Enhancement Techniques for Network Topologies in Information Technology

[*]G.Srinivasa Rao, [**]Dr.Prof.Arun Singh, [*]Dr.G.Appa Rao, [*]D.Kishore Kumar, [*]M.Prasada Rao
[*]D.Veerabhadra Rao, [*]P.Surya Chandra, [*]S.Venkata Lakshmi [*]D.Ramya Mallika, [*]N.Rajender

[*]Gitam University., [**]Magadh University

Abstract
Recently, several authors have proposed models based on power-laws to characterize Internet topologies. Most of these works use the BGP (Border Gateway Protocol) tables published by Oregon Route Views. The adjacency matrix containing AS (autonomous system) connectivity is built from a BGP table. Having access to BGP routing tables from several geographical sites gives a broader vision of ASs connectivity, since several ASs and links may be hidden for an AS due to routing processes policies, while they may be visible to other ASs. We compare BGP tables of different sizes and enrich the adjacency matrix with the union of them. This comparison is based on the AS degree connectivity, clustering coefficient and path length.

Key Points
Power-laws, BGP, Autonomous Systems.

1. Introduction

An autonomous system (AS) is a set of routers under a single technical administration, using one or various IGPs (Interior Gateway Protocol, i.e., RIP, OSPF, etc.) and common metrics to route packets inside the AS and using an EGP (Exterior Gateway Protocol, i.e., BGPv4) to route packets to other ASs. BGP comes from Border Gateway Protocol and is an inter-autonomous system routing protocol. BGPv4 is extensively used to connect ISPs (Internet Service Providers) and to interconnect enterprises to ISPs. ISPs usually are providers (provide connectivity) of other ISPs that at the same time are providers of smaller ISPs. In the periphery of the Internet there are end ISPs that usually give services to enterprises that are not ISPs. ISPs can be classified as transit ISPs when they offer transit of traffic, multihomed ISPs when they are connected to more than one ISP and do not offer transit of traffic, and stub ISPs when they are connected to other ISPs only. An ISP can have more than one AS number assigned and give services to other ISPs on large geographical areas. We will consider for simplicity that ISPs are single autonomous systems, and we will study topologies based on BGP interconnectivity. These assumptions are not far from reality since the Internet topology is based on inter-domain interconnectivity and routing policies handled by BGP.
The primary function of BGP is to exchange network reachability information with other BGP peers on neighbour ASs. ASs can apply different policies to the way they import/export routes when using BGP. Routing policies define how routing decisions are taken in the Internet. If we have two networks separated by two outers belonging to different ASs, the policy comes into play when one AS decides to announce networks (prefix routes) to the other AS. Import policies allow an AS to transform incoming route updates. For example, denying or allowing an update, assigning a local preference attribute to an incoming route depending on the AS origin, AS path, etc. ASs only send their best route to their neighbours. Export policies allow an AS to determine whether to send this route to a neighbour and, if it does, to send the route with or without hints such as the community attribute or the MED (Multi Exit Discriminator). A route is expressed as a prefix, i.e., a group of one or more networks. Routing policies are not applied to each prefix separately but to a group of prefixes defined at the AS. BGP allows reachability information exchange among BGP peers of the same AS or of different ASs. This information will allow us to construct a graph of ASs connectivity, this is to say, the Internet topology at autonomous systems level. A BGP peer builds a routing table database consisting of the set of all feasible aths and the list of networks (prefixes) reachable through each feasible path or AS path. An example of a BGP table isshown in Table 1, where the symbol ‘‘>’’ expresses the ‘‘best path’’. As BGP routers only send their best path to BGP peers, a BGP router will have a particular view of the Internet topology depending on where it is placed. In order to have a wider perspective of the Internet topology, it is necessary to study the union of several BGP tables of different places, as we do in this work. Oregon Route Views, is a repository that saves every two hours the BGP tables of ASs connected to the BGP route repository. Oregon Route Views uses AS6447 and it is currently connected to 60 neighbors.
Based on the work and making use of the Oregon Route Views BGP table, many researchers have recently investigated the Internet topology. These studies are relevant to the development of Internet topology generators, the deployment of content networks, the placing of web servers or in the development of inter-domain traffic engineering models, etc.

Manuscript received November 5, 2009
Manuscript revised November 20, 2009

Authors of argue that tables taken from only one point such as the Oregon Route Views give a poor vision of the Internet and they propose to use more than one point of vantage. Authors from CAIDA, propose to generate active probes to complete the AS interconnectivity. Bu and Towsley, propose using the CCDF (complementary cumulative distribution function) to better fit the power-law behaviour described by Faloutsos et al. Finally, Lakhina et al, describe sampling biases in IP topology measurements using traceroutes. Here, we compare several BGP tables from different geographical sites and their union, using classical network topology measures: AS path distribution, clustering coefficient and degree distribution together with information about the number of ASs and the number of edges seen in the different perspectives. We will compare tables of different sizes and from ASs with different interconnectivity and we will join the repository of Route Views with these BGP tables.Other authors have centred their study on the novel definition of complex networks. A complex network shows certain organization principles that are encoded in its topology. These works study €three main topology models: the classical Erdos -Renyi model for random graphs, the small-world model motivated by small paths between two nodes and high clustering coefficients and the scale-free model that presents power-law degree distributions.

