

A spectral analysis of the internet topology

Report**Author(s):**

Vukadinović, Danica; Huang, Polly; Erlebach, Thomas

Publication date:

2001-07

Permanent link:

<https://doi.org/10.3929/ethz-a-004256706>

Rights / license:

[In Copyright - Non-Commercial Use Permitted](#)

Originally published in:

TIK Report 118

A Spectral Analysis of the Internet Topology

Danica Vukadinović, Polly Huang, Thomas Erlebach

Abstract—In this paper we investigate properties of the Internet topology on the AS (autonomous system) level. Among techniques in spectral graph theory, we find the normalized Laplacian spectrum (*nls*) of AS graphs 1) unique in spite of the explosive growth of the Internet and 2) distinctive in setting AS graphs apart from synthetic ones. These properties suggest that *nls* is an excellent candidate as a concise fingerprint of Internet-like graphs.

Further analysis into the theory of *nls* leads us to a new structural classification of AS graphs with plausible interpretations in networking terms. Extensive analysis by AS-level data supports this claim. More importantly, along the way, new power-law relationships are unveiled, giving rise to a hybrid model encompassing both structural and power-law properties. We think that these new insights may have a profound impact on future protocol evaluation and design.

Keywords—Internet topology, network modeling, Laplacian eigenvalues, power-laws, AS domain connectivity.

I. INTRODUCTION

SIGNIFICANT research efforts have recently been invested in the analysis of the Internet topology. The current Internet is the result of rapid, distributed growth without controlled planning by a central authority. Therefore, its topology is not the product of a deliberate engineering attempt aimed at obtaining the best global solution possible, but rather reflects in great parts the choices and decisions made by individual organizations whose subnetworks form the Internet. As a consequence, the characteristics of the Internet topology can only be investigated by analyzing the available data about the current connectivity of routers or autonomous systems or snapshots of that connectivity taken at an earlier time.

Gaining additional knowledge about the properties of the Internet topology is important for several reasons. First, it can lead to an improved understanding of the Internet, e.g., behavior in the presence of overload or link failures. Then, it allows new algorithms, protocols, and applications to be designed and tuned so as to make the best possible use of the available topology. In particular, optimization problems related to resource allocation, call admission control and routing, that are provably difficult to solve for general topologies, might allow efficient solutions for a class of networks containing the real Inter-

net. Furthermore, topology characteristics have evolved in a stable manner over the last few years which allows more accurate predictions about what the Internet might look like a few years in the future. Besides, if certain aspects of the topology are found to be detrimental to the efficient operation of the Internet, this knowledge might encourage network providers to implement the required topology changes. Finally, a good understanding of the Internet topology can lead to improvements in network topology generators in order to generate “Internet-like” networks of various sizes for simulations. Network simulations with realistic topologies can again help to design, tune and evaluate new algorithms and mechanisms. It is well recognized [1] that the lack of an appropriate model has made it difficult to analyze and simulate new proposals in this field.

A. Existing models and generators

Progress in topology generation has been made from simple regular models such as “tree” or “mesh” to more sophisticated ones using randomness and hierarchy. Parameterized generators of more realistic topologies are now available, but they tend to emphasize either the hierarchical or the statistical aspect, thus missing out on the other.

The first wide-spread random generation model is due to Waxman [2]. It was created to fit topological properties of the Internet at the end of the eighties for the purpose of comparing minimum Steiner tree algorithms (for multi-point connections). In this model, a fixed set of nodes is placed in a square in the plane uniformly at random. The probability of adding a link between two nodes u and v is given by $P(u, v) = \alpha \exp(-\frac{d}{\beta L})$, where $0 < \alpha, \beta \leq 1$ are parameters, d is the Euclidean distance from u to v , and L is the maximum distance between any two nodes. The model has undergone a number of changes and extensions.

In 1997, Calvert et al. [3] proposed new eclectic models, Transit-stub and Tiers, combining the good characteristics of regular and random models and accentuating the existing hierarchy of the Internet. In the Transit-stub model, domains are classified into stub domains (domains through which traffic is routed only if its source or destination is within that domain) and transit domains (domains that route also traffic whose source and destination are not inside that domain). Subgraphs are repeatedly generated randomly according to the desired edge count, and the unconnected examples are discarded. In the end, extra-edges are added from randomly selected stub domains to tran-

sit nodes and from stub to stub domains respecting given parameters. This model is part of the Georgia Tech Internetwork Topology Models (GT-ITM) generator. In the Tiers model there are three levels of hierarchy: WAN, MAN and LAN. Connected subgraphs are produced by creating a minimum spanning tree in a single level. Redundant edges in and between levels are added respecting Euclidean distance (closer first). Zegura et al. [4] made a comparative study of the most popular models at that time and pointed out that some applications are very sensitive to the assumed topology.

A major new insight into properties of the real Internet topology was gained by Faloutsos et al. [5]. They found four power-laws¹ that appear to hold for various relations between popular graph metrics in the Internet: (*P1*) node degree vs. node rank; (*P2*) degree frequency vs. degree; (*P3*) the number of nodes within a certain number of hops vs. number of hops; (*P4*) 20 largest eigenvalues of the adjacency matrix vs. their ranks. The conjecture that these power-laws are invariants of the Internet topology has had a big impact on the research dealing with generators of realistic Internet topologies.

Medina et al. [6] have run experiments to check the existence of power-laws in different topologies and to explore their possible causes; their results showed that (*P3*) and (*P4*) are found also in grid graphs and Waxman topologies, with different parameters. They have given two reasons for the first two power-laws, *incremental growth* and *preferential connectivity*, and used these principles for the construction of a new topology generator called BRITE. In BRITE, nodes are distributed in a coarse grid either uniformly or using a bounded Pareto distribution to decide the number of nodes in each big square. Then the respective number of nodes are placed randomly in smaller squares avoiding collision. A small, randomly connected core is first generated and then other nodes are added incrementally together with their links. Links can be added using Waxman's probability function or using preferential connectivity. When preferential connectivity is selected, the destination of a link from the new node to a node in the set C of existing nodes is chosen to be node i with probability $\frac{d_i}{\sum_{j \in C} d_j}$, where d_i is the degree of node i . In addition, a hybrid probability can be used, i.e., $\frac{w_i d_i}{\sum_{j \in C} w_j d_j}$, where w_i is the link probability in Waxman's model.

Jin et al. [7] proposed a model called Inet. For a given number of nodes and percentage of nodes with degree 1, (*P1*) and (*P2*) are used to determine the degree distribu-

tion of the resulting graph. A spanning tree using only nodes with degree at least two is created. The degree 1 nodes are then attached to the tree with proportional probability. The remaining "free degrees" are connected randomly starting from the node with the largest degree. As a consequence of this generation procedure, Inet graphs resemble the Internet topology more than those produced by other models, at least considering power-law properties.

