THE EUROPEAN
PHYSICAL JOURNAL B

Regular Article

Evolution of the Internet AS-level ecosystem

S. Shakkottai[1], M. Fomenkov[2], R. Koga[2], D. Krioukov[2,a], and K.C. Claffy[2]

[1] Department of ECE, Texas A&M University, College Station, TX 77843-3128, USA
[2] Cooperative Association for Internet Data Analysis, University of California, San Diego, La Jolla, CA 92093-0505, USA

Abstract. We present an analytically tractable model of Internet evolution at the level of autonomous systems (ASs). We call our model the multiclass preferential attachment (MPA) model. As its name suggests, it is based on preferential attachment. All of its parameters are measurable from available Internet topology data. Given the estimated values of these parameters, our analytic results predict a definitive set of statistics characterizing the AS topology structure. These statistics are not part of the model formulation. The MPA model thus closes the "measure-model-validate-predict" loop, and provides further evidence that preferential attachment is a driving force behind Internet evolution.

1 Introduction

The Internet is a paradigmatic example of a complex network. Many researchers have used publicly available data on Internet topology and its observed evolution to test a variety of physical models of complex network structure and dynamics. The large-scale Internet topology represent the structure of connections between companies owing parts of the Internet infrastructure, each company roughly corresponding to an *autonomous system (AS)* [1]. An AS might be a transit Internet service provider (ISP) or customer, a content provider or sink, or any combination of these. Some ASs span multiple continents and are highly interconnected, while others are present at a single geographical location and have only a few links. In 1999 Faloutsos et al. [2] observed that despite all this diversity, the distribution of AS degrees obeys a simple power law. This observation remains valid today, after ten years of Internet evolution [3].

Many researchers have attempted to model the Internet as an evolving system [4–17], and have studied its properties [18–21]. However, questions regarding the main drivers behind Internet topology evolution remain [22]. In this paper our main objective is to create an evolutionary model of the AS-level Internet topology that simultaneously:

1. is realistic;
2. is parsimonious;
3. has *all* of its parameters measurable;
4. is analytically tractable; and
5. "closes the loop".

[a] e-mail: dima@caida.org

Parsimony implies that the model should be as simple as possible, and, related to that, the number of its parameters should be as small as possible. The fifth requirement means that if we substitute measured values of these parameters into analytic expressions of the model, then these expressions will yield results matching empirical observations of the Internet. However, most critical is the third requirement [22]: as soon as a model has even a few unmeasurable parameters, one can freely tune them to match observations. Such parameter tweaking to fit the data may create an illusion that the model "closes the loop", but in the end it inevitably diminishes the value of the model because there is no rigorous way to tell why one model of this sort is better than another since they all match observations. To the best of our knowledge, the multiclass preferential attachment (MPA) model that we propose and analyze in this paper is the first model satisfying all five requirements listed above.

A salient characteristic of our model is that we distinguish between two kinds of ASs: ISPs and non-ISPs. The main difference between these two types of ASs is that while both ISPs and non-ISPs can connect to ISPs, no new AS will connect to an existing non-ISP since the latter does not provide transit Internet connectivity. In Section 2 we analyze the effect of this distinction on the degree distribution. In Section 3 we account for other processes. ISPs can form peering links to exchange traffic bilaterally. They can also go bankrupt and be acquired by others. Finally, they can multihome, i.e., connect to multiple providers. Prior work has often focused on these processes as the driving forces behind Internet topology evolution. We show that in reality they have relatively little effect on the degree distribution. Using the best available Internet topology data, we measure the parameters

reflecting all the process above, and analytically study how they affect the degree distribution.

However, the degree distribution alone does not fully capture the properties of the Internet AS graph [23]. The dK-series formalism introduced in [23] defines a systematic basis of higher-order degree distributions/correlations. The first-order ($1K$) degree distribution reduces to a traditional degree distribution. The second-order ($2K$) distribution is the joint degree distribution, i.e., the correlation of degrees of connected nodes. The distributions can be further extended to account for higher-order correlations [23], or for different types of nodes and links, called annotations [24]. In the economic AS Internet, there are two types of links: links connecting customer ASs to their providers (customer-to-provider links), and links connecting ISPs to their peers (peer-to-peer links). Reproducing the $2K$-annotated distribution of AS topologies suffices to accurately capture virtually all important topology metrics [23,24]. An important feature of the MPA model is that, by construction, it naturally annotates the links between ASs by their business relationships, and that we can analytically calculate, in Section 4, the distributions for the number of peers, customers, and providers that nodes have. In Section 5 we perform a $2K$-annotated test. We generate synthetic graphs using the MPA model, and find that these graphs exhibit a startling similarity to the observed AS topology according to almost all the $2K$-annotated statistics. This validation, in conjunction with observations in [23,24], ensures that other important topology metrics also match well.

