Autonomous System Ranking by Topological Characteristics: A Comparative Study

Mehmet Engin Tozal
School of Computing and Informatics
University of Louisiana, Lafayette, LA 70504 USA
Email: metozal@louisiana.edu

***Abstract*—The Internet is a highly engineered, large scale complex system serving billions of people worldwide. The whole system is formed by tens of thousands of autonomous networks owned by different organizations. These autonomous networks are connected to each other through hundreds of thousands of relations which reflect the business partnerships among the network operators as well as the traffic routing in the Internet. Ranking ASes by their topological characteristics allow us to acquire immediate insight on the complex structure of the Internet and make decisions based on various criteria. In this study we compare and contrast six different AS ranking schemes based on the topological features of the ASes: customer degree, provider degree, peer degree, customer-cone size, alpha centrality and betweenness centrality. We report varying levels of agreement/disagreement among the ranking schemes and show that selecting multiple ranking schemes might be necessary to gain a diverse insight on the topology.**

I. INTRODUCTION

The Internet is a highly engineered, large scale complex system serving billions of people worldwide. The system is formed by tens of thousands of autonomous networks owned by different organizations including companies, network service providers, cloud providers, web hosting companies, universities and government agencies all around the world. A group of networks managed by one or more network operators under a well defined routing policy is called an *Autonomous System (AS)* in the Internet [9]. These ASes, identified by unique AS numbers, connect to each other in different forms to enable the "global" Internet communication. Individual users, small businesses and ASes located at the edge of the Internet participate in the global infrastructure by means of other ASes called Internet Service Providers (ISPs). Typically, ISPs are business entities providing Internet access service to their customers while getting the same service from one or more upstream ISPs. At the core of the Internet, a small number of ISPs peer with each other through settlement-free interconnections and attain the global communication infrastructure.

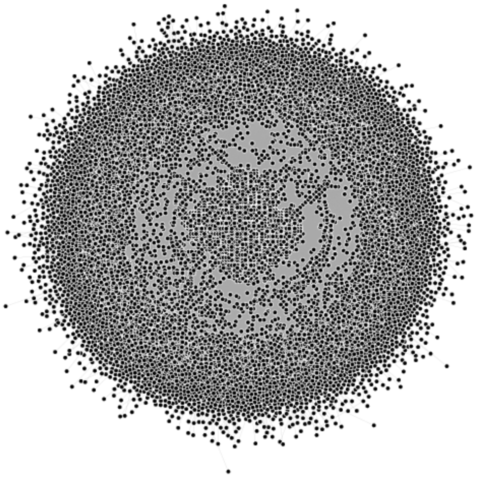


Fig. 1: AS-level Internet topology represented as a graph. Around 55 thousand ASes are connected to each other through more than 466 thousand relations.

The majority of the the ASes (around 85%) are located at the edge of the Internet and they are solely Internet access consumers. That is, they pay to ISPs to acquire global Internet access. Note that these ASes may be content or service providers, yet they are consumers in terms of the Internet access service. The ASes forming the communication infrastructure, on the contrary, are Internet access consumers and providers, simultaneously. Internet access service is provided and consumed with respect to business relations among the ASes. That is, ASes are connected to each other via business relations that define Internet access services. Moreover, the inter-AS traffic in the Internet is usually routed according to the business relations among ASes [11].

Traditionally, business relations are categorized as customer-to-provider (c2p), peer-to-peer (p2p) and sibling-to-sibling (s2s) [6]. In a c2p relation, the provider AS provides global reachability to its customer AS. In return, the customer pays to the provider for the traffic exchanged between them. In a p2p relation, two peer ASes provide mutual reachability to each other and their customer ASes. Peer ASes typically engage in settlement-free business agreements which means that neither party pays to the other for the traffic exchanged. In the less frequently observed s2s relation, two ASes provide full reachability to each other because they are operated by the same or sibling organization(s). More complex relations such as hybrid relations and partial relations are reported in the

Internet as well [7]. However, c2p and p2p relations abstract the majority of the business agreements between ASes for practical purposes [11].

Figure 1 shows the AS-level topology of the Internet obtained in June 2016 [3]. The figure consists of 54,140 ASes connected to each other through 466,190 relations (logical links). Among those ASes 45,796 (85%) are located at the edge of the Internet without having any customer ASes. Put in other words, 45,796 ASes are solely Internet access consumers and the remaining 8,344 (15%) ASes at the center provide Internet access service to organizations and individuals. Out of 466,190 relations among the ASes, 107,195 (23%) are c2p and 358,995 (77%) are p2p relations. In the figure black circles correspond to the ASes and gray links represent the relations among them regardless of the relation type.

Obviously, it is difficult to gain any visual insight from Figure 1. Often, researchers mine the topology of the Internet to extract useful information and transform the information into an understandable format. *Ranking* is a data mining technique which orders an object in relation to others based on some criteria. Ranking is a widely used technique because it presents complex data in an understandable format by reducing measurement(s) into a sequence of ordinal numbers; it helps us to acquire immediate insight on complex data; and it allows us to make quick decisions based on various criteria. In the same regard, ranking ASes in the Internet (i) serves as an auxiliary tool to help network practitioners to manage and troubleshoot their networks; (ii) provides a comparative guidance to chief information officers (CIOs) to negotiate new business relations among ISPs; and (iii) provides an immediate insight to network researchers to assess and study the connectivity of the Internet.

