# Is Power Law the Universal Law for Large Real Graph: Internet?

Pathik Sharma
Masters, Computer Science,
State University of New York, Stony Brook
New York, USA

## ABSTRACT

Power laws are observed in many large networks such as Internet, WWW, automobile networks, etc. First half of the paper captures some of the instance of internet over a period of time. Now after comparing topologies of Internet with power laws, it cannot be ignored by saying it a mere coincidence. So, real networks (Internet) thus follows power law. Determining the behavior of such a large network has always been curiosity. Owing to that, different parameters such as number of nodes and edges, average neighborhood size, etc. has been derived to better predict behavior. Second half of paper discuss about generating topologies nearly same as realistic graphs. Also the generated topology must adhere to power laws. Some factors are imposed on topologies to get finer results. To better predict the behavior of topologies some more parameters such as diameter, Incremental Growth and Preferential Connectivity are considered. At last some extension of proposed method is discussed along with its possible application.

## Keywords

Power Laws, Pareto distribution, Zipf distribution, Rank, Eigenvalue, Internet topology, Incremental Growth, Preferential Connectivity, Data Mining.

## 1. INTRODUCTION

Power-laws are observed in many natural, physical, biological, and human-related behaviors. Examples includes sizes of earthquakes, topography, solar flares, human respiratory system [10], urban population, wealth, automobile networks [11], etc. Recently, power laws have also been observed in software systems and computer networks in the area of virtual memory, file systems and database queries. Several characteristics of Internet and WWW [5] also exhibits power law. Some of them are the number of visitors to a website [12], the number of hyperlinks to a webpage [12], the size of web objects [13], the number of links to routers on the internet [2], and the number of friends of users on online social networks [14].

In statistics, a power law is a functional relationship between two quantities where one quantity varies as power of another. A power function is a scale invariant function, *f(x)*, of the form [7]:

$$f(x) = Cx^a$$

; Where *C* and *a* are some constant; and *a* is scaling exponent. Normally the graphical representation of this equation results in skewed distribution (depending upon constant *a*). So, log-log graph is generally plotted, which results in linear line (with slope of graph near to 1). Various quantitative results are thus derived from these graphs and can be used to know the real world topologies more accurately. Presence of power law indicates that arbitrarily large values can occur with a non-negligible probability, and therefore, rather than ignoring these values as "outliers", its statistical significance needs to be carried out.

Previously, the analysis of any topology was done on the basis of maximum, minimum or average based parameters. But real graphs were found to be sparse. Internet (the largest real graph) contained more than 80% nodes below average parameters and so the results derived from such parameters were unable to define the behavior of internet in near future. Now after comparing topologies of Internet with power laws, it cannot be ignored by saying it a mere coincidence.

Power law was first observed in Internet traffic measurement by Pareto and hence it is known as Pareto Distribution [6]. It is also known as 80-20 law according to which 20% of all people receive 80% of all income and 20% of most affluent receive 80% of that 80% income and so on. Another classical example is Zipf distribution [6], which was first used to model word frequencies from any book, movies or web article. This discrete distribution defines the rank of word on the frequency of that word coming in any book or article. As already said, the graphical representation of this distribution results in skewed graph; where the graph is divide into two parts: *body* (small to medium-sized values that are responsible for much of distribution) and *tail* (large sixed values that are responsible for the rest)[7]. The distribution can be said Heavy-Tailed distribution if the tail is not exponentially bounded. Hence the change in large-sized values (that occur less frequent) has greater impact than change in small-sized values (that occur more frequent); whereas a distribution is Long-Tailed if it contain more values that are less frequent (constituting majority of graph values) and form a long tail if seen in graphical representation.

In this paper, I will discuss about the 4 power laws and its relationship with Internet topology given by *Faloutsos et al.* [2] Then, the paper will discuss about generating topologies based on that power law. The remainder of this paper is organized as follows. In the next section, I will review the background and previous work in the field of power laws. Section 3 will contain all the power laws along with its strengths and weakness. Section 4 contains generating real topologies based on power laws. Section 5 discusses several methods that can be proposed to possibly explore further in getting real topologies from power laws and Section 6 contains the possible application of power laws. Concluding the paper; Section 7 discuss the scope of possible exploration in the field of power laws and its relation with large real graphs topologies.

