Flattening and Preferential Attachment in the Internet Evolution

Miao Li, Hui Wang, Jiahai Yang
1. Network Research Center, Tsinghua University
2. Tsinghua National Laboratory for Information Science and Technology(TNList)
Beijing , R.P. China
mileslmw87@126.com, {hwang, yang }@cernet.edu.cn

Abstract—**Understanding of the Internet evolution is important for many research topics, such as network planning, optimal routing design, etc. In this paper, we try to analyze CAIDA AS-level topology dataset from 2004 to 2010 to validate two conjectures on the Internet evolution, i.e., the Internet flattening trend and the preferential attachment rule. Our analysis shows that the evolvement of the Internet core is different from the edge of Internet. We classify the Internet into several layers using different layering methods, i.e., Rich Club coefficient based method, k-core decomposition method and SARK hierarchy model, and then study the changes of the features of these layers. Under all of these laying methods, we find that the boundaries between neighboring layers in the Internet core are more and more blurred; ASes in the core distribute more evenly and different layers are closer to each other in size, while the Internet edge still has a distinct hierarchical characteristic. It is more evident in Asia and Europe than North America. The other difference between Internet core and Internet edge is that link births/deaths in the Internet core follow the "Preferential Attachment/de-attachment" rule, while link births/deaths in the Internet edge follow a super linear preferential attachment/de-attachement rule. On the other hand, in both Internet core and Internet edge, link births caused by AS births present stronger preference than link rewiring.**

***Keywords-** topology evolution; flattening; preferential attachment*

I. INTRODUCTION

The Internet, as a network of networks consisting of thousands of ASes (Autonomous Systems), experiences AS births, AS deaths and changes of connectivity between ASes. It is an evolving entity with constituent network components being constantly added, upgraded and engineered [1]. The dynamics and evolution of the Internet is the optimization result of individual ASes. Understanding evolution of the Internet is necessary and meaningful. It provides a perception of the Internet topology which is important for network planning, routing protocol design, topology generation for simulations and new generation Internet architecture design.

Though an extensive literature has focused on this topic in recent years, there are still problems in many aspects of the Internet evolution. In this paper, we focus on two conjectures: the Internet flattening trend and the preferential attachment rule.

Traditionally, the Internet is viewed as a hierarchical network. We can split the Internet into several layers according to certain metrics since there are considerably significant boundaries between different layers. It is conventional wisdom that 10-15 ASes peer with each other forming a clique and occupy the highest layer in the core (or say Tier-1). Some more ISPs (Tier-2), such as national ISPs, large regional ISPs and large content providers, buy transit service from tier-1 ISPs while providing transit service to lots of small regional ISPs and small corporations in the edge of Internet (Tier-3). Besides, the hierarchical Internet is a Pyramid structure considering the size of each layer: the upper the layer is, the less ASes it contains. The edge of the Internet contains the majority of ASes. Many researches are based on the assumption that the Internet is typically hierarchical and the hierarchical property is one of the important metric to evaluate the network model [2].

However, recent 10 years witnessed enormous changes of the Internet structure. The rapid expansion of CP or CDNs and the appearance of IXP have reduced peering cost sharply in the past few years. Besides, multihoming and some complex relationships such as partial transit have become more common. Against this background, more and more evidences, both anecdotal and academic, show that the Internet is heading towards a flattening structure. First, the flattening has drawn attention of operator groups such as NANOG. In addition, more and more ISPs have declared that they have occupied the center of the Internet ecosystem, gained significant power in the Internet and the ability to offer more reliable and cheaper transit services; Second, some researches and measurements have directly noticed the flattening trends and they described this trend from different point of view, e.g. considering the change of traffic volume and the decrease of AS level route length.

Traffic migration and the shortening average AS path length are signals of flattening of the Internet. However, they have limits on describing the change of the Internet structure. Furthermore, we also need to study the evolution trend in the Internet core and edge separately. In this paper, we try to study this trend from the layer viewpoint, which is the core concept of the Internet hierarchy model, and demonstrate the flattening phenomenon occurring in the Internet from the following aspects: boundaries between neighboring layers are blurred; with better AS connectivity, ASes are distributed in different layers more evenly and layers are closer to each other in size. Under the same hierarchy model, flattening is manifested by the blurred boundaries between neighboring layers and the increasing AS proportion of the higher layers which implies the Internet is evolving from a pyramid structure to a better connected or even full-meshed structure.

We use three layering methods, i.e. rich club coefficient based method, k-core decomposition method and SARK hierarchy model, and study Internet topology evolution during the last 7 years. We study the changes of features of different layers and find that the boundaries between neighboring layers in the Internet core are more blurred over time; ASes in the core distribute more evenly and different layers are closer to each other in size, while the Internet edge still has a distinct hierarchical characteristic. It conveys the message that flattening is indeed taking place in the Internet core. We also find this trend is more evident in Asia and Europe than North America. This trend has significant impact on the design of a more robust routing architecture. This finding is also quite meaningful for ISPs` network planning.

In addition to observing the Internet evolution from the view of the whole Internet, we also study the evolution of each AS individually (node level). This is related closely to the researches of topology generation models. Preferential attachment of nodes to nodes with high degrees is coined in famous BA model, and many follow-up researches are based on the preferential attachment rule. However, there are always doubts whether the preferential attachment rule is still reasonable today, and whether it should be linear preference or super linear preference [6]. The common method adopted by modeling work is to compare the modeling result with just one snapshot on some static topological properties. However, examining the evolution of topology over time is a more direct and convincing way. Chen et al. [7] validated this rule in 2002 and stated the preferential attachment was super linear. But in the following nearly ten years, few works could give a convinced result based on real measurement. Researchers do not know whether this rule is still validated today.

