A Qualitative Comparison of Power Law Generators

Javier Martin Hernandez, Tom Kleiberg, Huijuan Wang and Piet Van Mieghem

***Abstract*— Several topology generator algorithms have been proposed to match the heavy-tailed behavior of the real Internet AS-level Topology. Although evaluation studies have been presented for each topology generator, each of these studies assumes a different framework. Therefore, it is difficult to compare the differences between the topology generator algorithms. The objective of this paper is to fill this gap by comparing representative families of topology generators. The resultant graphs are assessed based on metrics such as the degree distribution and the spectrum of the adjacency matrix.**

I. Introduction

One of the first widely used Internet topology generators was developed by Waxman [31]. The Waxman algorithm is a variant of the Erdös-Rényi random graph [18] based on the Euclidean distance. Later research claimed that the real Internet topology does not obey a random-structure, but instead possesses some kind of hierarchy, for instance a differentiation between transit and stub nodes. As a consequence, the structural generators as Transit-Stub [32], Tiers [33], and GT-ITM[1] appeared. These structural generators were considered *valid* until the appearance of a seminal paper by Faloutsos *et al.* [28] in 1999. In that paper, the nodal degree of the Internet AS-level topology was shown to closely obey a power law. The graphs generated by the structural generators do not exhibit this power law behavior, turning them into deficient Internet topology models [22]. The work by Faloutsos *et al.* [28] fueled the development of a new family of generators, such as the Barabási-Albert (BA [3]) and Power Law Random Graph (PLRG [1]). Here, we focus on a comparison between this last generation of power law algorithms.

The Internet topology at the granularity of Autonomous Systems (AS) is a heavy-tailed scale-free small-world network [21] [25] [28]. However, the question of how closely the Internet topology follows power laws remains open to debate [24] [26].

Power-Laws and scale-freeness are two important concepts to clarify. A random variable $X > 0$ is said to possess a power law if the probability density function (*pdf*) obeys

$$f_X(x) = cx^{-\beta} \ (x > 0, \beta > 1) \tag{1}$$

where c is a normalization constant, and β is the power law exponent. The corresponding power law density function is scale-free because it satisfies

$$f_X(ax) = g(a) f_X(x) \tag{2}$$

An increase by a factor a in the scale or units by which one measures x does not change the overall density $f_X(x)$, except for a multiplicative scaling factor. Scale-free can be interpreted as a synonym for *power law*, but most actual usages of "scale-free" appear to have a richer notion. Sometimes additional features are considered such as underlying self-similarity or fractal geometry. In the remainder of this paper, we will use the notions of scaling and power law distribution interchangeably and only insist that the right tail of the distribution satisfies property (1).

A large number of algorithms have appeared trying to emulate the Internet AS-level graph. All the algorithms have an evocative approach to the problem [27] in common: the algorithms can reproduce a metric of interest, but do not capture the underlying causal mechanisms. Usually a well-understood network metric (in most cases the nodal degree distribution) is chosen and an algorithm that matches the metric is developed. This approach presents several problems. First, it is hard to choose the single metric. Second, a method that matches the chosen metric, often does not fit other metrics of interest.

In heavy-tail distributions such as the Lognormal, Weibull and Pareto, the (right) tail decreases subexponentially. Thus, the deviation from the mean can vary by orders of magnitude, turning the mean into an uninformative and not representative measure. The higher moments $E[X^k]$ of the distribution function (1) only exist if $\beta > k$ [19].

This paper is organized as follows. In Section II, we present the power law algorithms: PLRG, Havel-Hakimi (in short Havel), Takao, Barabási-Albert, Barabási-Albert rewired (BA-r) and Generalized Linear Preference (GLP). In Section III, we introduce the metrics used to compare the output graphs: degree distribution, average hopcount, clustering coefficient, assortativity and spectrum of the adjacency matrix. Further in Section IV, we present the qualitative analysis of the different algorithms under study. Finally, we conclude in Section V.

II. Topology generators

We can classify the topology generators into two families: the curve fitting family and the preferential attachment family. The **curve fitting family** generators use an explicit scale-free degree distribution $\mathcal{D} = \{d_1, d_2, ..., d_N\}$ (the *curve*). Given $\mathcal{D}$, the algorithms interconnect the set of $\mathcal{N}$ nodes such that the resultant graph $G(N, L)$ with N nodes and L links has degree distribution D. Generators of this family are PLRG, Havel-Hakimi and Takao. The **preferential attachment family** combines the ideas of network growth and preferential attachment. The graph starts with a small fully connected graph, and divides its growth into timesteps. Every timestep adds one node and m edges. For the added edges, the probability to attach to an existing node is proportional to the degree of the latter. Generators of this family we consider are

Delft University of Technology, faculty of Electrical Engineering, Mathematics and Computer Science, P.O Box 5031, 2600 GA Delft, The Netherlands.

[1]http://www3.cc.gatech.edu/projects/gtitm/

Barabási-Albert, Barabási-Albert rewire and GLP. Finally, we also consider hybrid generators as inet3 that use both curve fitting and preferential attachment.

A. Degree distribution generation

The performance of the PLRG, Havel-Hakimi and Takao algorithms depends highly on the provided input degree sequence $\mathcal{D} = \{d_1, d_2, ..., d_N\}$. Each nodal degree d_j is a random variable with power law distribution (1). All the degrees are assumed to be independent, thus ignoring the basic law [19] of the degree $\sum d_j = 2L$ that eventually correlates all nodal degrees. Since the process followed to generate degree sequences is identical for PLRG, Havel and Takao, their probability density functions present the same behavior (while others parameters may differ).

B. PLRG

PLRG stands for Power Law Random Graph [1]. The algorithm first assigns the calculated degree sequence $\mathcal{D}$ to the N nodes in the graph. It then randomly matches degrees among all the nodes. The produced graphs may not be connected and may contain self loops and duplicated links. In our simulations, we delete self-loops, merge duplicated links, and extract the giant component. The connected components theory [14][16][20] states that for values of β between $2 < \beta < \beta_0$ ($\beta_0 = 3.478$ as derived in [1]) the random graph has a.s. a giant component, and the size of the second largest component is $O(\log N)$. For $1 < \beta < 2$ the second largest component is a.s. of size $O(1)$.

We use the notation from Cormen *et al.* [7] to display the algorithm.

```
PLRG(D)
1  vector V ← 0
2  for j ← 1 to length[D]
3    do append D[j] copies of j to V
4  randomize the position of elements in V
5  for j ← 1 to length[V]/2
6    do connect V[j] to V[N-j+1]
```

Our topologies have a size of around 3000 nodes, meaning that the second largest component size is ≤ 3 . Empirical results [10] show that 20% of the total number of nodes do not belong to the giant component.