B. Topology data

The Internet topology is usually represented as a graph. On the router level, individual hosts and routers are the nodes of the graph, and physical connections between hosts are the edges. It is difficult to obtain accurate topology data of the Internet on the router level.

On a more abstract level, one can investigate the AS-level topology of the Internet. Here, each node of the graph represents an autonomous system (AS, see RFC 1930), i.e., a subnetwork under separate administrative control. An edge between two nodes means that the two AS domains appear consecutively on some BGP path, indicating that there is a direct connection between the two AS domains. Reasonably accurate data about the AS-level topology of the Internet can be inferred from BGP routing tables and is available on the NLNR website. Therefore, we focus our research on the analysis of the AS-level topology. For our analysis, we used snapshots of the AS topology from November 8, 1997 to March 16, 2001 taken roughly every 3 months. We treat the graphs as simple, undirected graphs, i.e., we remove parallel links.

We note that an alternative approach to determining the AS-level topology using router-level path traces was recently proposed in [8]. They report that their method resulted in an AS-graph covering 62% of existing ASs (found in BGP routing tables) and about 61% of the AS-level connectivity of those ASs. On the other hand, their approach detects parallel links, discovers ASs that are hidden in BGP routing tables due to AS aggregation, and identifies AS border routers. Since we are interested in using real AS graphs with maximum coverage and do not analyze the additional details revealed by the approach in [8], we found it more appropriate to use the AS-level topologies obtained from BGP routing data as explained above.

In order to compare properties of the AS graphs with graphs produced by a state-of-the-art network topology generator, we selected Inet-2.1. For each of the AS graphs we generate an Inet graph with the same number of nodes. Inet-2.1 allows to specify the fraction of vertices with degree 1. We specified this fraction identical to the one measured for the corresponding AS graph. Nevertheless, the Inet-2.1 generator produces graphs with a small amount

¹A power-law holds between two properties y and x if y is roughly proportional to x^c for some constant exponent c . If (x, y) data pairs are plotted with both axes in logarithmic scale, the resulting points lie close to a straight line with slope c .

of parallel edges. We removed those parallel edges since we deal specifically with simple, undirected graphs. As an effect, the fraction of nodes with degree 1 in these normalized Inet graphs was slightly higher than specified.

We choose to compare AS only with Inet graphs for two reasons. First of all, Inet explicitly models a number of power-law properties and is very successful in that. Before pursuing work further in this direction, we need to verify if it implicitly captures spectral properties as well. Secondly, it is difficult to generate comparable graphs using BRITE or GT-ITM. Current implementation of BRITE generates graphs with integer average degree, and those graphs with average degree greater than 2 do not have any leaves. GT-ITM requires several levels of structural information that are not easily accessible, for example distribution of transit-stub or stub-stub links.

C. Our contribution

We analyze Internet AS graphs from 1997 to 2001 using normalized Laplacian spectral techniques as a tool and identify new structural properties of the AS graphs that are not captured by the previously known power-laws. These properties are stable in the AS graphs from 1997 to 2001 in spite of the explosive growth of the number of nodes and edges. We give intuitive explanations of these properties and present ideas how new and improved network generators could possibly be built based on these properties.

The outline of the paper is as follows: In Section II we give the definitions and basic properties of the normalized Laplacian spectrum of a graph. We derive a lower bound on the multiplicity of eigenvalue 1 that turns out to be close to the real value on the AS graphs. The quantities used in the computation of this bound form the basis of the new structural model of AS graphs that we explain and interpret in Section III. In Section IV, statistics and comparisons based on the structural model are presented, and a hybrid graph generation model is proposed. Finally, in Section V, we summarize our results and discuss future work.

II. NORMALIZED LAPLACIAN EIGENVALUES

Previous studies in the context of network models have considered the largest eigenvalues of the adjacency matrix of a graph, but it was noted in [6] that these eigenvalues seem to satisfy a power-law relationship for many different topologies so that they are not very useful in distinguishing graphs. Therefore, we propose to look not only at the largest eigenvalues, but at the (multi-)set of all eigenvalues, called the *spectrum*. In addition, we do not use the standard adjacency matrix, but the normalized Laplacian of the graph [9]. Among other reasons, this has the advantage that all eigenvalues are contained in the interval $[0, 2]$

so that it becomes easy to compare the spectra of different graphs even if the graphs have very different sizes.

Let $G = (V, E)$ be an undirected, simple graph, where V is the set of vertices and E is the set of edges. Let $\|V\| = n$, $\|E\| = m$, and d_v be the degree of node v .

Definition 1: The normalized Laplacian of the graph G is the matrix $\mathcal{L}(G)$ defined as follows:

$$\mathcal{L}(G)(u, v) = \begin{cases} 1 & \text{if } u = v \text{ and } d_v \neq 0, \\ -\frac{1}{\sqrt{d_u d_v}} & \text{if } u \text{ and } v \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

Note that if A is the adjacency matrix of the graph G (where $a_{ij} = 1$ if there is an edge between v_i and v_j , and $a_{ij} = 0$ otherwise) and D is a diagonal matrix having $d_{ii} = d_{v_i}$, then $\mathcal{L}(G) = D^{-\frac{1}{2}}(D - A)D^{\frac{1}{2}}$. The normalized Laplacian spectrum (*nls*) is the set of eigenvalues of $\mathcal{L}(G)$, i.e., all values λ such that $\mathcal{L}(G)u = \lambda u$ for some $u \in \mathbb{R}^n$, $u \neq 0$.

The spectrum of a graph does not determine the graph uniquely, i.e., there are non-isomorphic graphs with the same spectrum. However, it determines the number of vertices, edges and, as we will see, some other topological properties. A first natural question is: What is the normalized Laplacian spectrum for simple topologies such as stars, chains, grids and random trees of n nodes? The *nls* of a star S_n is 0, 1 (n-2 times), 2, and the *nls* of a chain P_n is $1 - \cos(\frac{\pi k}{n-1})$, $k = 0, \dots, n-1$. For grids and trees, plots of the numerically computed spectrum are shown in Fig. 1. To generate our *nls* plots, we compute all n eigenvalues with MATLAB, sort them in non-decreasing order, and plot them so that the i -th smallest eigenvalue λ_i , $1 \leq i \leq n$, is drawn at (x, y) with $x = (i-1)/(n-1)$ and $y = \lambda_i$. In this way, the plot is always within $[0, 1] \times [0, 2]$ and it becomes convenient to compare the *nls* of graphs with different numbers of nodes.