## 2. BACKGROUND & PREVIOUS WORK

The accurate characterization of Internet topologies is key to successful modelling and analysis of internet and its protocol [15]. Here, the first step is to draw a correct topology of internet to derive the most accurate conclusions. Now the topologies drawn should depict the true parameters of Internet through which the future behavior of Internet can be found out. *Faloutsos et al.* [2] derived four power laws (that I will discuss in

subsequent Section) based on Internet topology. For that, they decomposed Internet into sub networks called *domains* or *autonomous systems.* Furthermore, Internet was studied in different granularity viz. router level [16] (considering router as a node), and inter-domain level [17] (consider domain as a node and each edge as an interconnection). Also different results were derived based on different granularity resulting in different conclusions. This showed that power law also differentiate from topology to topology.

Through this paper, *Faloutsos et al.* [2] found that world's largest real network (Internet) follows power law. Based on this several parameters such as number of nodes, number of edges, effective diameter and average number of neighbors with in *h* hops are derived that are useful for analysis of protocols and for speculations of Internet topology in the future. However, causes for the appearance of such power laws have not been mentioned convincingly.

*Alberto Medina et al.* [1], thus came up with four factors that help better understand the formation of Internet topologies. These factors are (F1) preferential connectivity of new node to existing edges, (F2) incremental growth of network, (F3) distribution of nodes in space, and (F4) locality of edge connections. Now to better understand the graph, other parameters apart from [2] are diameter, average path length and clustering coefficient. Also, different network topology is generated that resembles real network and follows power laws.

## 3. POWER LAWS IN INTERNET TOPOLOGY [2] [3]

The primary aim was to study the behavior of Internet evolution over time. For this, Internet was divided into two different granularities: 1. Inter-domain level and 2. Router level.
The data for the Inter domain level was collected from National Laboratory for Applied Network Research [2]. So, the work was based on three Internet instances between November 1997 and December 1998 (three real graphs correspond to six month intervals approx.). During this period of 1 year (as shown in fig. 1) the size of network increased substantially (45%). And the data for Router level was collected from Pansiot and Grad [16].

*Faloutsos et al.* [2] wanted to find metric that can precisely quantify topological properties and describe concisely skewed data distribution. Previously, metrics were based on maximum, minimum and average values that were not good descriptors of real topologies; they miss a lot of information and probably the "interesting" part that needs to be captured [2].

In [2], probability distribution function is used while in [3] cumulative distribution function is used. The cumulative distribution is preferable since it can be estimated in a statistically robust way [3]. So some power law are slightly changes from [2] to get better results.
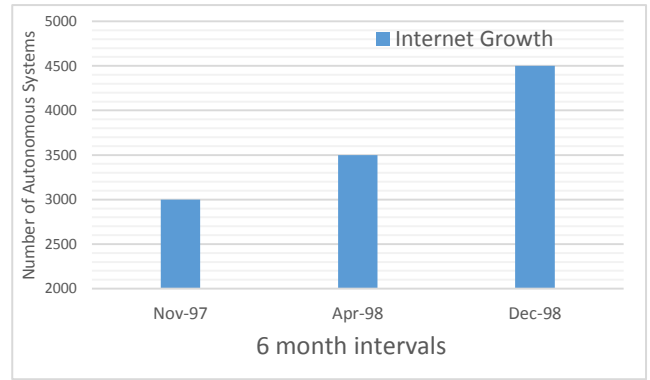


**Figure: 1[2] The growth of Internet over given span. (Data for inter-domain graph)**

Based on the above facts *Faloutsos et al.* [2] devised these power laws:

### 3.1 The Rank exponent R:

> **Power-Law 1 (rank exponent)** *Given a graph, the degree, $d_v$, of a node v, is proportional to the rank of the node, $r_v$, to the power of a constant R:*
>
> $$d_v \, \alpha \, r_v^{R}$$

The graphs of inter-domain instance and router level when plotted in log-log scale exhibits linear curve with co-relation coefficient 0.974 and 0.948 respectively. (The coefficient near to 1 can be consider optimally following power law with a straight line curve).

