

$F(8) = 21$, $L(8) = 47$
$F(19) = 4181$, $L(19) = 9349$
$F(22) = 17711$, $L(22) = 39603$
$F(29) = 514229$, $L(29) = 1149851$.

For example, when $F(8) = 21$ and $L(8) = 47$, S_k assumes the values 115, 120, 122, 119, etc.. The means μ from Step 4 corresponding to the pairs $F(n)$ and $L(n)$ are as follows (rounded to integer values): 112; 22,332; 94,603; 274,6752, respectively, whereas the corrected means μ' of the Step 5 are as follows: 128; 25,576; 108,342; 3,145,661 (rounded to integer values). The application of the Steps 6 and 7 yielded the following Power Laws for the four $F(n)$ and $L(n)$ pairs respectively:

$$f(x) = 10^{\frac{128}{21}} \cdot x^{-2.23} = 1{,}324{,}3171 \cdot x^{-2.23}$$

$$f(x) = 10^{\frac{25576}{4181}} \cdot x^{-2.23} = 1{,}309{,}903 \cdot x^{-2.23}$$

$$f(x) = 10^{\frac{108342}{17711}} \cdot x^{-2.23} = 1{,}309{,}904 \cdot x^{-2.23}$$

$$f(x) = 10^{\frac{3145661}{514229}} \cdot x^{-2.23} = 1{,}309{,}904 \cdot x^{-2.23}$$

It can be seen that the values obtained for the constant are fairly stable and compare well with that obtained in the experiment of the Appendix using linear regression: $10^{6.1043} = 1{,}271{,}452$.

6 High Degree Seeking Walk

In a high degree seeking algorithm (HDS) an arbitrary node is chosen first, then a node with a degree higher than the current node; once the highest degree node has been found, a node of approximately second highest degree will be chosen, and so on. In a peer-to-peer (P2P) system, like GNUTELLA (which obeys a power law), a query is iteratively sent to all the nodes in a neighborhood of the current node until a matching is found. This broadcasting is costly in terms of bandwidth. If every node keeps adequate information (e.g., file names) about its first and second neighbors, then HDS can be implemented. Because storage is likely to remain less expensive than bandwidth, and since network saturation is a weakness of P2P, HDS can be an efficient alternative to usual searching. Adamic et al. show [2] that the expected degree $E(\alpha, n)$ of the richest neighbor of a node having degree n is given by

$$E(\alpha, n) = \frac{n(\alpha-2)}{(1-N^{2/\alpha-1})^n} \sum_{x=0}^{\lfloor N^{1\alpha} \rfloor} x(1+x)^{1-\alpha}(1-(x+1)^{2-\alpha})^{n-1} \quad (25)$$

where N denotes the number of nodes in the graph, α is the power law exponent. Fig. 3 shows simulation results for the ratio $E(\alpha, n)/n$. It can be seen that for the power law exponent between 2 and 2.3, the chance to find a richer neighbor is higher than the degree of the node itself within a relatively large interval of degree values, which means that HDS can be applied non-trivially. In Web search engines and retrieval, crawlers implement different strategies, e.g., breadth-first-search, to crawl the Web graph. However, it is becoming increasingly difficult to cope with scalability limitations. One possible way is given by an HDS-based crawling strategy which exploits the power law property of link distribution. From eq. (9) we have that $2 < \alpha < 2.3$, which may be viewed as a theoretical justification for the application of HDS to GNUTELLA or to crawling.

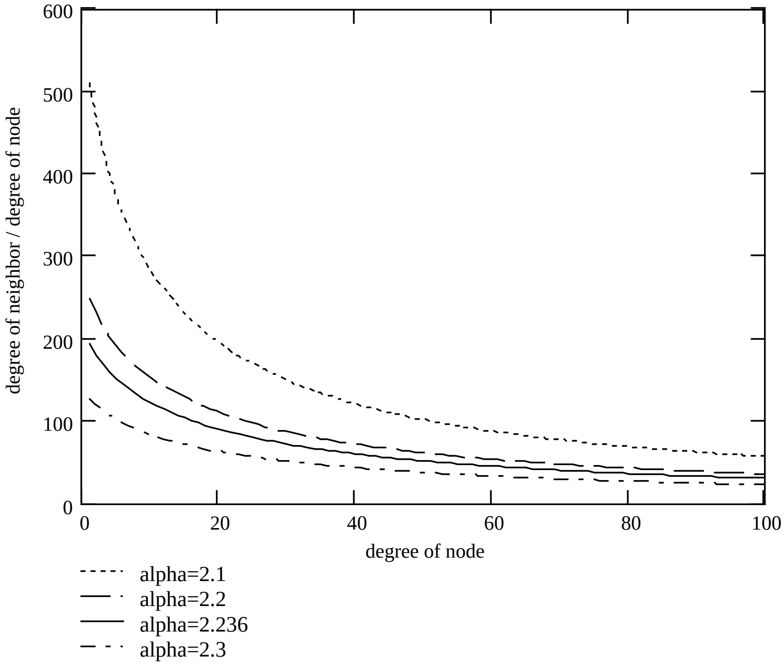

Fig. 3. Simulation of the ratio of the expected degree of the richest neighbor of a node with degree *n* for different values of the power law exponent alpha. The total number of nodes is equal to 100,000,000; and α = alpha = 2φ + 1= √5

