Supporting Text

We consider the RIPE WHOIS Internet data as characterized by the Cooperative Association for Internet Data Analysis (CAIDA, www.caida.org) [1, 2], and show that the Tempered Preferential Attachment (TPA) model (see main text) provides an excellent fit to this data. First, we define the complementary cumulative probability distribution (ccdf), and then derive the ccdf for a TPA graph. Next we discuss the ccdf for the WHOIS data. Finally we discuss the fit provided by the TPA model and by a power law with exponential decay (PLED).

Defining the Ccdf

The complementary cumulative probability distribution, ccdf(x):

$$\mathrm{ccdf}(x) = 1 - \sum_{j=1}^{x-1} p_j = \sum_{j=x}^{\infty} p_j. \tag{1}$$

The ccdf predicted by TPA with $A_1 \neq A_2$

First recall the recursion relations. The recursion relations defining the degree distribution for TPA graphs were derived explicitly in refs. 3 and 4. Here, we derive the corresponding ccdf. These are Eq's **16** and **17** in [3]:

$$p_i = \left(\prod_{k=2}^{i} \frac{k-1}{k+w} \right) p_1 = \left(\prod_{k=1}^{i-1} \frac{k}{k+w+1} \right) p_1, \quad \text{for} \quad i \leq A_2, \tag{2}$$

and

$$p_i = \left(\frac{A_2}{A_2+w} \right)^{i-A_2} p_{A_2} \;\; = \;\; q^{i-A_2} \; p_{A_2}, \quad \text{for} \quad i \geq A_2. \tag{3}$$

Note

$$p_{A_2} = \left(\prod_{k=1}^{A_2-1} \frac{k}{k+w+1} \right) p_1, \tag{4}$$

and, for convenience, we defined:

$$q \equiv \left(\frac{A_2}{A_2+w} \right). \tag{5}$$

We will first calculate the ccdf for $i \geq A_2$ as we will use that result to determine the ccdf for $i < A_2$.

Calculating the ccdf, for $x \geq A_2$. Recall the definition of the ccdf from Eq. **1**:

$$\begin{aligned} \mathrm{ccdf}(x) &= \sum_{j=x}^{\infty} p_j \\ &= p_{A_2} \sum_{j=x}^{\infty} q^{j-A_2} \\ &= p_{A_2} \sum_{j=0}^{\infty} q^{j+x-A_2} \\ &= p_{A_2} q^{x-A_2} \sum_{j=0}^{\infty} q^j. \end{aligned} \tag{6}$$

Since $q < 1$, the sum in Eq. **6** is a geometric series; $\sum_{j=0}^{\infty} q^j = 1/(1-q)$. Thus we can write:

$$\boxed{\mathrm{ccdf}(x) = \left(\frac{p_{A_2}}{1-q}\right) q^{x-A_2}, \quad \text{for} \quad x \geq A_2.} \tag{7}$$

Calculating the ccdf, for $x < A_2$. This is slightly more complicated, as we have different functional forms for $x < A_2$ and $x > A_2$.

$$\begin{aligned} \mathrm{ccdf}(x) &= \sum_{j=x}^{\infty} p_j \\ &= \sum_{j=x}^{A_2-1} p_j + \sum_{j=A_2}^{\infty} p_j \\ &= \sum_{j=x}^{A_2-1} p_j + \mathrm{ccdf}(A_2) \\ &= \sum_{j=x}^{A_2-1} p_j + \left(\frac{p_{A_2}}{1-q}\right). \end{aligned} \tag{8}$$

Plugging in the relation for p_i from Eq. **3**, we obtain:

$$\boxed{\mathrm{ccdf}(x) = p_{A_2} \left(\frac{1}{1-q} + \sum_{j=x}^{A_2-1} \prod_{k=j}^{A_2-1} \frac{k+w+1}{k} \right), \quad \text{for} \quad x < A_2.} \tag{9}$$

Standard Normalization. First we can check that Eqs. **7** and **9** give the same value for ccdf(A_2). They do:

$$\mathrm{ccdf}(A_2) = \frac{p_{A_2}}{1-q}. \tag{10}$$

And we can determine the value of p_{A_2} by the normalization condition that

$$\text{ccdf}(1) = 1 = p_{A_2}\left(\frac{1}{1-q} + \sum_{j=1}^{A_2-1}\prod_{k=j}^{A_2-1}\frac{k+w+1}{k}\right). \tag{11}$$

In other words,

$$\boxed{p_{A_2} = \left(\frac{1}{1-q} + \sum_{j=1}^{A_2-1}\prod_{k=j}^{A_2-1}\frac{k+w+1}{k}\right)^{-1}.} \tag{12}$$

Normalizing without degree $d = 1$ nodes. We may want to neglect nodes with degree $d < 2$ for various reasons (see main text and refs 1 and 2). In that case, the normalization would be:

$$\text{ccdf}(2) = 1 = p_{A_2}\left(\frac{1}{1-q} + \sum_{j=2}^{A_2-1}\prod_{k=j}^{A_2-1}\frac{k+w+1}{k}\right). \tag{13}$$

Thus

$$\boxed{p_{A_2} = \left(\frac{1}{1-q} + \sum_{j=2}^{A_2-1}\prod_{k=j}^{A_2-1}\frac{k+w+1}{k}\right)^{-1}} \tag{14}$$

with Eqs. **7** and **9** unchanged (except Eq. **9** now holds for $2 \leq x < A_2$, rather than for $1 \leq x < A_2$).