*Discussion:* Based on the rank of node, the degree can be found using:

$$d_v = \frac{m}{N^R} \, r_v^{R}$$

; m = minimum degree of the graph.

Also, number of edges can be found using:

$$E \approx \frac{1}{2\,(R+1)} \left(1 - \frac{1}{N^{R+1}}\right) N$$

A 5% increase (decrease) in rank exponent R results in 10% increase (decrease) in number of edges.

### 3.2 The Degree exponent O:

For a Complementary Cumulative Distribution Function (CCDF),

> **Power-Law 2 (degree exponent)** *Given a graph, the CCDF, $D_d$, of an degree d, is proportional to the degree to the power of a constant, D:*
>
> $$D_d \, \alpha \, d^{D}$$

The co-relation coefficient for inter-domain and router level are 0.968-0.99 and 0.966 (which again leads to linear graph when plotted in log-log scale).

*Discussion:* The intuition behind this power-law is that the distribution of the degree of Internet nodes is not arbitrary. The qualitative observation is that degrees range over several orders of magnitude in a scale-invariant way. If a graph does not follow Power-Law 2, or if its degree exponent is considerably different from the real exponents, it probably does not represent a realistic topology.

## 3.3 The Eigen exponent E:

Eigen values of the graph are calculated because many basic topological properties such as diameter, number of edges and nodes, connected components and number of walk can be derived [20].

> **Power-Law 3 (eigen exponent)** *Given a graph, the eigenvalues, $\lambda_i$, are proportional to the order, I, to the power of a constant, ε:*
>
> $$\lambda_i \, \alpha \, i^{\varepsilon}$$

Here, the Eigen exponent for inter-domain remain -0.50 whereas the Eigen exponent for router level graph is -0.17 which significantly varies from inter-domain level.

*Discussion:* Eigen exponent can spot differences between families of graph.

## 3.4 The Hop-plot exponent H:

Here, number of pairs of node *P(h)* with in *h* hops is defined as total number of pair of nodes with in less or equal to *h* hops, including self-pair, and counting all other pair twice. The graph plotted is approximated to first 4 hops in inter-domain graph.

> **Power-Law 4 (hop-plot exponent)** *The total number of pair of nodes, P(h), within h hops, is proportional to the number of hops to the power of a constant, H:*
>
> $$P(h) \, \alpha \, h^{H}, \quad h \ll \delta$$

Here, in router level graph, the number of nodes is far greater than inter-domain graph. So, plot for the router level is more refined than inter-domain graph (as only 4 hops are considered in inter-domain level).

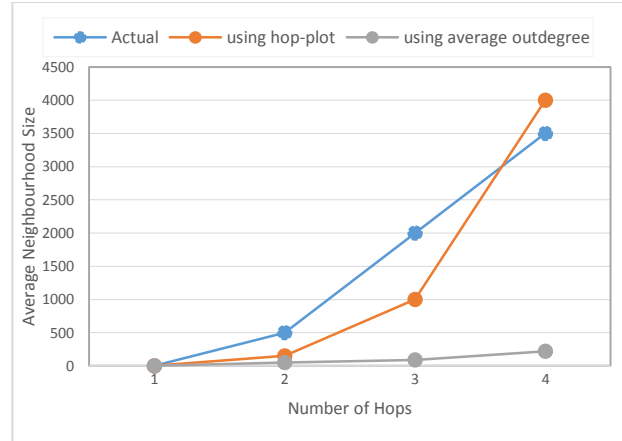The proximity of Hop-plot exponent to the actual graph is shown in Figure 2.



**Figure 2 [3]: Average neighborhood size VS number of hops for actual, Hop-plot and Average degree (Inter-domain)**

*Discussion:* The effective diameter D can be calculated as 80% of nodes in Inter-domain graph is covered for hop=4. Also as number of nodes for router level is more and so is the value for *h*, the effective diameter can be accurately define comparative to domain-level. So, hop-plot can give better estimation of effective diameter for different graph families.