A. Properties of normalized Laplacian spectrum

One can see in [9] that the spectrum of $\mathcal{L}(G)$, for any G , is contained in the closed interval $[0, 2]$ and that the smallest eigenvalue is always 0. The multiplicity of eigenvalue 0 (i.e., how many of the n eigenvalues are 0) is equal to the number of connected components of G . The largest eigenvalue is less than or equal to 2, where equality holds for a bipartite connected component. The spectrum of a graph is the union of the spectra of the connected components.

Looking at AS connectivity maps of the Internet during the period from 1997 to 2001, we observed that in spite of an increasing number of nodes and edges, the normalized Laplacian spectrum of the graphs stays the same. A closer look at the *nls* of these graphs has shown a large multiplicity of the eigenvalue 1. This has motivated our interest in the distribution of the spectrum of $\mathcal{L}(G)$, particularly in

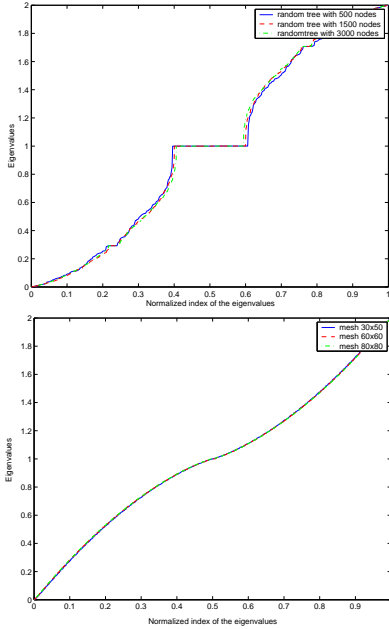


Fig. 1. The normalized Laplacian spectrum of grids and random trees. The trees were generated by starting with one node and then repeatedly adding a new node and making it adjacent to an existing node chosen uniformly at random.

the development of a lower bound for the multiplicity of eigenvalue 1 in terms of structural properties of graphs.

B. Multiplicity of the eigenvalue 1

We will use a technique similar to [10] to find a lower bound for the multiplicity of eigenvalue 1. Denote by $P(G) = \{v \in V \mid d_v = 1\}$ the set of leaves in G , called “**pendants**”, and by $Q(G) = \{v \in V \mid \exists w, (v, w) \in E, w \in P\}$ the set of the neighbors of the leaves, called “**quasi-pendants**”. Let $R(G) = V \setminus (Q(G) \cup P(G))$ be the set of nodes that are not leaves and that are not neighbors of leaves, called “**inners**”. Let p, q, r respectively be the cardinalities of the sets $P(G)$, $Q(G)$, and $R(G)$. We call the subgraph of G induced by $R(G)$ $Inner(G)$. By inn we denote the number of isolated vertices in $Inner(G)$. Let $m_G(1)$ denote the multiplicity of the eigenvalue 1. Then we obtain the following lower bound.

Theorem 1: The multiplicity of eigenvalue 1 of the normalized Laplacian is bounded from below by the sum of the number of pendants, the number of isolated inner nodes, and the negative of the number of quasi-pendants.

$$m_G(1) \geq p - q + inn \quad (1)$$

Proof: We can assume the following labeling of the nodes, because the eigenvalues are independent of the labeling: v_1, \dots, v_n where $v_1, \dots, v_r \in R(G)$, $v_{r+1}, \dots, v_{r+q} \in Q(G)$, and $v_{r+q+1}, \dots, v_n \in P(G)$. Also, we can assume that $(v_{r+i}, v_{r+q+i}) \in E, i = 1..q$.

Then, the structure of the normalized Laplacian is

$$\mathcal{L}(G) = \begin{pmatrix} R & rQ & 0 \\ rQ^T & Q & qP \\ 0 & qP^T & I_p \end{pmatrix}$$

Here R is an r -by- r matrix, rQ is r -by- q , Q is q -by- q , qP is q -by- p and I_p is the p -by- p identity matrix. From the basic equations $\lambda u = \mathcal{L}(G)u$, we obtain that $m_G(1) = \text{nullity}(\mathcal{L}(G) - I_n)$, where I_n is the n -by- n identity matrix.

Using the labeling assumptions, we observe that qP contains a principal submatrix D_q which is diagonal, having $-\frac{1}{\sqrt{d_{v_{r+i}}}}$, $i = 1..q$ on the diagonal. Now let $LI(G) = \mathcal{L}(G) - I_n$. Using D_q and elementary transformations that do not change the rank (adding a multiple of one row to another row, or the same for columns), we obtain a new matrix $LI'(G)$ from $LI(G)$:

$$LI'(G) = \begin{pmatrix} R - I_r & 0 & 0 & 0 \\ 0 & 0 & D_q & 0 \\ 0 & D_q^T & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Now it is enough to prove that $\text{nullity}(LI'(G)) = n - \text{rank}(LI'(G)) \geq p - q + inn$. We have that $\text{rank}(D_q) = q$, thus $n - \text{rank}(LI'(G)) \geq n - 2q - \text{rank}(R - I_r) = p - q + r - \text{rank}(R - I_r)$. Now, if inn is the number of isolated vertices in $Inner(G)$, each row that contains an isolated vertex will have 0 at the first r columns of $LI'(G)$, thus $\text{rank}(R - I_r) \leq r - inn$ and the statement follows. ■

Remark 1: The bound can be improved as follows:

$$m_G(1) \geq p - q + inn + s \quad (2)$$

where $s = \sum_{i=1}^l k_i - 1$, where k_i is the number of the mutually non-adjacent nodes with the same neighborhood in $Inner(G)$ and there are l such classes of nodes. In the matrix $(R - I_r)$ those nodes will have exactly the same entries, which will additionally decrease the rank.

III. A NEW STRUCTURAL MODEL

In this section, we present our new structural view of the Internet topology that is motivated by our observations concerning the normalized Laplacian spectra of AS and Inet graphs. Subsequently, we interpret the structural model in networking terms and provide evidences supporting the interpretations.

A. Theoretical derivation

Recall that we consider the Internet topology at the AS-level, so that we have an undirected simple graph

$G = (V, E)$ where vertices (nodes) are AS domains and an edge $\{u, v\} \in E$ means that the two AS domains u and v are connected. We found remarkably similar plots of the nls for all real Internet AS-level snapshots from November 1997 to March 2001. The same consistency was detected for Inet graphs with different numbers of nodes, but the nls of Inet graphs was clearly different from the nls of real AS graphs (see Fig. 2). Together with the facts and plots about the nls of grids and random trees presented in the previous section, this shows that the nls can be used as a kind of “fingerprint” for network topologies (even for arbitrary large graphs that are difficult to compare directly).

