# Traffic-Driven Model of the World Wide Web Graph

Alain Barrat[1], Marc Barthélemy[2], and Alessandro Vespignani[1,3]

[1] Laboratoire de Physique Théorique (UMR du CNRS 8627), Bâtiment 210, Université de Paris-Sud 91405 Orsay, France
[2] CEA-Centre d'Etudes de Bruyères-le-Châtel, Département de Physique Théorique et Appliquée BP12, 91680 Bruyères-Le-Châtel, France
[3] School of Informatics, Indiana University, Bloomington, IN 47408, USA

**Abstract.** We propose a model for the World Wide Web graph that couples the topological growth with the traffic's dynamical evolution. The model is based on a simple traffic-driven dynamics and generates weighted directed graphs exhibiting the statistical properties observed in the Web. In particular, the model yields a non-trivial time evolution of vertices and heavy-tail distributions for the topological and traffic properties. The generated graphs exhibit a complex architecture with a hierarchy of cohesiveness levels similar to those observed in the analysis of real data.

## 1 Introduction

The World Wide Web (WWW) has evolved into a huge and intricate structure whose understanding represents a major scientific and technological challenge. A fundamental step in this direction is taken with the experimental studies of the WWW graph structure in which vertices and directed edges are identified with web-pages and hyperlinks, respectively. These studies are based on crawlers that explore the WWW connectivity by following the links on each discovered page, thus reconstructing the topological properties of the representative graph. In particular, data gathered in large scale crawls [1, 2, 3, 4, 5] have uncovered the presence of a complex architecture underlying the structure of the WWW graph. A first observation is the *small-world* property [6] which means that the average distance between two vertices (measured by the length of the shortest path) is very small. Another important result is that the WWW exhibits a power-law relationship between the frequency of vertices and their degree, defined as the number of directed edges linking each vertex to its neighbors. This last feature is the signature of a very complex and heterogeneous topology with statistical fluctuations extending over many length scales [1].

These complex topological properties are not exclusive to the WWW and are encountered in a wide range of networked structures belonging to very different domains such as ecology, biology, social and technological systems [7, 8, 9, 10]. The need for general principles explaining the emergence of complex topological

S. Leonardi (Ed.): WAW 2004, LNCS 3243, pp. 56–67, 2004.

features in very diverse systems has led to a wide array of models aimed at capturing various properties of real networks [7, 9, 10], including the WWW. Models do however generally consider only the topological structure and do not take into account the interaction strength –the weight of the link– that characterizes real networks [11, 12, 13, 14, 15, 16]. Interestingly, recent studies of various types of weighted networks [15, 17] have shown additional complex properties such as broad distributions and non-trivial correlations of weights that do not find an explanation just in terms of the underlying topological structure. In the case of the WWW, it has also been recognized that the complexity of the network encompasses not only its topology but also the dynamics of information. Examples of this complexity are navigation patterns, community structures, congestions, and other social phenomena resulting from the users' behavior [18, 19]. In addition, Adamic and Huberman [4] pointed out that the number of users of a web-site is broadly distributed, showing the relevance and heterogeneity of the traffic carried by the WWW.

In this work we propose a simple model for the WWW graph that takes into account the traffic (number of visitors) on the hyper-links and considers the dynamical basic evolution of the system as being driven by the traffic properties of web-pages and hyperlinks. The model also mimics the natural evolution and reinforcements of interactions in the Web by allowing the dynamical evolution of weights during the system growth. The model displays power-law behavior for the different quantities, with non-trivial exponents whose values depend on the model's parameters and which are close to the ones observed empirically. Strikingly, the model recovers a heavy-tailed out-traffic distribution whatever the out-degree distribution. Finally we find non-trivial clustering properties signaling the presence of hierarchy and correlations in the graph architecture, in agreement with what is observed in real data of the WWW.

### 1.1 Related Works: Existing Models for the Web

It has been realized early that the traditional random graph model, i.e. the Erdös-Renyi paradigm, fails to reproduce the topological features found in the WebGraph such as the broad degree probability distribution, and to provide a model for a dynamical growing network. An important step in the modeling of evolving networks was taken by Barabási et al. [1, 20] who proposed the ingredient of preferential attachment: at each time-step, a new vertex is introduced and connects randomly to already present vertices with a probability proportional to their degree. The combined ingredients of growth and preferential attachment naturally lead to power-law distributed degree. Numerous variations of this model have been formulated [10] to include different features such as re-wiring [21, 22], additional edges, directionality [23, 24], fitness [25] or limited information [26].