We will discuss the use of scale-free models to generate Internet topologies. In particular we study the application of scale-free models to the whole BGP table and to part of the BGP table. Using well-known heuristics proposed by Gao in, we infer peering relationships between ASs and take away end customers and small ISPs from the BGP table. In this way we can analyze the use of scale-free models applied to the core of the Internet and identify the kind of peering of any AS.

In Section 2 we define the metrics selected to compare the BGP tables. Section 3 is devoted to the explication of the methodology used. Section 4 shows the obtained results applied to the whole BGP table. Sections 5 shows the results applied to parts of the BGP table. Section 6 finalizes with the conclusions of the work.

2. Metrics

In this section we define the metrics selected to compare our sources of data. For that purpose we consider the AS level topology as the graph G=(N,E), where N is the number of vertices or ASs and E the number of edges or links that connect the vertices. We define the adjacency matrix A as a symmetric matrix of size NxN with components $a_{ij}=1$ if node i has an edge joining node j and 0 otherwise. At each of the different BGP tables we will investigate the AS degree rank, the AS degree CCDF (complementary cumulative distribution function), the AS path length and the clustering coefficient.

2.1. Rank degree

This metric ranges ASs in decreasing order of degree d and plots the pair degree d versus rank r in log–log scale. The degree, d_i , of node i is the number of neighbours directly connected with it, of so $d_i=\sum a_{ij}$ for all j. This metric was introduced by M. Faloutsos, P. Faloutsos, C. Faloutsos with the name of Power-Law 1 (rank exponent) and establishes that the degree, d_i , of AS i is proportional to the rank of this AS, r_i , to the power of a constant η:$d_i \propto r_i^{\eta}$, where the symbol $\propto$ means proportional.

Table 1
Example of a BGP table

Network	Next hop	Metric	Weigh	Path
*> 0.0.0.0	195.66.225.254		0	5459 i
* 3.0.0.0	204.42.253.253		0	267 1225 701 80 i
*	206.157.77.11	70	0	1673 701 80 I
*	12.127.0.249		0	7018 701 80 I
*>	204.70.4.89		0	3561 701 80 I
*	205.158.2.126		0	2828 701 80 I
*	158.43.206.96		0	1849 702 701 80 i

2.2. Degree CCDF

Faloutsos et al,introduced the metric f_d as the fraction of nodes with degree d and demonstrated that it follows a power-law of the type: $f_d \alpha d^{\xi}$, the exponent ξ is obtained using a linear regression. Bu and Towsley, define the empirical complementary distribution(ecd) as the CCDF of the degree distribution. This metric plots the fraction of nodes with degree greater than or equal to d versus the degree of the AS. In a probabilistic sense the ecd is defined as F_d= Prob{D ≥d}=∑ f_i for d≤i<∞, where D is a random variable that indicates the number of incident neighbours upon an AS, i.e., its degree. F_d also follows a power-law with exponent α,F_d α d^{α} . Note that rank degree distributions emphasize the degree of the largest ASs while the degree CCDF emphasizes the degree of the smallest ASs.

2.3. AS path-length

To measure the AS path-length, l, we will obtain the CCDF of the AS path length, F_l . This metric plots the cumulative fraction of path-lengths greater than or equal to *l* versus *l*. In a probabilistic P sense Fl could be seen as F_l= Prob{L ≥l}=∑ f_i for l≤i<∞. L is a random variable, limited by 1 ≤ L ≤ L_m , where L_m is the longest AS path found in the BGP table, f_l is the fraction of paths with length l and l is the number of ASs traversed to reach the target.

2.4. Clustering coefficient

The clustering coefficient, C, is a concept used in the theory of small-world networks and is a metric that indicates the grade of connectivity of any node, by definition 0 ≤ C ≤ 1. The clustering coefficient C_i for any node i in the graph is defined as the ratio between the number of connections among the d_i neighbours of a given node i and its maximum possible value,d_i (d_i-1)/2where di is the degree of node i. C_i is the fraction of the edges that actually exists, and C(G) is the average value of C_i :

$$C(G) = \frac{1}{N^{\geq 2}} \sum_{i \in G} C_i$$

$$C_i = \frac{\#\text{ of edges in } G_i}{d_i(d_i - 1)/2}$$

In the above equations G_i is the subgraph of node i, and is defined by taking only the neighbours of node i into account. C_i is only defined for ASs with degree d_i ≥2. This is because for d_i =1, Ci is undetermined. $N^{\geq 2}$ is the set of ASs with degree d_i ≥ 2 and $N^{=1}$ is the set of ASs with d_i =1 and N =$N^{\geq 2}$+$N^{=1}$. The clustering coe□cient provides a measure of how well the neighbours of a node are interconnected. Fully connected networks have a clustering coefficient C=1. A network of isolated nodes has C= 0.