## 3.5 Conclusion
- Discovered 4 power laws with high co-relation coefficient.
- Proposed power-law exponent, instead of averages, that defines skewed distribution resembling more towards real graphs.
- Predictions can be carried out by derived parameters such as number of nodes and edges, average neighborhood size, and effective diameter.

## 3.6 Improvements

Author here has described power laws on the basis of given topologies and found out that real graph follows power law. But how to obtain a topology from the given power law is still a question? Again, the cause of appearance of power law is not given.

The author has taken a subset of Internet topology (the data for inter-domain and router level graph) and has derived power laws. It may or may not be true for the today's complete real graph.

The actual meaning of exponents derived still needs to be defined in order to get better and closer anticipated results. More parameter needs to be evaluated as only power law cannot predict the topology of real systems.

These are some of the questions and challenges that author has to overcome in order to come up with more realistic and deterministic graphs for future predictions.

# 4. GENERATING INTERNET TOPOLOG-IES [1]

Power laws has been defined. Now topologies needs to be constructed in such a way that it follows power law. *Alberto Medina et al.* [1] defined four factors that can be used to generate topology. Along with that, parameters such as diameter, average path length, and clustering coefficient is also derived for better results.

First, let me consider commonly used topology generation methods [18]. Such as

*Random Method:* Set of node is distributed in plane uniformly at random. A link is added between each pair of nodes with certain probability (Waxman probability is used). It cannot be used with large scale networks especially sparse, but connected like Internet.

*Regular Method:* This method generates regular topologies with specific and rigid structure. It is often used in analytic studies.

*Hierarchical Method:* In this method, topologies are created by connecting small random graphs together to form a large scale structure. It provide a balance between randomness and structure.

*Transit-Stub (TS) Method:* A random graph is connected first using Waxman method. Each node represents an entire *Transit domain* and then expanded to another random graph, representing backbone topology. Now for each node a random graph is generated i.e. *Stub domain*. Finally, some edges are added in to it. It is possible to generate real graph (sparsely connected Internet like topology) with typically low degree.

But the most common network topologies generator (Waxman and TS) fails to depict real networks and varies greatly when compared to real graphs. As a result power law arise. But according to power law, a topology cannot be derived given present parameters. So, a BRITE topology generator is used which contains 4 factors: Node placement, Connection Locality, Preferential Connectivity (PC) and Incremental Growth (IG).

## 4.1 BRITE Topology Generator:

Consider a plane *HS* x *HS* squares. Each high level square is then subdivided into *LS* x *LS* low level squares. Each low level square is assigned at most one node. Consider the factors:

*Node Placement*
A random placement of node in low level square avoiding collision. For each of high level square, *bounded Pareto distribution* is used.

*Connection Locality*
The parameter *m* controls the number of neighbor nodes to which a new node connects when it joins the network. The greater the value of *m*, the denser the generated topology.

*Incremental Growth (IG)*
It take one of the two values:

- *Inactive:* places all node at once before adding any link. Any node is randomly selected and *m* links are used to connect it from all other nodes.

- *Active:* places nodes gradually. A new node connects to other nodes only when other nodes are connected to some other nodes.

*Preferential Connectivity (PC)*
There are three possible values:

- *None:* Preferential connectivity is turned OFF. Newly connected nodes connects using Waxman probability function.

- *Only:* Node gets connected using some probability:

$$\frac{d_i}{\sum_{j \in C} d_j}$$

This process implies that a new node joining the network selects with high probability, those nodes with high Outdegree.

- *Both:* Combines preferential connectivity and connection locality. Combining Waxman probability and above probability:

$$\frac{w_i \, d_i}{\sum_{j \in C} w_j \, d_j}$$

Also, for generated topology the parameters are *diameter* (Average shortest length from any one node to the farthest node), *average path length* (average path length taken over all pair of nodes) and *clustering coefficient* (consider any node *v* and its set of neighbor $k_v$ contains at most $k_{v+1} C_2$ edges between these nodes. Clustering coefficient $C_v$ is the fraction is that set of edges which are actually present in graph).