A very interesting class of models that considers the main features of the WWW growth has been introduced by Kumar et al. [3] in order to produce a mechanism which does not assume the knowledge of the degree of the existing vertices. Each newly introduced vertex $n$ selects at random an already existing vertex $p$; for each out-neighbour $j$ of $p$, $n$ connects to $j$ with a certain probability $\alpha$; with probability $1 - \alpha$ it connects instead to another randomly chosen node. This model describes the growth process of the WWW as a copy mechanism in which newly arriving web-pages tends to reproduce the hyperlinks of similar web-pages; i.e. the first to which they connect. Interestingly, this model effectively recovers a preferential attachment mechanism without explicitely introducing it.

Other proposals in the WWW modeling include the use of the rank values computed by the PageRank algorithm used in search engines, combined with the preferential attachment ingredient [27], or multilayer models grouping web-pages in different regions [28] in order to obtain bipartite cliques in the network. Finally, recent models include the textual content affinity [29] as the main ingredient of the WWW evolution.

## 2 Weighted Model of the WWW Graph

### 2.1 The WWW Graph

The WWW network can be mathematically represented as a directed graph $\mathcal{G} = (V, E)$ where $V$ is the set of nodes which are the web-pages and where $E$ is the set of ordered edges $(i, j)$ which are the *directed* hyperlinks ($i, j = 1, ..., N$ where $N = |V|$ is the size of the network). Each node $i \in V$ has thus an ensemble $\mathcal{V}_{in}(i)$ of pages pointing to $i$ (in-neighbours) and another set $\mathcal{V}_{out}(i)$ of pages directly accessible from $i$ (out-neighbours). The degree $k(i)$ of a node is divided into in-degree $k^{in}(i) = |\mathcal{V}_{in}(i)|$ and out-degree $k^{out}(i) = |\mathcal{V}_{out}(i)|$: $k(i) = k^{in}(i) + k^{out}(i)$. The WWW has also dynamical features in that $\mathcal{G}$ is growing in time, with a continuous creation of new nodes and links. Empirical evidence shows that the distribution of the in-degrees of vertices follows a power-law behavior. Namely, the probability distribution that a node $i$ has in-degree $k^{in}$ behaves as $P(k^{in}) \sim \left(k^{in}\right)^{-\gamma_{in}^k}$, with $\gamma_{in}^k = 2.1 \pm 0.1$ as indicated by the largest data sample [1, 2, 4, 5]. The out-degrees ($k^{out}$) distribution of web-pages is also broad but with an exponential cut-off, as recent data suggest [2, 5]. While the in-degree represents the sum of all hyper-links coming from the whole WWW and can be in principle as large as the WWW itself, the out-degree is determined by the number of hyper-links present in a single web-page and is thus constrained by obvious physical elements.

### 2.2 Weights and Strengths

The number of users of any given web-site is also distributed according to a heavy-tail distribution [4]. This fact demonstrates the relevance of considering that every hyper-link has a specific weight that represents the number of users which are navigating on it. The WebGraph $\mathcal{G}(V, E)$ is thus a directed, weighted graph where the directed edges have assigned variables $w_{ij}$ which specify the weight on the edge connecting vertex $i$ to vertex $j$ ($w_{ij} = 0$ if there is no edge pointing from $i$ to $j$). The standard topological characterization of directed networks is obtained by the analysis of the probability distribution $P(k^{in})$ [$P(k^{out})$]

that a vertex has in-degree $k^{in}$ [out-degree $k^{out}$]. Similarly, a first characterization of weights is obtained by the distribution $P(w)$ that any given edge has weight $w$. Along with the degree of a node, a very significative measure of the network properties in terms of the actual weights is obtained by looking at the vertex incoming and outgoing strength defined as [30, 15]

$$s_i^{out} = \sum_{j \in \mathcal{V}_{out}(i)} w_{ij} \ , \ s_i^{in} = \sum_{j \in \mathcal{V}_{in}(i)} w_{ji} \ , \tag{1}$$

and the corresponding distributions $P(s^{in})$ and $P(s^{out})$. The strengths $s_i^{in}$ and $s_i^{out}$ of a node integrate the information about its connectivity and the importance of the weights of its links, and can be considered as the natural generalization of the degree. For the Web the incoming strength represents the actual total traffic arriving at web-page $i$ and is an obvious measure of the popularity and importance of each web-page. The incoming strength obviously increases with the vertex in-degree $k_i^{in}$ and usually displays the power-law behavior $s \sim k^{\beta}$, with the exponent $\beta$ depending on the specific network [15].