2 Two-class preferential attachment model

In the original preferential attachment (PA) model [25], there is only one type of nodes. Suppose that nodes arrive in the system at the rate of one node per unit time, and let them be numbered $s = 1, 2, 3, ...$ as they arrive. Then the number of nodes in the system at time t is equal to t. A node entering the system brings a link with it. One end of the link is already connected to the entering node, while the other end is loose. We call such an un-associated end a *loose connection*. According to the PA model, nodes attach to existing ones with a probability proportional to their degrees. Thus, the probability that a node of degree k is selected as a target for the incoming loose connection, is its degree divided by the total number of existing connections in the system, which is $\frac{k}{2t}$. The original PA model yields a power-law degree distribution $P(k) \sim k^{-\gamma}$ with $\gamma = 3$, but the linear preference function can be modified by an additive term such that the model produces power laws with any $\gamma > 2$ [25–27].

In the Internet, there are two fundamentally different types of ASs – ISPs and non-ISPs – that differ in whether they provide traffic carriage between the ASs they connect or not. No new AS would connect to an existing non-ISP since it cannot provide Internet connectivity. No other work has attempted to model this observation, which is fundamental to understanding the evolution of Internet
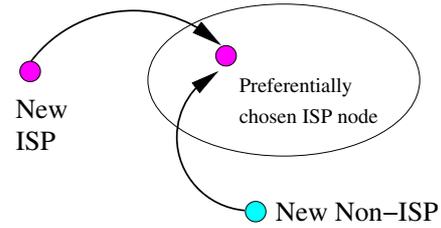

Fig. 1. (Color online) Two-class preferential attachment model. There are ISP nodes and non-ISP nodes.

AS-level topology. Thus, our first modification of the PA model is to consider these two classes of nodes (Fig. 1). New ISPs appear at a rate 1 and connect to other ISP-nodes with a linear preference. New non-ISPs appear at some rate ρ per unit time, implying that the ratio r of non-ISP nodes to the total number of nodes is

$$r = \frac{\rho}{1 + \rho}.$$

Non-ISPs attach to existing ISP-nodes with a linear preference. However, *no further attachments to non-ISPs can occur*. Thus, with respect to degree distribution, only the ISP-nodes contribute to the tail of the power law as the non-ISP nodes will all have degree 1. Given that a link between an ISP node and a non-ISP node has only one end that contributes to the degree of ISP nodes, and since we look only at ISP nodes to find the degree distribution, the links that have an ISP node on one end and a non-ISP node on the other should be counted only once, i.e., they are counted as contributing 1 connection.

Since each ISP node contributes 2 connections to the network and a non-ISP node contributes only 1, the total number of existing connections in the network at time t is $(2 + \rho)t$, which implies that the probability of any loose connection connecting to an ISP-node of degree k is

$$\frac{k}{(2 + \rho)t}. \tag{1}$$

We use the notation $p(k, s, t)$ to denote the probability that an ISP node s has degree k at time t. Then the average degree of an ISP node s at time t is

$$\bar{k}(s, t) = \sum_k k p(k, s, t). \tag{2}$$

Both entering ISPs and non-ISPs have one loose connection each, so the number of loose connections entering the system at time t is $1 + \rho$. Then, from (1), the continuous-time model of the system is

$$\frac{\partial \bar{k}(s, t)}{\partial t} = \frac{1 + \rho}{(2 + \rho)t} \bar{k}(s, t), \tag{3}$$

with boundary condition $\bar{k}(t, t) = 1$ for $t > 1$. Solving this equation yields

$$\bar{k}(s, t) = \left(\frac{s}{t}\right)^{-\frac{1+\rho}{2+\rho}}. \tag{4}$$

This model represents a deterministic system in which if ISP node s has degree $\bar{k}$, then ISP nodes that arrived before s (in the interval $[0, s)$) have degree at least $\bar{k}$. Thus, s also represents the number of ISP nodes that have degree at least $\bar{k}$. It follows from (4) that the number of ISP nodes that have degree $\bar{k}$ or higher is

$$\frac{t}{\bar{k}^{\frac{2+\rho}{1+\rho}}}.$$

Note that the number of ISP nodes that arrive in $[0, t]$ is just t. Hence, the *fraction* of ISP nodes that have degree $\bar{k}$ or higher is

$$\frac{1}{\bar{k}^{\frac{2+\rho}{1+\rho}}}.$$

Since this fraction is essentially a complimentary cumulative distribution function (CCDF), we differentiate it and multiply by -1 to obtain the density function

$$f(\bar{k}) = \frac{2+\rho}{1+\rho}\bar{k}^{-\left(2+\frac{1}{1+\rho}\right)}, \qquad (5)$$

which corresponds to the probability distribution

$$P(k) \sim k^{-\left(2+\frac{1}{1+\rho}\right)} = k^{-(3-r)}. \qquad (6)$$

Validation against observed topology 1.
Dimitropoulos et al. [28] applied machine learning tools to the best available data from the Internet registry (WHOIS) and routing (border gateway protocol (BGP)) systems to classify ASs into several different classes, such as Tier-1 ISPs, Tier-2 ISPs, Internet exchange points, universities, customer ASs, and so on. They validated the resulting taxonomy by direct examination of a large number of ASs. We use their results and divide ASs into two classes based on whether they are ISPs or not. According to [28] (the dataset is available at http://www.caida.org/data/active/as_taxonomy/), the number of ISPs is about 30% of ASs, while non-ISPs make up the other 70%, i.e., $r = 0.7$ and $\rho = 7/3$. The measured value of ρ yields $P(k) \sim k^{-2.3}$, with exponent close to the observed value between -2.1 and -2.2 [2,29,30].