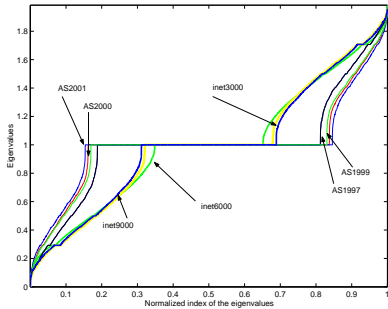


Fig. 2. The normalized Laplacian spectrum of selected graphs.

In particular, we found a significant difference between the multiplicity of the eigenvalue 1 in the nls of the real AS graphs and the Inet graphs with the corresponding number of nodes. Theorem 1 gives a lower bound on the multiplicity of eigenvalue 1 in terms of pendants, quasi-pendants, and isolated inner nodes. We found that this lower bound is close to the real multiplicity observed in the AS graphs and Inet graphs. Therefore, we classify the nodes of the graphs into sets P , Q , R and I as follows, and investigate their cardinalities. A node is

- . in P if its degree is 1 (i.e., if it is a leaf),
- . in Q if it has at least one P neighbor in G ,
- . in I if it is an isolated node in $Inner(G)$, the subgraph induced by nodes not being in P or Q , and
- . in R if it is contained in a connected component of $Inner(G)$ with at least 2 nodes.

B. Physical interpretation

The classes P , Q , R and I are defined in graph-theoretic terms motivated by the analysis of the Laplacian spectrum. To relate these notions to ASs in the real Internet, we now propose plausible interpretations of the four node sets and give some evidences obtained from AS name tables.

B.1 Q nodes, best-connected nodes of the Internet.

The class Q contains only a small number of nodes compared to the size of the whole graph and to the sizes of

the other classes, but the best-connected nodes (largest degree) belong to Q . The subgraph induced by Q nodes has a similar structure for all observed graphs: it contains a big connected component with a characteristic nls (see Fig. 3) and about 5% of isolated nodes.

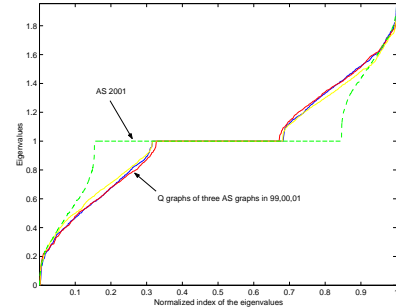


Fig. 3. The distinctive nls of Q subgraphs.

We interpret the nodes in the big connected component as core nodes (see TABLE I). Note that Q nodes have leaf neighbors by definition. The isolated nodes, which have no Q neighbors, can be explained as exchange points serving to connect P , R , and I nodes (see TABLE II).

TABLE I

EXAMPLE Q NODES AS PART OF THE CORE

AS number	AS name
209	ASN-ASN-QWEST Qwest
701	ASN-ALTERNET UUNET Technologies, Inc.
1239	ASN-ICM-INRIA Sprint
2548	ASN-ICIX-MD-AS Business Internet, Inc.
2914	ASN-VERIO Verio, Inc.
3257	Nacamar Global ASN
3356	ASN-LEVEL3 Level 3 Communications, LLC
3549	ASN-GBLX Global Crossing
3561	ASN-CWUSA Cable & Wireless USA
6453	ASN-TEGLOBE-AS Teleglobe Inc.
7018	ASN-ATT-INTERNET4 AT&T

TABLE II

EXAMPLE Q NODES AS EXCHANGE POINTS

AS number	AS name
8961	Emirates Internet Exchange MCI Peer router
7656	TWNAP-ADMIN-AS-AP Taiwan Internet Exchange Center

B.2 R nodes, core and alliances.

The subgraph induced by R consists of a larger number of connected components. Their size and frequency exhibit power-law relationships (see Section IV). The biggest connected component dominates by its size and nodes degrees. We interpret the connected components of R nodes, with the exception of the biggest one, as *AS alliances*. Alliances can be built on national, regional, commercial, or other grounds. TABLE III shows an example of an Austrian AS alliance, a connected component of R

with 7 nodes in the AS graph of March 16, 2001. The connections within that component are depicted in Fig. 4.

TABLE III
AN EXAMPLE NATIONAL R ALLIANCE

AS number	AS name
5385	TELEPORT Consulting und Systemmanagement GmbH
8447	Telekom Austria Autonomous System
8562	Datakom Informationsservice GmbH
8670	UTIC AS=20
9023	Oesterreichische Lotterien GmbH
12762	VTG
15824	DATAKOM-MEHRWERTDIENSTE

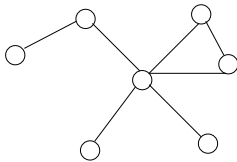


Fig. 4. Interconnection of Austrian AS alliance.

When examining the structure of the biggest R component, we concluded that a consistent interpretation of that component is not possible, because significant fluctuations in the members of that component are present during the observed three-year period (see Fig. 5).

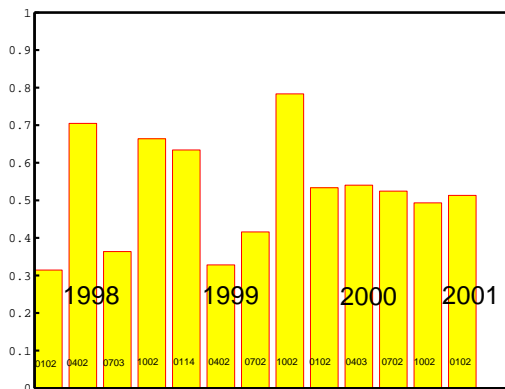


Fig. 5. The percentage of the nodes in the biggest R component that are still in the biggest R component in the direct successor AS graph.

In order to gain further insights, we investigated the k -cores of the AS graphs. The k -core of a graph is defined to be the subgraph obtained by recursively deleting all nodes with degree less than k . Intuitively, the deeper cores of an AS-graph (i.e., the k -cores for larger values of k , say $k \geq 5$) should correspond roughly to more well-connected “backbones” of the Internet at that time. We found that most of the nodes in the deeper cores are in Q and some are members of the biggest R component. This fact motivates an interpretation of the biggest component of R as being made up partly of AS domains belonging to the core

and partly of multi-homed stubs or alliances. Note that core nodes in R do not have any leaf neighbors. Examples of these two node types in R are given in TABLE IV and V. Nodes that were in the biggest component of R in all observed graphs are shown in TABLE VI.

