Spectral Analysis of Internet Topologies

Christos Gkantsidis, Milena Mihail, and Ellen Zegura
College of Computing
Georgia Institute of Technology, Atlanta, Georgia 30332
Email: {gantsich, mihail, ewz}@cc.gatech.edu

***Abstract*— We perform spectral analysis of the Internet topology at the AS level, by adapting the standard spectral filtering method of examining the eigenvectors corresponding to the largest eigenvalues of matrices related to the adjacency matrix of the topology. We observe that the method suggests clusters of ASes with natural semantic proximity, such as geography or business interests. We examine how these clustering properties vary in the core and in the edge of the network, as well as across geographic areas, over time, and between real and synthetic data. We observe that these clustering properties may be suggestive of traffic patterns and thus have direct impact on the link stress of the network. Finally, we use the weights of the eigenvector corresponding to the first eigenvalue to obtain an alternative hierarchical ranking of the ASes.**

I. Introduction

Studying and modeling network topologies is necessary for protocol performance evaluation and simulation of a variety of network problems. Early modeling efforts focused around random graphs with relatively regular degree distributions [6], [37], [40], [41]. With the rapid growth of the network and the persistent effort of network measurement [13], [14], [34], real topology data started becoming available, in particular at the AS (Autonomous System) level. Using such data Faloutsos et al. first observed that the degree distribution of the AS level topology is actually consistently highly skewed [11]. Consequently, the research community has shown considerable interest in obtaining topology models that better resemble the real data [2], [5], [18], [23], as well as understanding the impact of such network topologies on the performance of network protocols [27], [32].

This new generation of synthetic Internet topology models is strongly driven by the observed skewed statistics of the degree sequence and its evolution, and by even further observations of more detailed graph theoretic characteristics of the network. Most notably, following the natural intuition that, for example, geography must be relevant in the real Internet topology, [5] paid special attention to the "clustering" coefficient; the observation of the significance of geography has been also made in [39] and [21].

In this paper we revisit the issue of clustering. As opposed to previous work that has focused on the clustering coefficient, our starting point is the method of *spectral filtering*. This method examines the large eigenvalues of matrices related to the adjacency matrix, and looks for clusters in the eigenvectors associated with these eigenvalues. Indeed, the first reference to the large eigenvalues of the adjacency matrix of the AS Internet topology is the "eigenvalue power-law" which was reported together with the "degree power-law" in [11]. The connection between spectral filtering and graph connectivity, including clustering, has been extensively studied in discrete mathematics (e.g. see the books of [9], [29] and the further references that they point to), and has found very successful applications in information retrieval and data-mining where clusters represent groups of data with semantic proximity [3], [17], [20], [26], [28]. Practical experience suggests that spectral analysis might be better suited for data that lack regularity (thus it has been extensively used in computer science), while clustering coefficients are better suited for data that have stronger regularities (thus they have been extensively used by physicists who study lattices, crystals, etc.). Indeed, by definition, spectral filtering yields a large number of clusters, and it can be applied iteratively in subgraphs of a network. By contrast, it is not clear how to grow clusters around nodes with large clustering coefficient and this approach is not typical in information retrieval or data-mining[1].

Our contributions include:

- The observation that the eigenvectors related to the largest eigenvalues of the adjacency matrix of the AS topology examined in [11] do not express interesting clusters. This is an experimental validation of the result of [24] who showed that the eigenvalues power law is a consequence of the degree power law. We thus conclude that further normalizations are needed to retrieve non-trivial clustering properties.
- Adaptation of the spectral filtering method in the context of the AS Internet topology, by (a)performing inverse frequency normalization via stochastic matrices, (b)considering similarity transformations and (c)considering the entire topology as well as subgraphs of the topology. As a result, we get non-trivial groupings of ASes with clear semantic proximities, such as geography and business interests. We note that without this adaptation, i.e., by considering the eigenvectors corresponding to the eigenvalues of the adjacency matrix as in [11], we get trivial groupings corresponding to the large ISPs and their customers: this is indeed a restatement of the highest degrees.

The first and second authors were funded by NSF ITR-0220343; the third author was funded by NSF ANI-0081557. This work was also funded by a Georgia Tech Edenfield Faculty Fellowship.

[1] Though a related approach called "k-means" is quite common; but we do not expand further on it, since we do not use it in this work.

- The observation that the clustering properties (a)vary in the core and the edge of the network and across geographic areas, (b)persist over time, and (c)are not accurately matched by synthetic Internet topology generators, though the Power Law Random Graph (PLRG) model comes close [2].
- Study of the connection between the information retrieved by spectral filtering and link stress (link stress can be thought of as a first approach towards congestion). In particular, we argue that the eigenvectors associated with the largest eigenvalues are suggestive of non-trivial intracluster traffic patterns that cause significant decrease in the link stress. The decrease is much more notable in the Internet than in any synthetic topology. If on the other hand the traffic patterns become intercluster the link stress correspondingly increases. This reasoning is in line with [7], [10] which suggest that network characteristics should be studied in the context of the design problem they are trying to solve.
- A method to define intracluster and intercluster "traffic" patterns. These are patterns that deviate from uniform treatment of all pairs of nodes, and may represent "good" and "bad" test case for network performance.
- A detailed and efficient AS ranking method according to the first eigenvector of a suitably defined stochastic matrix, which has strong correlation with other known hierarchical assignments [31]. This approach is an adaptation of the pagerank used by Google [25]. An adaptation of the same method for ranking links between ASes, found that rankings are highly correlated with link stress under uniform traffic. A further adaptation of the method to obtain groups of ASes that correspond to seemingly highly stressed cuts.

The balance of the paper is as follows: In Section II we cover necessary primitives from linear algebra and highlight the intuition behind the spectral filtering method. We also introduce normalizations and similarity transformations, and discuss their suitability and necessity for graphs with skewed statistics, like the Internet topology. In Section III we describe the spectral filtering results for the AS Internet topology, and give the qualitative nature of the information retrieved by the eigenvectors. In Section IV we give an application of the information retrieved by the eigenvectors in terms of defining non-trivial traffic patterns that deviate from uniform traffic. In Section V we give a method of ranking ASes and links between ASes that is highly correlated with hierarchical assignments. We summarize in Section VI.

II. Spectral Analysis of Matrices arising from Graphs

In this Section we give a high level overview of the intuition and the primitives of spectral filtering. We discuss the basics of eigenvalues and eigenvectors of matrices, some useful transformations and normalizations, and why the eigenvectors corresponding to the large eigenvalues contain information relevant to clustering. This motivates the processing that we

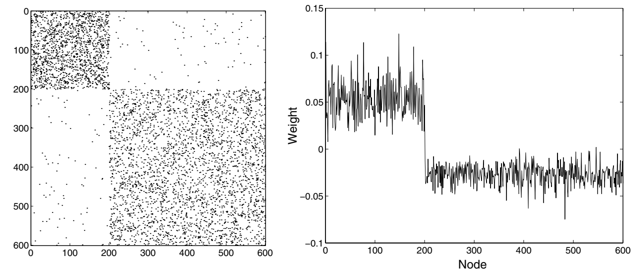

Fig. 1. The adjacency matrix (left) of a random graph on 600 vertices. There is a dot in position (i, j) iff there is a link between i and j. The first and second diagonal blocks correspond to subgraphs with high connectivity. Off-diagonal blocks represent sparse edges between the subgraphs. The second eigenvector (right) assigns positive weights to the nodes of the first block and negative weights to the nodes of the second block.

will do to the eigenvectors of the AS Internet topology in Section III. We also give a plausible explanation of the eigenvalue power-law of [11] as a restatement of the Zipf with exponent 1 rank-degree distribution; this serves as additional motivation for the processing in Section III, in the sense that without this processing the spectral method does not give non-trivial information.