In this study we compare and contrast six different AS ranking schemes based on the topological features of the ASes: customer degree, provider degree, peer degree, customer-cone size, alpha centrality and betweenness centrality. Instead of studying total degree centrality, we distinguish between different types of degree centralities because they are implemented and interpreted differently in the Internet. The degree centrality display a *local* aspect of an AS because the information is directly related to the AS itself. On the other hand, alpha and betweenness centralities demonstrate *global* aspects of an AS because their computations involve the AS's connectivity to other ASes in the Internet topology graph. Lastly, customer-cone size as well as the distinct degree centralities indicate the *architectural* aspects of an AS because they are related to the autonomous and perpetual network planning. We restrict our analysis to the ASes in the core of the Internet because those ASes together enable the global Internet communication while the ASes at the edge do not carry any transit traffic.

Our experimental results show that ranking ASes by their structural properties significantly varies in most cases. Hence, one should evaluate ranking results carefully and often look at multiple rankings together. Ranking by customer degree emphasizes the ASes that provide Internet access service mostly to the ASes at the periphery of the Internet. Ranking by provider degree gives priority to those ASes that are more resilient. That is, they have more provider ASes to distribute traffic over or switch over in case of a failure. Ranking by peer degree stresses the ASes that have lower operational costs due to peering agreements. Ranking by customer-cone size emphasizes the ASes located at the core of the topology. Ranking by alpha centrality gives higher ranks to the ASes connected to other highly ranked ASes. Ranking by betweenness centrality gives priority to those ASes carrying most of the traffic between pairs of ASes. We observed a high level of agreement between rankings by customer degree and by customer-cone size as well as moderate levels of agreement between rankings by peer degree and alpha centrality, betweenness centrality and customer degree and betweenness centrality and customer-cone size. Please see Section IV for a more detailed discussion of our experimental results.

The rest of the paper is organized as follows. Section II presents the related work. In Section III we introduce the structural ranking schemes used in this study. In Section IV we compare and contrast the ranking schemes. Finally, Section V concludes the paper.

II. RELATED WORK

AS customer-cone is a widely known measure to compare and rank the ASes in the Internet [11]. In general, the customer-cone of an AS denotes the ASes (including the self), IP addresses and IP address prefixes that can be reached following only the customer relations in an AS-level Internet topology map. Specifically, the customer-cone of an AS is the set of ASes consisting of the AS itself, its customer ASes and the customer-cones of those customer ASes. Customer-cone size of an AS roughly denote the other ASes that are expected to directly or indirectly pay to the AS for the transit traffic. This measure not only reflects the position of an AS in the semi-hierarchical structure of the Internet but also reflects the routing influence of the AS in the Internet. On the other hand, multi-homing and peering practices in the Internet introduces multiple paths that bypass upstream providers. Therefore, the routing influence of an AS in the Internet might be different from the one reflected by its customer-cone.

Zimmerli et al., [16] suggested an AS rating approach based on the traceroute collected [1] performance metrics. The authors collected end-to-end performance metrics including packet loss, round-trip delay and hop count from a mesh of 159 vantage points. They ranked ASes based on the network performance within the ASes and the performance of the neighboring ASes. This ranking scheme is highly volatile because it is sensitive to the real time changes in the Internet. Besides, it is difficult to scale the technique to the entire Internet.

Clérot and Nguyen [4] proposed an AS ranking heuristic based on the concept of alpha-centrality [2] in social network analysis. Their heuristic starts from an undirected graph of ASes and gradually introduce asymmetry by allowing directed edges reinforce the relationships between ASes. The rank score of an AS has two terms: the first term denotes the centrality of an AS by its own and the second term denotes the centrality of an AS that is inherited from its neighbors. The authors show that the ranking results of their heuristic is quite close to the results of the simple degree-based centrality. Their approach facilitates the examination of addition or removal of links between ASes. This method requires careful selection of parametric values and it may artificially rank the ASes with many neighbors higher.

In an indirectly related study, Wagner et al., [15] proposed an AS ranking method for detecting the ASes which provide transit services to the ASes that host malicious software and services. The authors use existing AS scores reflecting the malware hosting capacity of ASes to annotate AS graphs and use PageRank [13] to rank the ASes.

In this comparative study, we focus on the topological characteristics of the ASes that form the global communication infrastructure of the Internet. We rank the ASes with respect to their customer degrees, provider degrees, peer degrees, customer-cone sizes, alpha and betweenness centralities as well as analyze the similarities and dissimilarities among the ranking schemes.

III. TOPOLOGICAL RANKING SCHEMES

In complex systems analysis, structural centrality is a measure that defines the "importance" of a component within a network (graph). Typically, the higher the centrality, the more important a component is. However, the concept of "importance" is intrinsically versatile and depends on the application domain or context. Therefore, several centrality measures have been proposed in the complex systems literature. In this study, we use multiple centrality measures to rank the ASes forming the communication infrastructure of the Internet.

Vertex Centrality is a function, $f : V \to \mathbb{R}$, assigning a real value to each vertex, $v_i \in V$, in a graph, $G = (V, E)$, such that $f(v_i)$ reflects the importance of vertex v_i with respect to other vertices, $v_j \in V - \{v_i\}$, in the graph. In the following we focus on six vertex centrality measures that demonstrate the local (customer, provider and peer degrees), global (betweenness and alpha centralities) and architectural (customer cones and design-specific degree centralities) characteristics of the ASes in the Internet.