TABLE IV
EXAMPLE STUB R NODES IN THE BIGGEST COMPONENT

AS number	AS name
159	ASN-SONNET The Ohio State University
15896	Yahoo!DE

TABLE V
EXAMPLE CORE R NODES IN THE BIGGEST COMPONENT

AS number	AS name
2551	ASN-ICG ICG NetAhead, Inc
3220	kpnQwest Sweden
3333	RIPE NCC
3914	ASN-DX-NET DataXchange Network, Inc
5056	ASN-INS-NET-2 Iowa Network Services
5413	XO Communications European
5459	LINX-AS
6427	ASN-LIGHTNING Lightning Internet Services
8933	The TEN-155 IP Service
12179	ASNBLK-INTERNAP-2BLK InterNAP Network Services

TABLE VI
EXAMPLE PERSISTENT R NODES IN THE BIGGEST COMPONENT

AS number	AS name
559	ASN-SWITCH-AS SWITCH Teleinformatics Services
3333	RIPE NCC
5496	Wirehub! Internet global backbone

B.3 P and I nodes, stub domains.

P nodes are leaves (nodes with degree 1) by definition. Therefore, they must be stub nodes (nodes that do not forward traffic that neither originates in that node nor is destined for that node). I nodes, whose degree is small in most cases (i.e., they have just two or three neighbors in Q), are mostly multi-homed stub domains. The percentage of I nodes is increasing in the AS graphs over time, and the percentage of P nodes is decreasing. Currently, the I class is the biggest part of the Internet. It became bigger than the P class at the beginning of 2001. This positive trend in the number of I nodes and negative trend in the number of P nodes agrees with the fact that more and more leaf domains want to become multi-homed for better fault-tolerance [11].

The number of I nodes with high degree is rather small: for example, in the AS graph of March 16, 2001, there are only 28 nodes with degree greater or equal to 6, and the maximum degree of an I node is 11. These nodes are

mainly big companies or universities with multiple connections to the backbone, but not providing forwarding services. Examples of I nodes are shown in TABLE VII.

TABLE VII
EXAMPLE I NODES

AS number	AS name
33	ASN-DEC Digital Equipment Corporation
103	ASN-NWU Northwestern University
109	ASN-CISCO Cisco Systems, Inc.
151	ASN-IND-NTC Hewlett-Packard Company
1889	HP-EUROPE-AS Hewlett-Packard Company
3608	ASN-NBNIC-NCA National Computerization Agency
4983	ASN-INTEL-SC Intel Corporation
5779	ASN-Y-BR-SERV Yahoo! Broadcast Services, Inc.
6200	AS-UIC University of Illinois at Chicago
6201	ATEXT Architext
6245	ASN-INTERNIC InterNIC Registration Services
6716	NetKonec Communications USA
8153	ASN-NORTELCH2 Nortel (Northern Telecom)
10627	ASN-CRITICALPATH Critical Path, Inc.
13756	ASN-INFLOW-NET Inflow Inc.

In summary, the classes P and I represent the outskirts of the Internet, and in a sense they correspond to the stub domains in the Transit-Stub model. Although the smaller alliances in R are arguably also the outskirts of the Internet, for the sake of simplicity and since some nodes in the biggest R component are found in the inner part of the AS graphs, we refer to, for the detailed analysis below, Q and R as the core, and P and I as the edge of the Internet.

IV. ANALYSIS

Having unveiled a sound theory as a succinct way to identify Internet topologies and proposed a structural model with plausible explanation, we look into how the Internet evolves in this structural sense and how it compares to graphs generated by existing tools. Through our analysis, we find both structural and power-law properties important and in two counts observe a power-law in the internal structure of R and the inter-connectivity between P and Q . These give rise to a hybrid generator that encompasses both statistical and structural properties.

A. Methodology

The set of graphs undergone analysis has been described earlier in Section I-B. To observe how AS graphs evolve structurally and how they compare to Inet graphs, we look at the following three sets of metrics.

1. Ratio of nodes in P , I , Q , R
2. Ratio of links connecting PQ , IQ , QQ , RQ , RR
3. Degree of connectivity in core (Q , R) and edge (P , I)

It is not to our surprise that the numbers of nodes in each component and the numbers of links inter-connecting the components are increasing. Thus we focus our analysis on how the components expand or shrink relatively to the whole graph and omit discussions on absolute numbers.

Results are depicted in Fig. 6, 7, and 8 respectively. Each plot shows changes of the AS graphs in the corresponding metric (Y axis) in time (X axis). In the next two subsections, we highlight important trends in the evolution of AS graphs and significant differences to Inet graphs.

B. Evolution of AS graphs

Observation 1. We see from the top plots of Fig. 6 that ratio of nodes in P is decreasing while that of I is increasing. That means, applying our interpretation of P and I , the area of single-homed leaf nodes is shrinking while that of multi-homed stub nodes is rapidly increasing. This agrees with a big ISP's observation [11] on increasing demand for multi-home access from its clients.

Observation 2. The ratio of nodes in P and I combined has increased from approximately 67% to 74%. Q and R components, containing the core of the Internet, each remains stable and the combined ratio decreases from 33% to 26% (bottom plots of Fig. 6). In terms of ratio of nodes, the edge of the Internet is growing faster than the core.

Observation 3. Given that the ratio of nodes in P is decreasing, it is not surprising to see that the ratio of links interconnecting QP is decreasing and similarly for QI being increasing as I increases (top plots of Fig. 7). QP and QI combined increases from 54 to 60%. QQ , QR , and RR , each remains relatively stable and the combined ratio decreases from 46 to 40%. This once again shows that in terms of the ratio of links, the edge of the Internet is growing faster than the core.

Observation 4. More interestingly, the ratio of links in Q and R decreases by 6%, which is slightly lower than the 7% decrease in ratio of nodes. The ratio of links outgrowing the ratio of nodes in the core indicates that connectivity of the core (Q and R) is increasing. This can be confirmed by the middle plot of Fig. 8.

All four observations either agree with practical experience or can be linked to one another. These lead us to believe that we have a plausible interpretation of P , Q , R , and I . That makes abstract graph theory terms easier to understand and more accessible for network engineers.

C. Comparison to Inet graphs

Similarity. AS and Inet graphs are similar in size of P and size of PQ (top-left plot of Fig. 6 and 7). This is because Inet specifically models the ratio of degree 1 nodes which are in essence P nodes.