C. Havel-Hakimi

Similar to PLRG, Havel-Hakimi requires as input a given degree sequence. An important feature of Havel is that its behavior is deterministic: given a degree sequence $\mathcal{D} = \{d_1, d_2, ..., d_N\}$, the resultant graph G is always the same. The original Havel algorithm was designed to check whether a degree sequence is graphical[2] or not. By reversing the original Havel algorithm, we have a deterministic and efficient algorithm to generate a connected graph. Havel algorithm is based on the following theorem.

[2]A degree sequence $\mathcal{D}$ is called graphical if is possible to draw at least one graph with degree sequence $\mathcal{D}$.

Theorem 1: (Havel-Hakimi theorem) Let $\mathcal{D} = \{d_1, d_2, ..., d_N\}$ be a sequence of non-negative integers with $d_1 \geq d_2 \geq ... \geq d_N$. Let $\mathcal{D}'$ be a sequence $\{d'_1, d'_2, ..., d'_N\}$ obtained from D by setting $d'_1 = 0$, $d'_i = d_i - 1$ $(i = 2, .., d_1 + 1)$, and $d'_j = d_j (j = d_1 + 2, ..., n)$. Then $\mathcal{D}$ is graphical if $\mathcal{D}'$ is graphical.

In words, $\mathcal{D}$ is graphical if the following sequence is graphical: replace d_1 by 0 and substract 1 from the next d_1 degrees:$\{d_2, d_3, ..., d_{d_1+1}\}$. The recursion ends when the sequence consists of all zeros, or when the sequence has only one non-zero element. The sequence is then classified as graphical and non-graphical, respectively. The meta code implementation is

```
HAVEL_HAKIMI_THEOREM(D)
1  if NON_ZERO_ELEMENTS[D] = 1
2  then sequence is not graphical
3         terminate
4  else
5    i ← FIRST_NON_ZERO(D)
6    for j ← 1 to D[i]
7    do D[i]  ← 0
8       D[i+j] ← D[i+j] - 1
9       sort D from D[i+1] to length[D]
```

When a non-graphical sequence is detected the degree sequence is ignored (*terminate*), and a new one is requested. The Havel algorithm meta code is as follows.

```
HAVEL_HAKIMI(D)
1  matrix M ← 0
2  while D different from 0
3  do append D as row to M
4     HAVEL_HAKIMI_THEOREM(D)
5  for j ← rows[M] to 2
6  do k ← FIRST_NON_ZERO(M[j])
7     for i ← length[D] to k
8     do if M[j-1][i] > M[j][i]
9           then interconnect k-1 to i
```

The function FIRST_NON_ZERO($\mathcal{D}$) returns the position of the first element different from zero in $\mathcal{D}$.

The algorithm checks whether $\mathcal{D}$ is graphical or not in the fourth first lines, saving the generated degree sequences in the matrix M. In the last four lines the connectivity information is obtained, through comparing adjacent rows of M.

D. Takao

The Takao algorithm was introduced by Takao Asano in [23]. Similarly to PLRG and Havel, the Takao algorithm requires an explicit degree sequence as input. Takao is a deterministic algorithm, such that, for degree sequence $\mathcal{D}$ it will always produce the same output graph.

The main feature of the Takao and Havel algorithms is their speed, in the order of $O(\sum_{i=1}^{N} d_i/2)$ provided a graphical $\mathcal{D}$. The Takao algorithm first checks whether the given degree sequence is graphical by applying Theorem 2. Similarly to Havel, the connectivity information is obtained from the processed matrix M', but the connectivity processes are completely different. Takao theorem is defined as follows

Theorem 2: (Takao theorem) Given a degree sequence $\mathcal{D} = \{d_1, d_2, ...d_N\}$ of positive integers where $n > d_1 \geq d_2 \geq ... \geq d_n > 0$. Let $h = d_n$, $x = min\{j|d_j = d_h\}, y = max\{j|j \leq n-1, d_j = d_h\}$. Describe $\mathcal{C} = \{c_1, c_2, ...d_{N-1}\}$ as a sequence of positive integers where $c_1 \geq c_2 \geq ... \geq c_{n-1}$, andv

$$c_i = \begin{cases} d_i - 1, & \text{if } 1 \leq i \leq x-1 \text{ or } y-h+x \leq i \leq y \\ d_i, & \text{if } x \leq i \leq y \text{ or } y+1 \leq i \leq n-1 \end{cases}$$

$\mathcal{D}$ is graphical if C is graphical

In words, at each step we are removing from $\mathcal{D}$ the node with the lowest degree h, and substracting d_h units from the set of highest degrees.

E. Barabási-Albert

The two main concepts behind BA algorithm are the preferential attachment and Yule's process. The combination of a growth process with preferential attachment has been proved theoretically [5] and empirically [3] to lead to scale invariant distributions.

The algorithm starts creating a small fully connected core with m_0 nodes. Then, BA incrementally constructs a topology by continuously adding nodes. At each time step one node is added with m links. After t time steps, the model leads to a random network with $m_0 + t$ nodes and mt links. The probability that a new node will be connected to the node i is given by

$$\Pr[X = i] = \frac{d_i(t)}{\sum_{\forall j} d_j(t)} \tag{3}$$

The linear preferential attachment via (3) represents the *rich-get-richer* idea. The probability of a new node attaching to a node with high degree is high, while the probability of a new node attaching to a node with low degree is low. The BA algorithm is as follows.

```
BARABASI_ALBERT(N,m0,m)
1  CREATE_CORE(m0)
2  while topology has < N nodes
3  do for m times
4       do create a new node x
5            y ← P_ATTACH()
6            connect x to y
```

where the function P_ATTACH() returns a random node selected through Eq. (3). The probability that a node i has d_i edges, follows a power law with tail exponent $\beta = 3$. BA predicts correctly the emergence of power laws. However, for many real systems such as the Internet, β is somewhere between 2 and 3. The static $\beta = 3$ provided by BA algorithm is therefore not satisfactory.

The rewiring variant Barabási-Albert rewire (BA-r) [4] separates the growing process into three main events: addition of node, addition of m links, or rewire of m links. Each of these operations is randomly chosen with fixed probabilities p and q, where $p + q < 1$. BA-r algorithm is defined as follows.

```
BARABASI_ALBERT_REWIRE(N,m0,m,p,q)
1  CREATE_CORE(m0)
2  while topology has < N nodes
3  do if p
4       then for m times
5            do x ← P_ATTACH()
6               y ← P_ATTACH()
7               add a new edge from x to y
8       if q
9     then for m times
10           do select a random edge e
11              select a random end of e
12              new_y ← P_ATTACH()
13              rewire the selected end to new_y
14    else
15      create a new source node: x
16      y ← P_ATTACH()
17      add a new edge from x to y
```

F. GLP

The Generalized Linear Preference (GLP) is a variant of the BA model. GLP adds more flexibility than BA in specifying how nodes connect to other nodes [10]. It has been reported [24] that in the real Internet, new ASs have a much stronger preference to connect to high degree ASs than predicted by the linear preferential model . To achieve a stronger preference for high degre nodes, the probability that a new node will be connected to the node i is adapted as

$$\Pr[X = i] = \frac{d_i - \theta}{\sum_{\forall j}(d_j - \theta)} \tag{4}$$

where $\theta \in (-\infty, 1)$ is a tunable parameter. When θ is smaller, less preference is given to high degree nodes. The GLP algorithm is a mix between BA and BA-r. GLP decides at each time step between two operations with probability p: addition of a node, or addition of m links. The algorithm is defined as follows.

```
GLP(N,m0,m,theta)
1  CREATE_CORE(m0)
2  nodes ← m0
3  while topology has < N nodes
5  do if p
6       then for(m times)
7            do src ← P_ATTACH_GLP(theta)
8               dst ← P_ATTACH_GLP(theta)
9               ADD_EDGE(src,dst)
10    else
11      create a new source node: x
12      y ← P_ATTACH_GLP(theta)
13      add a new edge from x to y
```

The function P_ATTACH_GLP() returns a random node selected via (4).

