Improved Tradeoff-based Models of the Internet

Anthony Spatharis[1], Ilias Foudalis[1], Minas Gjoka[2], Panagiota Krouska[1],
Christos Amanatidis[1], Christos Papadimitriou[3], and Martha Sideri[1]

[1]Deparment of Computer Science
Athens University of Economics and Business
46 Patision, Str, GR 104 34, Athens, Greece*
E-mail: (spathari,sideri)@aueb.gr

[2]Donald Bren School of Information and Computer Sciences
University of California, Irvine, CA 92697
E-mail: mgjoka@uci.edu

[3]Deparment of Computer Science
University of California at Berkeley, USA
E-mail: christos@cs.berkeley.edu

Abstract: This paper introduces and evaluates several new models of the Internet graph, inspired by the model proposed by Fabrikant, Koutsoupias, and Papadimitriou (*FKP*), in which connections are chosen based on a tradeoff between a geometric objective and a topology objective. For each new node, the proposed models add at least two edges which optimize in addition to the original tradeoff, secondary criteria such as path independence, and then proceed to add, with some small probability, a new edge from the receiving node. In another version of the model, new edges can be added only to certain nodes, which are designated as fertile, an attribute that changes dynamically. These models are evaluated by comparing them to the real Internet AS graph (or network, interchangeably) with respect to a suite of many test parameters (such as power law exponent and local clustering rank) proposed in the literature.

Keywords: *Internet topology*, *multi-objective optimization*, *power-law, scale-free graphs, network modeling and simulation*, *complex networks.*

1. Introduction

Complex networks have recently attracted much attention in disciplines as diverse as sociology, biology, economics, and computer science; see [5] [6] [7] [10] [11] [16] [18] for several reviews and bibliographies from different perspectives. In computer science, interest in complex networks arose mainly in connection to the Internet and the worldwide web.

The Internet is a hierarchical network composed of communication devices (routers) interconnected with links, and belonging to tens of thousands of different administrative domains. Understanding the structure of the Internet is a major research frontier in Computer Science. In particular, it is of great importance to understand the quantitative properties of the Internet topology (the Internet graph), since proposed new protocols need to be tested on models that are as realistic as possible.

In this regard, the observation, due to [13], that the degrees of the Internet graph are power law distributed brought about a revolution in the area. It made researchers realize that sophisticated graph models are needed that go beyond the traditional $G[n,p]$ random graph models. Indeed, starting with the preferential attachment model [3], many such models have been proposed, see [1] [5] [14] for recent surveys.

One of the most influential models of the Internet topology is the so-called *FKP* model proposed in [12]. In this model, nodes arrive one after the other and each node is connected to the previous node that minimizes a linear combination of (intuitively, a tradeoff between) two objectives: Euclidean distance, and topological (hop) distance from the first node. While this model has been shown in [12] to produce power law-distributed degrees for large areas of the tradeoff, it has also several shortcomings. For example, the produced network is a tree; as a consequence, the average degree is 2 (less than half of that of the real Internet).

This paper extends the FKP-based model in several novel directions, in a way that addresses these shortcomings. The new models differ from the original FKP model in two respects: First, each new node is connected to the network via not one but several edges, and these edges are chosen in a way that optimizes the original tradeoff, but also a secondary criterion related to how independent these edges are (that is, how much different are the paths to the first node that they create). Second, for each of the old nodes that receive an edge, a new edge may be added, with some small probability, to the rest of the network. This reflects the real-life situation in which new connections force an already established node to increase its connectivity to the rest of the Internet. Third, not all old nodes can

---

* This research project is co-financed by E.U.-European Social Fund (75%) and the Greek Ministry of Development-GSRT (25%).

receive such edges. Only certain nodes are designated as *fertile* and are thus qualified to be connected to new nodes; fertility depends on connectivity and changes dynamically.

Perhaps more importantly, these new models are tested experimentally by calculating several parameters (including most of those proposed in the literature) and are compared with the same parameters of the real Internet. Many of the new models produce graphs that are in excellent fit with the Internet with respect to the test parameters. We consider this thorough experimental evaluation methodology as one of the main assets of this work. In a separate paper [17], it is proven analytically that several of these models produce graphs with power law-distributed degrees.