3. Methodology

In this work, we use besides the data of Route Views six public available BGP table. In order to get the adjacency matrix, $\{a_{ij}\}$, we need to analyze BGP routing tables, which provide us with AS paths and links contained in them. It is important to note that BGP is a protocol of peering relationships and not of physical connections. For that reason the local view of an AS located in Europe could be (and is) different of an AS located in Asia and so on. This fact motivated us to investigate the differences and similarities of the distinct BGP tables. We capture 6 BGP tables, 2 located in the USA, 3 in Europe and 1 in Asia.
Table 2 shows these data sources and their parameters. All BGP tables were collected over the same period of time and the difference in collection times between the first and the last one was 1 week. Of these tables Oregon and Ripe-rcc are remote route collectors, which have a repository where the complete data can be obtained via an anonymous ftp. Ripe-rcc is a set of 9 remote route collectors, 7 deployed in Europe, 1 in Japan and 1 in the USA. The rest of the data sources were downloaded using the CISCO or Zebra ''sh ip bgp'' command via telnet to the site. Swinog is a medium size AS with 41 neighbours and Exodus Comm. Europe, Exodus Comm. Asia and Opentransit are leaf ASs with only one neighbour. See Table 2.

Table 2
BGP tables analyzed

Source of data	Total # of ASs	Total # of links	Max. degree	AVG degree	Neighbours
Exodus Comm. Eur.	13801	17311	1616	2.5086	1
Opentransit	9413	12245	1075	2.6017	1
Swinog	14018	28653	2609	4.0880	41
Exodus Comm. Asia	11264	15819	2041	2.8087	1
Oregon Route View	14154	28918	2606	4.0861	61
Ripe-rcc	14053	27122	2601	3.8599	170

From Table 2, we can see also that Swinog, Ripe-rcc and Oregon are numerically (around 14000 ASs seen) very similar in spite of the difference of neighbours. The three leaf ASs capture less AS connectivity, where Opentransit is the weakest with only 9413 ASs seen. Another interesting fact is the column AVG degree (average degree), that is the ratio between the number of links (doubled) and nodes in the graph. For small ASs the average degree is around 2.8 while for Swinog and the repository tables, the average degree is 4.08. We can observe that a repository table has almost the same average degree like a medium AS such as Swinog.

4. The whole BGP table

In this section we will investigate the AS connectivity. We analyze each of the BGP Tables and their union.

4.1. Rank degree and degree CCDF

Fig. 1 shows the rank degree in a log–log plot for the BGP Tables of Opentransit and Oregon, that is the strongest and the weakest in a numerically sense, respectively. The parameters of the Power-Law 1 for every one of the data sources are shown in Table 3, where η is the pendent of the curve, c is a constant calculated as $d_{i=}cr^{\eta}$ when η=0 and R^2 is the correlation coefficient.
We observe from Fig. 1, that the curve shape of Opentransit is similar to Oregon, and this behaviour is followed by the other four data sources (for the sake of clarity, the other curves are not shown in the figure). Repositories such Oregon or RIPE-rcc and Swinog have more neighbours and therefore more complete BGP tables than leaf ASs. This fact makes that the degree of the larger nodes in Oregon and RIPE-rcc BGP tables is higher than the degree of the larger nodes in the leaf ASs BGP tables.
In other words, the adjacency matrix, A, is denser for BGP tables that contain more paths.

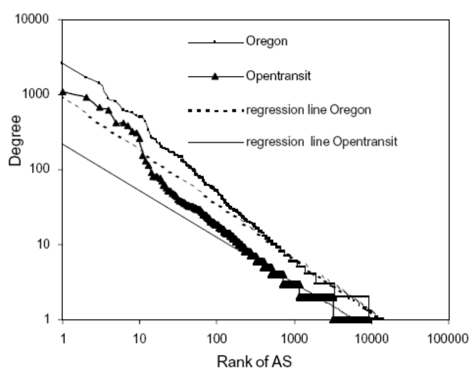


Fig. 1. Power-Law 1 (rank exponent).

This observation allows us to conclude that the union of all the BGP tables may offer better samples than individual ones. With respect to the correlation coefficient, R^2 , we can say that this is a low value for each table and that the regression line does not fit very well with the data except for high degree ranks. These high ranks correspond with the long tail of the data, that is, ASs with degrees lower than 10 for the Opentransit table and lower than 40 for the Oregon table. More than 80% of the ASs have a low degree. Furthermore, the denser the adjacency matrix A is, the better the power-law fits with higher degrees. As a conclusion, we can say that the rank (and therefore f_d) does not fit a power-law well due to the deviation of the low ranks (high degrees) with respect to the regression line.
The degree CCDF, F_d , seems to better fit a power-law than the rank degree or the density f_d . Fig. 2 plots the degree CCDF of Oregon and Opentransit. As in Fig. 1, the curve shape of Opentransit is similar to Oregon. This behaviour is followed by the other four data sources (not shown in the figure).
The parameters of the power-law are shown in Table 4. The power-law is calculated as $F_d=cd^{\alpha}$ with α and c being constants and R^2 is the correlation coefficient. We notice that the correlation coefficients, R^2, are higher than the ones obtained with power-law 1 indicating a better fitting. We observe that Ripe-rcc has the highest coefficient of correlation producing an exponent α=-1.197. Again repositories or ASs such Swinog have good correlation coefficients, however the repository ones have parameters that are not much better than those of a medium size AS such as Swinog.