Now consider the graphs for BRITE topology using random node replacement. Here, IG and PC variants is used to plot graph. IG = 0 means Incremental Growth is not used While PC = 1 means Preferential Connectivity is used. Here all topologies used are Waxman, Grids, Transit-Stub, 4 variants of IG and PC.
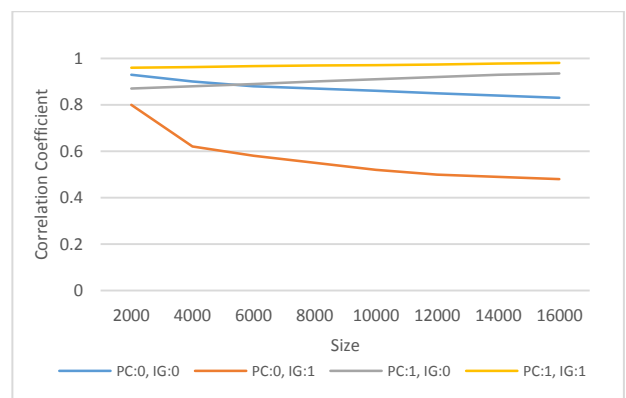


**Figure 3 [1]: Rank exponent for BRITE topology (Random Node Replacement)**

Topology generation model that includes both Preferential Connectivity and Incremental Growth will produce topology that highly resemble Internet topologies with respect to the rank
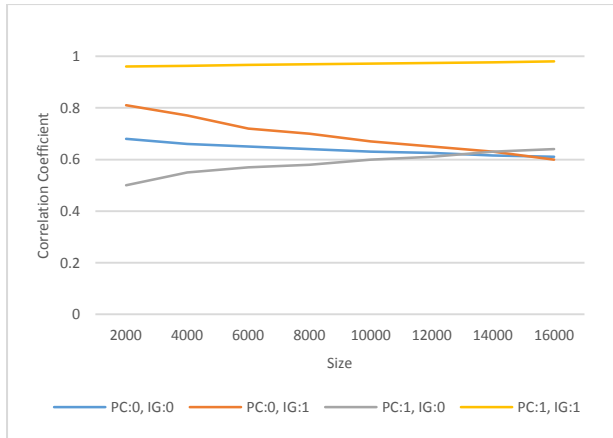
**Figure 4 [1]: Outdegree exponent for BRITE topology (Random Node Replacement)**

BRITE topologies generated with both preferential connectivity and incremental growth closely resemble Internet topologies. Thus, like the rank exponent, the Outdegree exponent can effectively distinguish different classes of topologies.
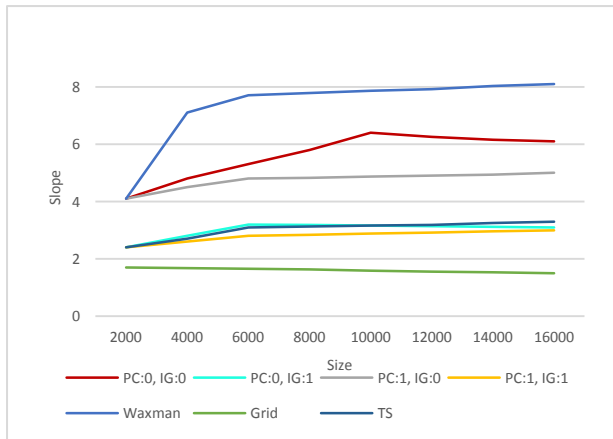