GLP is shown to yield graphs with power law degree distributions. In addition, Bu *et al.* [10] demonstrate how to choose parameters so as to produce a desired power law exponent. The implementation of BA, BA-r and GLP algorithms have been retrieved from the BRITE project[3].

[3]http://www.cs.bu.edu/brite/

G. Inet3

Inet3 [11] is based on empirical data extracted from the Oregon Route-Views project[4]. Thus the algorithm is based on the BGP Autonomous System (AS) topology.

Inet3 follows the next sequential steps to generate the final topology: first compute the number of months t that would take the 1997's Internet to reach N nodes (exponential growth of the number of nodes is assumed). Second, compute the new frequency (and rank) distributions using the calculated t. The degree distribution is calculated through the pdf $f_X(x) = e^{at+b}x^{-S}$ where a and b are known constants (extracted from Oregon Route-Views). For the 2% higher degree nodes, apply instead the ccdf formula $F_X(x) = e^{ct+d}x^{-R}$. Third, assign degree 1 to the $m\%$ of nodes. Fourth, form a spanning tree with nodes of degree higher than 1, creating G. Fifth, attach nodes with degree 1 to G using linear preferential attachment (3). Finally match the remaining nodes with G using linear preferential attachment.

Inet3 depends on the empirical constants a, b, c, d, S, R and m. The values for the constants have been extracted from November 1997's Internet snapshot[5].

III. Topological characteristics

In this Section, we introduce the metrics used to analyze and compare the different topology generators.

A. Degree Distribution

Applying logarithms to the both sides of equation (1) we obtain the equation of a line:

$$\log(y) = \log c - \beta \log x \tag{5}$$

The simplest way to obtain β is performing a linear regression of $f_X(x)$ when plotted in a log-log scale. To illustrate Eq. (5), we plot in Figure 1 the degree distribution of PLRG in log-log scale. The data set has been obtained from simulations.

The frequency distribution plot systematically underestimates β [27]. Even worse, the plotted pdfs have a tendency to suggest falsely that a scaling behavior exists. Figure 2, and Figures 1a-1d on [27] demonstrate how an exponential distribution can be wrongly interpreted as *scale-free*.

Figure 2 illustrates simulation results for BA-r. The frequency degree data (pdf) appears linear on log-log scale, which leads us to think *a priori* that the degree distribution behaves like a power law. But if we plot the same data as a ccdf, we can clearly observe how the power law relation fails for high degrees. In conclusion, the rank degree plots are more reliable than the frequency degree plots. Hereforth, we only use rank degree plots.

B. Hopcount

The hopcount or path length between two nodes is a characterizing property of a graph. It is defined as the minimal number of distinct links that forms a path connecting the two

[4]http://www.routeviews.org

[5]So far, the last version of Inet generator is *Inet-3.0*.

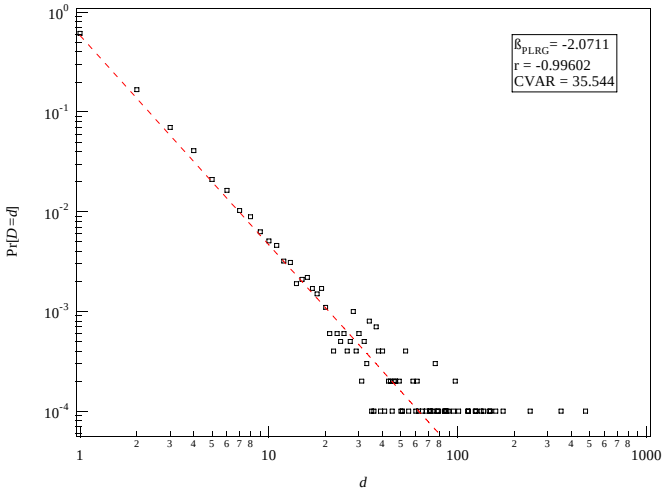

Fig. 1. pdf of the PLRG degree distribution. $N = 3050$ nodes. The right-hand of the distribution is noisy because of sampling errors. This can be solved increasing the number of simulations.

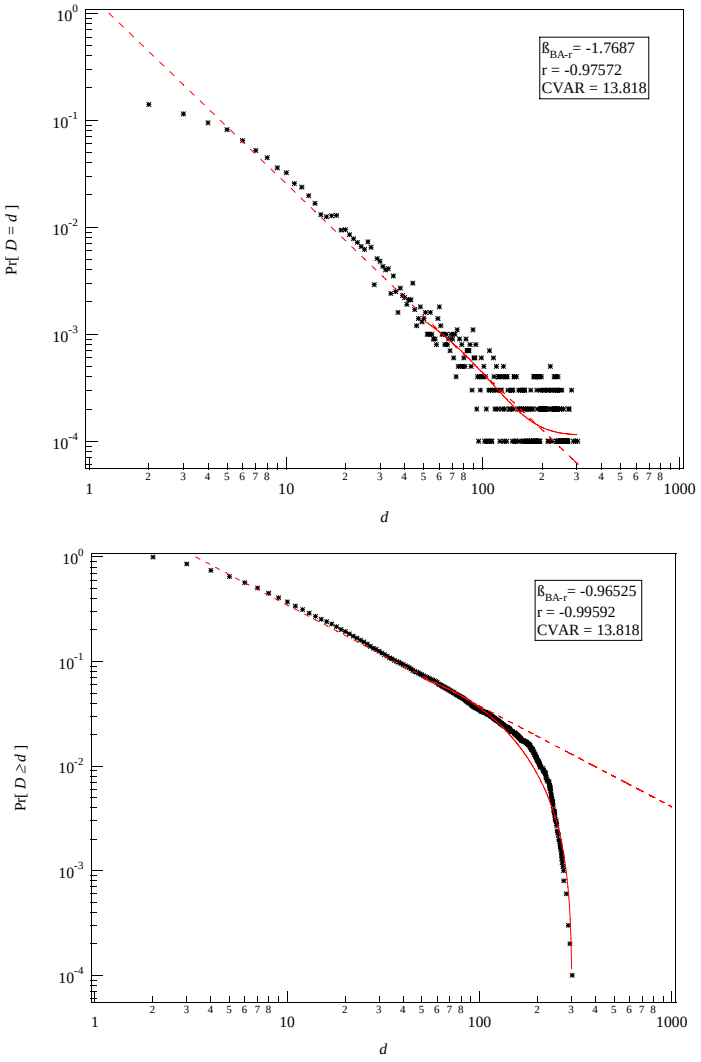

Fig. 2. pdf of BA-r in log-log scale (up) and ccdf of the same distribution in log-log scale (bottom). The pdf apparently behaves like a power law. The ccdf clearly demonstrates that the sequence is not obeying (1), at least for high degrees. Thus the degree sequence is not scale-free.

nodes. The mean hopcount is the mean of the shortest path lengths connecting each node $n \in \mathcal{N}$ to all other nodes in G.