Based on real measurement data, ranging from 2004 to 2010, a total of 73 monthly snapshots, we study the properties of AS/link births/deaths and preferential attachment rule at node level. We find link births in the Internet core follow the linear preferential attachment rule, while in the edge of the Internet, we observe evident super linear preferential attachment property on link births. Moreover, link births caused by AS births present stronger preference than link rewiring. Furthermore, we verify that the de-preferential attachment rule indeed exists and link deaths share lots of common properties with link births.

The rest of the paper is organized as follows. In Section Ⅱ, we discuss related work. In Section Ⅲ, we present definitions and our data preparation method. In Section Ⅳ, we analyze the flattening trend in recent years. And then in Section Ⅴ, we study the evolution at node level and analyze the preferential attachment and preferential de-attachment rules. Finally, in section Ⅵ we conclude our work.

II. RELATED WORK

In the last decade, a number of studies characterized AS-level topology of the Internet. Based on measurement results, some studies show that the Internet share lots of properties with other complex networks, such as “randomness”, “scale free”, “small world”, however, other studies highlight that the Internet has its typical properties and hierarchy is an important one [8,9]. Considering the hierarchical property, several works classify the ASes into different tiers or layers. Some of them [10, 11] are based on the topological structure, such as node degree, while other layering methods take the relationships between ASes into consideration. Ge et al. classify ASes into seven layers based on inferred customer-to-provider relationships [12]. X. Dimitropoulos et al. map all ASes into 7 levels using machine learning techniques based on metrics such as the number of inferred customers, IRR description, etc [13]. Subramanian et al. classify ASes into five layers based on inferred customer-to-provider as well as peer-to-peer relationships [2].

More and more recent studies examine the topology evolution of the Internet over time. Many literatures analyze real measurement data and try to find Internet evolution trend or rules. Magoni et al. find exponential growth in the number of ASes and links from 1997-2000. Leskovec et al. [14] measure the average degree and effective diameter of the Internet AS graph and conclude that the AS graph is getting denser. Dhamdhere et al. [15] measure the topology in the last decade and highlight a slower exponential growth of the Internet in terms of both ASes and inter-AS links which is mostly due to enterprise networks and content/access providers at the periphery of the Internet and find the average AS path length of the growing Internet remains almost constant mostly due to the increasing multihoming degree of transit and content/access providers.

As aforementioned, the Internet flattening trend has been noticed by researchers. C. Labovitz et al. [3] find the migration of a majority of Internet traffic away from Tier-1 to the direct links between large CP and customer networks. Considering the traffic migration is mainly caused by video traffic between the two video service giants, Google and Comcast, it was not enough for us to draw the flattening conclusion solely based on this traffic migration. Several researches study this trend from AS path length. Through monitoring and analyzing the inter-domain routing from BGP routing tables, Routeviews and RIPE RIS, Y Xiang et al. [5] find ASes close to Tier 1 contribute 36% to the decrease in route length and content providers contribute a lot to the flattening. Taking both traffic and route length into consideration, based on Traceroute method, Gill et al. show that CP brought their networks to users, bypassing Tier-1 ISPs on many paths which might flatten the Internet topology [4].

However, these observations and analysis based on routing path length or traffic volume could not depict the evolution of a hierarchical network sufficiently. Dhamdhere et al. studies this evolutionary transition with an agent-based network formation model which predicts several substantial differences between the Hierarchical Internet and the Flat Internet in terms of profitability, path lengths, etc [16].

There are also researches trying to validate some rules based on the change of topology at node level over time. After it is proposed in the BA model, preferential attachment rule has been used as a basic assumption in many modeling works, e.g. AB [17], GLP [18] with minor modifications. However, the linear preferential attachment mechanism is always doubted. From the view of modeling, Zhou et al. propose new variants of BA model with super linear preferential attachment mechanism [6] and show the better generative topology result on some static properties. There are also researches validate the rule based on the dynamic Internet instances: Siganos et al. [19]

observe the Internet evolution during 1997-2001 at different levels and find the link super-linear preferential attachment rule; Chen et al. validate BA model in 2002 and also find the Internet is indeed growing incrementally and new ASes have a much stronger preference to connect to high vertex degree ASes than predicted by the linear preferential model [7]. But in the following nearly ten years, few works could give a convinced result based on real measurement data.

III. DATASETS AND METHODOLOGY

A. Data Source

To study the evolution of the global Internet, both historical topological structure of the Internet and the contractual relationships between them are needed. However, ISPs are not willing to release data makes our study difficult. Fortunately, CAIDA publishes measurement and inference data of the Internet AS-level topology [20].

Researchers in CAIDA collect BGP routing tables from Routeviews, which provides historical routing information; then, after filtering backup and transient links, they infer AS relationships (customer-to-provider and peer-to-peer) using their MAX2SAT techniques [21] based on multi-objective optimization and infer sibling relationship using WHOIS information [22]. The main idea behind inference heuristics of c-p and p-p link is an optimally balanced trade-off between AS relationship information that can be extracted from AS degrees and maximization of the number of valid paths in the resulting annotated AS topology [23]. The accuracy of this data set has been validated by several works [23, 24, 25].

The CAIDA data set contains the Internet topology annotated with relationship from January 2004 to January 2010, with one snapshot per week. Considering one snapshot for each month is enough for our analysis on the Internet evolvement, we do the following to generate snapshots in our dataset to further avoid false path introduced by misconfiguration. We compare all the CAIDA snapshots in the same month and consider a link to be valid if it appears in the majority of snapshots in this month. Then, we generate a merged monthly snapshot comprising of all the valid links to represent topology in this month. For convenience, in the following analysis, we use number i ($i = 1\ldots73$) to mark these 73 snapshots and represent the time in the six years.