In a recent authoritative survey of the area [14] it is pointed out that, when it comes to models of the Internet, there is a tradeoff between simplicity and realism; that is, how close to the real Internet the model is, and how complex and unintuitive does it have to be to achieve such closeness. The models presented here seem to occupy a very attractive region in this tradeoff.

The rest of the paper is organized as follows. Section 2 provides a brief review of the background and related work. In Section 3 the new models are presented. In section 4, several metrics of the new models are computed and compared with those of the physical Internet graph as well as with other existing models. Finally, conclusions are given in section 5.

2. Background and Related Work

The oldest and most influential model of random graphs is the 1959 $G[n,p]$ model of Erdős and Rényi, in which edges among n nodes are selected independently and with probability p [1]; the degrees are thus concentrated by a Gaussian distribution, according to the central limit theorem. In 1999 it was observed [13] that the Internet topology exhibits a power law degree distribution asymptotically:

$$P(k_i > x) = (x / \beta)^{-\gamma} \propto x^{-\gamma} \qquad \text{Eq. (2.1).}$$

Here the symbol $\propto$ stands for "proportional to", k_i is the degree of the ith highest degree node, β is a scale parameter, and γ is the *shape* parameter or *tail exponent*. It was later pointed out that, in addition, Internet graphs are highly "clustered," among other properties.

Many evolutionary models (in which nodes arrive one by one and are somehow connected to the already existing nodes) have been proposed to explain these observations. Of these the *preferential attachment* models assume that the arriving node is connected to old ones with probability proportional to their degrees [1] [3], the *tradeoff-based* models [2] [9] [12] in which the connections are such that a multiobjective criterion is optimized. Both of these families of models have been shown to generate power-law degree distribution for the Internet AS graph topology, but are lacking with respect to other criteria. The next paragraph focuses on the latter model.

The original tradeoff model, known as FKP [12], is a simple evolutionary model. Each node i arrives at a uniformly random point in the unit square, and attaches itself to the node j that minimizes the weighted sum of these two objectives:

$$min_{j<i} \{\alpha \cdot d_{ij} + ecc(j)\} \qquad Eq.\ (2.2).$$

The first term is the Euclidean distance between the two nodes (the "last mile cost"), while the second is the eccentricity of j, its distance from the center (assume for simplicity to be node 1, the historically first node). α is a weight capturing the relative importance of last mile costs; it is known that, except for extremely small and extremely large values of this parameter, the degree sequence of the resulting graph is power law distributed [4] [12].

3. The Network Models

There are many reasons why, despite initial promise, the original FKP model fails to be a good model of the Internet topology. First, since each node attaches itself to the network by a single edge, the resulting graph is a tree – to put it otherwise, the average degree is two, whereas the Internet's average degree is 4.6 for the graph of the *autonomous systems* (*AS*). Naturally, there is no clustering in FKP.

3.1 The Best Two (BT) Model

As in FKP, in the new models an arriving node is located uniformly at random in the unit square. In the simplest and most basic modification, called *BestTwo*, instead of selecting one node j that minimizes (2.2), the two best such nodes are choosen, and connect the new node to both.

This immediately makes the graph nontree, and increases the average degree to 4 – closer to 4.6. One disadvantage of this method is that the two nodes that optimize (2.2) may be very close, resulting in connections and paths of low diversity.

3.2 The Independent Path (IP) Model

***Step* 1a**. A new node i is linked to the node j according to Eq. (2.2) where $ecc(j)$ is the hop distance from vertex j to the initial node.

***Step* 1b**. A second edge is attached from the new node i to an already existing different node k that minimizes the following quantity:

$$min_{k<i} \{\alpha \cdot d_{ik} + Com(j, k)\}, \text{ Eq. (3.1)}$$

where $Com(j,k)$ is the number of common nodes between the two shortest paths from both j and k to the center. In other words, this algorithm rewards independence between the two connections.

3.3 The Enhanced Models

In the real Internet, links are not created only when a new node arrives. As old nodes acquire more and more connections and traffic, they may seek new connections. This is captured by modifying the previous models with the following additional step.

Step **2**. After the new node has been connected to two hosts, one of the hosts is also connected, with probability p, with another old node, called the peering node. This node is the one which minimizes Eq. (2.2) for the host node.

These models are called *Enhanced BestTwo* (*EBT*) and *Enhanced Independent Path* (*EIP*) with parameter p. When $p = 0$, two unenhanced models are obtained, while for $p = .3$ the average degree of the resulting graph is 4.6, the same as that of the real Internet.