The key point of this section is that the observed value of the power-law exponent close to 2 finds a natural and simple explanation: it is due to preferential attachment and to a directly measured high proportion of non-ISP nodes to which newly appearing nodes cannot connect.

3 Multiclass preferential attachment: peering, bankruptcy, multihoming, and geography

In this section we add further refinements to our model and show that, contrary to common beliefs, none of these refinements have a significant impact on the degree distribution shape.

Relationships between ASs change over time, as ASs pursue cost-saving measures. If the magnitude of traffic flow between two ISPs is similar in both directions,
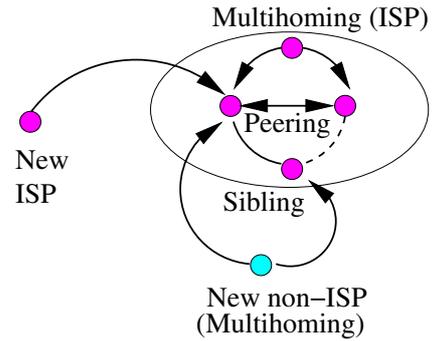

Fig. 2. (Color online) The multiclass preferential attachment model.

then reciprocal peering with each other allows each ISP to reduce its transit costs. Under the assumption that all customer ASs generate similar volumes of traffic, high degree ASs would exchange high traffic volume and rationally seek to establish reciprocal peering with other high degree ASs. We denote the rate at which peering links appear by c. The probability that a new peering link becomes attached to a pair of ISP-nodes of degree k_1 and k_2 is proportional to $k_1 k_2$.

When ISPs go bankrupt, their infrastructure is usually acquired by another ISP, which then either merges the ASs or forms a "sibling" relationship in which their routing domains appear independent but are controlled by one umbrella organization. Thus, in terms of the topology, bankruptcy means that a connection shifts from one ISP to another. Since high degree ISPs tend to be wealthier, they are more likely to be involved in such takeovers. We denote the rate of bankruptcy by μ per unit time.

A growing AS may decide to multihome, i.e., to connect to at least two Internet providers. One would expect that higher degree ISPs with a need for reliability would multihome to other higher degree ISPs. We model this phenomenon by assuming that multihoming links appear in the system at rate ν per unit time. The probability that a new link becomes attached to a pair of ISP-nodes of degree k_1 and k_2 is proportional to $k_1 k_2$. The links are directed from the customer to the provider, and we assume that the higher degree ISP is the provider. We also assume that non-ISPs multihome to an average of m providers each. The model is illustrated in Figure 2.

We analyze this complete MPA model, using techniques similar to that of Section 2. The relation we obtain for $\bar{k}(s, t)$ is

$$\bar{k}(s, t) = \left(\frac{s}{t}\right)^{-\alpha} + \frac{\mu}{\alpha}, \qquad (7)$$

where

$$\alpha = \frac{1 + 2\nu + m\rho + 2c + m\mu}{2 + 2\nu + m\rho + 2c}. \qquad (8)$$

Proceeding in a manner identical to Section 2 yields the relation

$$P(k) \sim k^{-\gamma}, \qquad (9)$$

where

$$\gamma = \frac{1}{\alpha} + 1 = 2 + \frac{1 - \mu}{1 + 2\nu + m\rho + 2c + \mu}. \qquad (10)$$

Validation against observed topology 2. We used the annotated Route Views data [31] from [32] (the dataset is available at http://www.caida.org/data/active/as_taxonomy/) in order to obtain the empirical distribution of number of ISPs to which ASs multihome. We find that the average number of providers that ISPs connect to is 2, meaning that $\nu = 1$. Indeed, since ISPs arrive at rate 1 and choose one provider initially, multihoming links entering at rate $\nu = 1$ yield the average number of providers per ISP of 2. The average number of providers that non-ISPs multihome to is 1.86, i.e., $m = 1.86$. Dimitropoulos et al. [32] also showed that roughly 90% of links are of customer-provider type, i.e., these links pertain to transit relationships, with payments always going to the provider ISP. They find (a lower bound of) 10% of links are peering, i.e., these links correspond to bilateral traffic exchange without payment. In the model, customer links appear in the system at a rate of $1 + m\rho$. We thus calculate $c = (1 + \nu + m\rho)/9 = 0.704$ peering links per unit time. The authors of [32] also estimate that the fraction of sibling links is too small to measure accurately and we take $\mu = 0$. Substituting these values into the exponent expression (10) results in $\gamma = 2.114$ that matches the observed values lying between 2.1 and 2.2 [2,29,30].