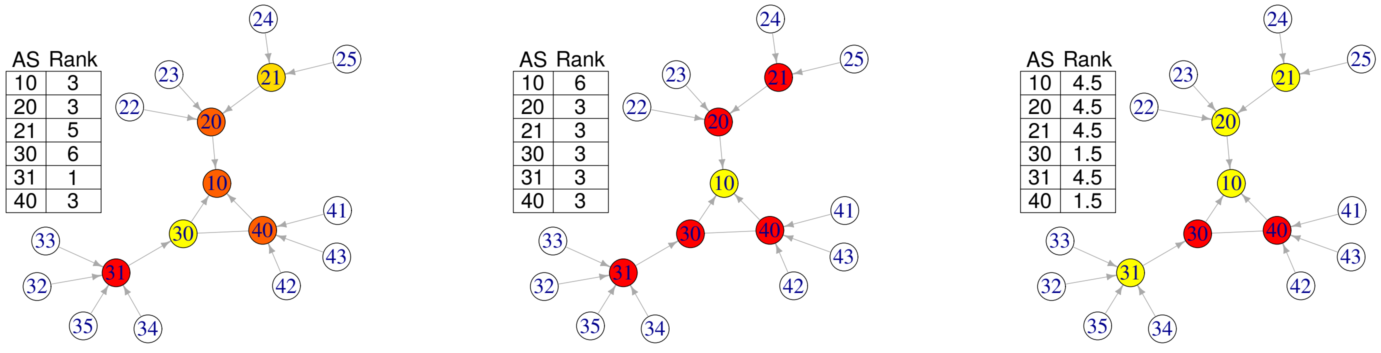
Degree centrality is a simple, yet powerful measure used in complex system analysis. It is defined as the number of edges of a given vertex in an undirected graph. For directed graphs, indegree and outdegree centralities specify the number of incoming and outgoing edges, respectively. An AS in an Internet topology map however, have different types of relations with its neighboring ASes. A c2p relation indicates an edge between a customer AS and provider AS whereas a p2p relation is indicates an edge between two peering ASes. Usually an AS engages in different relations with different ASes. We distinguish between three types of degree centralities: **customer degree**, **provider degree** and **peer degree** since they are implemented and interpreted differently in the Internet. To illustrate, an AS in the Internet may prefer to peer with ASes having a larger customer degree to reduce its operational costs. On the other hand, ASes having larger provider degree might be better choice to get Internet access service because they are more resilient to provider AS failures or misconfigurations.

A certain type of degree centrality assumes all neighboring ASes are equivalent to each other. However, having a relation to one AS might be more relevant than having the same relation to another AS. **Eigenvector centrality** [8] assigns a centrality score to a vertex based on the centrality scores of its neighboring vertices. More specifically, eigenvector centrality is the eigenvector associated with the leading eigenvalue (largest magnitude eigenvalue) of the adjacency matrix of a graph, i.e. $\boldsymbol{x}, \lambda_0 : \boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$ and $|\lambda_0| \geq |\lambda_i|$. However, eigenvector centrality may produce unreliable results for asymmetric graphs. **Alpha centrality** [12] is a close generalization of eigenvector centrality which involves structure-independent exogenous importance of vertices in addition to the structure-dependent endogenous importance of vertices. Formally, alpha centrality is defined as $\boldsymbol{x} = \alpha \boldsymbol{A^T} \boldsymbol{x} + \boldsymbol{e}$ where $\boldsymbol{e}$ denotes the exogenous importance of vertices and α represents the impact of endogenous importance in calculation of the centrality. Note that alpha centrality converges to eigenvector centrality as α approaches to the reciprocal of the leading eigenvalue, $1/\lambda_0$. In the context of the Internet, an AS having more prominent upstream providers and peers will receive higher alpha centrality score compared to the ASes having less prominent providers and peers. Hence, those ASes might be more attractive to get Internet access service for a customer AS.

Customer-cone size is a well known measure to compare the ASes in the Internet. In general, customer-cone [11] of an AS is recursively defined as a set consisting of the AS itself along with its customers' customer-cones. That is, the customer-cone of an AS is a set consisting of the AS itself and its customer descendants. The customer-cone of an AS corresponds to a sub-topology where the connected component is formed through c2p relations. Customer-cone size, the number of ASes in the customer-cone of an AS, may show the importance of the AS regarding the global traffic routing in the Internet. The inter-AS traffic is more likely to be kept within the sub-topology as the customer-cone size increases.

Betweenness centrality is a measure that quantifies the centrality of a vertex in terms of its involvement in connecting pairs of vertices in a graph. Formally, betweenness centrality of a vertex v_k is defined as $\beta(v_k) = \sum \sigma_{v_i v_j}(v_k)/\sigma_{v_i v_j}$ such that $\sigma_{v_i v_j}$ is the number of shortest paths between vertices v_i and v_j and $\sigma_{v_i v_j}(v_k)$ is the number of those paths that pass through v_k where $v_i \neq v_j \neq v_k$. In an AS-level Internet topology map, the betweenness centrality of an AS indicates the centrality of the AS with respect to the traffic it carries between pairs of ASes. ASes having higher betweenness centrality are the ones that play important roles in inter-AS transit traffic.