Ordinarily, a topology is assumed to be either completely regular or completely random. But many biological, technological and social networks lie somewhere between these two extremes. Systems can be highly clustered, like regular lattices, yet have small average hopcount [30], like random graphs. These are called *small-world* networks, by analogy with the small-world phenomenon [13]. The average hopcount distribution is an important tool to distinguish between *small world* and large diameter graphs, like lattices [19].

C. Assortativity

The assortativity coefficient $R(G)$ is introduced by Newman in [15], is

$$R(G) = \frac{\sum_{(i,j)\in\mathcal{L}} d_i d_j - \left(\sum_{i\in\mathcal{N}} \frac{1}{2} d_i^2\right)^2 / L}{\sum_{i\in\mathcal{N}} \frac{1}{3} d_i^3 - \left(\sum_{i\in\mathcal{N}} \frac{1}{2} d_i^2\right)^2 / L} \tag{6}$$

Assortative mixing ($R > 0$) is defined as "a preference for high-degree nodes to attach to other high-degree nodes" and disassortative mixing ($R < 0$) as the converse, where "high-degree nodes attach to low-degree ones". Assortative and disassortative mixing patterns indicate a generic tendency to connect to similar or dissimilar peers respectively.

If we plot the *average degree of the neighbors of a given node* i versus the node degree d_i we obtain a graphical representation of the assortativity coefficient. The slope of the linear fit is directly related to $R(G)$ (see Section IV-C for examples).

D. Clustering Coefficient

The clustering coefficient $c_G(i)$ characterizes the density of connections in the environment of a node i. The clustering coefficient is defined as the ratio of the number of links y connecting the d_i neighbors of i over the maximum possible $\frac{1}{2} d_i(d_i - 1)$,

$$c_G(i) = \frac{2y}{d_i(d_i - 1)} \text{ provided that } d_i \geq 2 \tag{7}$$

The clustering coefficient for the whole graph C_G is defined as the average of the clustering coefficient for all the nodes in the graph.

C_G expresses local robustness in the graph and thus has practical implications: the higher the local clustering of a node, the more interconnected are its neighbors. In addition, the small world graph shows high clustering coefficients in comparison to random networks, which represents an elevated *cliquishness* [30].

E. Spectrum

The adjacency matrix A of a graph G with N nodes is an $N \times N$ matrix with elements $a_{ij} = 1$ only if (i, j) is a link of G, otherwise $a_{ij} = 0$. We assume bidirectional links, if there is a link from i to j ($a_{ij} = 1$) then there is a link from j to i ($a_{ji} = 1$) for any $j \neq i$. Moreover, we exclude self-loops ($a_{jj} = 0$) or multiple links between two nodes i and j. The spectrum of G is the set of eigenvalues of its adjacency matrix A.

The spectrum is an important *global* characteristic of a topology. It yields tight bounds for a wide range of graph characteristics, such as distance-related parameters, expansion properties, and values related to separator problems estimating graph resilience under node/link removal. Most networks with high values for the largest eigenvalues have a small diameter, expand faster, and are more robust, which makes these sets of eigenvalues important.

Faloutsos *et al.* [28] showed that the higher eigenvalues of the Internet AS-level correspondent adjacency matrix follow an empirical power law, such that:

$$f_\lambda(x) \simeq x^{-\delta} \tag{8}$$

Separately, Dogorostev *et al.* [8] showed that the tail parameter β of the degree distribution, and the tail parameter δ for the highest eigenvalues are directly related through

$$f_\lambda(x) \simeq x^{1-2\beta} \tag{9}$$

Eq. (8) and (9) provide a relation between the tail exponent of the degree distribution, and that of the spectrum.

IV. Comparing Generators by Simulation

A. Methodology

All generated networks consist of $N = 3050$ nodes. The main reason for this number is that inet3 requires at least 3037 nodes, which is the number of nodes in 1997 Internet AS topology.

To generate the degree sequences, we choose $\beta = 2.18$ as tail exponent. This value of β is in agreement with results for the Internet AS-level topology [10] and [12]. The tail exponent β appears to be constant over time [11].

B. Degree Distribution

Figure 3 plots the pdf and ccdf in log-log scale of all the degree distributions obtained through simulation. The linear slope in the ccdf plot indicates the obtained distributions follow indeed a power law.

The linear regression has been calculated for both the pdf and ccdf data sets. Additionally, we calculate the Pearson's r linear correlation coefficient and the coefficient of variation ($C_v = \sigma^2/\mu$). The first estimates the goodness of the linear fits, the second quantifies the grade of variability. All these analytic results are summarized in Table I.

Figure 4 shows the difference between the β extracted from the pdf, and the β extracted from the ccdf. This *error* decreases if we consider the degrees with small probability as noise, thus ignoring them (i.e. degrees with $f_X(x) < 10^{-4}$). To avoid this estimation error, the nodes with $f_X(x) < 1\%$ have not been taken in account in the linear fit process of β for the pdf.

Table II assesses the effect of removing low probability degrees from the pdf. If we consider low-probability degrees as noise and we remove them, the β approximates to the tail exponent obtained through the ccdf.