The key point of this section is that the large ratio of non-ISPs to ISPs ρ is the dominating term in determining the value of γ, bringing it down from 3 to 2.3, while all the other parameters are less significant, decreasing γ further from 2.3 to 2.1, its observed value. However, the extensions considered in this section do strongly affect other network properties, such as clustering, as we will see in Section 5.

Our last comment in this section concerns geography. We could divide the world into different geographical regions, each growing at a different rate. Due to the self-similar nature of scale-free topologies [33], the resulting graph would still bear identical properties to the MPA model as long as the parameters ρ, c, and μ are the same in all regions. Evidence supporting this hypothesis is available in [30,34], where Chinese or European parts of the Internet are shown to have properties virtually identical, after proper rescaling, to the global AS topology.

4 Peer, customer, and provider distributions

In this section we calculate the distributions for the number of peers, customers, and providers that ISPs have in the MPA model.

The first two distributions are power laws with the exponents equal to γ, the exponent of the overall degree distribution. To see this, we first focus on the peer distribution. We denote the average number of peers that an ISP arrived at time s has at time t by $\bar{\zeta}(s,t)$. The dynamics of peer link formation is given by

$$\frac{\partial \bar{\zeta}(s,t)}{\partial t} = \frac{2c}{(2 + 2\nu + m\rho + 2c)t} \bar{k}(s,t), \qquad (11)$$

since $2c$ is the rate of arrival of loose peer connections, which attach to target nodes with probability proportional to the target node degree. If we define

$$\beta = \frac{2c}{(2 + 2\nu + m\rho + 2c)},$$

so that,

$$\frac{\partial \bar{\zeta}(s,t)}{\partial t} = \frac{\beta}{t} \bar{k}(s,t), \qquad (12)$$

then after the substitution of $\bar{k}(s,t)$ from (7) with $\mu = 0$, we have

$$\frac{\partial \bar{\zeta}(s,t)}{\partial t} = \frac{\beta}{t} \left(\frac{s}{t}\right)^{-\alpha}. \qquad (13)$$

We then solve the above for $\bar{\zeta}(s,t)$:

$$\bar{\zeta}(s,t) = \frac{\beta}{\alpha} \left(\frac{s}{t}\right)^{-\alpha}. \qquad (14)$$

Thus, the number s of ISP nodes that have $\bar{\zeta}$ or more peers is

$$s = \frac{t}{\left(\frac{\alpha}{\beta}\bar{\zeta}(s,t)\right)^{\frac{1}{\alpha}}}. \qquad (15)$$

Dividing the right side by t (the total number of ISP nodes in the system at time t) gives the cumulative distribution of the average number of peers of ISPs. Differentiating and multiplying by -1 yields the distribution

$$f(\bar{\zeta}) = \frac{\bar{\zeta}^{-\left(\frac{1}{\alpha}+1\right)}}{\alpha \left(\frac{\alpha}{\beta}\right)^{\frac{1}{\alpha}}}, \qquad (16)$$

which corresponds to $P(\zeta) \sim \zeta^{-\gamma}$ with the same power law exponent $\gamma = \frac{1}{\alpha} + 1$ as in (10). We can show in an identical fashion that the distribution for the number of customers that ISPs have follows the same power law distribution. We will check the validity of these results in Section 5.

We next show that the distribution for the number of providers that ISPs have is a random variable $1 + X$, where X approximately follows an exponential distribution with parameter ν. According to the model, multihoming links between ISPs are directed from the lower to higher degree ISP. Although the probability that the multihoming link connects to a particular ISP is proportional to the ISP's degree, the probability that the chosen ISP is the customer end of the multihoming link is inversely proportional to its degree. We can thus expect the probability that any particular ISP obtains an extra provider at any time to be approximately the same. Since upon its arrival each ISP always chooses a provider, the minimum number of providers of ISPs is 1. If we now denote by $1 + \bar{p}(s,t)$ the average number of providers that ISPs, appeared at time s, have at time t, then $\bar{p}(s,t)$ is a solution of the following equation

$$\frac{\partial \bar{p}(s,t)}{\partial t} = \frac{\nu}{t}, \qquad (17)$$
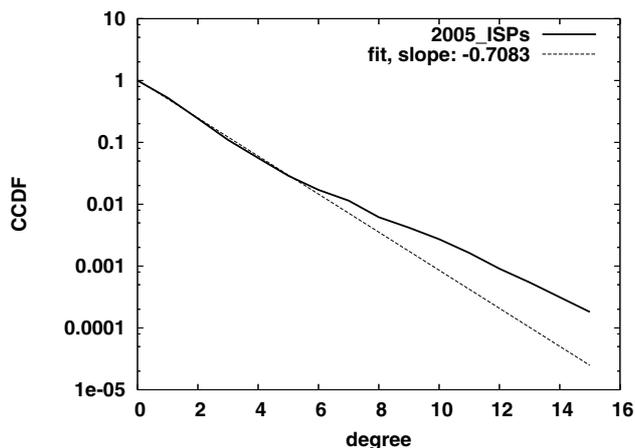
Fig. 3. Provider distribution.