Given a finite set of objects, $S = \{s_1, s_2, \ldots, s_{|S|}\}$, a *rank* is a function that maps the elements of S to ordinal numbers where $r(s_i)$ denotes the position of s_i when the objects are sorted by some criteria. Figure 2 shows vari-



AS	Rank
10	3
20	3
21	5
30	6
31	1
40	3

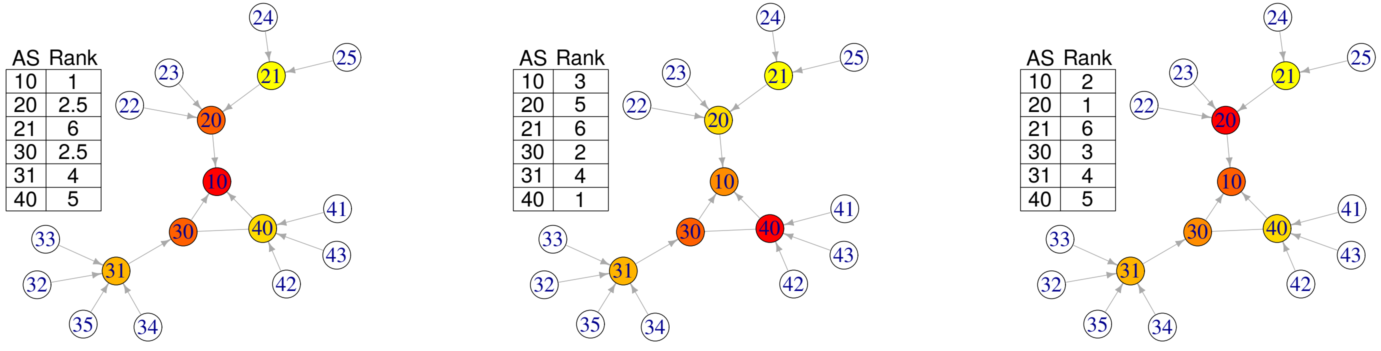
AS	Rank
10	6
20	3
21	3
30	3
31	3
40	3

AS	Rank
10	4.5
20	4.5
21	4.5
30	1.5
31	4.5
40	1.5

(a) Rank by Customer Degree ($\Re_C$)

(b) Rank by Provider Degree ($\Re_P$)

(c) Rank by Peer Degree ($\Re_R$)



AS	Rank
10	1
20	2.5
21	6
30	2.5
31	4
40	5

AS	Rank
10	3
20	5
21	6
30	2
31	4
40	1

AS	Rank
10	2
20	1
21	6
30	3
31	4
40	5

(d) Rank by Customer Cone ($\Re_O$)

(e) Rank by Alpha Centrality ($\Re_A$)

(f) Rank by Betweenness Centrality ($\Re_B$)

Fig. 2: Center ASes ranked by different topological ranking schemes. Nodes, arrows and lines represent ASes, c2p relations and p2p relations, respectively. The ranks in the accompanying tables are color coded in the graph from high (red) to low (yellow).

ous rankings of an example AS-level topology consisting of 17 ASes for illustration purposes. Each AS in the figure is assigned a unique AS number and represented by a circle in the topology. The ASes are connected to each other through 16 c2p and one p2p relations represented by arrows and lines, respectively. In the figure, 11 ASes are located at the edge of the topology whereas six ASes in the center,$\{AS10, AS20, AS21, AS30, AS31, AS30\}$, enable the overall communication infrastructure. Figures 2a thru 2f illustrates the rankings of the center ASes by six topological characteristics. In all ranking schemes the ties are resolved by *fractional ranking* where ASes having the same rank are assigned the mean of the ordinal ranks of the group.

Figure 2a ranks the center ASes by customer degree ($\Re_C$). In the figure $AS31$ is ranked first by four customers and $AS30$ is ranked the last by one customer. On the contrary, $AS31$ is a customer of $AS30$ i.e., $AS31$ gets Internet access service for its customers through $AS30$ tough it is ranked higher than $AS30$.

Figure 2b ranks the center ASes by provider degree ($\Re_P$). In the figure $AS20$, $AS21$, $AS30$, $AS31$ and $AS40$ share the same fractional ranking and $AS10$ is placed last since it does not have any providers. This ranking scheme underestimates the importance of the core ASes that fasten the topology together without having a provider AS, e.g., $AS10$ in Figure 2b.

Figure 2c ranks the center ASes by peer degree ($\Re_R$). In this small example $AS30$ and $AS40$ share the highest position since they are the only two ASes peering with each other. In the topology $AS30$ and $AS40$ reduce their operational costs without letting their customer-cone inter-AS traffic to pass through their provider AS, $AS10$.

Figure 2d ranks the center ASes by customer-cone size ($\Re_O$). $AS10$ has the largest customer cone, i.e., the entire topology. $AS21$ is placed the last because it has only two ASes (its direct customers) and itself in its customer-cone. This ranking scheme emphasizes the ASes in the core of the semi-hierarchical Internet topology.

Figure 2e ranks the center ASes by alpha centrality ($\Re_A$). In the figure $AS40$ is ranked top because it receives centrality scores from its direct customers as well as from two other prominent ASes, i.e., $AS10$ and $AS30$. Although $AS21$ have a larger customer cone compared to $AS31$, $AS31$ receives more centrality score from its direct customers. In other words, the shorter paths contribute more centrality score than the longer paths in the graph.

Figure 2f ranks the center ASes by betweenness centrality ($\Re_B$). In the figure $AS21$ is placed the last since it only appears on paths that are from/to $AS24$ and $AS25$. On the other hand, $AS20$ is placed first since it appears on all paths from/to its customer cone except the paths between $AS24$ and

$AS25$. Note that $AS10$ is placed second because the customer cones of $AS30$ and $AS40$ bypasses $AS10$ and exchange traffic via the peer link between them. Otherwise, $AS10$ would have been ranked first since it is at the core of the topology.