B. Layering Methods

We want to study how the Internet hierarchical structure changed based on the concept of layer. There are several methods to split the Internet into layers. The layering results should be consistent with intuitivelyon, i.e. customer ISP should be in the lower layer than its providers. To avoid the inaccuracy or error of one certain layering method, we apply three different well-tested layering methods to our dataset. In this section, we will introduce these three layering methods briefly, i.e., Rich Club connectivity (RCC) based method, k-core decomposition method and SARK hierarchy model.

Layering method based on RCC: The rich-club is an important concept in Internet topology modeling. The rich-club phenomenon introduced by [10] refers to that the rich nodes, which are a small number of nodes with large number of links, are very well connected to each other. Based on this phenomenon, we can split the Internet into two parts: rich club and other nodes. The rich-club is the center of the Internet core. There is an evident boundary between nodes in the rich club and other nodes concerning their rich club coefficient, which is defined as the ratio of the total actual number of links between members of the rich-club to the maximum possible number of links. The maximum possible number of links between n nodes is $n*(n-1)/2$. In each snapshot, we sort nodes in a decreasing order according to their degrees and calculate rich-club coefficient C(r) of each AS where rank r denotes the node position in the decreasing list. Our first layering method defines the rich club layer as nodes with rank less than r_{max}, where r_{max} equals to 0.2%.

K-core decomposition method: The k-cores are fundamental structures in graph theory and their study dated back to the 60’s [26]. The k-core of a graph is the sub graph obtained by the iterative removal of all nodes with degree less than or equal to k. The node coreness k of a given node is the maximum k such that the certain node is present in the k-core but removed in the $(k + 1)$-core. The maximum node coreness *k-max* in a graph is the graph coreness. Compared with degree, “*coreness*”, as one of network topology characteristics for node, has better ability to describe the hierarchy of network. It provides the depth information of a node in the graph.

During the decomposition process, we can classify the nodes with the same coreness into the same layer, denoted as coreness layer $L_{k\text{-coreness}}$. The high coreness layers, or say the inner layers, are viewed as the Internet core and the outside layers are the Internet edge. We use coreness of 5 to divide the core and the edge of the Internet based on our observation.

SARK Hierarchy model: The former two methods view the Internet as an undirected graph and split the graph using graph-theoretic metrics. In fact, taking the relationships between ASes into consideration is a better way to catch the essence of tiers in the hierarchy. We adopt the method introduced by [2]. It divides the Internet into five layers: customers, small regional ISPs, outer core, transit core and dense core, based on AS relationships and node connectivity boundaries.

In this model, Internet is abstracted as directed graph, a provider-customer relationship between A and B is represented by a directed edge from A to B and a peering relationship between A and B is represented by two directed edges. The leaves of this directed graph are classified as customers; the nodes removed by the process of iterative pruning the leaves of the graph are defined as small regional ISPs. The remaining set of ASes is regarded as the Internet core. They use a greedy heuristic method to further identify the dense core. They define the sub graph to be “almost a clique” if every node in it has the out-degree and in-degree of at least N / 2. Then they use an in-way cut method to find the transit core, considering the jump of out degree to the transit core and dense core as the layering boundary. The transit core consists of other large national ISPs and hosting companies that have peering relationships with each other and with some ASes in the dense core but do not tend to peer with other ASes in the outer core

---

This work is supported by the National Basic Research Program of China under Grant No. 2009CB320505, the National Science and Technology Supporting Plan of China under Grant No. 2008BAH37B05, and the National High-Tech Research and Development Plan of China under Grant No. 2008AA01A303 and 2009AA01Z251.

or in the edge of the Internet. The rest set of ASes is the outer core.

Based on these three layering results, we conduct an analysis to see the evolution of layer architecture. We expect to see the same conclusions on the Internet evolvement under all these three layering methods.

IV. INTERNET CORE FLATTENING TREND

In this section, we try to analyze the change of the Internet structure based on the concept of layer using the three different layering methods. From comparing the layering results under the three layering methods over time, we could observe the flattening in the Internet core reflected by better AS connectivity, blurred boundaries between layers in the Internet core and the phenomenon that ASes in the core distribute more evenly and different layers are closer to each other in size.

A. Flattening Reflected by blurred boundary between rich club and other nodes

We calculate and record the rich-club coefficient *C(r)* of each AS against the AS rank r for each snapshot in our dataset. Except for the boundary area, the RCC curves of all the 72 snapshots present smoothly. Figure 1 shows the curve of 2010.1 as an example on a log-log scale to present the general shape of the curve. No doubt a small set of nodes are very well connected between each other. About 30-40 nodes are very well connected as rich club layer and top 1% rich nodes have about 30% of the inter-AS links, which is consistent with the result in [10].

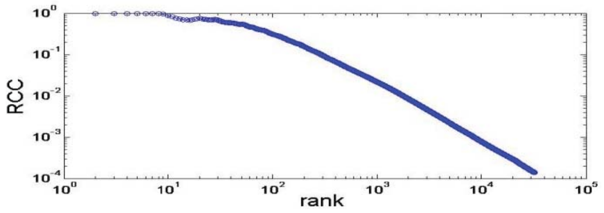


Figure 1. RCC against node rank

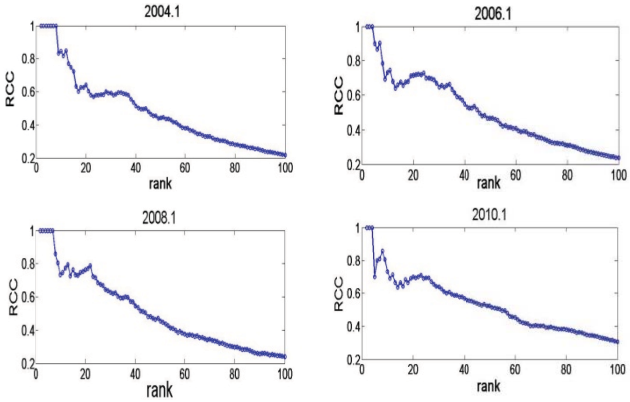


Figure 2. RCC of ASes the central part of the Internet in evolution view

However, in the boundary area of the rich club (r ranges from 20 to 60, containing around 0.2% ASes in the rich club), the curves turn less smoothly over time. In Figure 2, we take four snapshots as examples to illustrate the tendency, i.e. 2004.1, 2006.1, 2008.1, 2010.1, at intervals of 2 years and plot *RCC* of the top 100 nodes in each snapshot against to their node rank on a log-log scale. We can see double peaks or multiple peaks more and more clearly.