Difference 1. While we see the ratios of I and QI expand in AS graphs, they remain stable in Inet graphs (top-right plot of Fig. 6 and 7). This corresponds to the left plot of Fig. 8, where we observe connectivity of the edge components (P and I) in AS graphs increasing whereas that of

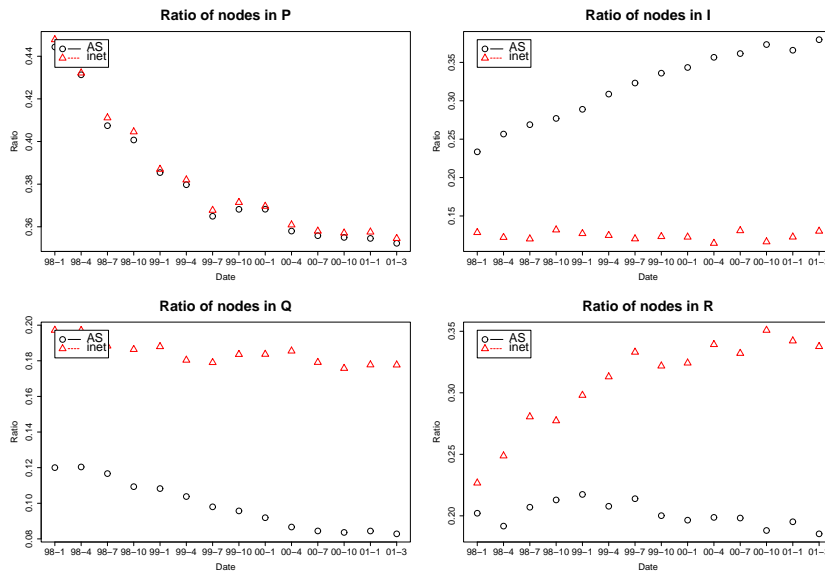


Fig. 6. Ratio of nodes in P , I (top), Q , R (bottom)

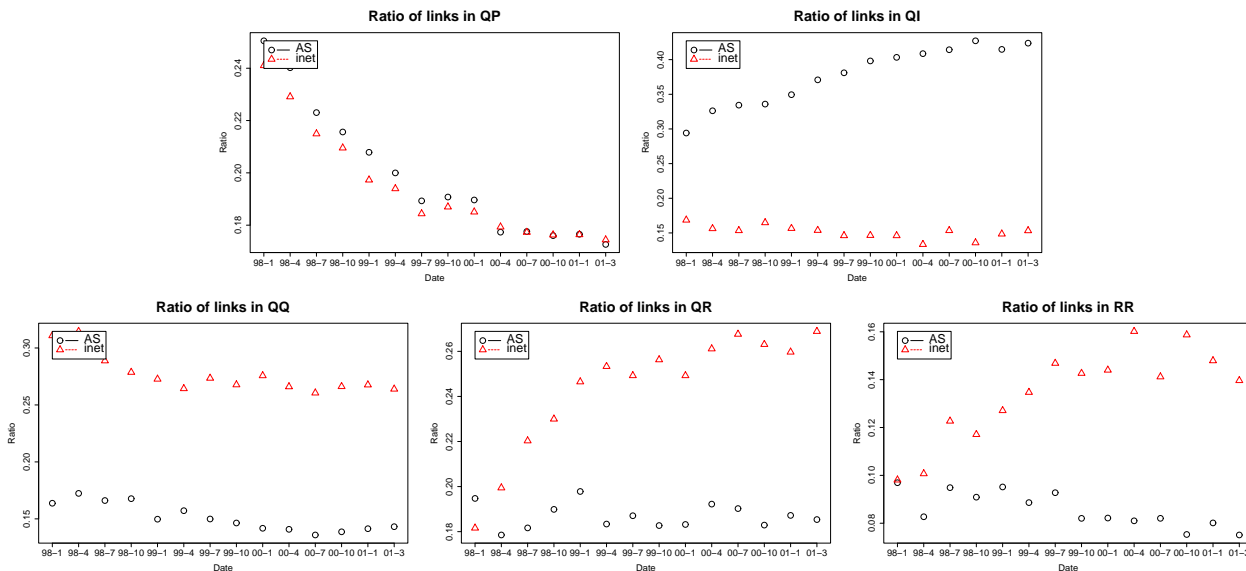


Fig. 7. Ratio of links in QP , QI (top), QQ , QR , RR (bottom)

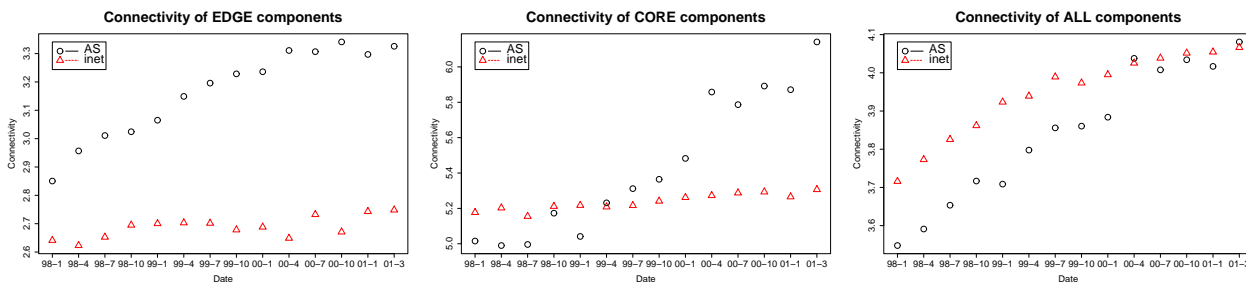


Fig. 8. Connectivity of edge, core, and all

Inet graphs remains stable.

Difference 2. The ratio of R remains stable in AS graphs while it expands significantly in Inet graphs (bottom-right plot of Fig. 6). Similar contrasts can be found in link statistics in Fig. 7: QR (bottom-middle) and RR (bottom-right).

Difference 3. Inet graphs, similar to AS graphs, have stable Q and QQ ratios. However, the ratio level is higher than that in AS graphs (bottom-left plot of Fig. 6 and 7).

Difference 2 and 3, contradicting to how AS graphs evolve, indicate that the core of Inet graphs is not only larger but also expanding while the edge losing its ground.

Difference 4. It appears that in the core of Inet graphs the ratio of links is growing just as fast as the ratio of nodes; thus resulting in a rather constant degree of connectivity (middle plot of Fig. 8).

These differences show that although successful in modeling the power-law properties, Inet fails to capture structural changes in AS graphs. In particular, the core of the Internet is becoming relatively smaller and the edge larger, but the evolution of Inet graphs shows the exact opposite. More interestingly, the Internet is structured so that the core and edge each has increasing connectivity. On the contrary, connectivity of the core and edge of Inet graphs each remains constant.