3.4 The Controlled Distance (CD) Model

The real Internet is not a tree, and has rather low power law exponent, while several of its nodes are leaves. The FKP model has low exponent but is a tree, with the vast majority of the nodes being leaves. Our models above are not trees, but have relatively high power law exponents (see Table 1), and no leaves. Our new model decreases a little this exponent while having high average degree, and creating several leaves. In many ways it is the best performing of our models, in terms of the test parameters of the next section. The intuition is that edges between nodes of intermediate centrality (not quite leaves, but neither very central) are needed, and our algorithm below achieves this.

The main idea is that the second edge is added not from the new node i, but from the node j to which node i was connected. The connection is chosen so that its hop distance from the center is not large:

Step **1**. As before, the new node i is linked to the node j that minimizes the quantity Eq. (2.2).
Step **2**. A second edge is attached from j to an another node k, that minimizes

$$min_k \{\alpha \cdot d_{jk} + ecc(k)\} \text{ Eq.(3.2)}$$

over all k such that the hop distance from j to k is at most a constant c.

To implement this model, one needs a fast incremental algorithm that maintains the hop distance between any two nodes, in the face of node and edge insertions. This was done by adapting a matrix representation of evolving distances [15].

3.5 The Fertile Node (FN) Model

In the real Internet not all nodes are likely to acquire new connections. As the Internet grows, a new node may or may not decide to invest on new connections depending on the current traffic needs.

Moreover, because of competition only some of the new nodes will get new customers. In the last model we take into account this insight. Some of the nodes are designated as *fertile* ones, and new nodes are connected to the fertile node that optimizes (2.2). A new node becomes fertile with probability proportional to the age of its parent (the number of iterations of the algorithm since that parent arrived, never to exceed some upper limit u). Besides, if a node has a number of children that exceeds another constant c, then one of these children, chosen at random, becomes fertile. More detailed results, both theoretical and experimental, concerning the fertile model are presented in [17].

4. Analysis of Internet Topology Models

We implemented the six models, including FKP, within the framework of a general and flexible topology generating *Java* program called *TopGen*, and evaluated them by implementing 11 important Internet-related graph metrics identified in [10]. The topology generator, its analysis evaluator and more experimental results can be found in *URL*: http://dias.aueb.gr/~p3030075/topgen.rar or by contacting the authors.

Table 1 displays the results of calculating these metrics for the output of the six models, and comparing them to the same metrics of the actual Internet graph obtained from [8]. To compare the models effectively, the generated graphs have the same number $n = 8000$ of nodes as the Internet graph (see Fig. 1), and the parameter α in all models is taken to be 90 (in the range of values predicted by *Theorem* 1 in [12] to yield a power law distribution). Also in *Appendix A* there are presented some plots from various metrics that where used (see Fig. 2 - 4).

4.1 Model Comparisons

With respect to the average node degree one can see that the BT and CD models are close to the average node degree of the autonomous systems (*AS*). The maximum node degree k_{max} of the AS graph is significantly smaller than those of the new models.

The *FN* model exhibits the smallest value of k_{max} which is slightly larger than the one of the Internet. On the other hand, the BT and EBT stochastic models produce substantially larger deviations.

Table 1. Internet AS Graph and Network Models' Simulation Results

	AS Graph	***FKP***	***BT***	***EBT***	***CD***	***IP***	***FN***
Avg. Node Degree $\langle k \rangle$	4.85	2	4.6	5.58	4	4	2
Max Node Degree	2498	2500	3900	3800	3500	3300	2900
Power-law Max Degree	4546	5298	2646	1550	2894	1005	7582
Frequent Exponent γ	2.14	2.3	2.61	2.78	2.586	2.9	2.21
Mean Local Clustering	0.29	0	0.6	0.42	0.59	0.33	0
Rank Exponent	0.8	0.76	0.67	0.7	0.67	0.69	0.763
CCDF Exponent	1.21	2.18	2.53	2.75	2.354	2.92	1.68
Characteristic Path Length	3.686	9.75	7	7.3	7.33	7.97	7.44
Avg. neighbor degree	0.0029	0.0005	0.001	0.001	0.002	0.0007	0.0006
Max neighbor degree	0.004	0.001	0.0015	0.002	0.002	0.001	0.001
Max degree ratio	0.14	0.003	0.006	0.008	0.004	0.004	0.0036

Concerning the average neighbour degree, it has been noticed that networks produced by the models presented here give much lower maximum degrees than those of real systems.