The appearance of multi peaks in the curve can be explained by the fact that boundary area nodes in the trough of the curve connect with lots of nodes, despite their fewer links to the center of rich club. To some extent, we can view these nodes as other centers of the rich club and the Internet. It provides an angel to analyze the flattening in the core from AS connectivity: ASes surrounding the rich club, such as large regional ISPs, in order to optimize their revenue, attempt to peer with each other; Meanwhile, with their increasing traffic and impacts, their dependence to the ASes in the center of rich club becomes weaker and this independence makes them more powerful. This is a strong signal that the core of the Internet is experiencing the flattening. It conforms to the fact that some tier-2 ISPs declare that they are in the center of the Internet.

B. Flattening Reflected by the Larger high coreness layers

In this subsection, we use the k-core decomposition method to classify ASes into different layers and explore the flattening under this layering methodology.

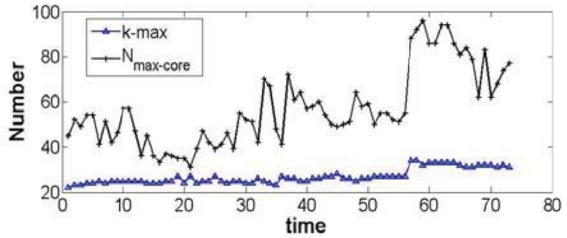


Figure 3. K-max and $N_{max\text{-}core}(i)$ curves over time

At first, for each snapshot *i*, we plot *k-max(i)* and the max-core size $N_{max\text{-}core}(i)$ against time in Figure 3. It presents a steady increase of the *k-max*, from 22 in 2004 to 31 in 2010 and records $N_{max\text{-}core}$ increasing from 45 in the 22-core in 2004 to 77 nodes in the 31-core in 2010. The increasing *k-max* and $N_{max\text{-}core}$ present an obvious densification in the Internet.

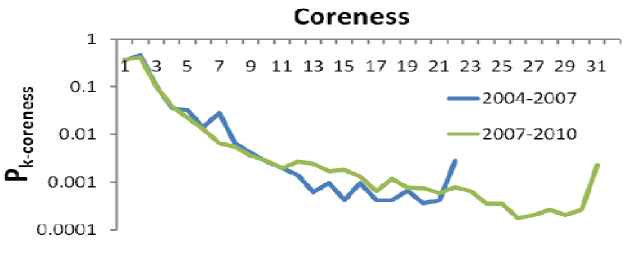


Figure 4. AS proportion of each coreness layer over time

The densification is just one of the evidence of flattening. To further reveal the trend, we consider the AS proportion of each coreness layer, $P_{k\text{-}coreness}$, defined as follows:

$$P_{k\text{-}coreness} = N_{k\text{-}coreness} / N_{as} \qquad (1)$$

where $N_{k\text{-}coreness}$ is the number of ASes in each layer and N_{as} is the total number of ASes in the Internet. In Figure 4, we plot the average $P_{k\text{-}coreness}$ in log scale against the coreness in 2004-2007, 2007-2010. We can see that there exists marginal variation between the two curves in the coreness range between 1 and 5. However, in the coreness range between 5 and *k*-max, the two curves have an intersection and the AS proportion of high coreness layer is higher over time. (Since *k*-max-coreness layer is the *k*-max core in fact and cannot compare the size directly with the other layers, we can ignore the increase in the tails of the curves.) This result reflects the fact that a larger proportion of ASes enters the higher layers and the high layers are larger. Actually, 5-core, only covering less than 10% ASes can be viewed as the core of the Internet and ASes in core-1 but not in core-5 are in the edge of the Internet. There is flattening in the core of the Internet, but no evidence of flattening appears in the edge of the Internet.

C. *Flattening Reflected by the change of Average Layer in SARK layering model*

As we mentioned in Section 2, the hierarchy model we choose takes the commercial relationships between ASes into consideration rather than only considering the connectivity between nodes. We apply this layering method to our dataset and present the sizes of different layers over time in Figure 5. We can see the size of each layer increases steadily these years. All these years, dense core contains less than 30 nodes. This figure coincides with the size of the rich club in the above RCC analysis. The size of transit core and dense core is approximate to the size of 6-core (6-core contains about 4% ASes) and the size of entire core is close to the size of 5 core (5-core contains about 7% ASes) in the k-core analysis.

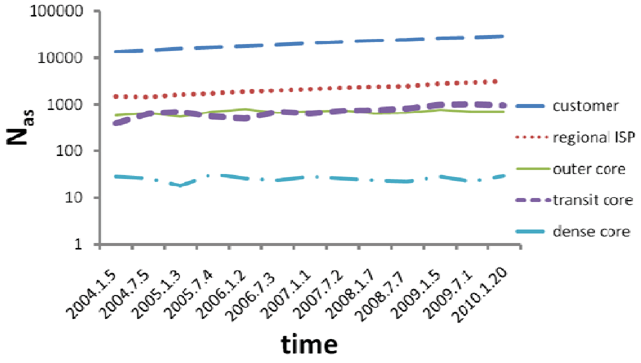


Figure 5. Evolution of layer size in SARK hierarchy model

Flattening in the Internet core: From Figure 5, we can see the transit core becomes larger and the size of the outer core is decreasing. For a clearer presentation of the flattening in the core, we use 1 to 5 to mark the layer from the inside to outside representing dense core, transit core, outer core, regional ISP and customer. We define average layer $\bar{L}^{core}$ and fraction of links R and demonstrate the trend based on these metrics:

$$\bar{L}^{core} = (N_{dense} + N_{transit}*2+N_{outer}*3) / (N_{dense}+N_{transit}+N_{outer}) \quad (2)$$

$$R^{core} = (E_{dt}+E_{do}+E_{to}) / (N_{dense}+N_{transit}+N_{outer}) \quad (3)$$

$$R^{internet} = (E_{dt}+E_{do}+E_{to}+E_{cr}+E_{ro}) / N_{as} \quad (4)$$

where N_{dense} is the dense core size, $N_{transit}$ is the transit core size, N_{outer} is the outer core size and E_{dt}, E_{do}, E_{to}, E_{cr}, E_{ro} represent the number of links between these different layers. $\bar{L}^{core}$ represents the average layer of the ASes in the core. To some extent, $\bar{L}$ convey the same message with the average AS path length. The decreasing of $\bar{L}$ tells that most ASes are getting closer to each other in distance from point view of layer and presents the flattening trend. And R represents the connectivity between different layers. In SARK layering model, the number of links between customer layer and the core (E_{co}, E_{ct}, E_{cd}), the number of links between regional ISP layer and dense core, transit core (E_{rt}, E_{rd}) are zero, so we do not need to take them into account when calculate the R^{core} and $R^{internet}$.

In the top panel of Figure 6, we plot $\bar{L}^{core}(i)$ of each snapshot and plot $R^{core}(i)$ in the bottom panel. We can see $\bar{L}^{core}$ keep dropping in recent years, which demonstrates the inflation of the transit core - as the result of a greater proportion of ASes enters into the transit core. This phenomenon can be explained by the narrowing gap between different layers. Since this hierarchy model classifies the ASes into transit cores based on the jump of their out degree to the dense core and other nodes in the transit core, with the out degree gaps between ASes in the Internet core narrower over time, more and more proportion of ASes enter into inner layers.

In the bottom panel of Figure 6, we observe that R^{core} presents a constant increase in recent years. The higher R^{core} presents denser links between different layers, which is the typical characteristic of the well connected network instead of a simple hierarchy structure. The increasing R^{core} further validate the better connectivity and the narrower gaps between different layers in the core.

From the above analysis, we can see the flattening in the core of the Internet as evidence of the smaller $\bar{L}^{core}$average layer in the core and the denser inter-layer links. At the same time, it is worthy to notice that in the bottom panel of Figure 6 $R^{Internet}$ stays around 5, which is about one eighth of R^{core}. The hierarchy structure is still dominant in the edge of Internet.

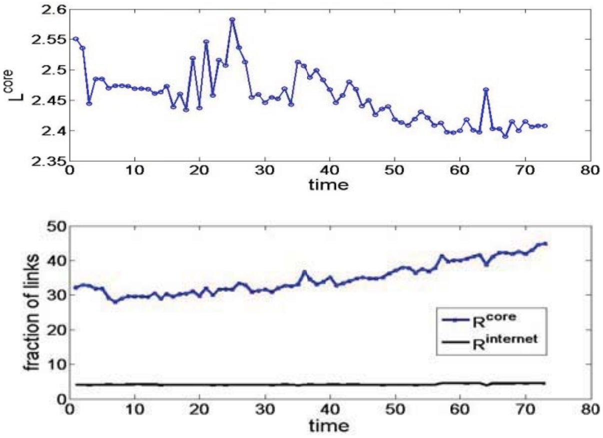


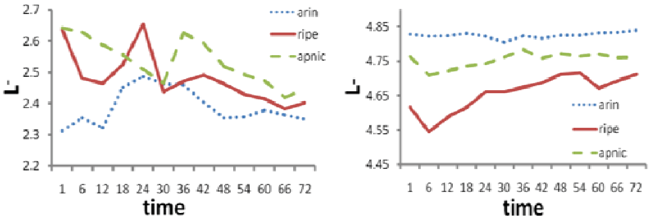
Figure 6. Average layer and connectivity between layers over time

Flattening in regional view: Then we apply the above analysis method to our dataset again, taking regional factor into account. We use the CAIDA geographic information [27]. We focus on the ASes in North America, Europe and Asia, which totally cover nearly 90% of ASes. We classify the ASes into 3 groups: ARIN, RIPE and APNIC, based on the registry which their ASNs are allocated from. Then based on the layering result above, we recalculate the average layer of each regional core and the entire region respectively, denoted by $\bar{L}^{core}_{arin}$,

$\bar{\mathrm{L}}^{\mathrm{core}}_{\mathrm{ripe}}$, $\bar{\mathrm{L}}^{\mathrm{core}}_{\mathrm{apnic}}$ and $\bar{\mathrm{L}}_{\mathrm{arin}}$, $\bar{\mathrm{L}}_{\mathrm{ripe}}$, $\bar{\mathrm{L}}_{\mathrm{apnic}}$. The $\bar{\mathrm{L}}^{\mathrm{core}}_{\mathrm{arin}}$, $\bar{\mathrm{L}}^{\mathrm{core}}_{\mathrm{ripe}}$ and $\bar{\mathrm{L}}^{\mathrm{core}}_{\mathrm{apnic}}$ are plotted against time in Figure 7 a, and Figure 7 b presents the curves of $\bar{\mathrm{L}}_{\mathrm{arin}}$, $\bar{\mathrm{L}}_{\mathrm{ripe}}$, $\bar{\mathrm{L}}_{\mathrm{apnic}}$.