Table 3
Power-Law 1

Source of data	c	η	R^2
Oregon	965.62	−0.7226	0.9318
Opentransit	216.68	−0.616	0.9098
Swinog	1062.8	−0.7365	0.9375
Exodus Comm. Europe	233.1	−0.6023	0.8867
Exodus Comm. Asia	324.77	−0.6443	0.9251
Ripe-rcc	829.46	−0.7083	0.9315

Table 4
Power-law CCDF parameters

Source of data	C	α	R^2
Oregon	1.0988	−1.2368	0.9735
Opentransit	0.2874	−1.0759	0.9273
Swinog	1.2773	−1.261	0.971
Exodus Comm. Europe	0.2156	−1.0502	0.9583
Exodus Comm. Asia	0.3094	−1.104	0.955
Ripe-rcc	0.7968	−1.1975	*0.9763*

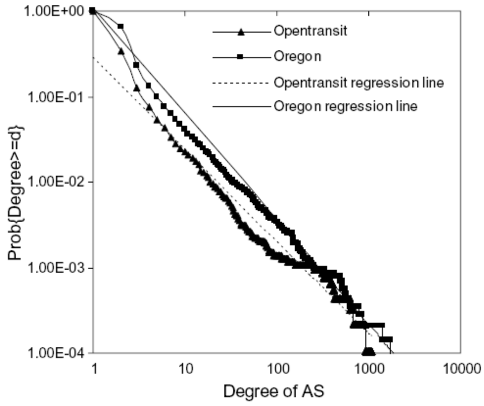


Fig. 2. Degree CCDF Prob{Degree} ⩾ d.

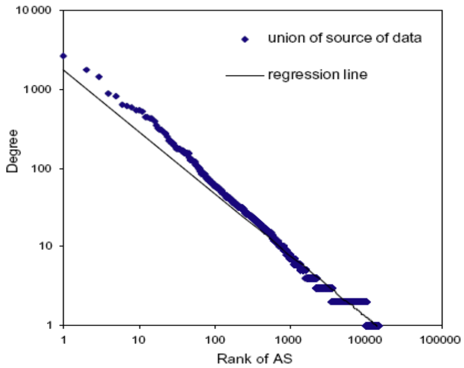


Fig. 3. Power-Law 1 of the six data sources.

4.2. Rank degree and CCDF of the Union

Fig. 3 shows the Power-Law 1 of the union of the six data sources. The parameters of its regression line are c =1735.2, η=0.7838 and R^2 =0.9402. The more complete the adjacency matrix A is, the better the power-law fits. That is, the correlation coefficient R^2 of the union is higher than that of any single source of data.
The last statement about the adjacency matrix is confirmed with the degree CCDF, when all the adjacency matrices of all the data sources are aggregated. Fig. 4 shows the degree CCDF of the union of the six data sources. The parameters of its regression line are c =2.1929 α= -1.3287 and $R^{2=}$ 0.9769.
The evolution of the union of the data sources is shown in Table 5. We observe an increment with respect to Oregon of 16.11% on the number of links (edges) with only an increment of 0.71% of ASs. This is an important observation, because due to the nature of BGP some ASs remain hidden from others, while some links do not appear due to the fact that only best paths are exported to other ASs. If we had more sources of data located in different sites we would have had a very much complete picture of the Internet topology at the AS level. Unfortunately there are few web sites where a complete BGP table is freely available.

Table 5
Evolution of the union of data sources

Source of data	Total # of ASs	Total # of links	Max. degree	AVG degree
Asia + Swinog + Opentransit = A	14056	28989	2619	4.12479
A + Exodus Europe = B	14073	29060	2620	4.12989
B + Oregon = C	14233	32441	2635	4.55856
C + Ripe-rcc	14256	33578	2639	4.7107

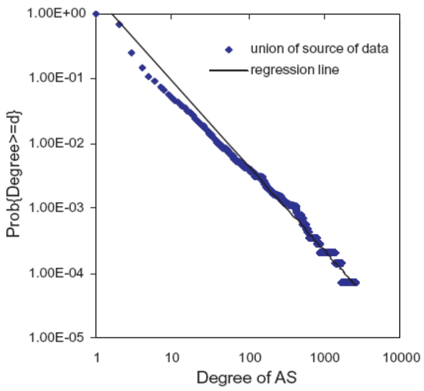


Fig. 4. CCDF of the union of the six data sources.