### 2.3 The Model

Our goal is to define a model of a growing graph that explicitly takes into account the actual popularity of web-pages as measured by the number of users visiting them. Starting from an initial seed of $N_0$ pages, a new node (web-page) $n$ is introduced in the system at each time-step and generates $m$ outgoing hyperlinks. In this study, we take $m$ fixed so that the out-degree distribution is a delta function. This choice is motivated by the empirical observation that the distribution of the number of outgoing links is bounded [5] and we have checked that the results do not depend on the precise form of this distribution as long as $P(k^{out}(i) = k)$ decays faster than any power-law as $k$ grows.

The new node $n$ is attached to a node $i$ with probability

$$Prob(n \to i) = \frac{s_i^{in}}{\sum_j s_j^{in}} \tag{2}$$

and the new link $n \to i$ has a weight $w_{ni} \equiv w_0$. This choice relaxes the usual degree preferential attachment and focuses on the popularity—or strength—driven attachment in which new web-pages will connect more likely to web-pages handling larger traffic. This appears to be a plausible mechanism in the WWW and in many other technological networks. For instance, in the Internet new routers connect to other routers with large bandwidth and traffic handling capabilities. In the airport network, new connections (airlines routes) are generally established with airports having a large passenger traffic [15, 31, 32]. The new vertex is assumed to have its own initial incoming strength $s_n^{in} = w_0$ in order to give the vertex an initial non-vanishing probability to be chosen by vertices arriving at later time steps.

The second and determining ingredient of the model consists in considering that a new connection $(n \to i)$ will introduce variations of the traffic across the

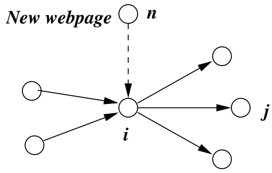


**Fig. 1.** Illustration of the construction rule. A new web-page $n$ enters the Web and direct a hyper-link to a node $i$ with probability proportional to $s_i^{in}/\sum_j s_j^{in}$. The weight of the new hyper-link is $w_0$ and the existing traffic on outgoing links of $i$ are modified by a total amount equal to $\delta_i$: $s_i^{out} \to s_i^{out} + \delta_i$

network. For the sake of simplicity we limit ourselves to the case where the introduction of a new incoming link on node $i$ will trigger only local rearrangements of weights on the existing links $(i \to j)$ where $j \in \mathcal{V}_{out}(i)$ as

$$w_{ij} \to w_{ij} + \Delta w_{ij}, \tag{3}$$

where $\Delta w_{ij}$ is a function of $w_{ij}$ and of the connectivities and strengths of $i$. In the following we focus on the case where the addition of a new edge with weight $w_0$ induces a total increase $\delta_i$ of the total outgoing traffic and where this perturbation is proportionally distributed among the edges according to their weights [see Fig. 1]

$$\Delta w_{ij} = \delta_i \frac{w_{ij}}{s_i^{out}} \; . \tag{4}$$

This process reflects the fact that new visitors of a web-page will usually use its hyper-links and thus increase its outgoing traffic. This in turn will increase the popularity of the web-pages pointed by the hyperlinks. In this way the popularity of each page increases not only because of direct link pointing to it but also due to the increased popularity of its in-neighbors. It is possible to consider heterogeneous $\delta_i$ distributions depending on the local dynamics and rearrangements specific to each vertex, but for the sake of simplicity we consider the model with $\delta_i = \delta$. We finally note that the quantity $w_0$ sets the scale of the weights. We can therefore use the rescaled quantities $w_{ij}/w_0$, $s_i/w_0$ and $\delta/w_0$, or equivalently set $w_0 = 1$. The model then depends only on the dimensionless parameter $\delta$. The generalization to arbitrary $w_0$ is simply obtained by replacing $\delta$, $w_{ij}$, $s_i^{out}$ and $s_i^{in}$ respectively by $\delta/w_0$, $w_{ij}/w_0$, $s_i^{out}/w_0$ and $s_i^{in}/w_0$ in all results.