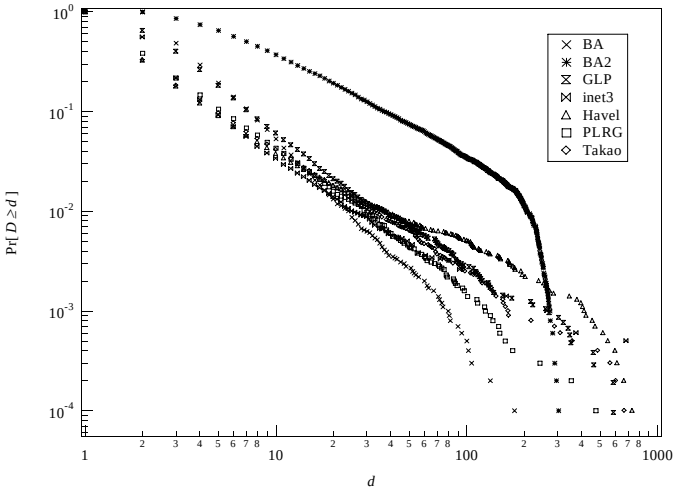

Fig. 3. ccdf graphics for all the 7 algorithms. It can be observed that the behaviour of all the algorithms (except BA-rewire) closely follows a power-law.

	pdf β	ccdf β	r	C_V
PLRG	2.07	**2.47**	0.99	35.3
Havel	2.22	**2.05**	0.97	21.9
Takao	2.10	**2.27**	0.99	38.1
inet3	2.27	**2.21**	0.99	31.3
BA	2.58	**2.96**	0.99	31.0
BA-r	1.76	**2.40**	**0.89**	13.8
GLP	2.12	**2.34**	0.99	25.5

TABLE I

LINEAR FIT OF THE BETAS FOR THE PDF AND THE CCDF GRAPHICS. WE ONLY SHOW THE CORRELATION COEFFICIENT FOR THE CCDF FIT. THE CORRELATION COEFFICIENT FOR THE PDF IS GREATER THAN 0.97 IN ALL THE CASES.

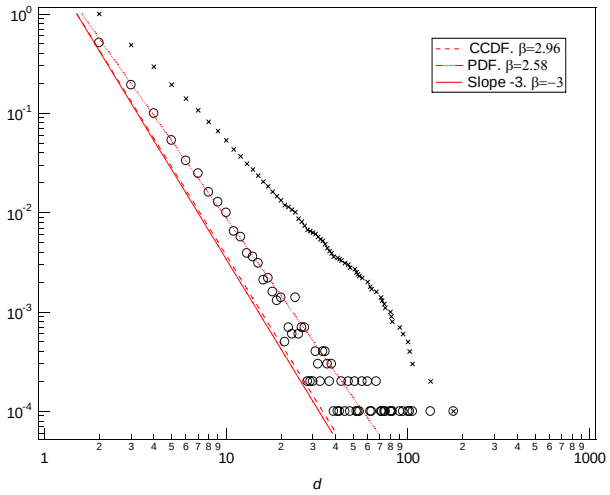

Fig. 4. This Figure illustrates how the pdf underestimates β. We know that BA produce graphs with $\beta = 3$. The dotted line plots the β estimated through the pdf ($\circ$), $\beta = 2.58$. The dashed line plots that for the ccdf ($\times$) $\beta = 2.96$.

	$f_X(x)$	$f_X(x) > 10^{-4}$	$f_X(x) > 10^{-3}$
β	1.12	1.71	2.22

TABLE II

HAVEL MEASURED TAIL EXPONENTS. THE FIRST COLUMN SHOWS THE TAIL EXPONENT CALCULATED FOR THE FULL SET OF DATA. THE SECOND AND THIRD COLUMNS SHOW THE TAIL EXPONENT CALCULATED IGNORING THE DEGREES WITH PROBABILITY LESS THAN 1E-4 AND 1E-3 RESPECTIVELY.

	$\mathbf{R(G)}$
PLRG	-0.13
Havel	-0.38
Takao	-0.23
inet3	-0.18
BA	-0.01
BA-r	-0.01
GLP	-0.11

TABLE III

AVERAGE ASSORTATIVITY COEFFICIENT R(G) OF THE CONSIDERED ALGORITHMS.

BA-rewire[6] generator has remarkable small C_V. Figures 2 and 3 illustrate that BA-r follows a power law distribution up to $d_i \approx 100$. However, the degree distribution becomes exponential for higher degrees, which explains the relatively small coefficient of variation. This critical exponential behavior cannot be easily observed from the pdf graphics, but it is clearly exposed in the ccdf as an abrupt bend (Figure 3).

As mentioned in Section II-B, the PLRG algorithm generates topologies with isolated nodes, duplicated links and self-loops. After these irregularities have been solved (delete isolated nodes, merge duplicated links, delete self-loops) the remaining topology has an obvious deficiency in the degree distribution, as several low degree nodes are removed from the graph. Consequently the tail exponent β is affected.

C. Assortativity

Figure 5 plots the average degree of the neighbors of a node with degree d for BA and PLRG. The top of Figure 5 illustrates the average degree of the neighbors for BA. The degree of the neighbors is fairly constant, which means that there is a weak correlation between the degrees of a node and that of its neighbors. As observed in Table III, the assortativity of inet3, PLRG, Havel, and Takao is clearly negative, which paints to disassortative behavior. This effect is represented as a negative slope in the down side Figure 5, meaning that low degree nodes tend to connect to high degree nodes, and conversely, high degree nodes tend to interconnect with lower degree nodes.

The assortativity coefficient for the Internet AS level graph is found to be $R(Internet) = -0.189$ [15], showing a signigficant disassortative behavior.

As an *expansion* of BA, GLP improves the assortativity results: GLP introduces a significant disassortative behavior.

[6]The family of Barabási-Albert generators does not have nodes with degree 1, since during the generation process each new node starts with initial degree $m = m_0 = 2$.

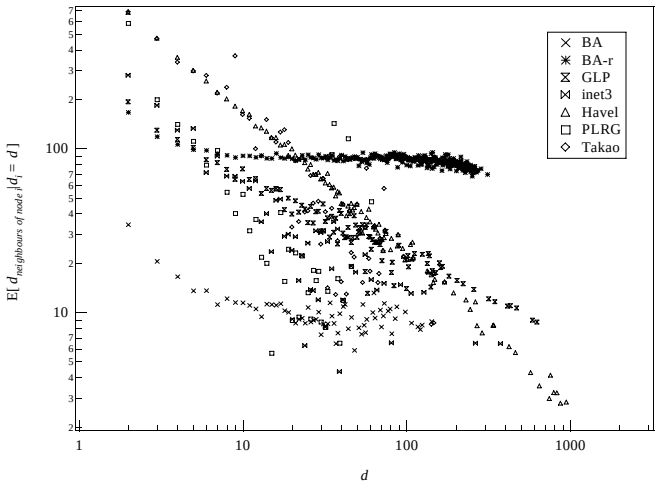

Fig. 5. Average degree of the neighbours (y-axis) given a degree d (x-axis).