4.3. AS path length

AS path length measures the separation between two ASs. The AS path length is the number of nodes traversed to reach the target AS. In practice BGP only considers the ‘‘best path’’, which is signalled with the symbol ‘‘>’’ in the BGP table, like we observe in Table 1.
In Fig. 5, we show the CCDF, F_1 , of the path length for the Opentransit BGP table and the union of all the tables. We notice that the probability of having small paths is greater than the probability of having long paths. All individual tables have similar AS path length distributions. We also show in the figure the regression line of the union of the tables. The curves fit an exponential-law: $y = 3.5359e^{-0:7335x}$

and a correlation coefficient R^2 =0.9911. This high value of R^2 confirms the validity of the exponential-law to model the AS path length.

4.4. Clustering coefficient

Table 6 shows the clustering coefficient for each data source and the union of all ones. In this table we note that leaf ASs have the highest number of ASs with degree 1. Oregon has the lowest number, 4812, of ASs with degree equal to 1. If we aggregate the tables, the adjacency matrix A is denser and the number of ASs with degree 1 decreases to 4506. That means, 6.35% less with respect to Oregon, 11.43% less with respect to Swinog and 27.84% less with respect to Opentransit.

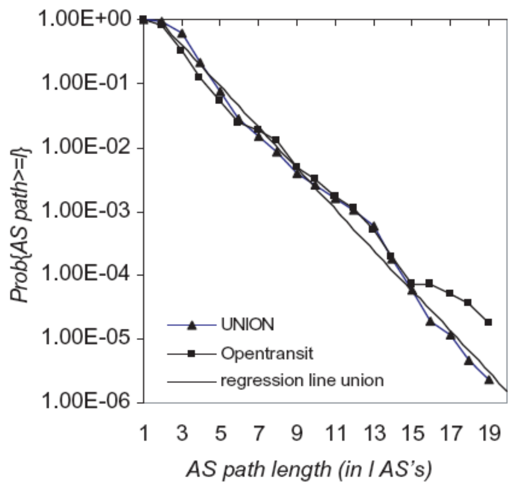


Fig. 5. CCDF of the AS path length, F_l.

74.88% of the ASs have degree 1 or degree 2. These are the leaf ASs that provide services to enterprises. Normally, leave AS are connected to two ISPs in order to obtain redundancy in case one ISP connection fails. Intuitively, one can imagine that the average AS path length has to be small, since one leaf AS will traverse a few number of medium or big AS to reach another leaf AS.

If we compare Opentransit with Exodus Comm Europe, two leaf ASs, we can note than Exodus sees a higher number of ASs, 13801 (see Table 2) than Opentransit, 9413 ASs. However, the clustering coefficient is higher for Opentransit. This observation is accentuated if we compare Exodus Comm Europe (13801 ASs and C = 0.179) with Exodus Comm Asia (11264 ASs and C=0.305). This means that a leaf AS that has a higher number of ASs in its BGP table sees less AS interconnectivity than a leaf AS with a lower number of ASs in its BGP table. Furthermore, a medium AS such as Swinog has a clustering coefficient similar to Oregon, a repository of 61 ASs. Joining tables from different geographical locations, even if they are of different sizes, improves the inter connectivity picture.

Finally, following the results we can talk about the duality of the topology of the Internet at the AS level, in which the Internet can be seen as a small-world network and a scale-free network. A small-world network is one in which the AS path length is small and the clustering coefficient is high (compared with a random graph). We can observe from Table 6 and Fig. 5 that the AS connectivity behaves as a small-world network with high clustering coefficients and small AS path lengths. However, a small-world network is not an indication of some type of organization principle (scale-free networks). This property is given for the presence of degree power-laws. From Fig. 2, we can observe that the degree CCDF follows a power-law, indicating scale-free behaviours.

Barabasi and Albert (AB model), propose a topological model that exhibits ''preferential attachments'' and ''constant growth''. A network exhibits a ''preferential attachment'' if the probability of connecting to a node epends on the node's degree. That means that a new AS in the Internet or an AS that wants to rewire a link would choose an AS with high degree.

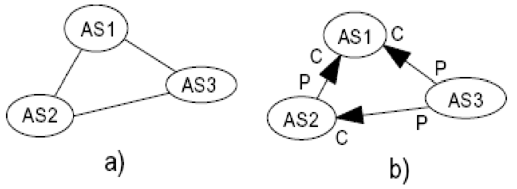
This model predicts a clustering coefficient following approximately a power-law $C \sim N^{-0.75}$, while for a random graph, C is expected to be C <d>/N , where <d> is the average degree. With N =14256 nodes, and <d>=4.71, see Table 5, the AB model predicts $C=7.66X10^{-4}$ and a random graph $C=3.3X10^{-4}$. The clustering coefficient obtained from the whole BGP table is C =0.46, much higher than predicted by the AB model or a random graph. A possible reason for this high clustering coefficient is that in the Internet around 75% of the nodes are end customers with a degree of 1 and 2. These nodes are not interconnected. The rest of the nodes are small ISPs and nodes that form the Internet core. These nodes are the ones that contribute to the clustering coefficient.