We can see from 2006, the 20th snapshot, the flattening in RIPE is clear, with $\bar{\mathrm{L}}^{\mathrm{core}}_{\mathrm{ripe}}$ dropping from 2.65 in 2004 to 2.4 in 2010 and APNIC also shows the decreasing trend. Roughly speaking, ARIN still occupies the innermost position in the core with the $\bar{\mathrm{L}}^{\mathrm{core}}_{\mathrm{arin}}$ around 2.4. Taking geographical factor into account, we can see the decreasing average layer of the Internet core $\bar{\mathrm{L}}^{\mathrm{core}}$ is mainly driven by Asia and Europe. In the entire regional view, $\bar{\mathrm{L}}^{\mathrm{entire}}_{\mathrm{arin}}$ and $\bar{\mathrm{L}}^{\mathrm{entire}}_{\mathrm{apnic}}$ do not show decrease sign, and $\bar{\mathrm{L}}^{\mathrm{Internet}}_{\mathrm{ripe}}$ even has an increasing trend.

From above analysis, we can see the Internet core presents flattening trend. ISPs in ARIN still occupy the central part of Internet. However, from business relationship and connectivity view, the impact of other large ISPs, represented by ISPs in Asia and Europe, are increasing and these ISPs contribute a lot to the flattening in the Internet core. Besides, the edge of the Internet still has strong characteristic of hierarchy.



a. Average layer of different region cores b. Connectivity between layers

Figure 7. Flattening trend in each geographical region

V. PREFERENTIAL ATTACHMENT IN THE INTERNET EVOLUTION AT NODE LEVEL

The importance of understanding the Internet topology evolution has been noticed and lots of topology modeling researchers follow the researches of evolutionary model. The Internet is experiencing many events every minute and researches of evolutionary model try to abstract rules from those events and then generate topology based on these rules. Generally speaking, the events mainly include AS birth, link birth, AS death, link death etc. And several attributes belong to these events. For example, whether it comes with an AS birth event and which ASes for the link to attach to are both the attributes of link birth event. We identify the events in the Internet in the following ways:

AS / link birth: It is identified as an AS birth or a link birth event if the AS or link in the *i*th snapshot does not appear in the earlier two consecutive snapshots, i.e. (*i-1*) and (*i-2*) snapshots. But if the born link in *i*th snapshot appears in the (*i-2*) snapshot but disappear in the (*i-1*) snapshot, we will regard it as error which may be caused by routing instability;

AS / link death: we identify it is an AS death or a link death event if the AS or link did not appear in the next two snapshots.

The link related events can be further classified into two categories: (1) link birth / death with one AS birth / death event, denoted as S_{ne};(2) link birth / death events between two existing ASes, denoted as S_{ee};

At present, preferential attachment rule is one of the most popular rules in the evolution researches, for its simplicity and universal in describing evolution of many different networks. Preferential attachment rule was coined in BA model. For a link birth event, with or without AS birth, BA model picks ASes for the links to attach to using a linear preferential rule. Typically, at each interval Δt, when a new link is added to the network, the probability to choose the existing node *j* (or say AS *j*) as link target is proportional to node *j*`s degree:

$$\prod(g_j,\, t+\Delta t) = g_j(t) \,/\, \sum\nolimits_h g_h(t) \qquad (5)$$

Where $g_i(t)$ is the degree of node *j* at time *t*. Since the denominator stands for the sum of the degrees of all current nodes, which is a fixed value, the probability is linearly correlated with its degree. This assumption is called as “linear preferential attachment rule”. Based on this rule, BA model can yield graphs with observable properties, such as power law, strong clustering and so on.

Preferential attachment rule has been supported and validated in many researches [26, 28]. However, Chen et al. stated in 2002 that the preferential attachment is super-linear and several researches doubt whether the linear preferential attachment is reasonable these years. In this section, we focus on the Internet evolution at node level and try to validate this assumption using most recent data. We take 13 snapshots from 2004.1 to 2010.1 to support this study since we find that the granularity of half a year can give a marginal percentage of error: the probability that two ASes appear or disappear in the same time slot is below 1%. We try to answer the following three questions:

1. Is the preferential attachment rule still reasonable after all these years?

2. Is the preferential attachment linear or super linear? Is it the same for all ASes, considering ASes in the Internet core and the edge?

3. Is it the same for link births/deaths caused by all related events?

A. *Preferential Attachment in the Internet*

We first study the probability PA_g that a node with degree *g* is involved into a link birth event in S_{ne}. We calculate PA_g in the following steps:

1. We group link birth events in S_{ne} by the degree of the link target ASes and count the number of link birth events in each group, denoted as E_g, where “*g*” is the degree of ASes in the group;
2. We count N_g , the number of ASes with degree *g*;
3. The probability that a node with degree *g* is involved in a link birth event in S_{ne} is

$$PA_g = E_g \,/\, N_g \qquad (6)$$

From the 13 snapshots, we get 12 results that depict probability PA_g against the degree. The general shape of the 12 curves are similar. To present the results more clearly, in the up panel of Figure 8, we plot PA_g against degree in a log-log scale between 2006.1 and 2006.7 as an example.

Then we study the property of link events in S_{ee}. The analyzing method is similar to the method used in S_{ne}. However, in S_{ee}, we will consider both nodes of a born link as the target nodes. The 12 results of S_{ee} is also coincident with each other well and we plot the result between 2006.1 and 2006.7 in the bottom panel of Figure 8 for a clear view. We can see the linear preferential attachment assumption can fit most part of the curve well and next, we will study the quantitative details from the curve fitting result.