Evidently, different structural ranking schemes order ASes differently. In the following section we apply those ranking schemes on a real world AS-level Internet topology map and shed light on the levels of agreement and disagreement between the ranking schemes.

IV. EMPIRICAL ANALYSIS

Although rankings help us to acquire quick insight on complex data, they reduce one or more measurements into a sequence of ordinal numbers. Hence, rankings of objects should be evaluated with care. In this section we experimentally compare and contrast six topological AS ranking schemes presented in Section III and discuss the results in detail.

Given a finite set of objects, $S = \{s_1, s_2, \dots, s_{|S|}\}$, *ranking* is a binary relation $\mathcal{R} = \{(s_i, s_j) \subset \mathcal{S} \times \mathcal{S}\}$ denoting the first element "precedes" (or "succeeds") the second element while satisfying irreflexivity, $(s_i, s_i) \notin \mathcal{R}$; antisymmetry, $(s_i, s_j) \in \mathcal{R} \Rightarrow (s_j, s_i) \notin \mathcal{R}$; and transitivity, $(s_i, s_j) \in \mathcal{R}, (s_j, s_k) \in \mathcal{R} \Rightarrow (s_i, s_k) \in \mathcal{R}$. While we use $\Re$ to denote a ranking scheme in general, we adopt the matrix notation to represent ranking relations among objects of a set. Traditionally, an $|S| \times |S|$ rank matrix, $\boldsymbol{R}$, over a set S is formulated as follows [10]:

$$\boldsymbol{R}[i,j] = \begin{cases} 1 & \text{if object } i \text{ precedes object } j \\ 0 & \text{if object } i \text{ ties with object } j \\ -1 & \text{if object } i \text{ succeeds object } j \end{cases} \tag{1}$$

However, Emond and Mason, in their relatively recent work [5], showed the shortcomings of Equation 1 (Kendall's representation) and suggested an alternative formulation as follows:

$$\boldsymbol{R}[i,j] = \begin{cases} 1 & \text{if object } i \text{ precedes or tied with object } j \\ 0 & \text{if } i = j \\ -1 & \text{if object } i \text{ succeeds object } j \end{cases} \tag{2}$$

Compared to the former one, the latter formulation allows the distances between ranking schemes abide by Kemeny-Snell axioms ensuring non-negativity, symmetry, triangle inequality and rank consistency [5]. In this study we adopt the matrix representation given in Equation 2. In the remainder we analyze multiple rankings of the non-edge ASes in the Internet using rank by customer degree ($\Re_C$), provider degree ($\Re_P$), peer degree ($\Re_R$), customer-cone size ($\Re_O$), alpha centrality ($\Re_A$) and betweenness centrality ($\Re_B$).

TABLE I: Top-16 ASes w.r.t the ranking schemes

	$\Re_C$	$\Re_P$	$\Re_R$	$\Re_O$	$\Re_A$	$\Re_B$
1	AS174	AS20940	AS6939	AS3356	AS24482	AS6939
2	AS3356	AS42473	AS24482	AS1299	AS25091	AS174
3	AS7018	AS36408	AS3549	AS174	AS6939	AS3356
4	AS4323	AS12772	AS36351	AS3257	AS20764	AS3549
5	AS209	AS15133	AS43531	AS2914	AS13335	AS7018
6	AS6461	AS12222	AS12989	AS6453	AS20940	AS1299
7	AS3257	AS12418	AS34224	AS4436	AS8075	AS209
8	AS2914	AS13335	AS25091	AS701	AS13238	AS4323
9	AS701	AS6621	AS48166	AS6762	AS12578	AS2914
10	AS1299	AS25152	AS57463	AS7018	AS44050	AS701
11	AS2828	AS16509	AS58511	AS6939	AS8400	AS3257
12	AS3549	AS714	AS10026	AS209	AS35297	AS6461
13	AS6939	AS4775	AS38880	AS3320	AS51028	AS9498
14	AS4436	AS36236	AS20764	AS5511	AS47541	AS20485
15	AS6453	AS40285	AS8220	AS1239	AS44020	AS2828
16	AS12389	AS6507	AS8492	AS5580	AS15169	AS8220

Table I shows the top-16 ASes listed by different ranking schemes. Obviously, the table demonstrates a great discrepancy among the ranking schemes. First of all, none of the ranking schemes have the exact same set of ASes in their top-16 lists. Second, the number of common ASes in the top-16 list varies between the ranking schemes. For example, rankings by customer degree ($\Re_C$) and betweenness centrality ($\Re_B$) have 13 ASes in common, whilst ranking by customer degree ($\Re_C$) and provider degree $\Re_P$ do not have any common ASes. Third, even though some ASes appear in multiple top-16 lists, they are ranked differently in all cases but one where AS6939 is ranked first in both rankings by peer degree ($\Re_R$) and by betweenness centrality ($\Re_B$). Among those observations, the last one has important implications because it reflects the degree of concordance/discordance among ranking schemes when applied to the entire set. To illustrate using a toy example, let $S = \{s_1, s_2, s_3, s_4\}$ be a set of objects. Let $\mathcal{O}_X = <s_1, s_2, s_3, s_4>$, $\mathcal{O}_Y = <s_2, s_1, s_3, s_4>$ and $\mathcal{O}_Z = <s_4, s_2, s_3, s_1>$ be three orderings of S based on different criteria. $\mathcal{O}_X$ and $\mathcal{O}_Y$ as well as $\mathcal{O}_X$, $\mathcal{O}_Z$ have a single element swap in their orderings; $s_1 \leftrightarrow s_2$ and $s_1 \leftrightarrow s_4$, respectively. However, the effects of these swaps are significantly different and become more visible when the rankings are represented as "precedence" relations: $\mathcal{R}_X = \{(s_1, s_2), (s_1, s_3), (s_1, s_4), (s_2, s_3), (s_2, s_4), (s_3, s_4)\}$, $\mathcal{R}_Y = \{(s_2, s_1), (s_2, s_3), (s_2, s_4), (s_1, s_3), (s_1, s_4), (s_3, s_4)\}$ and $\mathcal{R}_Z = \{(s_4, s_2), (s_4, s_3), (s_4, s_1), (s_2, s_3), (s_2, s_1), (s_3, s_1)\}$ Specifically, $\mathcal{R}_X$ and $\mathcal{R}_Y$ have one discordant pair while $\mathcal{R}_X$ and $\mathcal{R}_Z$ have five discordant pairs out of six pairs. That is, the level of disagreement between orderings $\mathcal{O}_X$ and $\mathcal{O}_Y$ is much smaller than the disagreement between $\mathcal{O}_X$ and $\mathcal{O}_Z$.