A. The Spectral Filtering Method

Let $G(V, E)$, $|V| = n$, be an undirected graph and let A be its adjacency matrix: $a_{ij} = 1$ if $(i, j) \in E$, $a_{ij} = 0$ otherwise. Since G is undirected, A is symmetric $a_{ij} = a_{ji}$. In general, the (i, j)-th entry of a symmetric matrix can be thought of as a measure of the correlation between parameters i and j. Let $\vec{e}$ be an n-dimension real vector; $\vec{e}$ can be thought of as a function on the vertices of G. We say that $\vec{e}$ is an *eigenvector* of A with *eigenvalue* λ if and only if $\vec{e}A = \lambda\vec{e}$. It is a well known fact of linear algebra that every $n \times n$ real symmetric matrix A has a *spectrum* of n orthonormal eigenvectors $\vec{e_1}, \vec{e_2}, \ldots, \vec{e_n}$ with real eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ [16], [38]. The eigenvectors are unique up to degeneracies related to equal eigenvalues. In general, the spectral filtering method can be applied with any matrix with real spectrum.

We demonstrate the essence of the spectral method with an example. The left panel of Figure 1 gives the adjacency matrix of a symmetric graph. A dot in position (i, j) in this graph corresponds to a link between i and j. There are two highly connected clusters in this graph; the first includes nodes 1 through 200 and the second all the other nodes. The two clusters are connected with a few links. The right panel of Figure 1 plots the weights assigned by the eigenvector which corresponds to the second largest eigenvalue. The nodes belonging to the first cluster were assigned positive weights and the nodes of the second cluster negative weights. Thus, an efficient heuristic to separate the two clusters is to examine the eigenvector.

In broad lines, the spectral filtering method for an $n \times n$ symmetric matrix A proceeds as follows:

STEP 1: Compute the k largest eigenvalues of A together with the corresponding eigenvectors. The parameter k depends on

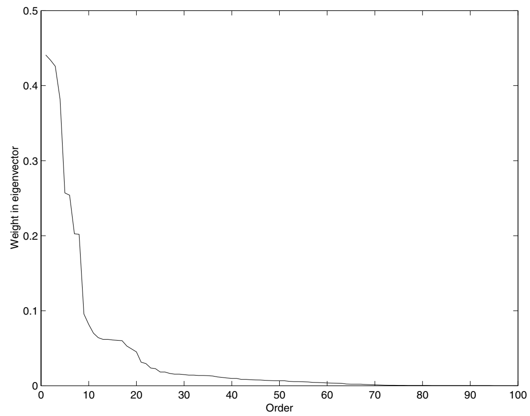

Fig. 2. Typical profile of the most positive weights assigned to nodes by the eigenvector corresponding to a large eigenvalue. This profile was taken from a principal eigenvector of the stochastic normalization of the AS topology.

the application and the instance, but it is always one to two orders of magnitude smaller than n.
STEP 2: For each i, $1 \le i \le k$, let $\vec{e_i}$ be the eigenvector associated with λ_i. Sort the vertices according to the weight assigned by $\vec{e_i}$. A typical profile of the sorted vertices is in Figure 2. Cut towards the most positive end (or towards the most negative end), with special preference to sharp jumps, if they exist (a good example of a sharp jump can be found in Table II). These groups are candidates for clustering and/or semantic proximity.

In general, the eigenvectors corresponding to large eigenvalues tend to capture global characteristics of the graph and its semantics, such as groups of nodes $S \subset V$ for which the ratio

$$\frac{\text{edges inside } S}{\text{edges incident to } S} = \frac{|\{(i,j) \in E : i \in S, j \in S\}|}{|\{(i,j) \in E : i \in S, j \in V\}|} \quad (1)$$

is large, indicating *clusters* of relatively high connectivity and, thus, presumably further semantic proximity, not necessarily otherwise expressed in the data (the deep theory of "expander" graphs supporting this claim can be found, for example in [9], [29]). In addition, because there is no polynomial time algorithm to find a set S minimizing the above ratio, the spectral method is an efficient heuristic. Eigenvectors corresponding to small eigenvalues tend to capture noise, or local characteristics that are explicit or can be easily computed from the data.

B. Algebraic Primitives of Spectral Filtering

More formally, we list a few technical facts which build the intuition behind the spectral filtering method (the statements are straightforward, though some of the proofs to which we point are quite technical).
(a) The largest eigenvalue λ_1 of a d-regular graph is d and the corresponding eigenvector assigns uniform weights to all vertices [9], [22]. All other eigenvalues λ_i, $2 \le i \le n$ are small, $|\lambda_i| \le O(\sqrt{d})$, almost surely [9].
(b) The eigenvalues λ_i, $1 \le i \le n$, of a graph with m edges and maximum degree d are bounded by $|\lambda_i| \le \min\{\sqrt{m}, d\}$ [22].
(c) The spectrum of the union of vertex disjoint graphs is the union of their spectra [9], [22].
(d) If A and B are the adjacency matrices of not necessarily disjoint graphs with eigenvalues $\alpha_1 \ge \alpha_2 \ge \ldots \ge \alpha_n$ and $\beta_1 \ge \beta_2 \ge \ldots \ge \beta_n$, then the eigenvalues of their union $C = A+B$ are $\gamma_1 \ge \gamma_2 \ge \ldots \ge \gamma_n$ with $\alpha_i + \beta_n \le \gamma_i \le \alpha_i + \beta_1$, $1 \le i \le n$ [16], [38]. In addition, the corresponding invariant subspaces of C follow from the invariant subspaces of A perturbed by no more than the maximum invariant subspace of B [16], [30].

The intuition behind the spectral filtering method is that, if we take the union of two vertex disjoint regular random graphs A_1 and A_2 and connect them with a few random edges B, then, combining Facts (a) through (e) above, the spectrum of $C = A_1 + A_2 + B$ will have $\gamma_1 \simeq \gamma_2 \simeq d$ (corresponding to the largest eigenvalues of A_1 and A_2) and $\gamma_i \simeq O(\sqrt{d})$, $3 \le i \le n$. Furthermore, we expect to identify the vertices of A_1 and A_2 by examining the eigenvectors corresponding to the first two eigenvalues. See Figure 1. Indeed, the second eigenvector assigns mostly large negative weights on A_1 and mostly large positive weights on A_2.

C. Similarity Transformation $\mathrm{SIM}(A) = A \cdot A^T$

Now suppose that $G(V,E)$, $|V| = n$, is a directed graph, and thus the adjacency matrix A is no longer symmetric. A is no longer guaranteed to have a complete real spectrum, and the notion of clustering is not well defined either. Let A^T be the transpose of A, i.e., $a^T_{ij} = a_{ji}$. Notice that the product $A \cdot A^T$ is a symmetric matrix. Notice further that its (i,j)-th entry is $\sum_{k=1}^{n} a_{ik} a_{jk}$, measuring the number of nodes that i and j point to in common. In the case where the nodes represent ASes and edges are directed from customers to their providers, the above sum relates i and j to the number of their common providers. Similarly, the product $A^T \cdot A$ relates i and j to the number of their common customers. The transformation $A \cdot A^T$ is very common in spectral analysis. Depending on the application, it is called self-adjoint, co-citation, co-variance, or *similarity transformation*. Here we shall use the notation $\mathrm{SIM}(A) = A \cdot A^T$.

D. Stochastic Normalization

The intuition behind the spectral filtering method that we gave in the previous paragraphs referred to regular graphs. Indeed, in practice, the spectral filtering method has been found to deteriorate rapidly when the frequencies of non-zero entries vary substantially [17], which is certainly the case with the very skewed degrees of Internet topologies. Inverse frequency normalization is a general approach to restore spectral filtering in such cases.

In its simplest form, inverse frequency normalization divides each entry a_{ij} with the sum $\sum_j a_{ij}$ of the entries of the corresponding row, thus obtaining a matrix where all the rows add up to 1. Notice that this is now a stochastic matrix, in the sense that it describes the transition probabilities of a Markov chain in the natural way. If, in addition, we make all diagonal entries $a_{ii} = 1/2$ and multiply all other entries by $1/2$ the range of the eigenvalues shifts to $(0, 1)^2$. Like symmetric matrices, such stochastic matrices have a complete spectrum of real eigenvalues and eigenvectors. For any matrix A, we denote its stochastic normalization $\mathrm{N}(A)$. In what follows, we may apply the stochastic normalization to either A or $\mathrm{SIM}(A)$, thus getting $\mathrm{N}(A)$ or $\mathrm{N}(\mathrm{SIM}(A))$.