5. Splitting the BGP table in three regions

Recently, some works have questioned the application of the AB model to Internet topologies. The major criticism made to the AB model is that the model does not take into account the dynamics of the BGP routing in the Internet and the business and geographic preferences. Some of these aspects could be studied identifying the end customers from the small provider ISPs and the core-transit ISPs and knowing whether ASs with low degree are connected with nodes with large degree. We also are interested in knowing whether power-laws arise in indegree and out degree adjacency matrices. This fact would allow us to create topology generators taking into account relationships. For that purpose, we need to classify the AS peering relationships in order to identify the kind of neighbours a node has. Source data used in this section is the union of the several BGP tables.

Table 6
Clustering coefficient and ASs with degree 1 and 2

Source of data	N^1	$N^{=2}$	$N^1 + N^{=2}$	C
Oregon	4812	6054	10866	0.4235
Opentransit	6245	1992	8237	0.1979
Swinog	5088	5775	10863	0.4222
Exodus Comm. Europe	9945	2285	12230	0.1792
Exodus Comm. Asia	6766	2968	9734	0.3050
Ripe-rcc	4754	6022	10776	0.3912
Union of source of data	**4506**	**6170**	**10676**	**0.4624**



$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \quad A^I = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad A^O = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

Fig. 6. Degree, indegree and outdegree adjacency matrices.

Gao and Rexford, give a heuristic to infer AS relationships into provider-to-customer (P2C), peer-to-peer (P2P), and sibling (SIB) relationships. A customer exports to a provider its routes and its customer routes. A provider exports to a customer its routes, its customer routes, its provider routes and its peer routes. A peer exports to a peer its routes and its customer routes but not its provider or peer routes. A sibling exports to a sibling its routes, its customer routes, its provider routes and its peer routes and thus allows transit. Using ao's heuristic we can obtain a directed provider-to-customer graph from which we define two more adjacency matrices: the indegree adjacency matrix A^I and the outdegree adjacency matrix A^O . The indegree adjacency matrix A^I has components $\{a^I_{IJ}\}=1$ when node i is a customer of node j or node i is a peer or sibling of node j and $\{a^I_{IJ}\}=0$ otherwise. The outdegree adjacency matrix A^O has components $\{a^O_{IJ}\}=1$ when node i is a provider of node j or node i is a peer or sibling of node j and $\{a^O_{IJ}\}=0$ otherwise. Fig. 6 shows anij example of peering relationships.

The indegree, d_i^I , of node i is the number of providers directly connected with it, so $d_i^I = \sum a^I_{ij}$ O for all j. The outdegree, d^O_i , of node i is the number of customers directly connected with it, so $d^O_I = \sum a^O_{ij}$ for all j. We can note that a node such as AS1 has degree $d_i = 2$, indegree d^i_I =2 and outdegree d_i^O =0. Knowing the peeringship will allow us to obtain the core-transit nodes.

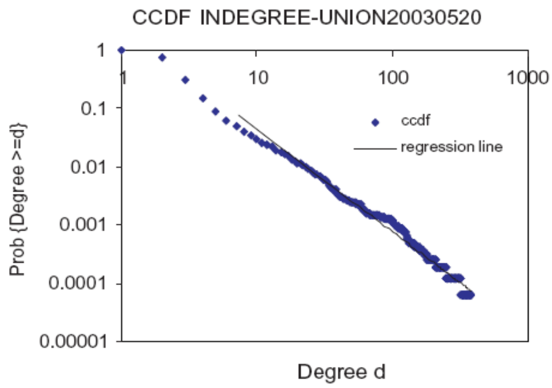


Fig. 7. Indegree CCDF ($R^2 = 0.983$).

Removing ASs with outdegree d_i^O =0, end customers that do not participate in the routing process are eliminated. Iterating this process, the Internet AS topology can be split in an end customer area, a small regional ISP area and a core area. Figs. 7 and 8 show the indegree and outdegree CCDF of the total BGP table. We can see that both follow a power-law with high correlation coefficients. However it is interesting to observe that the maximum indegree is nine times higher than the maximum outdegree.

Table 7 shows a classification of the relationships inferred from the union of data sources. The table shows the total number of ASs and link relationships and its distribution as end customers, small ISPs and the core of the Internet. From this table we can see that around 79.7% of the ASs are end customers with 61.6% of the links. These ASs may have degree higher than 1 but they do not participate in the routing process. 9% of the ASs are small provider ISPs with 8% of the relationships. These ASs act as transit between the end customers and the core of the Internet but they do not have any peer- to-peer relationships among themselves.

Finally, the core of the Internet has 11.18% of the ASs and 30.29% of the links. These AS are well interconnected among them forming a mesh topology. The Internet core can further be divided in more sets. However, it is not the purpose of this work to reproduce the same data. Our interest here is to know whether we can use the concept of scale-free networks in the core of the Internet.