because the rate of arrival of multihoming links is ν. Solving the above expression with the initial condition $\bar{p}(s, s) = 0$, we obtain

$$s = te^{\frac{-\bar{p}}{\nu}}. \tag{18}$$

Since the number of ISPs at time t is just t, this expression implies that the distribution of the number of ISPs that have $\bar{p}$ or more multihoming links is exponential $e^{\frac{-\bar{p}}{\nu}}$. Since all ISPs (except the first) have at least one provider, the distribution for the number of providers that ISPs have should approximately be a random variable $1 + X$, where X is exponentially distributed with parameter ν. Given that the argument above is rather heuristic, we can hardly expect this distribution in the real Internet to be exactly exponential. However, the most important consequence of this argument is that this distribution is quite unlikely to be heavy-tailed.

Validation against observed topology 3. Using the same data from [32], the complementary cumulative distribution function (CCDF) for the number of providers that ISPs have (after subtracting the one initial provider) versus the ISP degree is shown in Figure 3 plotted in the semi-log scale. The exponential curve fit to the initial part of the graph has a slope of -0.7, i.e., the average number of providers is $1 + 1/0.7 = 2.4$, which is close to our empirically measured mean value of 2 in Validation 2. The purpose of this validation is to show not that the distribution is exactly of the form $1 + X$ where $X \sim \exp(1/\nu)$, but that it is definitely not a power law.

5 Model validation by simulation

We have developed a model that describes the evolution of the AS-level topology and validated the analytical results using measured parameters. We now simulate the MPA model using all of the measured parameters.

The MPA model generates *annotated* graphs, with links connecting either customers to providers (c2p links)

or peers to peers (p2p links). Therefore the total node degree is a sum of the degrees of three types – the numbers of customers, providers, and peers attached to a node. Dimitropoulos et al. [24] have shown that the $2K$-annotated distribution of the Internet essentially defines its structure. In other words, if one randomizes the Internet preserving its $2K$-annotated distribution, then the randomized topologies will be almost identical to the original Internet topology. The $2K$-annotated distribution is a generalization of the joint degree distribution for graphs with links annotated by their types. These types can be abstracted by colors, and the traditional scalar node degree becomes a vector of colors specifying how many links colored by what colors are attached to the node. In the Internet case, these colors are *customer*, *provider*, and *peer*. The $2K$-annotated distribution is then the joint distribution for the vectors of colored degrees of nodes connected by differently colored links.

Given the findings in [24], in order to show that our model reproduces the Internet structure, it suffices to compare the $2K$-annotated distributions in simulated networks and the Internet. Unfortunately, the $2K$-annotated distribution is too multi-dimensional and sparse. Therefore, we can work only with its projections. The reasonable and informative projections include [24]: (i) the degree distribution (DD): the traditional distribution of total node degrees, i.e., the number of links of all types attached to a node; (ii) the annotated distributions (ADs): the distributions of the number of customers, providers, and peers that nodes have, i.e., the distribution of degrees of each type; (iii) the annotated degree distribution (ADD): the joint distribution of customers, providers, and peers of nodes, measuring the per-node correlations among these three degree types; and (iv) the joint degree distributions (JDDs): the JDDs measure the correlations of total node degrees "across the links" of different types.

Therefore, in our validation we compare all these metrics between the graphs that the MPA-model produces and the Internet topology annotated with AS relationships using [32] on Jan. 1, 2007. The specific dataset used is http://as-rank.caida.org/data/2007/as-rel.20070101.a0.01000.txt, which is a part of [35] providing publicly available weekly snapshots of the annotated Internet. These snapshots are based on the Route Views BGP data [31]. The values of the parameters we use in our simulations are $\rho = 2.3$, $\nu = 1$, $c = 0.704$, and $m = 1.86$, which we recall are the ratio of the numbers of non-ISPs to ISPs, ratio of ISP multihoming links to ISPs, ratio of peering links to ISPs, and the average number of providers to which non-ISPs multihome, respectively. We run the simulation with deterministic link arrivals based on their arrival rates. The total number of ISP and non-ISP nodes are 7200 and 16 800. These are the numbers of nodes that we were able to classify in the dataset using [28]. We do not model bankruptcy since as mentioned earlier, the rate at which it occurs is too small to get an accurate estimate of our bankruptcy ratio μ.