Figure 3 shows the amount of concordance/discordance among the six ranking schemes applied to the AS-level topology map of the Internet as of June 2016. In the figure ranking schemes on the vertical axis of the grid are compared to the ranking schemes on the horizontal axis. The counter-diagonal line of the grid shows perfect agreement because each ranking scheme is compared to itself. Each heat map in the figure shows the degree of agreement/disagreement between two ranking schemes. AS pairs preserving their relative order in both ranking schemes are considered to be concordant and shown as gray in the heat maps. On the other hand, AS pairs fail to preserve their relative order are considered to be discordant and shown as black in the heat maps. Note that the AS names on the axes of the heat maps are omitted to reduce the clutter in the images. Each heat map is symmetric around its counter-diagonal because if AS_i and AS_j are concordant (discordant) so are AS_j and AS_i. The grid is also symmetric by the same argument. The amount of gray (black) reflects the degree of agreement (disagreement) in the heat maps.

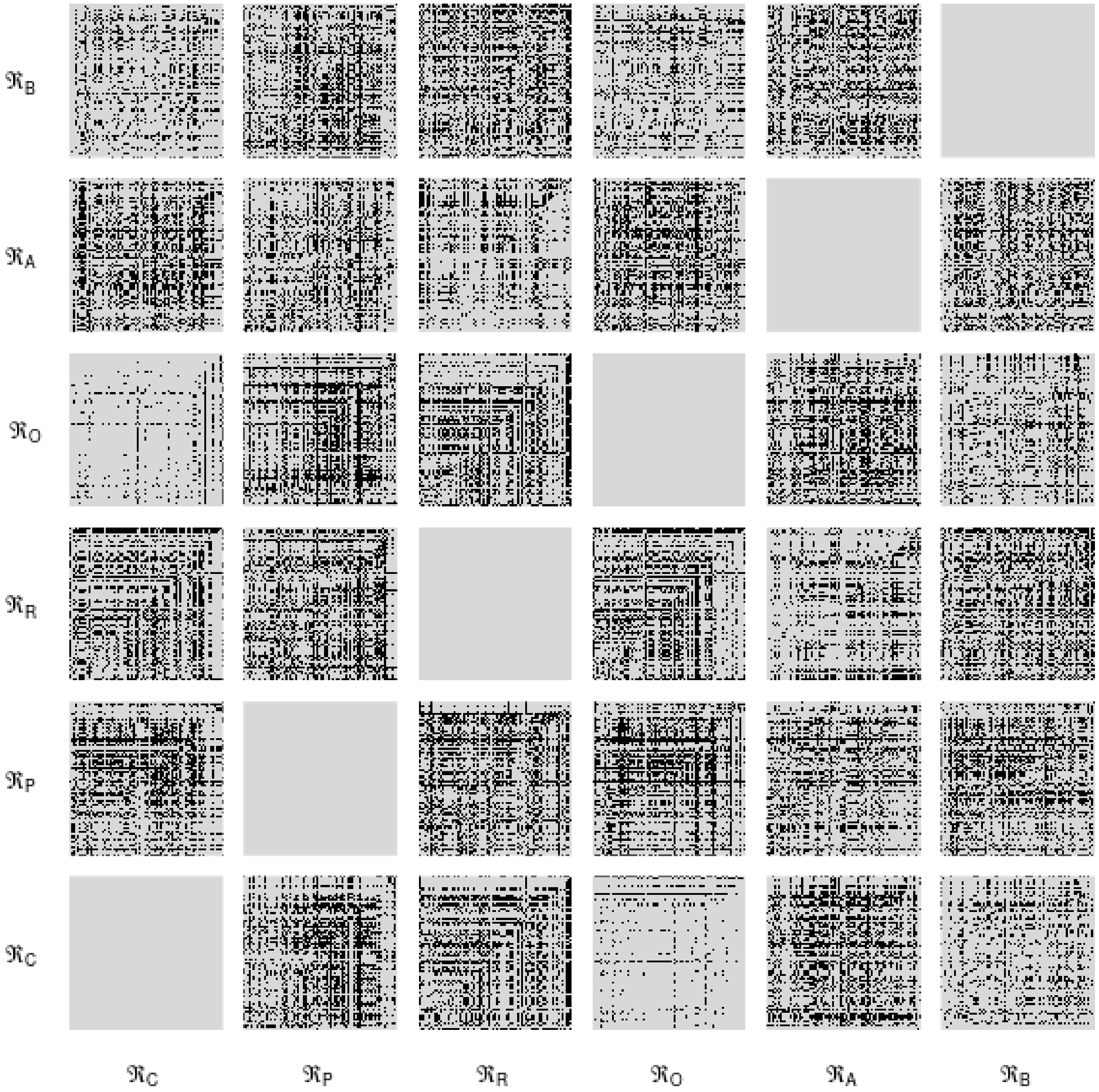


Fig. 3: Levels of agreement/disagreement between pairs of ranking schemes. Gray represents concordant pairs of ASes and black denotes discordant pairs of ASes. AS names on the axes of the heat maps are omitted to reduce the clutter in the images.