D. Towards a hybrid model

These structural differences are potentially critical when studying properties of routing protocols, for example quality of path, route convergence, and fault tolerance. To be more concrete, one can expect global effectiveness of an alternative route computation algorithm to be different when evaluated using graphs with a higher ratio of multi-homed stub nodes (I) and better connected core (Q and R). Below we outline a structural model that will allow us to verify this conjecture in the future.

Our premise is to model AS graphs encompassing both statistical and structural properties. By that, we mean to 1) generate a degree sequence that obeys power-law properties observed earlier, 2) form P , I , Q , and R clouds based on the size dynamics, and finally 3) inter-/intra-connect these clouds with analysis on their connectivity.

First results show that a power-law exists not only in the overall degree sequence but also within the structure. For instance, while investigating ways to form the R cloud, we find that the sizes of the connected components in R have a power-law distribution. Fig. 9 shows the sizes of the connected components in R ranked from the largest to the smallest (rank-size plot), and occurrences of different sizes (size-occurrence plot). The linear relationship of the two plots in log-log scale reveals a power-law property in

the internal structure of R .

A power-law is also present in the way P and Q clouds are inter-connected. The left plot in Fig. 10 shows, for each node in Q , its rank on node degree (Y axis) and rank on number of degree 1 or P neighbors (X axis). Those Q nodes with one degree-one neighbor are ranked the 411th (lowest) in X axis; those with 2, 3, 4, and 5 degree 1 neighbors are ranked 275th, 191st, 147th, and 109th, gradually improving. Each column in the plot shows that a variety of Q nodes, with different node degrees, may have the same amount of P neighbors. The middle plot of Fig. 10 illustrates the distribution of such Q nodes. The linear relationship in the log-log scale plot hints on a power-law distribution for Q nodes having one P neighbor.

This power-law property persists across different degree 1 neighbor ranks, until data points are too few to show a clear linear relationship in log-log plots. The right plot of Fig. 10 is another example of such a power-law existing for Q nodes having five P neighbors. Further analysis verifies that the same power-law property between P and Q clouds exists in other snapshots of the Internet.

These power-law properties within or between sub-structures of the AS graphs could be the missing links between the state-of-the-art generators, i.e., Inet and GT-ITM. Each of them identifies one important aspect of the Internet topology — Inet for the statistical aspect and GT-ITM for the structural aspect — but unfortunately misses out on the other. We think that a realistic model cannot do without either. Our analysis, although primitive, gives rise to a hybrid model that will lead to a promising AS graph generator. This might in turn have a profound impact on how routing mechanisms are evaluated and how Internet routing or other protocols evolve in the future.

V. CONCLUSIONS AND FUTURE WORK

In this section, we first summarize our work and explain the significance of our contribution. Then we give an outlook on interesting open questions for future research.

A. Contribution

We have applied normalized Laplacian spectrum analysis to Internet topology graphs. It turned out that the nls can be used as a concise fingerprint of a graph. Real AS graphs from 1997 to 2001 produced nearly identical nls plots in spite of the significant difference in the number of nodes (from 3015 in 1997 to 10515 in March 2001). Similarly, graphs generated with the Inet-2.1 generator had a characteristic nls independent of the number of nodes, which was different from the nls of real AS graphs, in particular with respect to the multiplicity of eigenvalue 1.

We adapted a lower bound on the multiplicity of eigen-

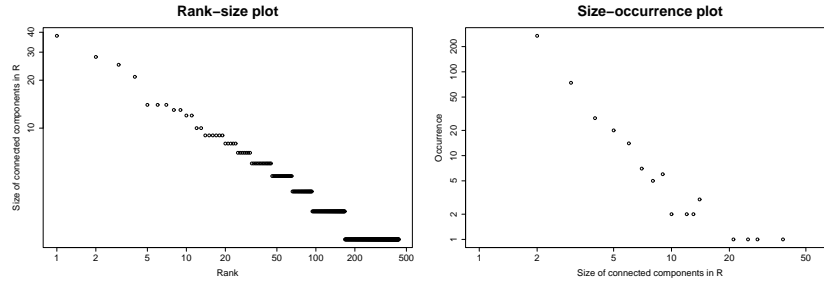


Fig. 9. Power-law in sizes of connected components in R

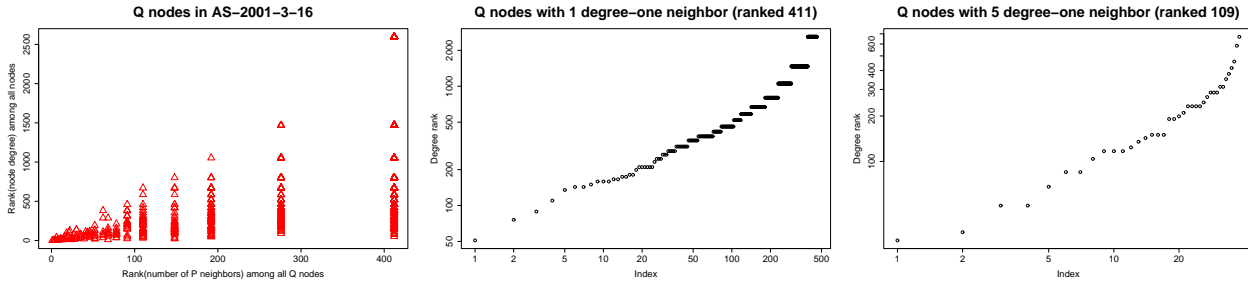


Fig. 10. Power-law in PQ inter-connection

value 1 in terms of the cardinalities of different node sets P , Q , R , and I to the normalized Laplacian of a graph. For the real AS graphs, we found plausible interpretations for the nodes in P , Q , R , and I . In particular, the classification of nodes into these four types provides a structured view of the graph, featuring leaf domains, multi-homed stub domains, alliances, and core nodes.

Our extensive numerical analysis of the sizes of P , Q , R and I and their connectivity characteristics in real AS graphs over time and in corresponding Inet graphs provides quantitative data that is useful for several purposes:

- It allows to observe trends and thus make better predictions about the future appearance of the Internet topology on the AS-level.
- It helps us to identify characteristics of graphs generated by topology generators that do not match the characteristics of real AS graphs sufficiently well.
- It provides insight into structural characteristics of the current Internet with quantifiable properties, as opposed to abstract, qualitative views of the hierarchy in the Internet.
- It allows us to improve existing network generators (like Inet) as well as to design new network generators based on the structural properties we observe in real AS graphs.

Besides, while analyzing the subgraphs induced by P , Q , R and I and their interconnections we obtained new, previously unobserved power-law relationships with respect to the sizes of connected components in R and to the degree rank of Q nodes with the same number of P neighbors.