## 2.4 Analytical Solution

Starting from an initial seed of $N_0$ nodes, the network grows with the addition of one node per unit time, until it reaches its final size $N$. In the model, every new vertex has exactly $m$ outgoing links with the same weight $w_0 = 1$. During the growth process this symmetry is conserved and at all times we have $s_i^{out} = m w_{ij}$. Indeed, each new incoming link generates a traffic reinforcement $\Delta w_{ij} = \delta/m$, so that $w_{ij} = w_0 + k_i^{in}\delta/m$ is independent from $j$ and

$$s_i^{out} = m + \delta k_i^{in} \ . \tag{5}$$

The time evolution of the *average* of $s_i^{in}(t)$ and $k_i^{in}(t)$ of the $i$-th vertex at time $t$ can be obtained by neglecting fluctuations and by relying on the continuous approximation that treats connectivities, strengths, and time $t$ as continuous variables [7, 9, 10]. The dynamical evolution of the in-strength of a node $i$ is given by the evolution equation

$$\frac{ds_i^{in}}{dt} = m\frac{s_i^{in}}{\sum_l s_l^{in}} + \sum_{j\in\mathcal{V}_{in}(i)} m\frac{s_j^{in}}{\sum_l s_l^{in}}\delta\frac{1}{m} \ , \tag{6}$$

with initial condition $s_i^{in}(t = i) = 1$. This equation states that the incoming strength of a vertex $i$ can only increase if a new hyper-link connects directly to $i$ (first term) or to a neighbor vertex $j \in \mathcal{V}_{in}(i)$, thus inducing a reinforcement $\delta/m$ on the existing in-link (second term). Both terms are weighted by the probability that the new vertex establishes a hyperlink with the corresponding existing vertex. Analogously, we can write the evolution equation for the in-degree $k_i^{in}$ that evolves only if the new link connects directly to $i$:

$$\frac{dk_i^{in}}{dt} = m\frac{s_i^{in}}{\sum_l s_l^{in}} \ . \tag{7}$$

Finally, the out-degree is constant ($k_i^{out} = m$) by construction.

The above equations can be written more explicitly by noting that the addition of each new vertex and its $m$ out-links, increase the total in-strength of the graph by the constant quantities $1 + m + m\delta$ yielding at large times $\sum_{l=1}^{t} s_l^{in} = m\left(1 + \frac{1}{m} + \delta\right)t$. By inserting this relation in the evolution equations (6) and (7) we obtain

$$\frac{ds_i^{in}}{dt} = \frac{1}{\delta + 1 + \frac{1}{m}}\left(\frac{s_i^{in}}{t} + \frac{\delta}{mt}\sum_{j\in\mathcal{V}_{in}(i)} s_j^{in}\right) \quad \text{and} \quad \frac{dk_i^{in}}{dt} = \frac{1}{\delta + 1 + \frac{1}{m}}\frac{s_i^{in}}{t} \ . \tag{8}$$

These equations cannot be explicitly solved because of the term $\sum_{j\in\mathcal{V}_{in}(i)} s_j^{in}$ which introduces a coupling of the in-strength of different vertices. The structure of the equations and previous studies of similar undirected models [31, 32] suggest to consider the *Ansatz* $s_i^{in} = Ak_i^{in}$ in order to obtain an explicit solution. Using (5), and $w_{ji} = s_j^{out}/m$, we can write

$$s_i^{in} = \sum_{j\in\mathcal{V}_{in}(i)} w_{ji} = k_i^{in} + \sum_{j\in\mathcal{V}_{in}(i)} \frac{\delta}{m}k_j^{in}, \tag{9}$$

and the Ansatz $s_i^{in} = Ak_i^{in}$ yields

$$\sum_{j\in\mathcal{V}_{in}(i)} s_j^{in} = \frac{m}{\delta}(A-1)s_i^{in} \ . \tag{10}$$

This allows to have a closed equation for $s_i^{in}$ whose solution is

$$s_i^{in}(t) = \left(\frac{t}{i}\right)^{\theta} \ , \text{ with } \ \theta = \frac{A}{\delta + 1 + 1/m} \tag{11}$$

and $k_i^{in}(t) = s_i^{in}(t)/A$, satisfying the proposed Ansatz. The fact that vertices are added at a constant rate implies that the probability distribution of $s_i^{in}$ is given by [10, 31, 32]

$$P(s^{in}, t) = \frac{1}{t + N_0} \int_0^t \delta(s^{in} - s_i^{in}(t)) di, \tag{12}$$

where $\delta(x)$ is the Dirac delta function. By solving the above integral and considering the infinite size limit $t \to \infty$ we obtain

$$P(s^{in}(i) = s) \sim s^{-\gamma^s_{in}} \ , \text{ with } \gamma^s_{in} = 1 + \frac{1}{\theta} \tag{13}$$