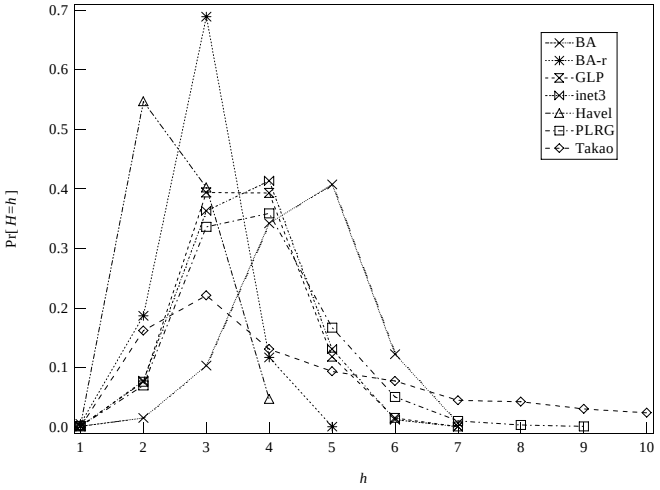

Fig. 6. Hopcount histogram for all the algorithms considered.

	theory	μ	σ^2	max
PLRG	5.51	3.8	0.83	9
Havel	1.39	2.4	0.54	4
Takao	3.18	7.3	5.19	243
inet3	2.66	3.6	0.72	7
BA	102	4.5	0.77	7
BA-r	4.54	2.9	0.36	5
GLP	3.86	3.6	0.72	7

TABLE IV

THEORETICAL MEAN, EMPYRICAL MEAN, VARIANCE, AND MAXIMUM VALUE OF THE HOPCOUNT FOR THE STUDIED GENERATORS. TAKAO MAXIMUM HOPCOUNT CLEARLY STANDS OUT OF THE REST. BA THEORETICAL VALUE IS NOT CORRECTLY APPROXIMATED, AS THE FORMULA IS ONLY VALID IN THE RANGE $2 < \beta < 3$.

Empirical data shows that 20% of the PLRG nodes do not belong to the Giant Component, and thus they are erased. As these erased nodes represent a significant fraction of nodes with low degree, substracting them from the graph implies that the higher degree nodes are *losing* low degree connectivity. This produces an increase on the *average degree of neighbors* for the high degree nodes**.**

Intuitively we might infer that that the assortativity of Havel is be positive, as the algorithm uses the lemma "*Higher degree connect to higher degree*". But the empirical results ($R(G)_{HAVEL} = -0.38$) contradict this interpretation. The essential reason for this apparent conflict is that $R(G)$ is normalized against a set of graphs containing self-loops and isolated nodes. So Havel could be thought of as disassortative when compared with *all* graphs [27].

D. Hopcount

Figure 6 shows the histogram of the hopcount distribution for all the proposed generators. The average hopcount of all the generators lies under 5 hops: the networks have tendency to present a small world graph. Takao is a notable exception, it internconnects long chains of nodes making inappropiate use of the input degree sequence. Table IV presents the simulation results.

Theory [20] [6] states, for large N, that the average hopcount of power law graphs is

$$E[H_N] = 2\frac{\log\log N}{|\log(\beta-2)|}(1+o(1))$$

yielding for $N = 3050$ and $\beta = 2.18$ that $E[H_N] \approx 2.42$.

The hopcount of the algorithms belonging to the same family of generators present remarkable differences. For instance, BA with mean value of 4.5 lies far from BA-r with mean of 2.4. These difference between generators of the same family implies that the hopcount distribution can be easily altered with simple algorithm modifications.

Havel and BA-rewire show a very narrow distribution that centers around two hops. Havel's algorithm (Section II-C) systematically interconnects high-degree nodes with high-degree nodes, creating a highly interconnected core, through which almost all nodes can be reached within 3 hops.

E. Clustering Coefficient

Figure 7 shows the histogram of the clustering coefficient for the evaluated generators, with the Takao clustering coefficient on top, and a vertical compound of all the histograms at the bottom.

BA, BA-r, GLP and inet3 present low variance in comparison to the rest of the algorithms. Havel and Takao posses higher variance and irregular distributions (bottom of Figure 7). Table V shows a summary of the results obtained.

The clustering coefficient of Barabási-Albert is very low ($\mu_{BA} = 0.012$), which implies that BA is organized as a tree. The clustering coefficient for inet3 ($\mu_{inet3} = 0.1255$) appears deterministic, with zero variance. This dissimilarity is a consequence of the construction process (Section II-G), where systematically the same parameters are used.

Both Havel and Takao present graphs with high C_G, while their average hopcount strongly differs. This difference arises because Takao algorithm interconnects *long* tails of nodes with low degree. These tails dramatically increase the average hopcount, but still presenting a central clustered core.

	μ	σ^2
PLRG	0.139	0.070
Havel	0.698	0.190
Takao	0.687	0.150
inet3	0.126	0
BA	0.012	0.002
BA-r	0.123	0.005
GLP	0.127	0.018

TABLE V
MEAN AND STANDARD DEVIATION FOR THE CLUSTERING COEFFICIENT.

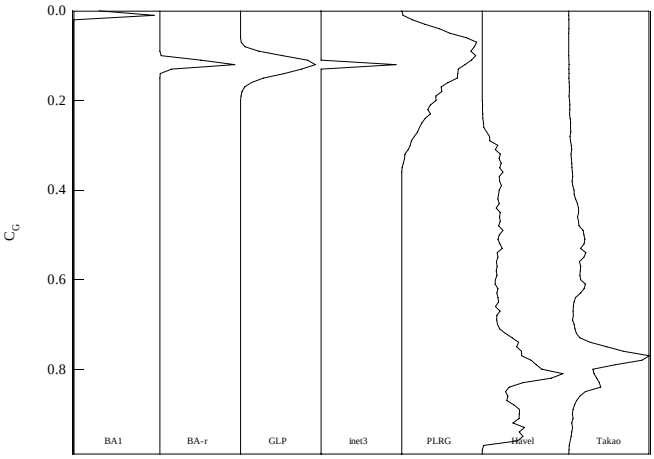

Fig. 7. Histograms of clustering coefficients. In the top we can see the spread histogram of Takao. In the bottom we plot a vertical compilation of all the clustering coefficients. From left to right: BA, BA-r, GLP, inet3, PLRG, Havel and Takao.

As the average Takao hopcount is too large, this algorithm does not present the small world graph properties. As remarked in [27] a power law degree sequence is not warranty for a small world graph. BA presents a very small clustering coefficient, neither representing a small world. PLRG, BA-r, GLP and Havel follow the small world property, as they combine low average hopcount with high clustering coefficients when compared [19] to the random graph $C_{Gp(N)} \approx 10^{-3}$.

F. Spectrum

As pointed by Faloutsos *et al.* [28], there exists an eigen exponent δ, such that the spectral density decays as a power law (8) for *large* eigenvalues. However, the linear regression is not trivial, as it depends highly on what we consider a *large* eigenvalue. This problem is illustrated in Figure 9.