7 Formal Relationships Between the Golden Section, Degree and Probabilities in Generative Models

Based on results presented by Bollobás[3], we can establish the following formal relationships between the Golden Section and the degree distributions and link probabilities in generative models.

In the LCD model, it is shown that in a graph with n vertices and m edges the fraction $F = \#d/n$ of vertices having degree d is bounded as follows:

$$(1-\varepsilon)\alpha \leq F \leq (1+\varepsilon)\alpha,$$

where $\alpha = 2m(m+1) / [(d+m)(d+m+1)(d+m+2)]$. Based on (9) and part 4.1, we take $F = d^{-(2\varphi+1)}$, and thus we obtain (taking the logarithm of both sides and appropriately re-arranging) the following relationships between the Golden Section and degree:

$$1/[(1-\varepsilon)\alpha d] \leq d^{2\varphi} \leq 1/[(1+\varepsilon)\alpha d].$$

In the Bollobás model, the number $x_i(t)$ of nodes having in-degree i at step t is given by:

$$x_i(t) = Ci^{-(1+1/c)}, \quad c = (\alpha+\beta)/[1+\delta(\alpha+\gamma)], \ \gamma \in \mathbb{R}_+$$

where α is the probability to add a new vertex together with an edge from it to an old vertex, β is the probability to add an edge between two existing vertices, and γ is the probability to add a new vertex and an edge from an old vertex to it (obviously $\alpha + \beta + \gamma = 1$). Based on (9) and part 4.1, we take $1 + 1/c = 2\varphi+1$, thus we obtain:

$$\varphi = [1+\delta(\alpha+\gamma)] / [2(\alpha+\beta)].$$

8 Conclusions

After a brief summary of the numeric values obtained experimentally for the degree exponent in the Web Power Law, experiments are reported that assessed the Power Law, and confirmed earlier results. Then, using hypothesis testing, it was shown that the mean value of the degree exponent could be taken as being equal to $\sqrt{5}$ with a 95% confidence. The direct relationship between $\sqrt{5}$ and the Golden Section φ yielded to considering also the Fibonacci and Lucas numbers. Formal relationships were derived, which made it possible to express the Web Power Law using Fibonacci and Lucas numbers. Based on this result, a method was proposed, called F-L-method, to construct a specific Power Law for a real Web portion under focus. Experimental evidence was given to support the applicability of the method proposed. Also, it was shown, that the results obtained in this paper may serve as a theoretical

[3] Bollobás, B. (2003). Mathematical results on scale-free networks. http://stat-www.berkeley.edu/users/aldous/Networks (downloaded December, 2004)

underpinning for the application of high degree seeking walks in crawling and peer-peer-to searching.

The computationally useful and mathematically interesting relationship between degree frequencies, Golden Section, Fibonacci and Lucas numbers contained in this property can open up further possibilities to involve number theory into the computational study of Web topology. Because in this paper statistical evidence was given that supported the possibility that the mean value of the degree exponent in the Web Power Law can be expressed in terms of the Golden Section, it might be conjectured that, at very large scales, the link topology in the World Wide Web evolves in such a way as to exhibit a Golden Section-based behaviour. What we do not know is whether the **Golden Section** characteristic shown in the Web is a **cause** (i.e., the robustness of the value of the degree exponent — in other words: the evolution of link topology — stems from the Golden Section ultimately), a **goal** (i.e., the link topology evolves in such a way that a Golden Section-based degree exponent be reached), or a **principle** (i.e., the link topology develops according to some Golden Section-based rule 'hidden' in the structure somewhere).