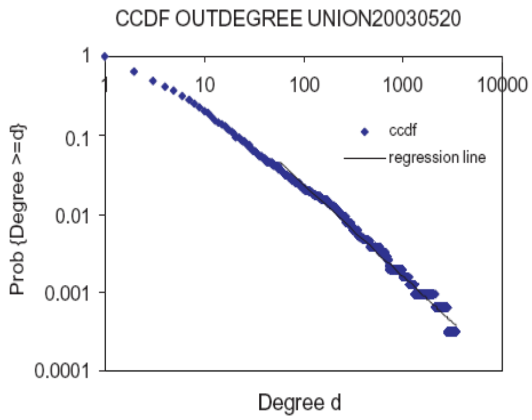


Fig. 8. Outdegree CCDF ($R^2 = 0.973$).

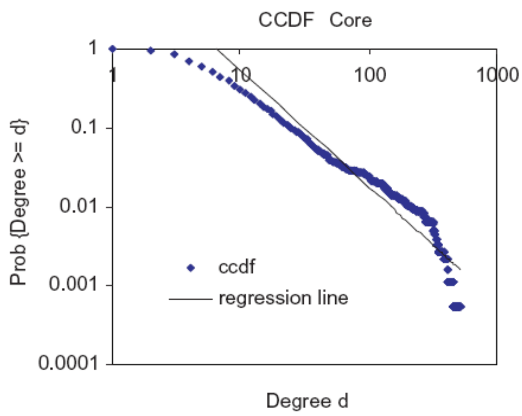


Fig. 9. Degree CCDF of the core-transit ISPs ($R^2 = 0.85$).

Fig. 9 shows the degree CCDF of the coretransit ISPs. We can see that the degree CCDF does not fit a power-law. Using a power-law, the correlation coefficient is R^2 =0.851, while using an exponential the correlation coefficient is R^2 = 0.935.

Table 8 shows some parameters obtained for the core-transit ISPs. The core-transit Internet has a clustering coefficient C = 0.48. This high clustering coefficient is due to the high connectivity of the 1817 nodes of the core-transit Internet. End customers and small ISPs with a total of 14427 ASs do not contribute to this clustering coefficient. Furthermore, the highest degree corresponds to an AS with 516 core-transit neighbours. The average degree of the core is 14.5 in comparison with average degrees around 4 for the whole BGP table.

All this data indicates a highly interconnected core-transit that should be modelled with different distributions than power-laws.

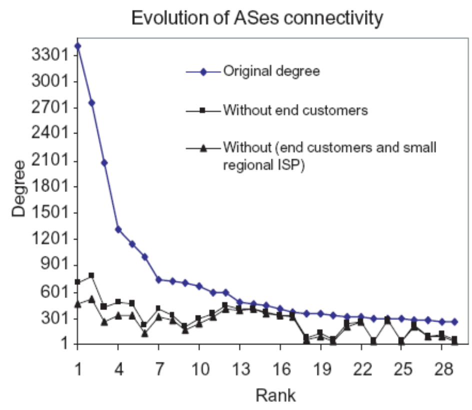


Fig. 10. Degree of the top 30 ASs with presence in the core.

Table 7
Number of end customers, small ISPs and ASs in the core and number of link relationships (P2C, P2P, Sibling)

	Number (#) of ASs/(#) links	% of ASs/links	P2C, P2P, SIB (#)	P2C, P2P, SIB (%)
Total number of ASs/links	16244/43483	100/100	38601, 4318, 564	88.71, 9.93, 1.3
End customers ASs/links	12954/26823	79.74/61.6	26823, 0, 0	61.6, 0, 0
Small provider ISPs ASs/links	1473/3483	9.06/8.01	3486, 0, 0	8.01, 0, 0
Core-transit ISPs ASs/links	1817/13174	11.18/30.29	8292, 4318, 564	19.06, 9.93, 1.3

Furthermore it is interesting to point out that nodes with very high degrees have a high component of end customers. Fig. 10 plots the degree of the top 30^0 ASs. The three curves show the degree of each AS with all their neighbours, the degree after removing end customers, and the degree after removing end customers and small ISPs.

We can see for instance that the node with the highest degree in the core with a degree of 512 (19%) nodes has a total degree of 2762 from which 1989 (72%) are end customers and 261 (9%) are small ISPs. That indicates that end customers and small ISPs could be added using a preferential attachment algorithm based on power-laws.