E. Faloutsos' Eigenvalue Power-Law

[11] examined the spectrum of the adjacency matrix of the AS Internet topology, without performing any normalization or other transformation. They reported a power-law on the twenty or so largest eigenvalues of this matrix with exponent between .45 and .5.

[24] observe that Faloutsos' eigenvalue power-law is a direct consequence of the degree sequence power-law along the lines of Facts (d) and (e) of Section II.A, in the following sense (see also Figure 3):

STEP 1: Decompose an undirected AS topology A as $A = F + E$, as follows. Initially F is the set of vertices that have the k highest degrees, and let $d_1, d_2, \ldots, d_k$ be these degrees. Initially F contains no edges. Let E be the entire AS topology graph. Now we will remove some edges of E and add them to F, so as to create k disjoint stars in F. We do this by the following process: For each vertex v that is not in F, if v is incident to k_v vertices in F, pick one of these vertices u with probability proportional to the degree of u in the entire graph, make the edge $\{v, u\}$ incident to the vertex $u \in F$ and remove the edge $\{v, u\}$ from E. Notice that F is now a set of vertex disjoint stars with degrees $d'_1, d'_2, \ldots, d'_k$, and E is the initial AS topology where all edges belonging to the stars have been removed.

STEP 2: Notice that the eigenvalues of a star of degree d are $\pm\sqrt{d}$ and 0 with multiplicity $d - 1$ [22]. Thus, by Fact (d) of Section II.A, the largest eigenvalues of F are $\sqrt{d'_1}, \sqrt{d'_2}, \ldots, \sqrt{d'_k}$. Also, by Fact (e) of Section II.A, the largest eigenvalues of $A = F + E$ cannot be perturbed by more than the largest eigenvalues of E.

STEP 3: For typical AS topologies, we have found experimentally that the above procedure, for $k = 100$, gives $d'_i \simeq d_i$, $1 \leq i \leq k$, hence the largest eigenvalues of F are close to $\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_k}$, and the largest eigenvalues of E are, in the worst case strictly smaller than $\sqrt{d_1}$ and on the average 1/5 of $\sqrt{d_1}$. Now by Fact (e) of Section II.A, the largest eigenvalues of $A = F + E$ can be understood to be close to $\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_k}$. Hence, for graphs where the largest degrees follow Zipf with exponent close to 1, as [11] reported for the AS Internet, the largest eigenvalues follow a power-law with exponent close to .5, also as [11] reported for AS the Internet. We also refer to [12], [19], [24] for formal analysis of these results in stochastic models of power-law random graphs.

[2] On the other hand, the eigenvectors of stochastic matrices are not necessarily orthogonal, and sometimes additional normalizations that rectify orthogonality are necessary for good results. In our analysis this did not turn out to be necessary. We also note that there are many further normalization methods, including so-called Laplacians and divisions by logarithmic or other functions of $\sum_j a_{ij}$, but, again, we did not use them in our analysis.

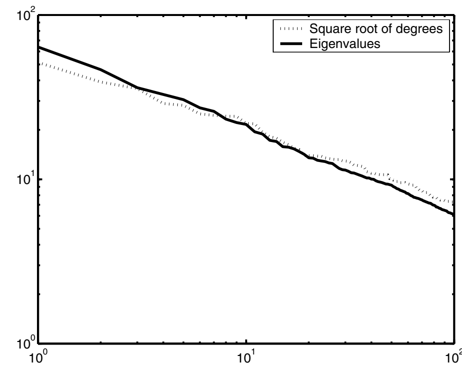

Fig. 3. We plot the 100 largest eigenvalues of the undirected AS topology and compare them to the square roots of the 100 largest degrees. The eigenvalue power-law follows the degree power-law. Both axes are in log scale.

We may now conclude that by looking at the eigenvectors corresponding to the largest eigenvalues examined in [11] we should not hope to get information beyond the ASes of highest degree and their customers. Indeed, in experiment, we have found these eigenvectors to be highly concentrated on the large ISPs. Therefore, to obtain more interesting clusters, we will need the processing discussed in Section II-C.

III. Spectral Analysis of AS Internet Topology

In this Section we describe the spectral analysis that we performed on AS Internet topologies. We discuss the used data, the processing, the behavior of large eigenvalues, and the resulting groups of ASes from the corresponding eigenvectors. We show that clustering varies in the core and the edge of the network, as well as across different geographic areas. On the other hand, the clustering is consistent over time. Finally, we compare the spectral characteristics of the real AS topologies to synthetic topologies.

A. Data Used, Transformations and Normalizations

We have used topology data from two sources. The first source is the data of [1] who collect BGP routing information from many routers in the Internet and combine all the routing tables to reconstruct the undirected AS topology. Using the heuristics in [31], they also provide the information whether an edge of the undirected topology corresponds to a customer-provider or a peering relationship. Finally, [31] give a heuristic to assign the ASes to the levels of a 5-level hierarchy. The most important ASes, such as big ISPs in the core of the Internet, are assigned to level 1. The smallest ASes are assigned to level 5. The topological data from this effort dating on April 6, 2002 are the ones used most in our study[3].

The second set of data is from [13]. Though this data is far less complete, it has the advantage that it spans the time period of 1997 to date. We have thus used this data to study the evolution of clustering over time. [13] does not contain

[3] We should note that perhaps the most complete set of data is in [8]. It was difficult to annotate these data with the AS hierarchy information of [1], and thus we did not use them.

information about the relationships between the ASes. We have used the algorithm of [15] to infer AS relationships[4].

The data in both [1] and [13] are not perfectly accurate. We do not believe though that this affects the results of our study, in the sense that missing links would quite likely strengthen the clustering findings.

An AS topology without AS relationships corresponds to an undirected graph with a symmetric adjacency matrix A, in the natural way. For such a topology we perform spectral analysis on the stochastic normalization $\mathrm{N}(A)$. An AS topology with customer-provider or peer relationships corresponds to a directed graph A', where a'_{ij}=1 and a'_{ji}=0 if and only if i is a customer of j and j is a provider of i, and $a'_{ij} = a'_{ji} = 1$ if and only if i and j are peers (in all other cases the entries are 0). For such a topology we perform spectral analysis on the stochastic normalization $\mathrm{N}(\mathrm{SIM}(A'))$.

If we perform spectral analysis starting from the entire undirected graph A or directed graph A' we find that the clusters indicated by the eigenvectors associated with the large eigenvalues correspond to groups of nodes assigned levels 3, 4 and 5 of the hierarchy of [31], thus are away from the core of the network. This is intuitive, since we expect the edge of the network to have more areas with higher connectivity inside the area and relatively lower connectivity to the rest of the network, along (1) of Section II. Similarly, we expect that the core of the network is better connected, and thus the ratio (1) of Section II is consistently higher in the core.

To capture the clustering properties of the core of the network we have to explicitly isolate the core from the edge and analyze the core alone. We have used two methods to isolate the core. When information about the AS hierarchy is available, such as in [1], we define the core to be the subgraph that contains only the ASes assigned to levels one through four. When the hierarchical information is not available, as in [13], we iteratively prune all the nodes in the graph that have degree one or two. The graph whose core we wish to find can be either directed or undirected. We denote the core as $\mathrm{Core}(A)$ and $\mathrm{Core}(A')$ depending on whether the original graph was undirected or directed respectively. As above, we perform spectral analysis to $\mathrm{N}(\mathrm{Core}(A))$ and $\mathrm{N}(\mathrm{SIM}(\mathrm{Core}(A')))$.

B. Results for the Entire AS Topology

Figure 4 shows the largest eigenvalues of the AS topology of [1]. We have considered the adjacency matrices of the topology with and without AS relationships, for both the entire network and the core. The point to notice in this graph is that the eigenvalues are quite high, indicating the existence of clusters in the underlying topology. Another interesting observation is the drop in the eigenvalues between the entire topology and the core of the network. This is expected because the core was constructed by removing small ISP's which tend to cluster more.

[4]In addition to customer-provider and peering, [15] includes sibling relationships; to be consistent with our first set of data, we replace sibling relationships with peering relationships.