Degree distribution (DD): Figure 4a shows the DD of the graphs generated by the MPA model, and its
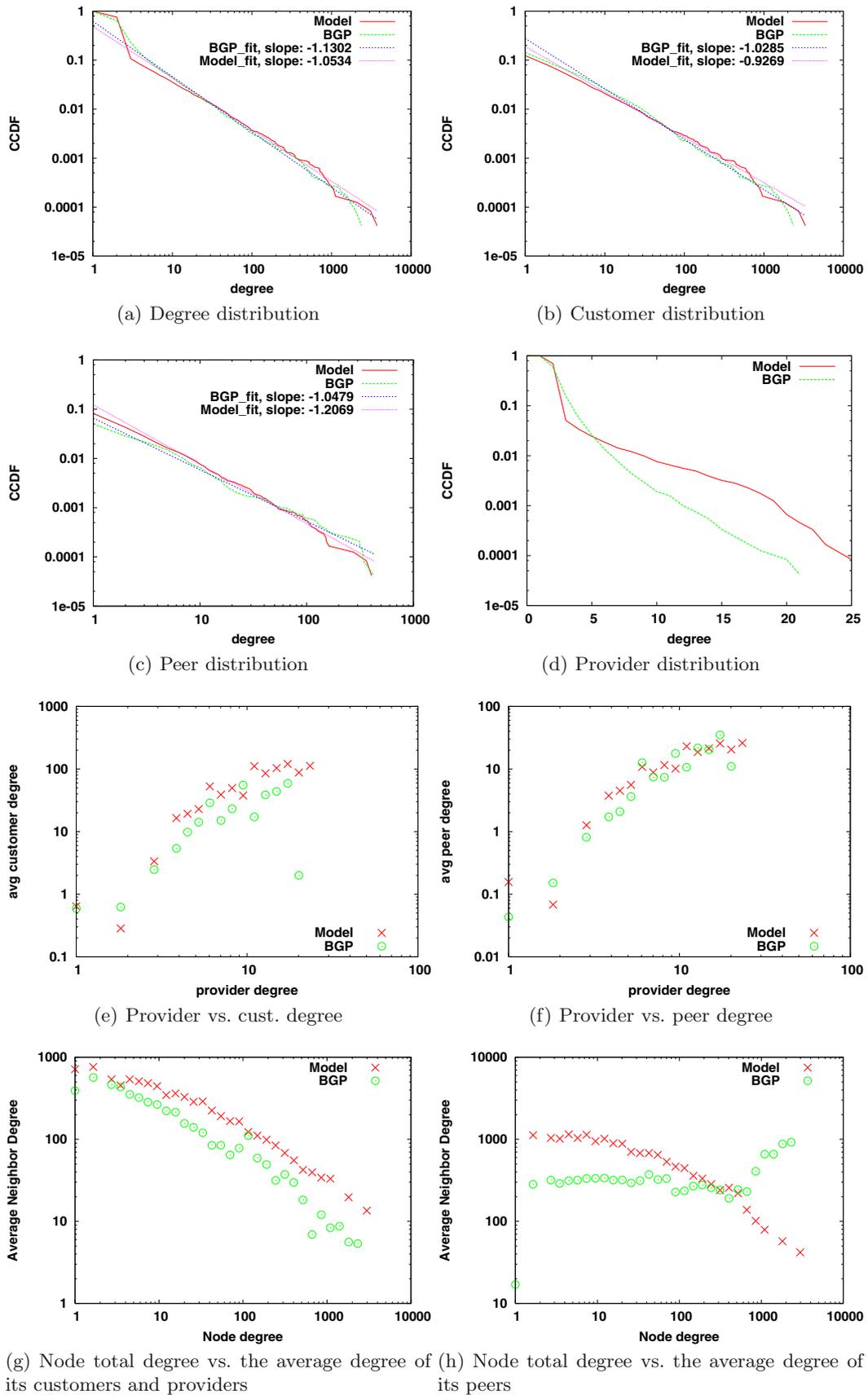
(a) Degree distribution

(b) Customer distribution

(c) Peer distribution

(d) Provider distribution

(e) Provider vs. cust. degree

(f) Provider vs. peer degree

(g) Node total degree vs. the average degree of (h) Node total degree vs. the average degree of its customers and providers   its peers

Fig. 4. (Color online) Validation of the MPA model by simulation.

(a) Average neighbor degree
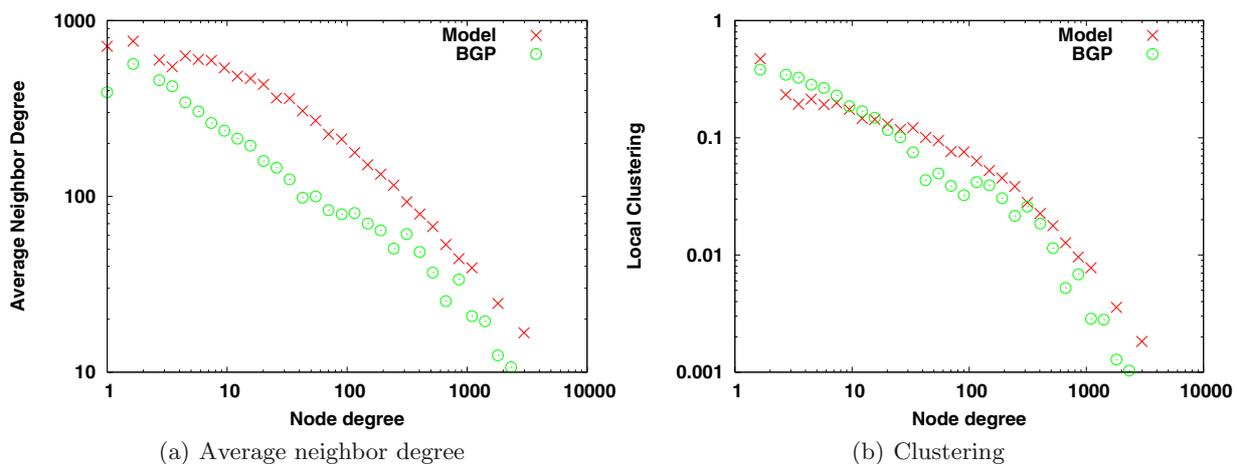

(b) Clustering

Fig. 5. (Color online) Average neighbor degree and clustering in simulations vs. real data.

comparison with the observed topology. As predicted in Section 3 the MPA model produces a power law DD, and the exponent of the CCDF matches well the BGP data.