Table 8
Core-transit ISP data

#ASs	#links	Max. degree	Avg. degree	Cluster. coeff
1817	13174	516	14.5	0.485

6. Conclusion

In this paper, we compare BGP tables from different sites and of different sizes. We have chosen six complete BGP tables. We have obtained the adjacency matrix A of these tables and of the union of all the tables. Since the degree rank and the degree density function do not fit a power-law 1 very well, we have chosen the degree CCDF. We have shown that the degree CCDF follows a power-law $F_d \alpha d^{-\alpha}$ that fits better than the degree density function and that the more complete the data sources are, the better the regression line fits the model. We have taken into account the union of more than a hundred ASs. We have also compared BGP tables of different sizes. The results show that repository collectors capture more of the AS connectivity. However, medium ASs also capture most of the AS connectivity since they are also connected to several ASs. Taking the union of the BGP tables means seeing a denser adjacency matrix and therefore improves the observed AS interconnectivity. Also remark that the number of leaf ASs practically remains invariant, around a 75% of the ASs. Our result shows that to get a more complete picture of the AS connectivity, it is necessary to have access at more BGP tables. Having more tables means an increment in the number of links seen although the number of ASs seen remains practically invariant. Since the degree CCDF confirms a power-law degree distribution, this means that the AS topology can be seen as a scale-free network. Any network to be considered as a small-world network must comply two conditions: a high value of clustering coefficient and a small path length. Our results show that the AS connectivity of the Internet covers these considerations. Specially with respect to the clustering coefficient we can say that the more complete the data, the more the number of ASs with degree 1 tends to decrease and with this the coefficient tends to augment, which means an improvement in the connectivity of the Internet at the AS level. Furthermore, models such as the AB model do not model very well the AS connectivity using the whole BGP table. Splitting the BGP table in end customers, small ISPs and coretransit Internet nodes, allows us to see that the core-transit nodes are highly interconnected, contribute to the clustering coefficient and do not follow a power-law. However, the end customers and small ISPs that do not contribute to the clustering coefficient could be added according to a power-law. These results can help to produce better Internet topology generators that could take into account peering relationships.

References

[1] Public route server and looking glass list'', Available from <http://www.traceroute.org>.
[2] RIPE Routing Information Service, Available from <http://data.ris.ripe.net>.
[3] Route Views ''University of Oregon Route Views Project''. Available from <http://www.antc.uoregon.edu/routeviews/>.
[4] R. Albert, A.-L. Barab_asi, Statistical mechanics of complex networks, Reviews of Modern Physics 74 (2002) 47.
[5] A. Broido, K.C. Claffy, Internet expansion, refinement and churn, ETT January 2002.
[6] T. Bu, D. Towsley, On distinguishing between Internet power-law topology generators, in: Proceedings of INFOCOM, 2002.
[7] H. Chang, R. Govindan, S. Jamin, S.J. Shrenker, W. Willinger, On inferring AS-level connectivity from BGP routing tables, in: Proceeding of INFOCOM, 2002.
[8] Q. Chen, H. Chang, R. Govindan, S. Jamin, S.J. Shenker, W. Willinger, The origin of power-laws in Internet topologies revisited, INFOCOM 2002.
[9] M. Faloutsos, P. Faloutsos, C. Faloutsos, On power-law relationships of the Internet topology, in: Proceeding of ACM SIGCOMM'99, August 1999.
[10] L. Gao, J. Rexford, On inferring Autonomous system relationships in the Internet, IEEE/ACM Transactions on Networking 9 (6) (2001).

1.Mr.G.Srinivasa Rao, M.Tech,(Ph.D), Sr.Asst.Professor. He has Submitted Ph.D thesis in M.U., Over 7 Years of teaching experience with GITAM University, handled courses for B.Tech, M.Tech. Research areas include Computer Networks And Data Communications. published 4 papers in various National and International Conferences and Journals.

2. Dr.Prof Arun Singh,Ph.D., MAGADH UNIVERSITY , Dept of Computer Science and Engineering, over 15 years of teaching experience. Published 10 papers in various National and International Conferences and Journals.

3.Dr. G.Appa Rao., M.Tech., M.B.A.,Ph.D., in computer science and Engineering form Andhra Universiy. Over 12 Years of teaching experience with GITAM University, handled courses for B.Tech, M.Tech. Research areas include Data Mining and AI. Published 8papers in various National and International Conferences and Journals

4. Mr.D.Kishore Kumar M.Tech in computer science and engineering. Over 4 Years of teaching experience with GITAM University, handled courses for B.Tech.

5. Mr.P.Surya Chandra., M.Tech in Information Technology. Over 3 Years of teaching experience with GITAM University, handled courses for B.Tech and M.Tech. One research paper was published in international journal and one conference.

6. Mr..D.Veerabhadra Rao., M.Tech., in Information Technology. Over 7 Years of teaching experience with GITAM University, handled courses for B.Tech and M.Tech .One research paper was published in international journal and one conference.

7.Mr. M. Prasada Rao., M.Sc., M.Tech(IT)., (Ph.D) area in Bio Informatics. Over 6 years of teaching experience with Chaitanya Engg college and GITAM University.Two International conferences from Andhra University. Handled courses for B.Tech. and M.Tech.

8. Mrs. S. Venkata Lakshmi M.Tech in Information Technology from Andhra University. Asst.Prof in GITAM University. Over 2 years of teaching experience with GITAM University and Andhra University handled courses for B.Tech, and M.C.A. and 2 years of industry experience as a software engineer.

9. Ms. D.RamyaMallika, M.Tech(IT) from GITAM University

10. Mr. N.Rajender M.Sc.,M.Tech(IT) from GITAM University