The quantities $s_i^{in}$, $k_i^{in}$ and $s_i^{out}$ are thus here proportional, so that their probability distributions are given by power-laws with the same exponent $\gamma^s_{in} = \gamma^s_{out} = \gamma^k_{in}$. The explicit value of the exponents depends on $\theta$ which itself is a function of the proportionality constant $A$. In order to find an explicit value of $A$ we use the approximation that on average the total in-weight will be proportional to the number of in-links times the average weight in the graph $< w >= \frac{1}{tm} \sum_l s_l^{out} = (\delta + 1)$. At this level of approximation, the exponent $\theta$ varies between $m/(m+1)$ and 1 and the power-law exponent thus varies between 2 ($\delta \to \infty$) and $2 + 1/m$ ($\delta = 0$). This result points out that the model predicts an exponent $\gamma^k_{in} \simeq 2$ for reasonable values of the out-degree, in agreement with the empirical findings.

## 3 Numerical Simulations

Along with the previous analytical discussion we have performed numerical simulations of the presented graph model in order to investigate its topological properties with a direct statistical analysis.

### 3.1 Degree and Strength Distributions

As a first test of the analytical framework we confirm numerically that $s^{in}$, $k^{in}$, $s^{out}$ are indeed proportional and grow as power-laws of time during the construction of the network [see Fig. 2]. The measure of the proportionality factor $A$ between $s^{in}$ and $k^{in}$ allows to compute the exponents $\theta$ and $\gamma$, which are satisfactorily close to the observed results and to the theoretical predictions obtained with the approximation $A \approx < w >$. Fig. 2 shows the probability distributions of $k^{in}$, $w$, $s^{in}$, and $s^{out}$ for $\delta = 0.5$. All these quantities are broadly distributed according to power-laws with the same exponent. It is also important to stress that the out-traffic is broadly distributed even if the out-degree is not.

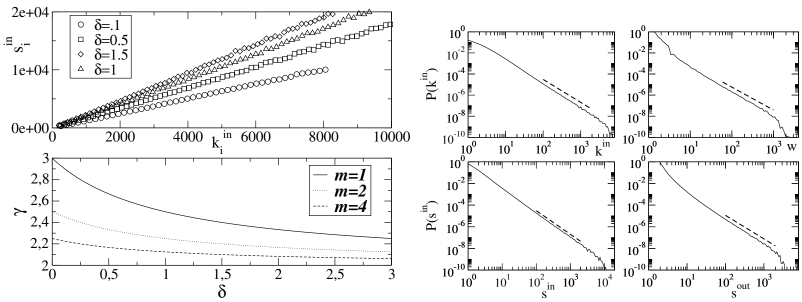


**Fig. 2.** Top left: illustration of the proportionality between $s^{in}$ and $k^{in}$ for various values of $\delta$. Bottom left: theoretical approximate estimate of the exponent $\gamma_{in}^{s} = \gamma_{out}^{s} = \gamma_{in}^{k}$ vs. $\delta$ for various values of $m$. Right: Probability distributions of $k^{in}$, $w$, $s^{in}$, $s^{out}$ for $\delta = 0.5$, $m = 2$ and $N = 10^5$. The dashed lines correspond to a power law with exponent $\gamma = 2.17$ obtained by measuring first the slope $A$ of $s^{in}$ vs. $k^{in}$ and then using (11) and (13) to compute $\gamma$

### 3.2 Clustering and Hierarchies

Along with the vertices hierarchy imposed by the strength distributions the WWW displays also a non-trivial architecture which reflects the existence of well defined groups or communities and of other administrative and social factors. In order to uncover these structures a first characterization can be done at the level of the undirected graph representation. In this graph, the degree of a node is the sum of its in- and out-degree ($k_i = k_i^{in} + k_i^{out}$) and the total strength is the sum of its in- and out-strength ($s_i = s_i^{in} + s_i^{out}$). A very useful quantity is then the clustering coefficient $c_i$ that measures the local group cohesiveness and is defined for any vertex $i$ as the fraction of connected neighbors couples of $i$ [6]. The average clustering coefficient $C = N^{-1} \sum_i c_i$ thus expresses the statistical level of cohesiveness by measuring the average density of interconnected vertex triplets in the network. Further information can be gathered by inspecting the average clustering coefficient $C(k)$ restricted to vertices with degree $k$ [33, 35]