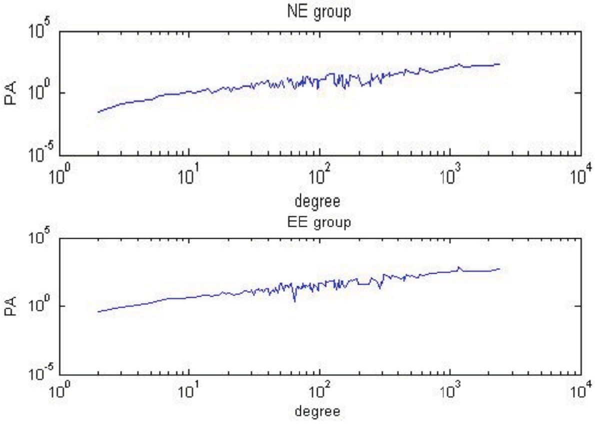


Figure 8. Link birth preferential attachment

B. *Different preferential attachment in the core and edge*

We apply linear curve fitting to Figure 8 and find the slope factor on the log-log scale in the high degree range is less than that in the low degree range. The degree 5 is the location of "knee" point. This threshold is in accordance with our previous hierarchy analysis, especially K-core decomposition in Section 3, in which 5 is the threshold to distinguish the Internet core and edge.

Then, using this threshold, we do piecewise fitting to the curve in Figure 8 and plot the fitting results in Figure 9 on a log-log scale (base=*e*). To the result of S_{ee}, we can see the average edge preferential factor is 1.69 and the core part preferential factor is around 0.88, while to S_{ne}, the average edge preferential factor is around 2.55, and the average preferential factor is around 1.1. The result conforms to our intuitive understanding:

- First, link births (in S_{ne}) caused by AS births present stronger preference than link rewiring (in S_{ee}). When an AS is just born, considering the better performance, it has to buy transit service from large ISPs to access the Internet directly. To the links of new ASes, the preference to high degree nodes is stronger than rewiring links between existing ASes, since existing ASes usually change links to optimize their cost, do free peering and so on;

- Second, link attractiveness of ASes in the edge of the Internet increase sharply and there is evident super linear preferential attachment property on link births in the edge Internet. Since the geographic scope of large ISPs is larger than that of small ISPs, especially ISPs in the edge of the Internet, they have more chances to connect with others. Furthermore, with the little amount of traffic, small ISPs` peering attractiveness is weaker from economic view, so the possibility of other ASes to buy their transit or peer with them is small. The geographical and traffic gaps between ASes in the Internet edge are wider.

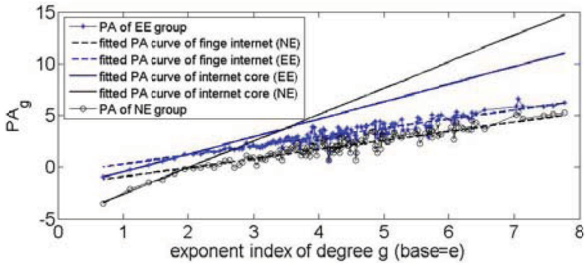


Figure 9. Piecewise fitting to the preferential attachment property

C. *Evolutional view of preferential attachment*

Next, we take an evolutional view to compare our piecewise fitting results over these years. As Figure 10 presented, preference in the core is stable with a slight decline in these years, while the preferential factor of the Internet edge shows an evident decrease, especially in S_{ee}, in the right panel of Figure 10. These observations conform with the flattening we introduced in chapter 3. Since characteristic of hierarchy structure in the core of the Internet is weaker and the gaps between large ISPs and small ISPs become narrower, it is more common and convinient for ISPs, both large and small, to do traffic engineering. It is the flattening that contributes to the decrease of preferential factors.

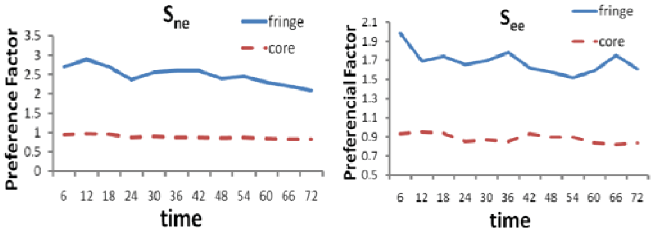


Figure 10. Preferential attachment tendency evolution

D. *Preferential De-attachment*

In recent years, some researches figure out that besides preferential attachment, an equally strong preferential de-attachment is also observed [15]. Using the same way as analysis on preferential attachment, we define PA_g^d to represent the preferencial de-attachment tendency(DPA):

$$PA_g^d = \overline{E}_g^d \;/\; \overline{N}_g^d \tag{7}$$

where $\overline{E}_g^d$ and $\overline{N}_g^d$ are the link death number and AS death number in group *g*, where *g* is the degree of ASes in the group. We take the result between 2006.1 and 2006.7 as an example and plot PA_g^d against their degree on a log-log scale in Figure 11. We can see the shapes of the curves are similar to the curves in our preferential attachment analysis; to S_{ee}, the average edge de-preferential factor equals to 1.58 while core de-preferential factor is 0.92; to S_{ne}, the edge factor equals to 1.91 while core average slope is 0.98. It is indeed that the preferential de-attachment rule indeed exists and link deaths share lots of common properties with link births.

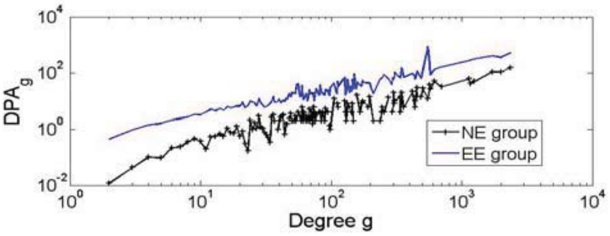


Figure 11. Link death preferential de-attachment

VI. CONCLUSION

In this paper, we analyze the Internet topology snapshots over the last 7 years to study the evolution trend. We focused on two conjectures on the Internet evolution, i.e., the Internet flattening trend and the preferential attachment. Particularly, we study the evolution in the core and the edge of the Internet separately and find the core and the edge present different evolution features.