The IP model resembles most accurately the real AS graph with respect to the mean local clustering, while the CD and BT models exhibit the largest deviations.

Our models achieve values of the exponents that are quite realistic. Most scale-free networks are known to have *CCDF* and frequent exponent values that differ by exactly one. This property is not true for the proposed optimization-based models.

The values of characteristic path length obtained for the models presented here are substantially larger than the AS graphs. But the new models, when compared to the FKP model, are much more accurate. Moreover, the *FN* model achieves a value of the frequent exponent that is very close to the one of the AS graph. Also the CD model achieves by far the best value of the frequent exponent when compared with the other two-edged models.

The *CD* model has been tested with various walk path lengths. The results (not shown in the table for brevity) indicate that by varying the path lengths we can get almost any frequent exponent γ in the interval [2, 3], and also variable values of clustering (some results are shown in Fig 2 and Fig. 4).

In the Fertile Node (*FN*) model the number of leaves is about 44% of the total number of nodes, while in the (Internet) AS-graph the number of leaves is 26% approximately. This, however, is a substantial improvement over the original FKP model, in which the number of leaves is about 87%.

Concluding, the new models are in reasonably good agreement with the real Internet in regard to these metrics. In particular, both Enhanced *EBT* and *EIP* models with $p = 0.3$ have the right average degree, while the latter has a mean local clustering that is very close to the real measurement. The exponent of the Controlled Distance (CD) model is the best match (that is, the lowest) of the three. However, the best exponent is that of the Fertile Node (*FN*) model.

5. Conclusions and Future Work

The FKP model is a primitive way of taking into account the complex decision making involved in the evolution of the Internet; its power law behavior can be shown analytically, and its power law exponent is, empirically, quite accurate. On the negative side, the FKP model generates trees with no clustering and too many leaves. The extensions of the FKP model introduced and evaluated in this paper occupy a very interesting point in the tradeoff, being quite simple and intuitive, while at the same time the producing graphs are rather Internet-like.

Furthermore, the experimental methodology (generic and flexible topology generator and standardized metric calculation) used in this paper may be useful in future work. The models proposed are simple enough that some of the crucial metrics may be possible to calculate analytically. We have some preliminary results on the exponent of the fertile node model [17], but much work needs to be done on this front.

Finally, it would be interesting to incorporate in our models *deletions* of nodes and edges, certainly an element of the real Internet which, to our knowledge, has not been considered in the literature.

References

[1] Albert Réka and Albert-László Barabási. Statistical Mechanics of Complex Networks. *Rev. Mod. Phys. 74*, pp. 47, 2002.

[2] Alvarez-Hamelin J. I. and N. Schabanel. An Internet Graph Model Based on Trade-off Optimization. *European Physical Journal B*, Vol. 38 (2), pp. 231 – 237, April 2004.

[3] Barabási Albert-László and Réka Albert. Emergence of Scaling in Random Networks. *Science*, volume 286, pages 509–512, 1999.

[4] Berger N., B. Bollobas, Chr. Borgs, J. Chayes, and O. Riordan. Degree distribution of the FKP network model. In *Proc. ICLP*, pp. 725-738, LNCS 2719, 2003.

[5] Bonato Anthony. A Survey of Models of the Web Graph. In *Proc. of Combinatorial and Algorithmic Aspects of Networking*, 2004.

[6] Bornholdt R. L. and H. G. Schuster, Editors. *Handbook of Graphs and Networks*: *From the Genome to the Internet*. Wiley–VCH, Berlin, 2003. Chapter: Mathematical Results on Scale-Free Random Graphs. By Bollobas B. and O. Riordan.

[7] Brandes U. and T. Erlabach (Eds.): Network Analysis. *LNCS 3418*, Spinger-Verlag, Berlin 2005.

[8] CAIDA – The Cooperative Association for Internet Data Analysis. In http://www.caida.org/. BGP: http://bgp.potaroo.net/. Internet Data Catalog: http://imdc.datacat.org.

[9] Chen Q, H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger. The Origin of Power Laws in Internet topologies Revisited. *Proc. IEEE INFOCOM*, N.Y, pp. 608-617, 2002.