In particular, we found out that Inet graphs, which seem to be the “best match” with real AS graphs among all topology generators available, consistently display a sig-

nificant difference in the nls with respect to the multiplicity of eigenvalue 1. This indicates that Inet graphs lack some structural properties of AS graphs, related to the way how stub nodes are connected to core nodes. Therefore, our results suggest that Internet topology models and topology generators must take into account structural properties (as in the transit-stub model) as well as power-law relationships (as in the Inet generator) in order to model real AS graphs well. We have already begun to experiment with Inet-2.1 modifications that improve the similarity of generated graphs with real AS graphs, and first results are encouraging.

Note that our structural explanation of P , Q , R and I nodes is roughly related to the hierarchical classification of nodes in the transit-stub model [3], to the degree-based classification in [12], and to the classification into core, subcore, regional, and stub nodes in [8]. In [12], AS nodes are classified into national or international backbones (degree ≥ 28), large regional providers (10-27), large MAN providers (4-9), and multi-campus corporate or academic networks (1-3). This classification is becoming outdated by the commercialization of the Internet over the past few years. We find, for example in recent snapshots of the Internet, that the best-connected AS nodes are often the large regional providers (TABLE I). We also note that, while [8] emphasizes methodology for topology discovery, we are interested in the evolution of the Internet topology in a structural sense, which leads to a more realistic model for random topology generation. In addition to the structural explanations, we obtain extensive data that provides a much more precise view of the structure of AS

graphs and quantitative relationships between the various substructures. While the transit-stub model lacks the ability to generate the power-law relationships observed in real AS-graphs, our results allows us to combine structure and power-laws into a single model.

Our work on the analysis of AS-level topology graphs complements research carried out at various research labs worldwide, e.g. [5], [4], [6], [13]. In particular, we think that our proposal to consider the normalized Laplacian spectrum as a “topology fingerprint” will be useful in detecting the similarity and/or differences of graphs obtained from various sources, even beyond the domain of communication networks. Note that some previous researchers who considered the eigenvalues of topology graphs have looked only at the 20 largest eigenvalues of the adjacency matrix [5]. Contrary to that approach, we study the whole spectrum instead of only the 20 largest values, and we use the normalized Laplacian of a graph instead of the standard adjacency matrix (motivated by results from spectral graph theory [9]). We think that the normalized Laplacian spectrum should become one of the standard metrics used in the comparison of network topology graphs, in addition to standard graph parameters like diameter, average degree, and average path length, power-law relationships as those observed in [5], and routing-related metrics like expansion, resilience and distortion [13].

B. Future Work

Our results provide several immediate starting points for future work. First of all, we intend to use the insights we obtained from our analysis for improving the quality of the Inet-2.1 generator (with respect to the similarity to real AS graphs) and to explore the potential of a newly designed random topology generator based on our structural view in terms of P , Q , R and I nodes.

Besides that, many aspects of the AS graphs are yet to be investigated thoroughly, for example the structure of the connections within Q and those between Q and R . On the first look, Q is in some sense a smaller version of the original graph. It can be further classified into P , Q , R and I , and the process can be continued recursively. This can be interpreted as a form of self-similarity in AS graphs. However, the quantitative measures are different in different scales so that the self-similarity applies only to the structural view of the graph, not to the exact parameters.

Furthermore, it will be interesting to identify additional characteristics of the nls (in addition to the multiplicity of eigenvalue 1) that can be interpreted in terms of graph theory or networking concepts.

Another open question is to explain how the local behavior of network administrators (who must determine to

which other ASs their network should be connected) leads to the global properties observed in the AS graphs, in particular to the various power-laws. First steps along this direction have been made in [6], where it was observed that incremental growth and preferential connectivity can explain the power-laws to some extent. If the effects of local behavior on global properties of the resulting network are understood, these might on the one hand lead to more natural network generators (that simulate the local behavior of the nodes while generating the network) and on the other hand help to determine changes in local behavior that might result in improved global structure of the network.

Finally, it is important to investigate the differences with respect to the *behavior* of a communication network (in terms of performance metrics such as throughput, delay, fault-tolerance) that arise from differences observed in the corresponding topology graphs. One means for this line of research are network simulations. Thorough case studies could help to identify which of the graph properties have substantial effects on network performance and which are only of theoretical interest and do not affect performance.

REFERENCES

- [1] V. Paxson and F. Sally, “Why we don’t know how to simulate the internet,” in *Winter Simulation Conference*, December 1997.
- [2] B. M. Waxman, “Routing of multipoint connections,” *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 9, pp. 1617–1622, 1988.
- [3] K. Calvert, M. Doar, and E. W. Zegura, “Modeling internet topology,” *IEEE Communications Magazine*, June 1997.
- [4] E. W. Zegura, K. L. Calvert, and M. J. Donahoo, “A quantitative comparison of graph-based models for Internet topology,” *IEEE/ACM Trans. Networkings*, vol. 5, no. 6, pp. 770–783, 1997.
- [5] M. Faloutsos, P. Faloutsos, and C. Faloutsos, “On power-law relationships of the Internet topology,” in *SIGCOMM’99*, 1999.
- [6] A. Medina, I. Matta, and J. Byers, “On the origin of power laws in Internet topologies,” *ACM Computer Communications Review*, vol. 30, no. 2, 2000, April 2000.
- [7] C. Jin, Q. Chen, and S. Jamin, “Inet topology generator,” Technical Report CSE-TR-433-00, EECS Department, University of Michigan, 2000.
- [8] H. Chang, S. Jamin, and W. Willinger, “Inferring AS-level Internet topology from router-level path traces,” in *Proceeding of SPIE ITCOM 2001*, Denver, CO, August 2001.
- [9] F. R. K. Chung, *Spectral Graph Theory*, Conference Board of the Mathematical Sciences, Providence, Rhode Island, 1997.
- [10] R. Grone, R. Merris, and V.S. Sunder, “The Laplacian spectrum of a graph,” *SIAM Journal on Matrix Analysis and Applications*, vol. 11, no. 2, pp. 218–238, April 1990.
- [11] C. Diot, “Sprint tier 1 IP backbone: architecture, traffic characteristics, and some measurement results,” 2001, <http://talk.epfl.ch/talkdetail.php?talkId=42>.
- [12] R. Govindan and A. Reddy, “An analysis of internet inter-domain topology and route stability,” in *Infocom’97*, 1997.
- [13] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, “Network topologies, power laws, and hierarchy,” Tech. Rep. 01-746, Computer Science Department, University of Southern California, June 2001.