That is, the higher (lower) the amount of gray the higher the degree of agreement (disagreement) between the ranking schemes. Figure 3 displays the highest amount of agreement between rankings by customer degree ($\Re_C$) and by customer-cone size ($\Re_O$). Moreover, rankings by customer degree ($\Re_C$) and betweenness centrality ($\Re_B$) as well as rankings by peer degree ($\Re_R$) and alpha centrality ($\Re_A$) show moderate levels of agreement. The rest of the heat maps do not indicate any strong agreement or disagreement between the ranking schemes.

In order to shed more light on the agreement/disagreement levels given in Figure 3, we analyze the rank correlation coefficients between the pairs of ranking schemes. A rank correlation coefficient is a statistic for measuring the strength of ordinal association between two ranking schemes. τ_x, an extension to Kendall's τ_b, is a non parametric statistic that shows the strength of association between two variables measured on ordinal scale. τ_x is congruent with Kemeny-Snell distance and handles ties and weak orderings better than τ_b [5]. Compared to Spearman's ρ, τ_x is less sensitive to errors and discrepancies in the data. Given two rank matrices $\boldsymbol{R}_A$ and $\boldsymbol{R}_B$ (Equation 2) over a set S, τ_x is defined as:

$$\tau_x = \frac{\sum_{i=1}^{|S|}\sum_{j=1}^{|S|}\boldsymbol{R}_A[i,j]\boldsymbol{R}_B[i,j]}{|S|(|S|-1)} \tag{3}$$

where $|S|$ denotes the cardinality of S. The numerator of Equation 3 increases as the rank matrices $\boldsymbol{R}_A$ and $\boldsymbol{R}_B$ are concordant on the relative orderings of pairs of objects. Similarly, the numerator decreases for discordant object pairs under the two ranking schemes. τ_x takes values between -1 and 1 such that -1 denotes perfect disagreement and 1 denotes perfect agreement between the two ranking schemes.

TABLE II: τ_x rank correlation coefficient

	$\Re_C$	$\Re_P$	$\Re_R$	$\Re_O$	$\Re_A$	$\Re_B$
$\Re_C$	1.00	0.28	0.29	0.85	0.26	0.59
$\Re_P$	0.28	1.00	0.28	0.26	0.37	0.29
$\Re_R$	0.29	0.28	1.00	0.28	0.59	0.28
$\Re_O$	0.85	0.26	0.28	1.00	0.26	0.53
$\Re_A$	0.26	0.37	0.59	0.26	1.00	0.33
$\Re_B$	0.59	0.29	0.28	0.53	0.33	1.00

Table II shows the rank correlation coefficient matrix of the six ranking schemes used in this study. The table is consistent with Figure 3. Rankings by customer degree ($\Re_C$) and by customer-cone size ($\Re_O$) have the highest amount of correlation, 0.85. The ASes ranked higher in terms of the customer-cone size ($\Re_O$) are naturally located in the core of the semi-hierarchical Internet topology. However, analyzing the data further reveals that those ASes also accommodate more customers compared to the ASes located further down in the topology. Hence, the high rank correlation between $\Re_C$ and $\Re_O$ naturally arises in the topology. This observation is compatible with the reported short distance between the ASes at the periphery of the Internet and the ASes at the core of the Internet [14].

In addition, Table II demonstrates a moderate level of agreement, 0.59, between rankings by peer degree ($\Re_R$) and alpha centrality ($\Re_A$). Peering allows two ASes to exchange the traffic between each other and their customers without crossing their provider ASes. Typically, the ASes in the middle of the Internet engage in settlement free peering relations to reduce their overall cost of operations. In alpha centrality, on the other hand, each node has an amount of exogenous centrality score and it transfers some of its centrality score to its neighboring ASes. ASes having large number of peers receive more centrality score, hence those ASes ranked higher in terms of peer degree are also ranked higher in terms of alpha centrality in the topology.

A moderate level of rank correlation also appears between ranking by betweenness centrality ($\Re_B$) and ranking by customer degree ($\Re_C$), 0.59. Betweenness of an AS reflects its centrality in the topology in terms of the traffic that it transfers between pairs of ASes. ASes having more customers naturally have higher betweenness centrality scores because they directly connect their customers to the rest of the Internet. As a result, ASes ranked higher in terms of customer degree are also ranked higher in terms of betweenness centrality in the Internet topology.

Rankings by betweenness centrality ($\Re_B$) and ranking by customer-cone size ($\Re_O$) also demonstrate a moderate level of correlation, 0.53. ASes having large number of descendants in the Internet appear more frequently on the paths between pairs of ASes. Because, unless there is a peering relation among the upstream providers of an AS pair, a path between them goes all the way up from the source and then all the way down to the destination. Therefore, those ASes ranked higher in terms of customer-cone size are also ranked higher in terms of betweenness centrality. Moreover, the customer degree of an AS contributes into the customer-cone set of an AS. Those ASes having high customer degree also have high customer-cone size and they appear more frequently on the paths from/to their customers.