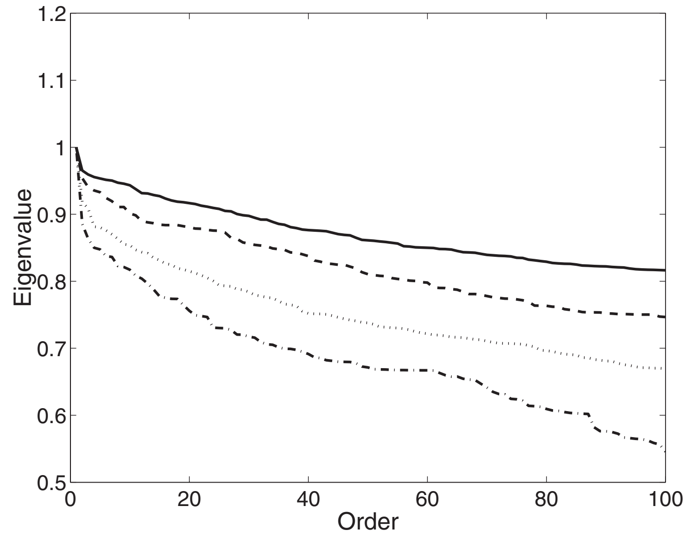

Fig. 4. The largest eigenvalues of the AS topology. The top line corresponds to the entire topology without AS relationships N(A). The second line corresponds to the entire topology with AS relationships N(SIM(A')). The third line corresponds to the core without AS relationships N(Core(A)). The bottom line corresponds to the core with AS relationships N(SIM(Core(A'))).

TABLE I

A SAMPLE OF A CLUSTER FOUND IN THE $N\,(\mathrm{Core}\,(A'))$ TOPOLOGY.

AS	Weight	Level	Description	Country
3257	0.1096	2	Tiscali Intl Network	DE
3303	0.1071	2	Swisscom Ltd	CH
293	0.1032	3	ESnet	US
5511	0.0986	1	France Telecom , Worldwide IP Backbone	FR
3549	0.0986	1	Globalcrossing	US
3582	0.0983	3	University of Oregon	US
4513	0.0972	1	Globix Corporation	US
6461	0.0967	3	Primary AS for Abovenet	US
1668	0.0917	2	AOL Transit Data Network	US
1299	0.0916	1	TeliaNet Global Network	SE
3356	0.0907	1	Level 3 Communications North America	US
701	0.0897	1	Alternet	US
3561	0.0896	1	Cable & Wireless (CW)	US
6395	0.0889	2	Broadwing Communications	US
8918	0.0884	2	Carrier1 Autonomous System	GB
4565	0.0869	2	Epoch Internet	US
1239	0.0867	1	SprintLink Backbone	US
6079	0.0866	3	RCN Backbone AS	US
6259	0.0862	3	Fiber Network Solutions, Inc.	US
2497	0.0852	2	IIJNET	JP
2914	0.0846	1	Verio	US
2828	0.0846	2	XO Communications, Inc.	US
2548	0.0842	2	DIGEX-AS	US
5459	0.0840	3	London Internet Exchange Ltd.	GB
5650	0.0840	2	Electric Lightwave, Inc.	US

Note: This cluster is taken using the eigenvector which corresponds to the highest eigenvalue. The ASes in this group are big ISP providers, mostly in North America and Europe. The weights of the eigenvector did not show a sharp jump.

Next we give some representative groups of nodes corresponding to the highest weights assigned by eigenvectors corresponding to large eigenvalues. The first example was taken using the $\mathrm{N}(\mathrm{SIM}(\mathrm{Core}(A')))$. The group corresponds to the largest eigenvalue, which is 1.0. In Table I, we list the members of the group that take the highest weights in the eigenvector.

In Table II, we give a group of ASes that belong to Chinese ISP providers. This was taken from the eigenvector of $\mathrm{N}(\mathrm{SIM}(\mathrm{Core}(A')))$ that corresponds to the 6th largest eigenvalue with value 0.8363. Notice that the clusters of relatively big ASes in Tables I and II (levels 1 through 3 of

TABLE II

A SAMPLE OF A CLUSTER FOUND IN THE $N\left(\text{Core}\left(A'\right)\right)$ TOPOLOGY

AS	Weight	Level	Description	Country
9810	0.5091	3	China Netcom	CN
9805	0.4463	3	SIEMENS LTD. CHINA	CN
9305	0.3141	4	Beijing Feihua Communication Technology Co,Ltd	CN
17969	0.3086	4	AS OF VLINE	CN
7467	0.2798	4	21VIANET(CHINA),INC	CN
17620	0.2622	3	China Netcom	CN
7549	0.1799	3	The North China regional network of CEInet	CN
4808	0.1743	4	Chinanet Beijing Site AS	CN
17431	0.1689	4	Beijing TONEK Information Technology Development Company	CN
9394	0.1553	3	CHINA RAILWAY Internet(CRNET)	CN
17622	0.1465	3	China Netcom	CN
9929	0.1076	3	China Netcom	CN
4799	0.0910	3	Golden Bridge Network of China	CN
10212	0.0524	3	Optic Communications Co., Ltd.	CN
4774	0.0510	3	Abone	JP
4813	0.0495	4	China Telecom GUANGDONG PROVINCE BACKBONE NETWORK	CN
4812	0.0495	4	China Telecom (Group) , Shanghai Telecom Company	CN
17444	0.0475	3	NWT IP Network	HK
7474	0.0451	2	Optus Communications	AU
11608	0.0426	2	Accretive Networks, Inc.	US
6993	0.0422	2	InterNAP	US
6939	0.0416	2	HE.net	US
4134	0.0407	2	Data Communications Bureau,MPT	CN
5650	0.0378	2	Electric Lightwave, Inc.	US
1668	0.0358	2	AOL Transit Data Network	US
4058	0.0058	3	LinkAGE Online Ltd.	HK

Note: This group was found in the eigenvector corresponding to the 6th largest eigenvalue. The last entry (AS 4058) does not belong to the group. We have included it to indicate a typical sharp jump suggestive of where to cut a group.

TABLE III

A SAMPLE OF A CLUSTER FOUND IN THE $N\left(\text{SIM}\left(\text{A}'\right)\right)$

AS	Weight	Level	Description	Country
15536	-0.2472	5	CEDEFOP	GR
15948	-0.2102	5	ICE/HT fundamental and technological research	GR
20813	-0.2102	5	Hellenic Open University	GR
6802	-0.1868	4	National Educational and Research Information Network	BG
3268	-0.1766	5	CYNET , Cyprus Academic Network , Cyprus	CY
13092	-0.1765	5	Univerzitet u Beogradu	YU
2546	-0.1707	5	ARIADNE NETWORK	GR
3323	-0.1707	5	National Technical University of Athens	GR
5470	-0.1707	5	Aristotle University of Thessaloniki	GR
5489	-0.1707	5	T.E.I. of Thessaloniki	GR
6744	-0.1707	5	Computer Technology Institute	GR
6867	-0.1707	5	University of Crete	GR
8248	-0.1707	5	Greek High-School Internet Network	GR
8253	-0.1707	5	Democritus University of Thrace Network	GR
8278	-0.1707	5	Technical University of Crete	GR
8617	-0.1707	5	University of the Aegean	GR
8618	-0.1707	5	Ionion University	GR
8643	-0.1707	5	ATHENAnet	GR
8700	-0.1707	5	T.E.I. OF LARISSA	GR
8762	-0.1707	5	T.E.I of Crete	GR

Note: The whole group contains several more ASes related mostly to academic institutions in Greece, Cyprus, and occasional ASes from other Balkan countries. There was a sharp jump (not indicated in the figure for lack of space) after which the entries were clearly outside the Balkans. This group was found in the eigenvector corresponding to the 2nd largest eigenvalue.

the hierarchy) appear in prominent positions when we examine the core of the topology. As we shall see below, such clusters do not appear when we examine the entire topology.

In Table III we give a group of ASes that belong to Greek academic institutions. This was taken from the eigenvector of $\text{N}(\text{SIM}(A'))$ that corresponds to the 2nd eigenvalue with value 0.9539. Notice that this cluster of rather small ASes (levels 4 and 5 of the hierarchy) appears in prominent position when we examine the entire topology.

We should note that the three examples presented here are typical. We chose to include the particular examples wanting to give one cluster from each continent.