There is a rich literature on the eigenvalues of graphs and their relation with topological properties. The eigenvalues of random graphs tend to the *semicircle law* [8] [19], whereas for scale-free networks the spectrum resembles a bell shape representing a tree-like graph [9]. Moreover, the shape of the spectrum at lower eigenvalues is directly related to the randomness used by the algorithms to generate the topologies. Stochastic topology generating algorithms create graphs that present bulk-shape forms for small eigenvalues [9]. This behavior can be observed for BA (top of Figure 8), BA-rewire,GLP and PLRG. On the other hand, the deterministic

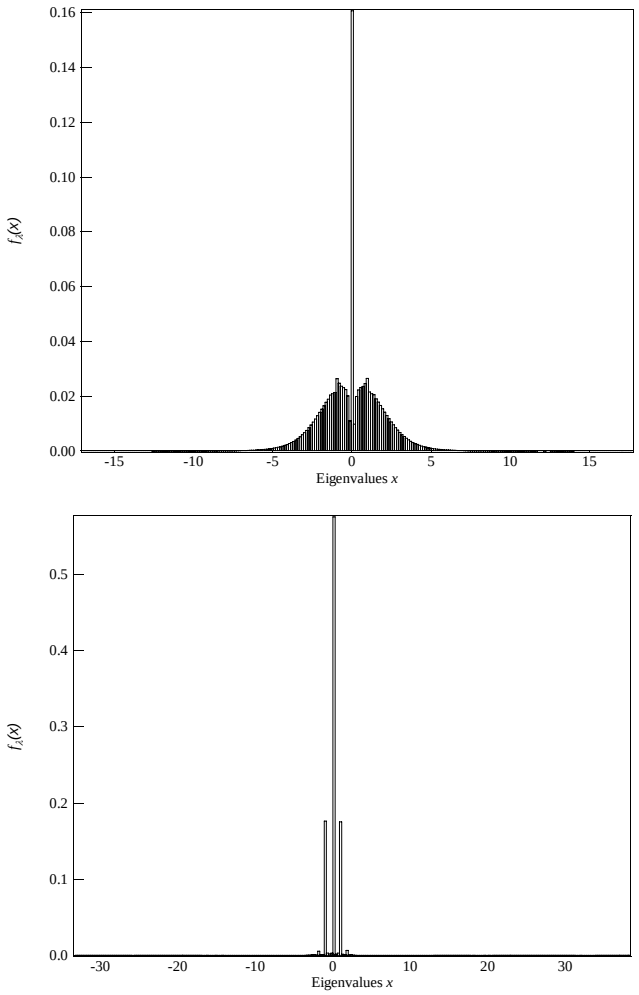

Fig. 8. Top: BA spectrum. The bell shape of the lower eigenvalues reveal a stochastic topology generator. Bottom: Havel spectrum. The spectrum of the deterministic algorithms present very sharp shapes, grouping the 87% of the total density in 3 eigenvalues.

algorithms Havel (bottom of Figure 8) and Takao concentrate their densities in the zero-eigenvalue λ_0. The Inet3 spectrum is qualitatively between the deterministic and random behavior.

Eq. (8) and (9), indicate that real Internet AS-level graphs present power law features in spectrum. The generators under study correctly emulate this scaling behavior: the spectrum decays as a power law for large eigenvalues. Figure 9 illustrates an example of this fitting process with GLP.

Table VI shows the relation between the tail exponent of the spectrum δ and the tail exponent of the degree distribution β stated in Section III-E. The difference is expressed with $|\Delta| = \delta - 2\beta + 1$.

In the experiments made by Dorogovstev *et al.* [8], the slope of the spectrum for BA model is $\delta_{BA} \approx 5$, while in our experiments we obtain $\delta_{BA} \approx 4.8$. The main difference between [8] and ours is the size of the networks.

Dorogovstev *et al.* [8] used $N = 2 \cdot 10^4$ nodes, whereas we use only $N = 3050$ nodes. We have performed simulations for BA with increasing N. Figure 10 shows the evolution in the spectrum as we increase N from 200 to 3050. As depicted in

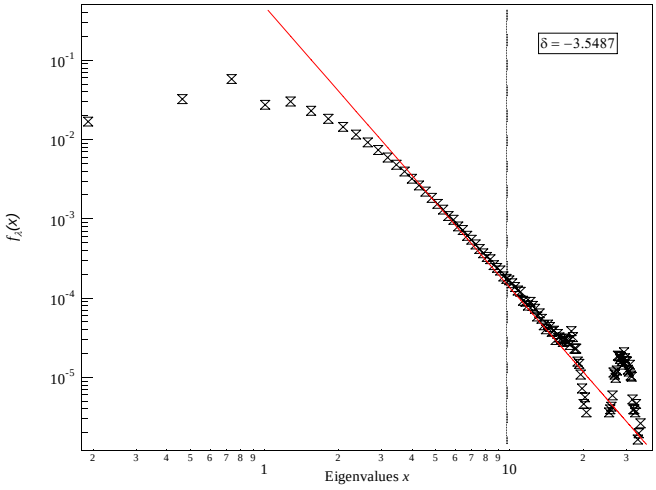

Fig. 9. Right side of the spectrum of GLP on log-log scale. The linear fit is expressed with the flat line. Only eigenvalues greater than 10 have been considered.

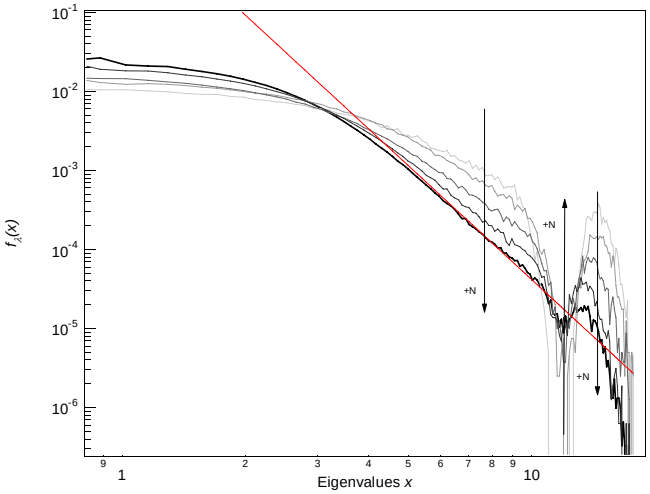

Fig. 10. The different lines represent the positive eigenvalues for N= 200, 400, 800, 1600 and 3050 respectively. The spectrum tends to a straight line with slope -5.

	β	δ	$\lvert\Delta\rvert$
PLRG	2.47	3.95	0.01
Havel	2.05	1.88	1.22
Takao	2.27	3.55	0.01
inet3	2.20	2.39	1.01
BA	2.95	4.78	0.12
BA-r	2.40	1.69	2.11
GLP	2.34	3.54	0.14

TABLE VI

COMPARISON BETWEEN THE TAIL EXPONENTS OF THE SPECTRUM (δ) AND THAT OF THE DEGREE DISTRIBUTION (β). DOGOROSTEV ET AL. FOUND THAT THE EMPYRICAL RELATION $\delta = 2\beta - 1$ HOLDS FOR THE REAL INTERNET AS GRAPH.

Figure 10, the *large* eigenvalues tend to a straight line as N increases. The value where δ tends to can not be empirically obtained because of network size limitations, but we may expect $\delta_{BA\ N\to\infty} \approx 5$ as shown in [8]. This demonstrates that the topologies we are using need a higher N to obtain more precise results, but we still can extract qualitative conclusions.