**Figure 5 [1]: Linear Fit slopes for Hop-plot exponent for all the topologies**

In summary, with respect to the hop-plot exponent, both Transit-Stub and BRITE topologies with incremental growth resemble Internet topologies.
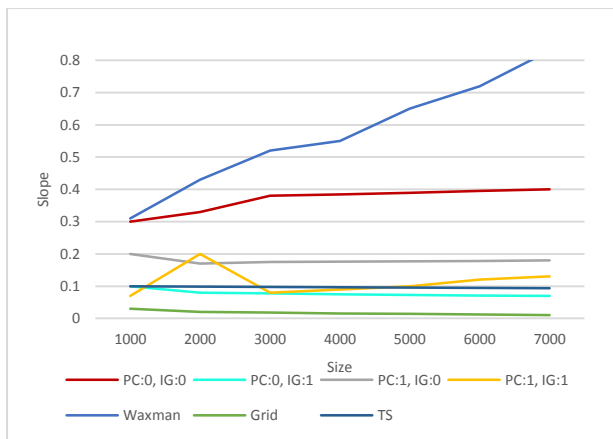


**Figure 6 [1]: Linear Fit slopes for Eigenvalue exponents for all topologies**

Again, BRITE topologies generated without preferential connectivity nor incremental growth behave similar to Waxman topologies.

## 4.2 Conclusion

- Topologies that strike good balance between randomness and structure are found to have metrics close to those measured for the internet.
- Hop-plot and Eigenvalue exponents hold for every topology discussed.
- Topologies generated using Preferential Connectivity and Incremental Growth shows best co-relation coefficient.
- Some topologies doesn't obey power law defined by Rank and Outdegree exponent.

## 4.3 Improvements

Two parameters IG and PC can bring the topologies nearer to real networks. Similarly if more parameters are attached with topology generators than a more smoother and finer graph can be formed.

As of now, topologies are considering only power laws to be a check for resemblance with Internet topologies but more parameters and semantics of these parameter can provide better results.

Topology generator can be made more efficient if combination of Power-Law Out-degree generator [4] and Recursive topology generator [4] can be used.

## 5. FURTHER RESEARCH

Till now, we have seen that some of the largest networks such as Internet topologies, WWW, Social Network follows power laws. *Faloutsos et al.* [2] derived power laws related to computer networks considering three instance of inter-domain graphs. After a decade, they reviewed similar results using Oregon routeviews project [3] that span across 1600 days containing 1253 daily instances. They compared an instance from *Oregon* (of a previous year) and more complete topology *Multi (*a complete graph after some years Oregon was taken*)* to confirm whether behavior of *Oregon* predicts as *Multi.* Also, they modified probability distribution function with cumulative distribution function to get more realistic graphs. These graph uses the phenomena of "rich-gets-richer [9]" and as a result skewed distribution is obtained.

Now to overcome the drawbacks of generating topologies from power laws, *Alberto Medina et al.* [1] devised 4 factors and some parameters such as clustering coefficient, average path length and diameter. Topologies generated followed power law to certain amount. But to get the best prediction from topologies some parameters [8] such as Resilience, Maximum clique size, Distortion, delay between links can be used. Combination of parameters versus coefficient can be calculated to get different results.

## 6.  APPLICATION OF POWER LAWS

Possible application where power laws can be used or is used in the field of Information Technology and computer networks are:

## 6.1 Database Caching

As per the power laws, one already know that there are some queries that are requested most frequent and some least frequent queries that are requested majority of times. If one can possibly predict these queries then, a better database management system along with the caching system can be formed.

## 6.2 Search Schemes [7]

Power law node connectivity distribution can help improve search in web. Some Web pages may have a high out-degree and others may have high in-degree [19]. Web pages that work as information aggregators typically have high out-degree and are referred to as hubs, whereas Web pages with high in-degree are typically referred to as authorities. The idea of hubs and authorities was used for ranking Web pages [19] and similar can be used in peer to peer network and other relevant systems.

## 6.3 Long Tail and Business Practice [7]

Online business can also take advantage of power laws. Sites like EBay and Amazon could possibly know most frequent items required to customer. At the same time there are lots of items that are least frequently ordered but constitute a larger domain.

## 6.4 Long Tail and System Design [7]

The long tail possess challenges in peer-peer file sharing such as Torrentz. In particular, mildly popular files may not have popularity and so it may not remain active for longer period. In this case, combine some mildly popular files in to one torrent so that the probability of this torrent getting popular will increase and will result in active torrent.