C. Results specific to Geography

Is the Internet topology homogeneous across the entire globe? Do the same connectivity patterns apply everywhere? The first synthetic models of Internet topologies which emphasized the principle of preferential connectivity [18], [23] were implicitly making such homogeneity assumptions. Recently, these assumptions have been challenged, most notably in [21], [39] who show strong correlation between the placement of ASes and routers with geography as well as economic development. We second and strengthen these findings, by observing that different geographic parts of the network exhibit different connectivity patterns.

We have used the data of [33] to assign ASes to continents. We constructed three graphs for the continents of North America (NA), Europe (EU) and Asia (AS)[5]. We included AS relationships, thus obtaining non-symmetric adjacency matrices A'_{NA} for North America, A'_{EU} for Europe and A'_{AS} for Asia. In Figure 5 we give the largest eigenvalues of $\text{N}(\text{SIM}(A'_{\text{NA}}))$, $\text{N}(\text{SIM}(A'_{\text{EU}}))$ and $\text{N}(\text{SIM}(A'_{\text{AS}}))$. We also give the plots for the spectrum of the corresponding cores $\text{N}(\text{SIM}(\text{Core}(A'_{\text{NA}})))$, $\text{N}(\text{SIM}(\text{Core}(A'_{\text{EU}})))$ and $\text{N}(\text{SIM}(\text{Core}(A'_{\text{AS}})))$. The point to notice is that, both in the entire topology and in the core, North America exhibits less clustering than Europe and Asia. This can be understood intuitively by thinking of the network in North America as being at a later evolutionary stage, and hence is more connected.

D. Spectrum Consistency over Time

Is the spectral behavior of the Internet topology consistent over time? See Figure 6. We have used snapshots from [13] taken one year apart and found consistent behavior of the largest eigenvalues of $\text{N}(A)$. This confirms the intuitive belief that the spectrum is a robust characteristic of a topology. Figure 6 refers to the entire AS topology without AS relationships. We have observed similar behavior in the evolution of the AS topology with AS relationships, as well as the core of the topology, and when restricted to specific continents.

E. Synthetic topologies

In Figure 7 we give the largest eigenvalues of the AS Internet topology, as well as similar graphs generated by Inet [18], Waxman, growth with preferential connectivity according to Barabasi-Albert and the improved GLP heuristics [5], [23] which explicitly tries to capture better clustering (all the above for the same number of nodes as the Internet topology), and the power law random graph (PLRG) model of [2] (for the specific degree sequence of the Internet topology). We give the spectrum of both the entire AS topology and the core (recall that the core of synthetic topologies where there is no other indication of hierarchy is obtained by iterative pruning).

For the entire Internet topology, all synthetic generators, except for the Power Law Random Graph (PLRG) [2], have

[5]It is possible that some ASes are present in more than one continents. We treated such ASes as belonging to only one continent. However, their number is very small, and the results are not affected.

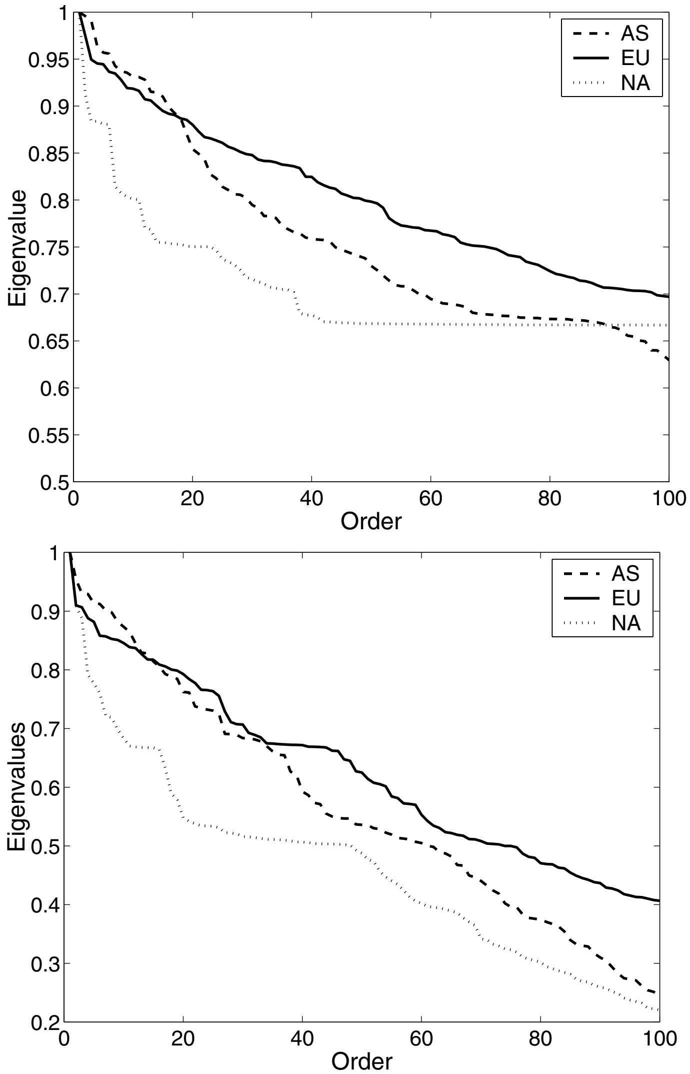

Fig. 5. The spectrum of different continents. The top graph is for the entire topology of each continent, while the bottom graph is for the core of the topology of each continent.

smaller eigenvalues. This means that they do not contain as strong clusters as the real Internet. This could have been expected since no synthetic generator attempts to capture such explicit notions as geography and business interests. But, why is PLRG an exception? Note that PLRG does not even generate a connected graph [2]. So, the same random principles that generate several isolated connected components in the entire graph, generate several badly connected subgraphs within the giant connected component.

For the core of the topologies, the WAXMAN and BA models produce higher eigenvalues. We believe that this is a pathological byproduct that these topology generators do not attempt to simulate any notion of core. Therefore, the behavior of the spectrum after pruning small degree vertices is the same as the entire topology.

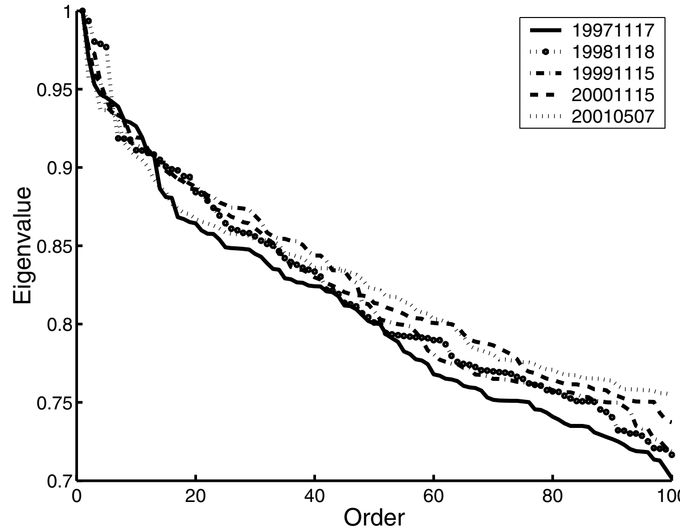

Fig. 6. The evolution of the largest eigenvalues of the AS topology. This data is from NLANR.

IV. Impact of Spectral Analysis on Performance and Traffic Primitives

What is the significance of the information retrieved by the spectral analysis of Section III? What is the significance of the eigenvectors associated with the large eigenvalues? The main difficulty in answering this question is in deciding which metric to pick and examine its correlation with clustering. In general, there is no consensus on the metrics by which Internet topologies should be evaluated. One approach is to include detailed graph properties [5], [18], [23], while another approach is to use metrics that distinguish graphs with heavy tailed degree sequences as opposed to more regular topologies and may be correlated with further coarse characteristics of the network [27], [32]. Our approach is closer to the latter, and influenced from the proposal of [7], [10] that *topology properties should be studied in connection to the functionality of the network*. In particular, we shall study the correlation of the information retrieved from the eigenvectors of Section III to the performance of a primitive experiment that studies the "congestion" in the network.

For an undirected (without AS relationships) topology, suppose that we send one unit of traffic along a minimum hop (shortest) path from each node to every other node[6]. This induces a *stress for each link* defined as the total number of paths going though the link. We study the maximum link stress, which can be thought of as an indicator of congestion.