Our work is based on the concept of layer. We use three different well-tested layering methods, i.e. Rich Club coefficient based method, k-core decomposition method and SARK layering model, to validate the flattening trend. We see the same conclusion under all three layering methods: in the core of the Internet, the boundaries of different layers are blurred; ASes distribute more evenly and different layers are closer to each other in size. It means the hierarchical characteristic of the Internet core is becoming weaker. And this flattening trend is more evident in Asia and Europe than North America. However, in the edge of the Internet, there is no obvious flattening evidence.

Our analysis on the Internet evolution at node level validates the preferential attachment assumption. The link births still follow the preferential attachment rule. To be more specific, there is evident super linear preferential attachment property on link births in the Internet edge, and link births caused by AS births present stronger preference than link rewiring. Moreover, the preferential de-attachment rule indeed exists and link deaths share lots of common properties with link births.

REFERENCES

[1] Chang, H, Jamin, S, Willinger, W. "To Peer or Not to Peer: Modeling the Evolution of the Internet's AS-Level Topology," *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings* , vol., no., pp.1-12, April 2006

[2] Subramanian, L.; Agarwal, S.; Rexford, J.; Katz, R.H.; , "Characterizing the Internet hierarchy from multiple vantage points," *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE* , vol.2, no., pp. 618- 627 vol.2, 2002

[3] C. Labovitz, D. McPherson, and S. Iekel-Johnson. Internet Observatory Report. 2009. Available: http://www.nanog.org/meetings/nanog47

[4] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "The flattening internet topology: Natural evolution, unsightly barnacles or contrived collapse?"in *Proc. of Passive and Active Network Measurement (PAM`08)*, 2008, pp. LNCS, vol. 4979.

[5] Yang Xiang; Xia Yin; Zhiliang Wang; Jianping Wu; , "Internet Flattening: Monitoring and Analysis of Inter-Domain Routing," *Communications (ICC), 2011 IEEE International Conference on* , vol., no., pp.1-6, 5-9 June 2011

[6] S. Zhou and R. J. Mondragon. Accurately modeling the internet topology. Physical Review E, 2004

[7] Qian Chen; Hyunseok Chang; Govindan, R.; Jamin, S.; , "The origin of power laws in Internet topologies revisited," INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE, vol.2., pp. 608- 617 vol.2, 2002.

[8] E. Zegura, K.L. Calvert, and M.J. Donahoo. A quantitative comparison of graph-based models for Internet topology. IEEE/ACM Transactions on Networking, Vol. 5, No. 6, 1997.

[9] L Li, D Alderson, W Willinger. A First-Principles Approach to Understanding the Internet's Router-level Topology. ACM SIGCOMM 2004.

[10] Shi Zhou; Mondragon, R.J. , "The rich-club phenomenon in the Internet topology," *Communications Letters, IEEE* , vol.8, no.3, pp. 180- 182, March 2004

[11] Xin Zhang; Hai Zhao; Lifei Wang; Chao Li; , "On the internet AS-level topology evolution basing on hierarchy characteristic," *Control and Decision Conference, 2008. CCDC 2008. Chinese* , vol., no., pp.4470-4475, 2-4 July 2008

[12] Z. Ge, D. Figueiredo, S. Jaiwal, , and L. Gao. On the hierarchical structure of the logical Internet graph. In SPIE ITCOM, 2001..

[13] X Dimitropoulos, D Krioukov, G Riley. Revealing the Autonomous System Taxonomy:The Machine Learning Approach. Arxiv preprint cs/0604015, 2006.

[14] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph Evolution: Densification and Shrinking Diameters. ACM Transactions on Knowledge Discovery from Data (ACM TKDD), 2007.

[15] Dhamdhere, A.; Dovrolis, C.; , "Twelve Years in the Evolution of the Internet Ecosystem," *Networking, IEEE/ACM Transactions on* , vol.19, no.5, pp.1420-1433, Oct. 2011

[16] A. Dhamdhere, C. Dovrolis, "The Internet is flat: modeling the transition from a transit hierarchy to a peering mesh," in Proceeding Co-NEXT '10 Proceedings of the 6th International Conference ACM New York, NY, USA, 2010.

[17] R. Alhn and A. BuraMsi, "Topology of evolving network Local cveou and universslity," Physical Revlew Letters. vol. 85, DO. 24, pp. 523G5237. December 2000.

[18] T Bu and D. Towsley, "On distinguishing ktween Internet power law topology geoerators:' in Proceedings of INFOCOM. 2002.

[19] G. Siganos, M. Faloutsos, and C. Faloutsos. The Evolution of the Internet: Topology and Routing. University of California, Riverside technical report.

[20] The CAIDA AS Relationships Dataset, 2004-2010, http://www.caida.org/data/active/as-relationships/

[21] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, kc claf fy, and G. Riley. AS relationships: Inference and validation. Comput Commun Rev, 37(1), 2007.

[22] RISwhois: IPv4/IPv6 address to origin mapping. [Online].Available:http://www.ripe.net/projects/ris/riswhois.html

[23] Claffy, K.; Young Hyun; Keys, K.; Fomenkov, M.; Krioukov, "Internet Mapping: From Art to Science," *Conference For Homeland Security, 2009. CATCH '09. Cybersecurity Applications & Technology* , vol., no., pp.205-211, 3-4 March 2009.

[24] Routeviews. University of Oregon Route Views Project. http://www.routeviews.org.

[25] H. Tangmunanmkit. R. Govindan. S. Jamin. S. Shenker, and W. Willinger. "Network topology generators: Degree-based VS. structural." in SIGCOMM, 2002.

[26] S. Shakkottai, M. Fomenkov, R. Koga, D. Krioukov, and K. Claffy, Eur Phys J B 74, 271 (2010).

[27] The CAIDA AS Information, 2011.3, http://www.caida.org/data/active/as-relationships/.

[28] H. Jeong, Z. Neda, and A. L. Barabasi, Europhys Lett 61, 567 (2003).