Finally, there is no significant level of rank correlation among other pairs of ranking schemes.

In summary, despite the existence a few high/moderate levels of agreement between structural AS ranking schemes, the ranking schemes display significant variability. We believe that it is necessary to carefully select a ranking scheme depending on the problem at hand. Moreover, one may need multiple ranking schemes to gain a diverse insight on the topology. Specifically, ranking by customer degree ($\Re_C$) emphasizes the ASes that provide Internet access service mostly to the ASes at the periphery of the Internet. Ranking by provider degree ($\Re_P$) gives priority to those ASes that are more resilient. That is, they have more provider ASes to distribute traffic over or switch over in case of a failure. Ranking by peer degree ($\Re_R$) stresses the ASes that have lower operational costs due to peering agreements. Ranking by customer-cone size ($\Re_O$) emphasizes the ASes located at the core of the topology. Ranking by alpha centrality ($\Re_A$) gives higher ranks to the ASes connected to other highly ranked ASes. Ranking by betweenness centrality ($\Re_B$) gives priority to those ASes carrying most of the traffic between pairs of ASes.

V. CONCLUSIONS

The Internet is one of the largest man-made complex systems. As of June 2016 it consists of 54,140 autonomous systems connected to each other through 466,190 relations. Extracting useful information from the AS-level topology of the Internet and transforming the information into an understandable format is of utmost importance. Ranking the ASes in the Internet serves as an auxiliary tool to help network practitioners to manage and troubleshoot their networks; provides a comparative guidance to chief information officers (CIOs) to negotiate new business relations among ISPs; and provides an immediate insight to network researchers to study the connectivity of the Internet.

In this work we compare and contrast six different AS ranking schemes based on the topological features of ASes: customer degree, provider degree, peer degree, customer-cone size, alpha centrality and betweenness centrality. Those six centrality measures demonstrate the local (degree centralities), global (betweenness and alpha centralities) and architectural (customer-cone size and design-specific degree centralities) characteristics of the ASes in the Internet.

We observe a high amount of agreement between rankings by customer degree and by customer-cone size. Our analysis show that the ASes at the edge of the Internet tend to connect to the ASes located more toward the core of the Internet. Rankings by peer degree and alpha centrality show a moderate level of agreement because ASes having large number of peers receive more alpha centrality score from the peering neighbors. Rankings by customer degree and betweenness centrality also demonstrate a moderate level of agreement. Our analysis show that those ASes having a large customer base also provide the primary Internet access service and they appear more frequently on the paths between pairs of ASes. Rankings by betweenness centrality and customer-cone size display a moderate level of agreement because as the customer-cone set gets larger for an AS the chance to appear on a path from/to its customers also increase.

Our results show that it is necessary to carefully select a ranking scheme depending on the problem at hand. Furthermore, one usually needs to select multiple ranking schemes to gain a diverse insight on the topology of the Internet.

REFERENCES

[1] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira. Avoiding traceroute anomalies with Paris traceroute. In *Proceedings of ACM Internet Measurement Conference*, Rio de Janeiro, Brazil, Oct 2006.

[2] P. Bonacich and P. Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social networks*, 23(3):191–201, 2001.

[3] CAIDA. Dataset. http://data.caida.org/datasets/as-relationships/serial-2/20160501.as-rel2.txt.bz2.

[4] F. Clérot and Q. Nguyen. A social network approach for the ranking of the autonomous systems of the internet. July 2005.

[5] E. J. Emond and D. W. Mason. A new rank correlation coefficient with application to the consensus ranking problem. *Journal of Multi-Criteria Decision Analysis*, 11(1):17–28, 2002.

[6] L. Gao. On inferring autonomous system relationships in the internet. *IEEE/ACM Transactions on Networking*, 9(6):733–745, Dec 2001.

[7] V. Giotsas, M. Luckie, B. Huffaker, and K. Claffy. Inferring Complex AS Relationships. In *ACM IMC*, Nov 2014.

[8] P. R. Gould. On the geographical interpretation of eigenvalues. *Transactions of the Institute of British Geographers*, (42):53–86, 1967.

[9] J. Hawkinson and T. Bates. Guidelines for creation, selection, and registration of an Autonomous System (AS). RFC 1930, Mar 1996.

[10] M. Kendall and J. D. Gibbons. *Rank Correlation Methods*. A Charles Griffin Title, 1 edition, 1948.

[11] M. Luckie, B. Huffaker, K. Claffy, A. Dhamdhere, and V. Giotsas. AS Relationships, Customer Cones, and Validation. In *Internet Measurement Conference (IMC)*, pages 243–256, Oct 2013.

[12] B. P. and L. P. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23(3):191 – 201, 2001.

[13] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Nov 1999.

[14] M. E. Tozal. The Internet: A system of interconnected autonomous systems. In *IEEE Systems Conference*, Orlando, FL, USA, Apr 2016.

[15] C. Wagner, J. François, R. State, A. Dulaunoy, T. Engel, and G. Massen. Asmatra: Ranking ass providing transit service to malware hosters. In *Integrated Network Management, IFIP/IEEE*, May 2013.

[16] L. Zimmerli, B. Tellenbach, A. Wagner, and B. Plattner. Rating autonomous systems. In *Internet Monitoring and Protection (ICIMP)*, May 2009.