Intuitively, we expect that there is more traffic between ASes that have geographic or business relationships. We use the following spectral-filtering based heuristic to group ASes into clusters:
(a) If n is the size of the topology, consider the $\alpha \cdot n$, where $\alpha = .5$ in our experiments, largest eigenvalues of $N(A)$, and the eigenvectors associated with each such eigenvalue.

[6] In case of many shortest paths, we pick one of them arbitrarily.

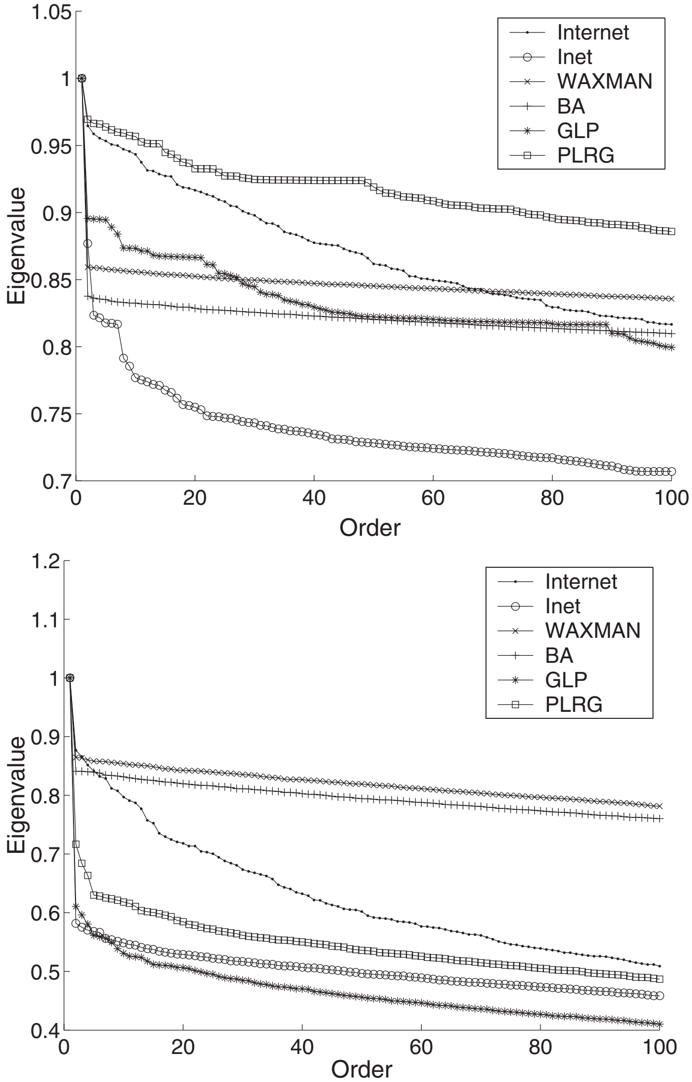

Fig. 7. Spectrum of real and synthetic Internet topologies. The top graph corresponds to the entire topology. The bottom graph corresponds to the core.

(b) Consider the nodes H_1 and H_2 that are assigned the highest $\beta{\cdot}n$ positive and the highest $\beta{\cdot}n$ negative weights in each such eigenvector. The parameter β is set to .25 in our experiments.
(c) Each AS which appears in H_1 or H_2 for at least one examined eigenvector will be assigned to the cluster of the positive or negative end of the first eigenvector in whose H_1 or H_2 it appeared. In this way we assign ASes to at most one cluster.

We say that a *traffic pattern is* $\epsilon\%$ *intraclustered* if each node sends $\epsilon\%$ of its traffic exclusively inside the cluster that it belongs, and $1-\epsilon\%$ of its traffic uniformly to all nodes (thus uniform traffic is 0% intraclustered).

We are interested in studying the change in the max link load as the traffic shifts from *uniform* to *intracluster* (and, *intercluster*). It is reasonable to expect that, in general, topologies with higher principal eigenvalues, and thus worse cuts (in the sense of (1) of Section II), should tend to exhibit worse link stress behavior. Thus, as we shift traffic from uniform to intracluster (resp. intercluster), we expect the maximum link stress to drop (increase) significantly, since we are increasing (resp. decreasing) the traffic that stays inside the cluster and reducing (resp. increasing) the traffic that crosses the bad cut.

TABLE V

DROP IN MAX LINK STRESS AND AVERAGE EXPECTED HOP DISTANCE, AS THE TRAFFIC SHIFTS FROM UNIFORM TO INTRACLUSTERED.

	Internet Max Link Stress	Internet Avg. Exp. Hop Dist	Inet Max Link Stress	Inet Avg. Exp. Hop Dist
0%	100.0%	3.3744	100.0%	2.7499
20%	91.5%	3.2855	97.7%	2.7151
40%	83.0%	3.1965	95.4%	2.6802
60%	74.4%	3.1076	93.1%	2.6454
80%	65.9%	3.0187	90.8%	2.6106
100%	57.4%	2.9297	88.5%	2.5757

Note: The same trend applies to the other synthetic topologies.

Indeed, the AS Internet topology is exhibiting sharper shift in link stress behavior than several synthetic topologies from Brite (BA,GLP,Waxman [4], [5], [23], [37]) Inet [18], and PLRG [2]. The results are given in Table IV. Assume for example that the traffic is 20% intraclustered. Then, the maximum link stress for the AS topology dropped to 91.5% of that in uniform traffic. For the same intracluster traffic, the max link stress in the topology generated by Inet dropped to 97.7%. Thus, the maximum link stress decreased by a factor of 8.5% in the case of the AS topology and by 2.3% in the case of Inet. At the extreme of 100% intraclustered traffic the max link stress in the Internet drops by more than 40%, while in every synthetic topology the drop was less than 23%, with the exception of Waxman, in which case the drop was around 30%.

We therefore propose that the information retrieved from the eigenvectors associated with the largest eigenvalues may be suggestive of intracluster traffic patterns. We propose to use the clusters suggested by these eigenvectors as one meaningful way to generate traffic patterns that deviate from uniform traffic. One additional remark is due. It may be thought that the decrease in link stress under intracluster traffic patterns is a straightforward consequence of shorter min-hop paths that would be used in an intraclustered traffic pattern, See Table V. For each node, define its *expected hop distance* as the expected hop distance of the node from every other node under a specific traffic pattern. Notice that both in the Internet and in the synthetic topology produced by Inet, the drop in the average expected hop distance is not nearly as striking as that of the max link stress. We therefore conclude that the drop in the link stress is a result of a *better distribution* of the weighed shortest paths rather than a mere decrease of their length. Thus the intracluster traffic pattern is indeed non trivial. Similar observation apply to intercluster traffic patterns.

TABLE IV

DROP OF MAX LINK STRESS AS THE TRAFFIC SHIFTS FROM UNIFORM TO INTRA-CLUSTER AND INTER-CLUSTER.

A. Intra-cluster

	Internet	Inet	PLRG	GLP	Waxman	BA
0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
20%	91.5%	97.7%	95.6%	95.8%	94.1%	96.4%
40%	83.0%	95.4%	91.2%	91.6%	88.2%	92.9%
60%	74.4%	93.1%	86.9%	87.3%	82.3%	89.3%
80%	65.9%	90.8%	82.5%	83.1%	76.3%	85.8%
100%	57.4%	88.5%	78.1%	78.9%	70.4%	82.2%

B. Inter-cluster

	Internet	Inet	PLRG	GLP	Waxman	BA
0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
20%	108.5%	102.4%	102.5%	103.9%	107.6%	102.8%
40%	116.9%	104.7%	105.1%	107.8%	116.1%	104.7%
60%	125.4%	107.1%	107.6%	111.7%	124.6%	107.1%
80%	133.8%	109.5%	110.1%	115.6%	113.2%	109.5%
100%	142.3%	111.8%	112.7%	119.5%	141.7%	111.8%

Note: The AS Internet exhibits more drop than any synthetic topology (almost twice as much with the exception of the Waxman model). We note that these numbers refer to the core of the network. The behavior was similar when we did the same experiment in the whole network, and in each specific continent.