Annotated distributions (AD): Figures 4b–4d show the ADs generated by the MPA model. We compare the customer, peer, and provider degree distributions of the simulated graph with that of the BGP tables. As predicted, the ADs of number of customers and peers are both power law graphs with the same exponent as the DD.

We plot the CCDF of the number of providers that ISPs multihome to (on linear x-axis and logarithmic y-axis) in Figure 4d. They are approximately of form $1 + X$, where X is exponentially distributed. The curves show a discrepancy in slope. We believe that it arises due to the fact that almost all the distribution mass is concentrated at small degrees, as the mean is 2, and the number of ISPs with high multihoming degree is small.

Annotated degree distribution (ADD): Each ISP has some numbers of providers, peers, and customers. The ADD is the joint distribution of these numbers across all observed ISPs. We illustrate these correlations in Figures 4e and 4f. To construct these plots, we first bin the ISP nodes by the number of providers that they have (the x-axis), and then compute the average number of customers or peers that the ISPs in each bin have (the y-axis). We observe that the MPA model approximately matches the BGP data against these metrics as well.

Joint degree distributions (JDDs): While the ADD contains information about the correlations between the numbers of different types of nodes connected to an ISP, it does not reveal information about the degree correlations between the parameters of different ISPs connected to each other, i.e., whether higher degree ISPs are more likely to peer with each other, etc. This information is contained in the average neighbor connectivity, which is a summary statistic of the joint degree distributions in Figures 4g and 4h. Specifically, let the probability that a node of degree k has a c2p link to a node of degree k' be called $P_{c2p}(k'|k)$. Then the average degree of the provider ISPs of ISPs that have degree k is $\bar{k}_{c2p}(k) = \sum_{k}' k' P_{c2p}(k'|k)$. In

a full mesh graph with n nodes and undirected links, since all nodes have degree $n - 1$, the value of this coefficient is simply $n-1$. We show the normalized value $k_{c2p}(k)/(n-1)$ in Figure 4g. The similarly normalized values of $\bar{k}_{p2p}(k)$ are shown in Figure 4h. These functions exhibit similar behaviors for the MPA model and BGP data.

Coupled with observations in [24] that the real Internet topology is accurately captured by its $2K$-annotated distribution, the results in this section provide evidence that the MPA model reproduces closely the Internet AS-level topology across a wide range of metrics. As an example, we show in Figure 5 two standard topology metrics: the average neighbor degree and clustering as functions of the total node degree. We observe that even though two-class preferential attachment in Section 2 produces tree networks, its multiclass extensions in Section 3, implemented in our simulations, closely reproduce clustering observed in the real Internet, which is a consequence of the $2K$-annotated randomness of the Internet, and the $2K$-annotated distribution match in Figure 4. Finally, the average total node degree in the real Internet and simulations is 4.1 and 4.2 respectively.

6 Conclusion

We constructed a realistic and analytically tractable model of the Internet AS topology evolution that we call the multiclass preferential attachment (MPA) model. The MPA model is based on preferential attachment, and we believe it uses the *minimum number of measurable parameters* altering standard preferential attachment to produce annotated topologies that are remarkably similar to the real AS Internet topology. Each model parameter reflects a realistic aspect of AS dynamics. We measure all parameters using publicly available AS topology data, substitute them in our derived analytic expressions for the model, and find that it produces topologies that match observed ones against a definitive set of network topology characteristics. These characteristics are projections of the

second-order degree correlations, annotated with AS business relationships. Matching them ensures that synthetic AS topologies match the real one according to all other important metrics [23,24].

The model parameter that has the most noticeable effect on the properties of generated topologies reflects the ratio of ISP to non-ISP ASs. Contrary to common beliefs, the parameters taking care of AS peering, bankruptcies, multihoming, etc., are less important, as far the degree distribution is concerned, although they do affect other network properties, such as clustering. No other parameters or complicated mechanisms appear to be needed to explain the Internet topology annotated with AS business relationships. In other words, preferential attachment, with the MPA modifications, appears to explain the complexity of the AS-level Internet abstracted as an annotated graph. An interesting open question concerns the origins of preferential attachment in the Internet. Given that the vast majority of AS links connect customer ASs to their providers [32], this question reduces to finding how customers select their providers. The popularity of providers, their "brand names", may be an important factor explaining the preferential attachment mechanism acting in the Internet.