$$C(k) = \frac{1}{N_k} \sum_{i/k_i=k} c_i \;, \tag{14}$$

where $N_k$ is the number of vertices with degree $k$. In real WWW data, it has been observed that the $k$ spectrum of the clustering coefficient has a highly non-trivial behavior with a power-law decay as a function of $k$, signaling a hierarchy in which low degree vertices belong generally to well interconnected communities (high clustering coefficient) while hubs connect many vertices that are not directly connected (small clustering coefficient) [34, 35].

We show in Fig. 3 how the clustering coefficient $C(k)$ of the model we propose increases with $\delta$. We obtain for $C(k)$ a decreasing function of the degree $k$, in agreement with real data observation. In addition, the range of variations spans

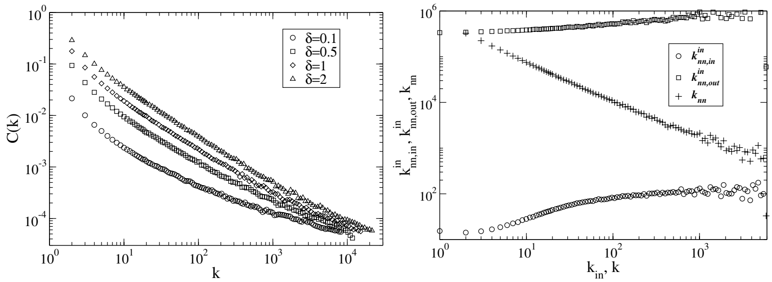


**Fig. 3.** Left: Clustering coefficient $C(k)$, for various values of the parameter $\delta$. Here $m = 2$ and $N = 10^5$. The clustering increases with $\delta$. Right: Correlations between degrees of neighbouring vertices as measured by $k_{nn}(k)$ (crosses), $k^{in}_{nn,in}(k^{in})$ (circles) and $k^{in}_{nn,out}(k^{in})$ (squares); $m = 2$, $\delta = 0.5$ and $N = 10^5$

several orders of magnitude indicating a continuum hierarchy of cohesiveness levels as in the analysis of Ref. [35].

Interestingly, the clustering spectrum can be qualitatively understood by considering the dynamical process leading to the formation of the network. Indeed, vertices with large connectivities and strengths are the ones that entered the system at the early times, as shown by (11). This process naturally builds up a hierarchy among the nodes, the "older" vertices having larger connectivities. Newly arriving vertices attach to pre-existing vertices with large strength which on their turn are reinforced by the rearrangement of the weights, as well as their neighbours. The increase of $C$ with $\delta$ is directly related to this mechanism: each time the extremity $i$ of an edge is chosen by preferential attachment, $i$ and its $m$ out-neighbours are reinforced, thus increasing the probability that, at a following step, a new node connects to both $i$ *and* one of its $m$ neighbours, forming a trangle. Triangles will therefore typically be made of two "old" nodes and a "young" one. This explains why $C(k)$ is large for smaller degree and also why $C(k)$ increases faster for smaller $k$ when $\delta$ increases. In contrast, increasing $\delta$ does not affect much the "older" nodes which implies that for large degrees, the clustering coefficient is not significantly affected by variation of $\delta$. These properties can be simply reconciled with a real dynamical feature of the web. Newly arriving web-pages will likely point to well known pages that, on their turn, are mutually pointing each other with high probability, thus generating a large clustering coefficient. On the contrary, well known and long established web-pages are pointed by many less popular and new web-pages with no hyperlinks among them. This finally results in a small clustering coefficient for well known high degree pages. The structure of the clustering coefficient is therefore the mathematical expression of the structure imposed to the web by its community structure that generally forms cliques of "fans" web-pages pointing to sets of interconnected "authority" web-pages.

Another important source of information about the network structural organization lies in the correlations of the connectivities of neighboring vertices [36].