[10] Costa L. F. Da, G. Travieso, and P. R. Villas Boas. Characterization of Complex Networks: A Survey of Measurements. In *Adv. Phs.* 56(1), 2007; *http://arXiv:cond-mat/0505185* v3 30, June 2005.

[11] Dorogovtsev S. V. and J. F. F. Mendes. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, 2003.

[12] Fabrikant F., E. Koutsoupias, and C. Papadimitriou. Heuristically Optimized Trade-offs: A new paradigm for power laws in the Internet. In *Proceedings of the 29th ICLP*, 2002.

[13] Faloutsos M, P. Faloutsos, and C. Faloutsos. On Power Law Relationships of the Internet Topology. *Computer Commun. Rev. 29, 251- 262*, 1999. Also, *ACM/IEEE Trans. on Networking,* vol. 11, no. 4, pp. 514-524, 2003.

[14] Krioukov D., Fan Chung, Kc Claffy, M. Fomenko, A. Vespignani, and W. Willinger. *The Workshop on Internet Topology* (*WIT*) *Report*, 2006.

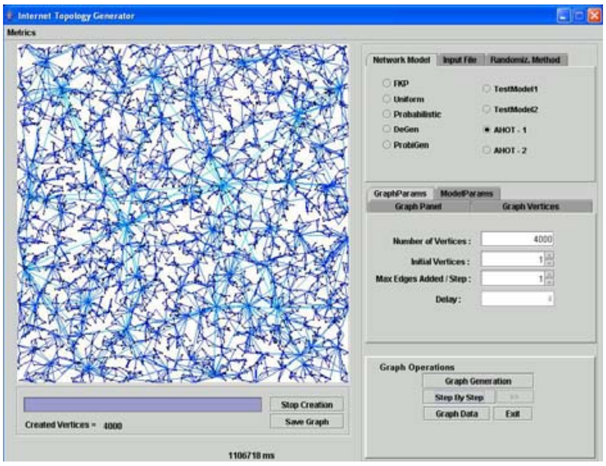
[15] Malarz K. and K. Kulakowski. Matrix Representation of Evolving Networks. *ACTA Physica Polonica B*, Vol. 36, pp. 2523-2536, 2005.

[16] Mitzenmacher M., G. Siganos, M. Faloutsos, P. Faloutsos, Chr. Faloutsos. A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Math.*, 1(2):226-251, 2004.

[17] Spatharis A., I. Foudalis, and M. Sideri. *Competition-Induced Tradeoff-based Modeling of the Internet*. Technical Report 2007.

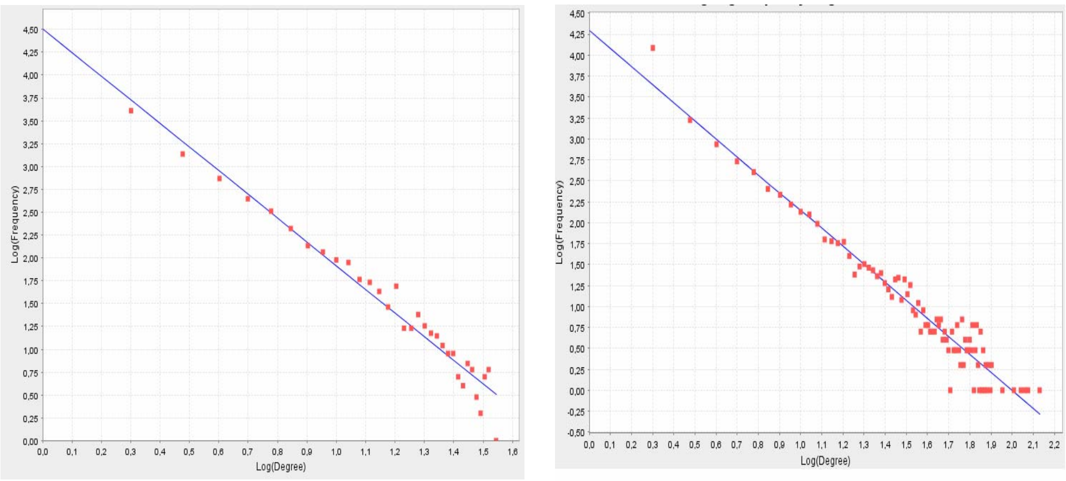
[18] Strogatz S. H., Exploring Complex Networks. *Nature 410*, pp.268 - 276, 2001.