V. RANKING BY THE FIRST EIGENVECTOR

The "significance" of an AS, or its position in a hierarchy, is a subjective matter, in the sense that ASes are never explicitly or implicitly assigned such rankings. There is relatively good agreement about the "top" and "bottom" of a hierarchy. For example, an ISP that has only peers and no provider is almost surely very big, while an AS that has no customers or peers and only one or two providers is almost surely very small. In two separate efforts, [15] and [31] gave heuristics to assign hierarchical levels to ASes, after inferring AS relationships and taking into account several non-trivial further characteristics.

In this Section we observe that a different heuristic, based on the weights assigned to the ASes by the first eigenvector of a suitably defined modification of the directed AS graph (i.e., after AS relationships have been inferred), is highly correlated with the hierarchy of [31].

The proposed heuristic is an adaptation of the *pagerank* method used by Google to infer quality of Web pages. The analogy is natural. Both the directed AS topology and the WWW are directed graphs. In the WWW, a hyperlink pointing from a page i to a page j indicates an endorsement of importance from i to j. In the Internet, an edge pointing from a customer i to a provider j can be thought of as a similar endorsement of importance, while in peers the endorsement becomes mutual.

The ranking method is the following. Let A' be the directed adjacency matrix. For each node i define the outdegree of i as $d_{\text{out}}(i) = |\{j : a_{ij} = 1\}|$. Now consider the stochastic matrix

$$\mathrm{P}(A)\ :\ p_{ij} = \begin{cases} = \frac{\alpha}{d_{\text{out}}(i)} + \frac{1-\alpha}{n} & \text{if } a_{ij} = 1 \\ = \frac{1-\alpha}{n} & \text{if } a_{ij} = 0 \end{cases}$$

The above stochastic matrix represents a random walk on the directed graph A', where with probability α we go to a provider or peer chosen uniformly at random, and with probability $1-\alpha$ we jump to a uniformly random node from the set of all nodes (the latter step is a standard correction to avoid degeneracies pertaining to sinks).

Let $\pi(v)$ be the stationary probability of the stochastic matrix $P(A)$. Google assigns to Web pages pagerank quality $\pi(v)$. By analogy, we assign to each AS hierarchical weight $\pi(v)$. In Figure 8 we compare the hierarchy of [31] to our hierarchical weight $\pi(v)$. We have used $\alpha = .95$; the results are similar for any $.9 \leq \alpha \leq .99$. To plot the graph, we have grouped the ASes by their level in the hierarchy. Then, we sort the ASes in each group by their weight in $\pi(v)$ and plot the weights in decreasing order. Observe that we use logarithmic scale for both axes.

There is notable correlation between the weights assigned to the ASes and their level in the hierarchy. Nodes assigned by [31] in high levels have higher values in $\pi(v)$. Also, the weights assigned to the ASes of a group are in general higher than the weights assigned to ASes that belong in groups of lower level. One noticeable exception is the weights assigned to levels 4 and 5. ASes in these levels have very small degree and they cannot be easily separated by the page rank method. At first glance it seems that there is an "anomaly" in the figure, since there are some ASes that are assigned larger weights than ASes which belong to higher levels. We argue that this could be a problem of the subjective nature of hierarchical assignment, and/or the heuristic used by [31] to assign ASes to levels. We will discuss two examples to make this point. The largest weights in levels 2 and 3 have a very high value which is comparable to the weights assigned to nodes in level 1. These weights correspond to the ASes of Tiscali Intl Network (AS number 3257) and of Abovenet (AS number 6461) respectively. We believe that they had to be assigned in the highest level. This is justified by their degrees in the adjacency matrix, which are 330 and 585 respectively, and by

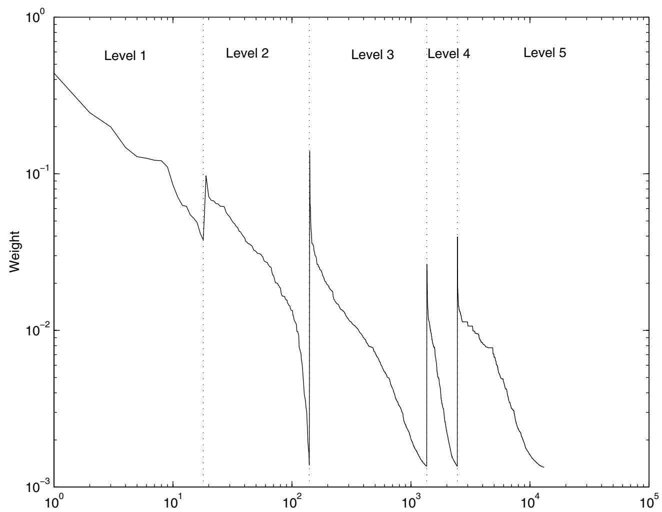

Fig. 8. Comparison of hierarchy with the first eigenvector.

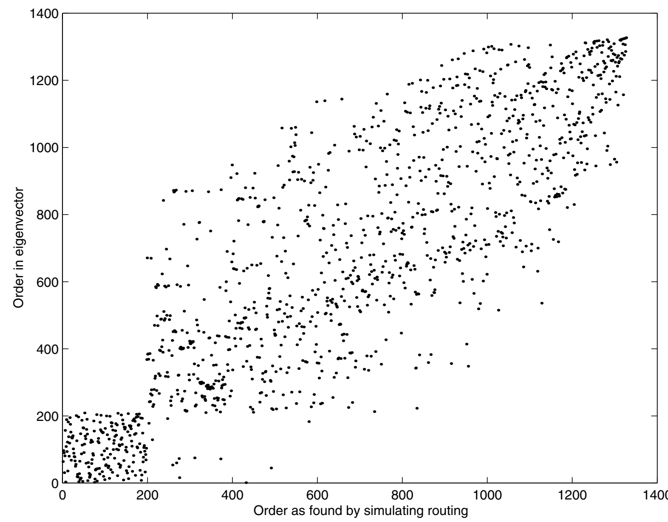

Fig. 9. Correlation between link importance as assigned by the left eigenvector of the SVD of the traffic matrix with the load of the link. Correlation coefficient is 0.8594.

the reputation they have as big ISP providers.

We extend the above method to obtain an assignment of significance to links. If n is the number of ASes and m is the number of links of the undirected AS topology, let $N=n^2$ be the number of pairs of ASes and associate with each such pair a shortest path between their endpoints. We may now consider the $m\times N$ traffic matrix T, where each row corresponds to a shortest path and there is a 1 on the columns of the links used by the path. Using the SVD method, which is a generalization of the decomposition into eigenvalues and eigenvectors for non-square matrices, we can compute the left eigenvector of T that corresponds to the largest eigenvalue. Just like pagerank, this eigenvector gives an order of importance to links. Links that get higher values are associated with links that accept more traffic and thus are candidates to be places of congestion. Observe that this statement was made without making any assumption about the traffic between any two ASes.

To find the correlation between the importance assigned to links and the amount of traffic they receive we did the following experiment[7]. We assumed that between each pair of ASes there is some amount of traffic flowing drawn from a uniform distribution that takes values between 0 and 2 traffic units[8]. After performing shortest path routing and assigning loads to links, we have ordered the links by their load. We are interested to find the relation between this ordering and the ordering given by the weights in the eigenvector. In Figure 9 we depict this relationship. There is a point in (i, j) when a link is in i-th position sorted by the load and in j position sorted by the weight in the eigenvector. Indeed it is easy to observe that there is strong correlation between the importance of the link and the amount of traffic it receives. The correlation coefficient in this case is 0.8594 indicating this strong correlation.

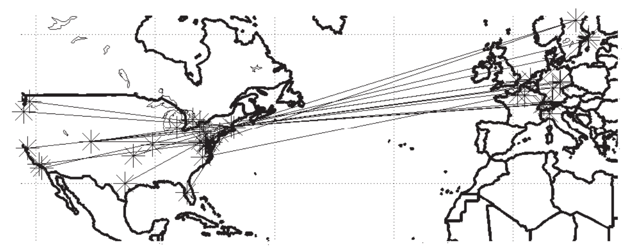

Fig. 10. An example of a link cluster.