References

1. Adamic, L. A. (2003). Zipf, Power-laws, and Pareto – a ranking tutorial. http://ginger.hpl.hp.com/shl/papers/ranking/ranking.html (visited: December 18, 2003)
2. Adamic, L., Lukose, R.M., Puniyami, A.R., Huberman, B.A. (2001). Search in power-law networks. *Physical Review*, The American Physical Society, vol. 64, pp: 046135(8)
3. Adamic, L.A., and Huberman, B.A. (2000). Power-Law Distribution of the World Wide Web. *Science*, vol. 287, March, p.2115a
4. Albert, R. (2000). PhD Dissertation, University of Notre Dame (visited December, 2003).
5. Barabási, A-L. Albert, R., and Jeong, H. (2000). Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A*, 281, pp: 69-77.
6. Broder, A., Kumar, R., et.al. (2000). Graph structure in the Web. *Computer Networks*, vol. 33, pp: 309-320.
7. Faloutsos, M., Faloutsos, P., and Faloutsos, Ch. (1999). On Power-Law Relationship of the Internet Topology. *Proceeedings of the ACM SIGCOMM*, Cambridge, MA, pp: 251-262.
8. Friedman, E., Uher, M., Windhager, E. (2003). A magyar Web. (The Hungarian Web). *Híradástechnika*, vol. 58, no. 3., pp: 25-31 (in Hungarian)
9. Gil, R, Garcia, R, Delgado, J. (2004). Are we able to charaterize Semantic Web behaviour? http://dmag.upf.es/livingsw (downloaded December, 2004)
10. Guilmi, C., Gaffeo, E., and Gallegati, M (2003). Power Law Scaling in the World Income Distribution. *Economics Bulletin*, vol. 15, no. 6, pp: 1-7.
11. Kahng, B., Park, Y., and Jeong, H. (2002). Robustness of the in-degree exponent for the World Wide Web. *Physical Review*, vol. 66, 046107 1-6.
12. Kumar, R., Raghavan, P., Rajagopalan, S., and Tomkins, A. (1998). Trawling the web for emerging cyber-communities. *Proceedings of the WWW8 Conference*. http://www8.org/w8-papers.
13. Menczel, F. (2002). Growing and navigating the small world Web by local content. *Proceedings of the PNAS*, vol. 99, no. 22, pp: 14014-14019.

14. Pandurangan, G., Raghavan, P. (2002). Using PageRank to charaterize Web Structure. *Lecture Notes In Computer Science*, vol. 2387, Springer Verlag, pp: 330-339
15. Pennock, D.M., Flake, G.W., Lawrence, S., Glover, E., and Giles, L. (2002). Winners don't take all: characterizing the competition for links on the Web. *Proceedings of the New York Academy of Sciences*, April 16, vol. 99, no. 8, pp: 5207-5211.
16. Shiode, N., and Batty, M. (2000). Power Law Distribution in Real and Virtual Worlds. http://www.isoc.org/inet2000/cdproceeedings/2a/2a_2.htm (downloaded on December 18, 2003.
17. Spiegel, M.R. (1961). *Theory and Problems of Statistics*. McGraw Hill.
18. Zipf, G. K. (1949). *Human Behaviour and Principle of Least Effort*. Addison Wesley, Cambridge, Massachussetts.

Appendix

Experiment 1. Using the "Barabási-data"[4], we repeated the fitting of a Power Law curve to out-degree distribution. The data was provided as a zipped file; after unzipping it the result was a text file which contained two numbers in each line: the leftmost number was the sequence number of Web pages (0; 1; 2; …; 325,729), the other number was the sequence number of the Web page pointed to by the page represented by the leftmost number. A noteworthy observation is that the exponent of the Web Power Law is slowly increasing from 1 with the number of pages (from a few hundred up to several ten thousand pages), and is starting to stabilise around the value $\alpha = 2.5$ if the number of Web pages involved is fairly high, above 100,000. Thus, for example, for 30,000 pages, the correlation — at a log scale — r between out-degree and frequency was only $r = -0.892$, and the fitting of a Power Law curve $C \cdot x^{-\alpha}$ using Mathcad's in-built curve fitting command *genfit* resulted in $\alpha = 0.867$ with an approximation error of the sum of the absolute values of differences of 3.7×10^6 at 10^{-4} convergence error, whereas using linear regression yielded $\alpha = 1.47$ with an approximation error of 1,589,104 at 10^{-4} convergence error. Fig. 1 shows our results (see text) for a number of 256,062 Web pages — involving 1,139,426 links — selected at random from the provided 325,729 pages. After processing this file the X data consisted of the out-degrees of Web pages, whereas the Y data consisted of the corresponding frequencies. For example, there were 2,206 pages having out-degree 13, and the out-degree 14 had its frequency equal to 1,311. The empirical correlation coefficient — taking log scale data — r between out-degree and frequency was $r = -0.94$. The linear regression yielded the following values: $\alpha = 2.5$ for the exponent; and $C = 10^{6.1043}$ for the constant. The computation was performed using Matchcad's in-built *line* command; the numeric computation used in this command as well as the fact that we used 69,667 pages less may account for the difference of 0.05 in the exponent value compared to the value reported in [4]. Because of the strong correlation (see above) and power-like behaviour, and also due to inherently present numeric approximation errors, we believe that the difference of 0.05 is not an important one.

[4] Provided at http://www.nd.edu/~networks/database/index.html; downloaded January 2, 2004