APPENDIX A



Number of Vertices : 4000
Initial Vertices : 1
Max Edges Added / Step : 1
Delay : 4
Created Vertices = 4000
1106718 ms

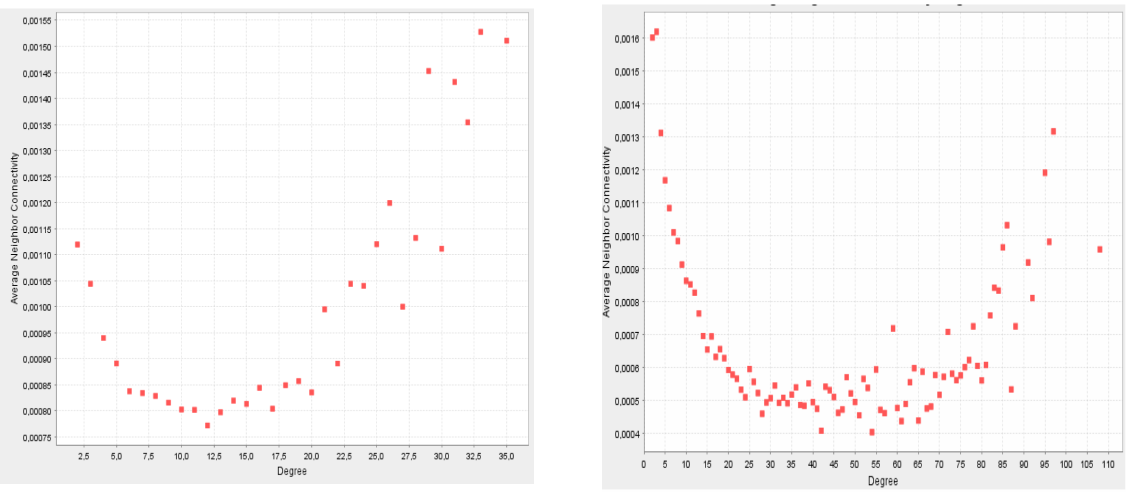
Fig.1. Visualization of a *CD* Graph generated for $\alpha = 90$ and $n = 8000$ using *TopGen* Program.



(a) CD Model

(b) Independent Path (*IP*) Model

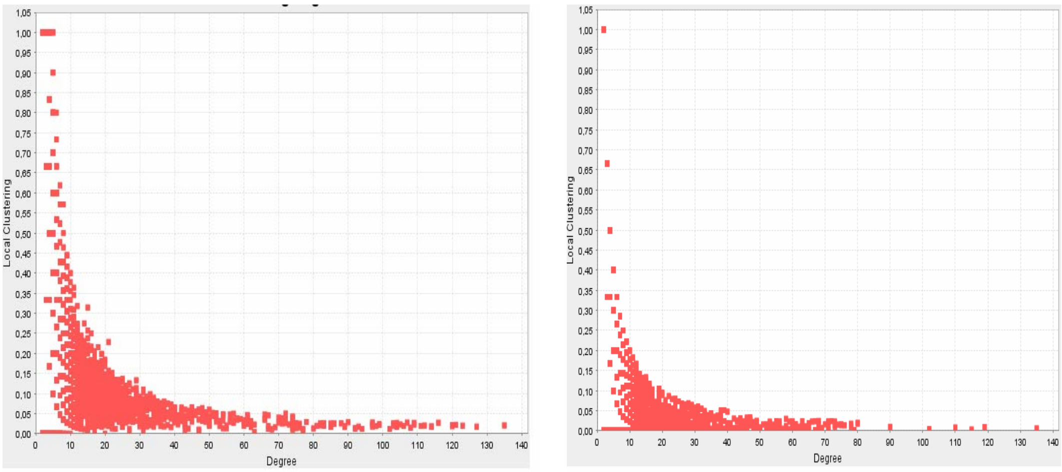
Fig. 2. Log-log Plot of Degree vs. Frequency. The solid line is obtained by least-square fitting.



(a) BestTwo Model

(b) CD Model

Fig. 3. Plot of Degree vs. Average Neighbor Connectivity



(a) Independent Path (*IP*) Model

(b) Enhanced Model

Fig. 4. Plot of Degree vs. Local Clustering