In addition, it is possible to use the left eigenvectors to identify *clusters* of related links that form a cut in the original adjacency matrix. The links in the cut carry traffic between areas in the Internet that are not well connected and thus they are candidates to be points of congestion. As a simple example we give Figure 10, where we draw a cluster of links (cluster in the same sense as the clusters defined earlier for ASes) taken from the left eigenvector which corresponds to the second largest eigenvalue. Intuitively, we expect that indeed the trans-atlantic links to carry a lot of traffic and thus be points of congestion as indicated. We have observed similar clusters using the other eigenvectors, and also in positions that seem intuitively natural (across Central and Eastern Europe, across the Pacific, e.t.c). It is still an open question to us how the clusters observed in the AS topology related with the link clusters.

VI. SUMMARY

Spectral filtering is a well known information retrieval method. We studied the adaptation of spectral filtering in the AS Internet topology. We found that the information retrieved corresponds to groups of nodes with semantic proximities. We found that the clustering behavior varies in the core and in the

[7] For this example we have used an induced graph of the real topology which includes all the ASes in levels 1 and 2 as assigned by [31]. Memory and processing limitations did not allow us to work with bigger matrices.

[8] Setting the traffic to 1 for each pair gave the same results.

edge of the network, and across different geographic areas. We gave two applications of spectral filtering. The first is to identify non-trivial intacluster and intercluster traffic patterns. Such traffic patterns affect the stress on elements of the network. The second application is an adaptation of Google's PageRank to obtain an alternative detailed characterization of hierarchy. Our study proves that spectral filtering methods can be successful in processing Internet topologies.

Beyond information retrieval, spectral methods have found great applicability in information compression, via the technique of low rank approximations [3], [26], [28]. Examining if such low rank approximations apply to the networking context (e.g. speed-up simulations) is a very important open question. We believe that our study is a first step in this direction.

Finally we should mention that the first reference to the high end of the spectrum of Internet topologies is due to [11], and another interesting study can be found in [36] who discuss properties of the entire spectrum and relate them to certain structural properties of the Internet graph.

ACKNOWLEDGMENT

We thank Lixin Gao for providing her scripts, Lakshmi Subramanian for providing his data. We thank Thomas Erlebach, Amin Saberi, and the anonymous referees for insightful suggestions on the presentation.

REFERENCES

[1] S. Agarwal, L. Subramanian, J. Rexford, and R. Katz. Characterizing the internet hierarchy from multiple vantage points web page. *http://www.cs.berkeley.edu/ sagarwal/research/BGP-hierarchy/*, July 2002.

[2] William Aiello, Fan R. K. Chung, and Linyuan Lu. A random graph model for massive graphs. In *ACM Symposium on Theory of Computing*, pages 171–180, 2000.

[3] Yossi Azar, Amos Fiat, Anna R. Karlin, Frank McSherry, and Jared Saia. Spectral analysis of data. *ACM Symposium on Theory of Computing*, pages 619–626, 2001.

[4] Albert-László Barabasi and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[5] Tian Bu and Don Towsley. On distinguishing between internet power law topology generators. *Infocom 2002*.

[6] K. Calvert, M. Doar, and E. Zegura. Modeling internet topology. *IEEE Communications Magazine*, June 1997.

[7] J.M. Carlson and J. Doyle. Highly optimized tolerance: A mechanism for powerlaws in design systems. *Physics review E, 60(2)*, pages 1412–1427, 1999.

[8] Chen Chang, Jamin Govindan, and Willinger Shenker. The origins of power-laws in internet topologies revisited. *Infocom 2002*.

[9] Fan R.K. Chung. Spectral graph theory. *American Mathematical Society Book Series*, 1997.

[10] A. Fabrikant, E. Koutsoupias, and C.H. Papadimitriou. Heuristically optimized tradeoffs: A new paradigm for powerlaws in the internet. *ICALP 2002*, 2002.

[11] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *Proc. of ACM SIGCOMM*, 1999.

[12] I.J. Farkas, I. Derényi, A. Barabási, and T. Vicsek. Spectra of real-world graphs: Beyond the semi-circle law. *e-print cond-mat/0102335*, 2001.

[13] National Laboratory for Applied Retwork Research. Route views archive. *http://moat.nlanr.net/Routing/rawdata*.

[14] Cooperative Association for Internet Data Analysis. Caida. *http://www.caida.org*.

[15] L. Gao. On inferring autonomous system relationships in the internet. *IEEE Glogal Internet*, 2000.

[16] G.H. Golub and C. F. Van Loan. Matrix computations. *Johns Hopkins University Press*, 1989.

[17] P. Husbands, H. Simon, and C. Ding. On the use of the singular value decomposition for text retrieval. *1st SIAM Computational Information Retrieval Workshop*, 2000.

[18] C. Jin, Q. Chen, and S. Jamin. Inet: Internet topology generator. (CSE-TR-433-00), 2000.

[19] K.I. Kahng and D. Kim. Spectra and eigenvectors of scale-free networks. *Physical Review E., Vol 64*, 2001.

[20] J. Kleinberg. Hubs, authorities, and communities on the www. *ACM Computing Surveys, 31(4es):5*, 1999.

[21] Anukool Lakhina, John Byers, Mark Crovella, and Ibrahim Matta. On the geographical location of internet resources. *Boston University*, 2002.

[22] L. Lovasz. Combinatorial problems and exercises. *North Holland, Amsterdam*, 1979.

[23] A. Medina, I. Matta, and J. Byers. On the origin of power-laws in internet topologies. *ACM Computer Communications Review*, April 2000.

[24] M. Mihail and C.H. Papadimitriou. On the eigenvalue power law. *Random 2002*, 2002.

[25] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. *Stanford Digital Library Technologies Project*, 1998.

[26] C.H. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala. Latent semantic indexing: A probabilistic analysis. *Proc. Principles of Database Systems (PODS)*, 1999.

[27] P. Radoslavov, H. Tangmunarunkit, H. Yu, R. Govindan, S. Shenker, and D. Estrin. On characterizing network topologies and analyzing their impact on protocol design. *USC-CS-TR-00-731, http://citeseer.nj.nec.com/radoslavov00characterizing.html*, March 2000.

[28] P. Raghavan. Information retrieval algorithms: A survey. *Proc. 8th SIAM Symposium on Discrete Algorithms (SODA)*, 1997.

[29] A. Sinclair. Algorithms for random generation and counting: A markov chain approach. *Springer-Verlag*, 1997.

[30] G.W. Stewart and J. Sun. Matrix perturbation theory. *Academic Press*, 1990.

[31] L. Subramanian, S. Agarwal, J. Rexford, and R.H. Katz. Characterizing the internet hierarchy from multiple vantage points. In *IEEE Infocom*, June 2002.

[32] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Network topology generators: Degree-based vs structural. *Sigcomm 2002*.

[33] NetGeo tool. Netgeo tool. *http://www.caida.org/tools/utilities/netgeo*, 2002.

[34] Traceroute.org. Public route server and looking glass site list. *http://www.traceroute.org*.

[35] V.V. Vazirani. Approximation algorithms. *Springer-Verlag*, 2001.

[36] Danica Vukadinovic, Polly Huang, and Thomas Erlebach. A spectral analysis of the internet topology. *Dimacs Workshop on Internet and WWW Measurement, Mapping, and Modeling, citeseer.nj.nec.com/vukadinovic01spectral.html*, 2001.

[37] B.M. Waxman. Routing of multipoint connections. *IEEE Jour. Selected Areas in Communications (Special Issue: Broadband Packet Communications)*, 6(9):1617–1622, 1988.

[38] J.H. Wilkinson. The algebraic eigenvalue problem. *Oxford Science Publications*, 1965.

[39] Soon-Hyung Yook, Hawoong Jeong, and Albert-Laszlo Barabasi. Modeling the internet's large-scale topology. *http://arxiv.org/abs/cond-mat/0107417*, 2002.

[40] E. Zegura, K. Calvert, and S. Bhattacharjee. How to model an internetwork. In *Proceedings of IEEE INFOCOM*, 1996.

[41] E. Zegura, K. Calvert, and M. Donahoo. A quantitative comparison of graph-based models for Internet topology. *Transactions on Networking, Vol 5, No 6*, pages 770–783, 1997.