Correlations can be probed by inspecting the average degree of nearest neighbor of a vertex $i$

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in \mathcal{V}(i)} k_j \ , \tag{15}$$

where the sum runs on the nearest neighbors vertices of each vertex $i$. From this quantity a convenient measure to investigate the behavior of the degree correlation function is obtained by the average degree of the nearest neighbors, $k_{nn}(k)$, for vertices of degree $k$ [33]

$$k_{nn}(k) = \frac{1}{N_k} \sum_{i/k_i=k} k_{nn,i}. \tag{16}$$

This last quantity is related to the correlations between the degree of connected vertices since on the average it can be expressed as $k_{nn}(k) = \sum_{k'} k' P(k'|k)$, where $P(k'|k)$ is the conditional probability that, given a vertex with degree $k$, it is connected to a vertex with degree $k'$. If degrees of neighboring vertices are uncorrelated, $P(k'|k)$ is only a function of $k'$, i.e. each link points to a vertex of a given degree with the same probability independently on the degree of the emanating vertex, and thus $k_{nn}(k)$ is a constant. When correlations are present, two main classes of possible correlations have been identified: *Assortative* behavior if $k_{nn}(k)$ increases with $k$, which indicates that large degree vertices are preferentially connected with other large degree vertices, and *disassortative* if $k_{nn}(k)$ decreases with $k$ [37].

In the case of the WWW, however, the study of additional correlation function is naturally introduced by the directed nature of the graph. We focus on the most significative, the in-degree of vertices that in our model is a first measure of their popularity. As for the undirected correlation, we can study the average in-degree of in-neighbours :

$$k^{in}_{nn,in}(i) = \frac{1}{k^{in}(i)} \sum_{j \in \mathcal{V}_{in}(i)} k^{in}(j) \ . \tag{17}$$

This quantity measures the average in-degree of the in-neighbours of $i$, i.e. if the pages pointing to a given page $i$ are popular on their turn. Moreover, relevant information comes also from

$$k^{in}_{nn,out}(i) = \frac{1}{k^{out}(i)} \sum_{j \in \mathcal{V}_{out}(i)} k^{in}(j) \ , \tag{18}$$

which measures the average in-degree of the out-neighbours of $i$, i.e. the popularity of the pages to which page $i$ is pointing. Finally, in both cases it is possible to look at the average of this quantity for group of vertices with in-degree $k^{in}_i$ in order to study the eventual assortative or disassortative behavior.

In Fig. 3 we report the spectrum of $k_{nn}(k)$, $k^{in}_{nn,in}(k^{in})$ and $k^{in}_{nn,out}(k^{in})$ in graphs generated with the present weighted model. The undirected correlations display a strong disassortative behaviour with $k_{nn}$ decreasing as a power-law.

This is a common feature of most technological networks which present a hierarchical structure in which small vertices connect to hubs. The model defined here exhibits spontaneously the hierarchical construction that is observed in real technological networks and the WWW. In contrast, both $k^{in}_{nn,in}(k^{in})$ and $k^{in}_{nn,out}(k^{in})$ show a rather flat behavior signaling an absence of strong correlations. This indicates a lack of correlations in the popularity, as measured by the in-degree. The absence of correlations in the behaviour of $k^{in}_{nn,out}(k^{in})$ is a realistic feature since in the real WWW, vertices tend to point to popular vertices independently of their in-degree. We also note that $k^{in}_{nn,out}(k^{in}) \gg k^{in}_{nn,in}(k^{in})$, a signature of the fact that the average in-degree of pointed vertices is much higher than the average in-degree of pointing vertices. This result also is a reasonable feature of the real WWW since the average popularity of webpages to which any vertex is pointing is on average larger than the popularity of pointing webpages that include also the non-popular ones.

Finally, we would like to stress that in our model the degree correlations are to a certain extent a measure of popularity correlations and more refined measurements will be provided by the correlations among the actual popularity as measured by the in-strength of vertices. We defer the detailed analysis of these properties to a future publication, but at this stage, it is clear that an empirical analysis of the hyperlinks traffic is strongly needed in order to discuss in detail the WWW architecture.

## 4 Conclusion

We have presented a model for the WWW that considers the interplay between the topology and the traffic dynamical evolution when new web-pages and hyperlinks are created. This simple mechanism produces a non trivial complex and scale-free behavior depending on the physical parameter $\delta$ that controls the local microscopic dynamics. We believe that the present model might provide a general starting point for the realistic modeling of the Web by taking into account the coupling of its two main complex features, its topology and its traffic.

*Acknowledgments.* A.B and A. V. are partially funded by the European Commission - Fet Open project COSIN IST-2001-33555 and contract 001907 (DELIS).