Table VI shows that BA-r, inet3 and Havel do not behave like scale-free graphs, as the measured δ do not match Eq. (9). The slope of the higher eigenvalues of Takao fits relation (9).

V. CONCLUSIONS

The metrics analyzed here are a grasp of the full range of characteristics that can be computed, but they seem sufficient to differentiate between all the algorithms. On the other hand, one single metric is still not enough to classify an algorithm.

We may summatize as follows:

- **inet3**, is not *reliable*. Trying to mimic the power law distribution (using constant parameters), inet3 provides undesirable secondary effects, for instance, the static clustering coefficient.
- **Barabási-Albert**, is a good example on how linear preferential attachment is a simple and intuitive process yielding power laws. But the lack of input parameters makes it useless to generate real-Internet topologies.
- **Barabási-Albert rewire**, the simulations provide a defect in the tail of the degree distribution as it decays clearly exponentially. This invalidates the model.
- **GLP** was designed to improve Barabási-Albert by softening the hopcount distribution, introducing weak disassortative behavior and closer tail exponent in the degree distribution. However, the input parameters still depend highly on real Internet AS-level data.
- **PLRG** starts with the disadvantage that we can not explicitly decide the number of nodes in the topology due to the randomness of interconnections. However, PLRG has demonstrated to be one of the most stable algorithms: none of its features deviates considerably from the (estimated) real Internet AS results.
- **Havel**, its design purpose is not the interconnection of power law degree distributions, but to deterministically interconnect a given degree sequence. Although the results are consistent, the high clustering coefficients and the non-fitting spectrum indicate that the Havel topologies remain still far from reality.
- **Takao**, again its purpose is not the interconnection of power law degree distributions but only interconnect a given degree sequence. This algorithm fails on the extremely high hopcount (hence in the diameter of the network).

REFERENCES

[1] W. Aiello, F. Chung, L. Lu. *A Random Graph Model for Massive Graphs*. 2000.
[2] R. Albert, H. Jeong, and A.-L. Barabási. *Attack and error tolerance of complex networks*, Nature 406, 378-382 (2000)
[3] A.-L. Barabási, R. Albert. *Emergence of Scaling in Random Networks*. Science vol. 286, October 1999.
[4] A.-L. Barabási, R. Albert. *Topology of Evolving Networks: Local Events and Universality*. Physical Review vol. 85 n24
[5] B.Bollobas, O. Riordan, J. Spencer, and G. Tusanady, *The degree sequence of a scale-free random graph process*, Random Structures and Algorithms, 18, pp 279-290, 2001.

[6] F.Chung, L.Lu, *The average distance in a random graph with a given expected degrees*, PNAS 99 (2002) 15879-15882.

[7] T.H. Cormen, C.E. Leiserson, and R.L. Rivest, *An introduction to Algorithms*. MIT Press. Boston. 1991.

[8] S.N. Dorogovstev, A.V. Golstev, J.F.F.Mendes, A.N.Samuknin. *Spectra of complex networks*. Physical Review E 68 046109, 2003.

[9] I.J. Farkas, I. Derenyi, A.-L. Barabási, T. Vicsek. *Spectra of "real-world" graphs: Beyond semicircle law.* Physical Review E 64 026704, 2001.

[10] T. Bu and D. Towsley, *On Distinguishing between Internet Power Law Topology Generators*, in Proc. Infocom 2002, June 23–27, New York.

[11] J. Winick, S. Jamin. *Inet3.0: Internet Topology Generator*. University of Michigan Technical Report CSE-TR-456-02.

[12] B. Zhang, R. Liu, D. Massey, L. Zhang. *Collecting the Internet AS-level Topology*. http://irl.cs.ucla.edu/topology/

[13] S. Milgram, *The Small World problem*. Psycol. Today, 2, 60-67, 1967.

[14] M. Molloy, B. Reed, *The Size of the Giant Component of a Random Graph with a given degree sequence.* Random Structures and Algorithms, Vol. 6, no.2 and 3 (1998), 295 - 305

[15] M.E.J. Newman, *Assortative Mixing in Networks*, Physical Review Lett. 89, 208701 (2002)

[16] M.E.J. Newman, S.H. Strogatz, and D.J. Watts. *Random graphs with arbitrary degree distribution and their application*, Phys. Rev. E, 64, 026118, 1-17.

[17] C. Labovitz, A. Ahuja, R. Wattenhofer, S. Venkatachary. "The Impact of Internet Polocy and Topology on Delayed Routing in the Internet". IETF Draft.

[18] P. Van Mieghem, *Paths in the simple Random Graph and the Waxman Graph*, Probability in the Engineering and Informational Sciences (PEIS), vol. 15, pp. 535-555, 2001.

[19] P. Van Mieghem, *Performance Analysis of Communications*, Cambridge University Press, 2006

[20] G. Hooghiemstra, R. van der Hofstad and P. Van Mieghem, *Random graphs with finite variance degrees*, Random Structures Algorithms, 26, 76-123, 2005.

[21] G. Siganos, M. Faloutsos, P. Faloutsos, and C. Faloutsos, *Power-laws and the AS-level Internet topology*, ACM/IEEE Transactions on Networking, vol. 11, no. 4, pp. 514-524, 2003.

[22] H. Tagmunarunkit, R. Govindan, S. Jamin, *Network topology Generators: Degree-based vs. Structural*, SIGMCOMM'02, August 19-23, 2002.

[23] Takao Asano. An *O(n log log n) Time algorithm for Constructing a Graph of Maximun Connectivity with Prescribed degrees*. Journal of Computer and System Sciences, 51, 503-510 (1995)

[24] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, *The origin of power laws in Internet topologies revisited*, in IEEE INFOCOM, 2002.

[25] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, *Internet topology at the router and Autonomous System level*, Physical Review E 65, 066130 (2002)

[26] W. Willinger, R. Govindan, S. Jamin, V. Paxson, and S. Shenker,*Scaling phenomena in the Internet: Critically examining criticality*, PNAS, vol. 99, no. Suppl. 1, pp. 2573-2580, 2002.

[27] L. Li, D. Anderson, J. C. Doyle, W. Willinger. *Towards a Theory of Scale-Free Graphs: Definition, Properties and Implications*, 2005.

[28] M. Faloutsos, P. Faloutsos, C. Faloutsos. *On Power-Law relationships of the Internet Topology*, 1999

[29] A. Capocci, F. Coliaori, *Mixing properties of growing networks nad the Simpson's paradox*, 2006.

[30] D. Watts, S. Strogratz. *Collective Dynamics of 'small world' networks*. NAture, 393, 440-442, 1998.

[31] B.M. Waxman. *Routing of multipoints connections. IEEE Journal of selected areas in communication*, 6(9):1617-1622, December 1988.

[32] K. Calvert, M. Doar, and E. Zegura. *Modelling Internet Topology*. IEEE Communications Magazine. June 1997.

[33] M. Doar. *A better model for generating test networks*. In *Proeeding of IEEE Global Telecimmunications Conference* (GLOBECOM), November 1996.