References

1. R. Pastor-Satorras, A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, 2004)
2. M. Faloutsos, P. Faloutsos, C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999)
3. A. Dhamdhere, C. Dovrolis, *Ten years in the evolution of the Internet ecosystem*, in *IMC '08: Proceedings of the 8th ACM SIGCOMM conference on Internet measurement* (ACM, New York, NY, USA, 2008), pp. 183–196, ISBN 978-1-60558-334-1
4. R. Albert, A.L. Barabási, Phys. Rev. Lett. **85**, 5234 (2000)
5. S.H. Yook, H. Jeong, A.L. Barabási, Proc. Natl. Acad. Sci. USA **99**, 13382 (2002)
6. K.I. Goh, B. Kahng, D. Kim, Phys. Rev. Lett. **88**, 108701 (2002)
7. H. Chang, S. Jamin, W. Willinger, *Internet connectivity at the AS-level: An optimization-driven modeling approach*, in *MoMeTools '03: Proceedings of the ACM SIGCOMM workshop on Models, methods and tools for reproducible network research* (ACM, New York, NY, USA, 2003), pp. 33–46, ISBN 1-58113-748-8
8. S. Zhou, R.J. Mondragón, Phys. Rev. E **70**, 066108 (2004)
9. S. Zhou, Phys. Rev. E **74**, 016124 (2006)
10. M.Á. Serrano, M. Boguñá, A. Díaz-Guilera, Phys. Rev. Lett. **94**, 038701 (2005)
11. M.Á. Serrano, M. Boguñá, A. Díaz-Guilera, Eur. Phys. J. B **50**, 249 (2006)
12. H. Chang, S. Jamin, W. Willinger, *To Peer or Not to Peer: Modeling the Evolution of the Internet's AS-Level Topology*, in *INFOCOM 2006, 25th IEEE International Conference on Computer Communications*, 23–29 April 2006, Barcelona, Catalunya, Spain (IEEE, 2006)
13. X. Wang, D. Loguinov, *Wealth-Based Evolution Model for the Internet AS-Level Topology*, in *INFOCOM 2006, 25th IEEE International Conference on Computer Communications*, 23–29 April 2006, Barcelona, Catalunya, Spain (IEEE, 2006)
14. R. D'Souza, C. Borgs, J. Chayes, N. Berger, R. Kleinberg, Proc. Natl. Acad. Sci. USA **104**, 6112 (2007)
15. J. Corbo, S. Jain, M. Mitzenmacher, D. Parkes, *An Economically-Principled Generative Model of AS Graph Connectivity*, in *NetEcon+IBC 2007: Proceedings of the ACM Joint Workshop on The Economics of Networked Systems and Incentive-Based Computing* (ACM, 2007)
16. S. Bar, M. Gonen, A. Wool, Comput. Netw. **51**, 4174 (2007)
17. P. Holme, J. Karlin, S. Forrest, Comput. Commun. Rev. **38** (2008)
18. R. Pastor-Satorras, A. Vázquez, A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001)
19. E. Ravasz, A.L. Barabási, Phys. Rev. E **67**, 026112 (2003)
20. J. Park, M.E.J. Newman, Phys. Rev. E **68**, 026112 (2003)
21. G. Bianconi, G. Caldarelli, A. Capocci, Phys. Rev. E **71**, 066116 (2005)
22. D. Krioukov, F. Chung, kc claffy, M. Fomenkov, A. Vespignani, W. Willinger, Comput. Commun. Rev. **37** (2007)
23. P. Mahadevan, D. Krioukov, K. Fall, A. Vahdat, Comput. Commun. Rev. **36**, 135 (2006)
24. X. Dimitropoulos, D. Krioukov, G. Riley, A. Vahdat, ACM T Model Comput. S **19**, 17 (2009)
25. A.L. Barabási, R. Albert, Science **286**, 509 (1999)
26. S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000)
27. P.L. Krapivsky, S. Redner, F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000)
28. X. Dimitropoulos, D. Krioukov, G. Riley, K.C. Claffy, *Revealing the Autonomous System Taxonomy: The Machine Learning Approach*, in *PAM 2006: Proceedings of the 7th International Conference on Passive and Active Network Measurement, Adelaide*, Australia, March 30–31 (2006)
29. H. Chang, R. Govindan, S. Jamin, S.J. Shenker, W. Willinger, Comput. Netw. **44**, 737 (2004)
30. P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, K.C. Claffy, A. Vahdat, Comput. Commun. Rev. **36**, 17 (2006)
31. *University of Oregon Route Views Project*, http://www.routeviews.org/
32. X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, K.C. Claffy, G. Riley, Comput. Commun. Rev. **37** (2007)
33. M.Á. Serrano, D. Krioukov, M. Boguñá, Phys. Rev. Lett. **100**, 078701 (2008)
34. S. Zhou, G. Zhang, G. Zhang, IET Commun. **1**, 209 (2007)
35. CAIDA, *AS Relationships Data*, Research Project, http://www.caida.org/data/active/as-relationships/