## References

1. A.-L. Barabasi and R. Albert, Science **286**, 509 (1999)
2. A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, S. Stata, A. Tomkins and J. Wiener, *Graph structure in the Web*, in *Proceedings of the 9th WWW conference* (2000).
3. R. Kumar, et al, *Stochastic models for the Web graph*, in 'Proceedings of the 41th IEEE Symposium on Foundations of Computer Science (FOCS)', pp. 57-65 (2000).

4. L.A. Adamic and B.A. Huberman, in 'Communications of the ACM', **44**, ACM Press New York, NY, USA, pp. 55-60 (2001).
5. L. Laura, S. Leonardi, S. Millozzi, U. Meyer, and J.F. Sibeyn, 'Algorithms and experiments for the Webgraph'. European Symposium on Algorithms (2003).
6. Watts D. J. and Strogatz S. H., Nature **393**, 440 (1998).
7. R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2000).
8. L.A.N. Amaral, A. Scala, M. Barthélemy, and H.E. Stanley, Proc. Natl. Acad. Sci. *USA* **97**, 11149 (2000).
9. S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
10. R. Pastor-Satorras and A. Vespignani, *Evolution and structure of the Internet: A statistical physics approach* (Cambridge University Press, Cambridge, 2004).
11. M. Granovetter, American Journal of Sociology, **78** (6) 1360-1380 (1973).
12. M. E. J. Newman, Phys. Rev. E **65**, 016131 (2001); M. E. J. Newman, Phys. Rev. E **65**, 016132 (2001).
13. A.-L. Barabasi, H. Jeong, R. Ravasz, Z. Neda, T. Vicsek, and A. Schubert, Physica A **311**, 590 (2002).
14. R. Guimera, S. Mossa, A. Turtschi, and L.A.N. Amaral, submitted (2003).
15. A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. (USA), **101**, 3747 (2004).
16. C. Quince, P.G. Higgs, and A.J. McKane, arXiv:q-bio.PE/0402014 (2004).
17. D. Garlaschelli et al, submitted (2003). Cond-mat-0310503.
18. B.A. Huberman and R. Lukose, Science **277**, 535 (1997).
19. B.A. Huberman, P. Pirolli, J. Pitkow, and R. Lukose, Science **280**, 95 (1998).
20. A.-L. Barabasi, R. Albert, and H. Jeong, Physica A **281**, 69 (2000)
21. B. Tadic, Physica A **293**, 273 (2001).
22. P.L. Krapivsky, G. Rodgers, and S. Redner, Phys. Rev. Lett. **86**, 5401 (2001).
23. S.N. Dorogovtsev and J.F.F. Mendes, Europhys. Lett. **52**, 33 (2000).
24. C. Cooper and A. Frieze, *A General Model of Undirected Web Graphs*, Proceedings of the 9th Annual European Symposium on Algorithms, Lecture Notes in Computer Science 2161, p. 500-511 (2001).
25. G. Bianconi and A.-L. Barabasi, Europhys. Lett. **54**, 436 (2001).
26. S. Mossa, M. Barthélemy, H.E. Stanley, and L.A.N. Amaral, Phys. Rev. Lett. **88**, 138701 (2002).
27. G. Pandurangan, P. Raghavan and E. Upfal, *Using PageRank to Characterize Web Structure*, In 'Proceedings of the 8th Annual International Computing and Combinatorics Conference (COCOON)', LNCS 2387, p. 330-339 (2002).
28. L. Laura, S. Leonardi, G. Caldarelli and P. De Los Rios, *A multi-layer model for the webgraph*, in 'On-line proceedings of the 2nd international workshop on web dynamics' (2002).
29. F. Menczer, Proc. Natl. Acad. Sci. *USA* **99**, 14014 (2002).
30. S.H. Yook, H. Jeong, A.-L. Barabasi, and Y. Tu, Phys. Rev. Lett. **86**, 5835 (2001).
31. A. Barrat, M. Barthélemy, and A. Vespignani, Phys. Rev. Lett. **92**, 228701 (2004).
32. A. Barrat, M. Barthélemy, and A. Vespignani, cond-mat/0406238.
33. A. Vázquez, R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* **65**, 066130 (2002).
34. J.-P. Eckmann and E. Moses, Proc. Natl. Acad. Sci. (USA), **99**, 5825 (2002).
35. Ravasz, E. & Barabási, A.-L. *Phys. Rev. E* **67**, 026112 (2003).
36. R. Pastor-Satorras, A. Vázquez and A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001).
37. M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).