## 7. CONCLUSION

Power law has been observed in many natural, man-made and biological phenomenon. It is also observed in computer networks esp. in Internet topology and WWW. Based on topologies available, four power laws were described. Then parameters such as diameter, average neighborhood and number of edges and nodes were calculated. These parameter helps topology to predict near future behavior. Then, topologies were generated using Power law as a check for large real networks. More parameters such as clustering coefficient, average path length, etc. were observed which brought topologies more near to real graphs. Also more realistic graphs can be obtain if more and more parameters are observed. Possible exploration can be done to find out the actual meaning associated with exponents. In the end, application where power laws can be fruitful has been discussed. The answer to the question asked above (in the title) is left for the readers to deduce.

## 8. REFERENCES

[1] Medina, Alberto, Ibrahim Matta, and John Byers. "On the origin of power laws in Internet topologies." ACM SIGCOMM computer communication review 30.2 (2000): 18-28.

[2] Faloutsos, Michalis, Petros Faloutsos, and Christos Faloutsos. "On power-law relationships of the internet topology." ACM SIGCOMM Computer Communication Review. Vol. 29. No. 4. ACM, 1999.

[3] Siganos, Georgos, et al. "Power laws and the AS-level internet topology." IEEE/ACM Transactions on Networking (TON) 11.4 (2003): 514-524.

[4] Palmer, Christopher R., and J. Gregory Steffan. "Generating network topologies that obey power laws." Global Telecommunications Conference, 2000. GLOBECOM'00. IEEE. Vol. 1. IEEE, 2000.

[5] Adamic, Lada A., and Bernardo A. Huberman. "Power-law distribution of the world wide web." Science 287.5461 (2000): 2115-2115.

[6] Newman, Mark EJ. "Power laws, Pareto distributions and Zipf's law." Contemporary physics 46.5 (2005): 323-351.

[7] Mahanti, Aniket, et al. "A tale of the tails: Power-laws in internet measurements." Network, IEEE 27.1 (2013): 59-64.

[8] Winick, Jared, and Sugih Jamin. Inet-3.0: Internet topology generator. Technical Report CSE-TR-456-02, University of Michigan, 2002.

[9] Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." science 286.5439 (1999): 509-512.

[10] Mandelbrot, Benoit B. The fractal geometry of nature. Macmillan, 1983.

[11] Faloutsos, Christos, and Ibrahim Kamel. "Beyond uniformity and independence: Analysis of R-trees using the concept of fractal dimension." Proceedings of the thirteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems. ACM, 1994.

[12] Clauset, Aaron, Cosma Rohilla Shalizi, and Mark EJ Newman. "Power-law distributions in empirical data." SIAM review 51.4 (2009): 661-703.

[13] Mitzenmacher, Michael. "A brief history of generative models for power law and lognormal distributions." Internet mathematics 1.2 (2004): 226-251.

[14] Mislove, Alan, et al. "Measurement and analysis of online social networks." Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. ACM, 2007.

[15] Paxson, Vern, and Sally Floyd. "Why we don't know how to simulate the Internet." Proceedings of the 29th conference on Winter simulation. IEEE Computer Society, 1997.

[16] Pansiot, Jean-Jacques, and Dominique Grad. "On routes and multicast trees in the Internet." ACM SIGCOMM Computer Communication Review 28.1 (1998): 41-50.

[17] Govindan, Ramesh, and Anoop Reddy. "An analysis of Internet inter-domain topology and route stability." INFOCOM'97. Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE. Vol. 2. IEEE, 1997.

[18] Zegura, Ellen W., Kenneth L. Calvert, and Michael J. Donahoo. "A quantitative comparison of graph-based models for Internet topology." IEEE/ACM Transactions on Networking (TON) 5.6 (1997): 770-783.

[19] Kleinberg, Jon M. "Authoritative sources in a hyperlinked environment." Journal of the ACM (JACM) 46.5 (1999): 604-632.

[20] Cvetkovic, Dragos M., Michael Doob, and Horst Sachs. Spectra of graphs: Theory and application. Vol. 413. New York: Academic press, 1980.