The WHOIS ccdf, for $d > 1$

We renormalize the WHOIS data to remove $d < 2$ nodes. By definition:

$$\sum_{j=1}^{\infty} p_j = 1.$$

Thus:

$$\sum_{j=2}^{\infty} p_j = 1 - p_1.$$

We want to renormalize ($p'_j = \eta p_j$) such that:

$$\sum_{j=2}^{\infty} p'_j = \eta \sum_{j=2}^{\infty} p_j = 1,$$

Thus $\boxed{\eta = 1/(1-p_1).}$ For the WHOIS data, $p_1 = 0.0573$. and $\eta = 1.0608$. See Fig. 4.

The ccdf for the renormalized probabilities:

$$\text{ccdf}'(\mathrm{x}) = \sum_{\mathrm{j=x}}^{\infty} \mathrm{p}'_{\mathrm{j}} = \eta \sum_{\mathrm{j=x}}^{\infty} \mathrm{p_j} = \eta\ \text{ccdf}(\mathrm{x}).$$

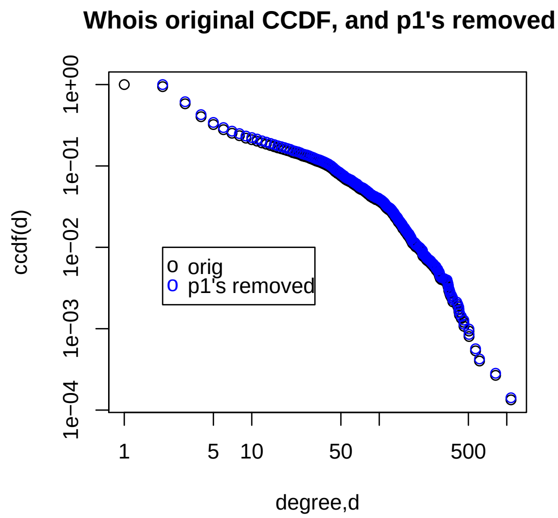

FIG. 4: Original ccdf of WHOIS data, and the renormalized ccdf'(x) = η ccdf(x).

Fitting TPA to WHOIS with $d \geq 2$

The WHOIS $d \geq 2$ distribution is discussed above. The TPA distribution with $d \geq 2$ is the same as with $d \geq 1$ except the value of p_{A_2} is defined as in Eq. **14**, in terms of $d = 2$ instead of $d = 1$. See Fig. 5 for details of the fit.

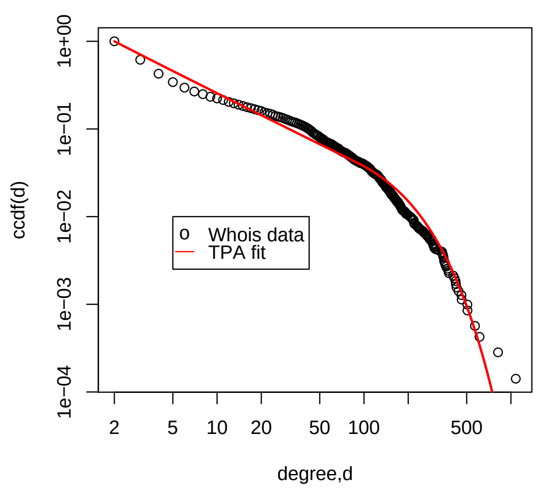

FIG. 5: WHOIS ccdf for $d \geq 2$. Data points are from the WHOIS tables. The solid line is the fit to TPA for $d \geq 2$ with $A_1 = 187$ and $A_2 = 90$ (and thus $\gamma = 1.83$). With this fit, $R = 0.986$, thus $R^2 = 0.972$.

Fitting PLED to WHOIS with $d \geq 2$

Assuming a PLED: $p(x) = Ax^{-b}\exp(-x/c)$. The normalization constant, A, is determined by the relation:

$$\sum_{x=2}^{\infty} p(x) = 1 = A \sum_{x=2}^{\infty} x^{-b}\exp(-x/c).$$

Then the ccdf is as follows. See Fig. 6 for details of the fit.

$$\text{ccdf}(x) = A \sum_{j=x}^{\infty} x^{-b}\exp(-x/c).$$

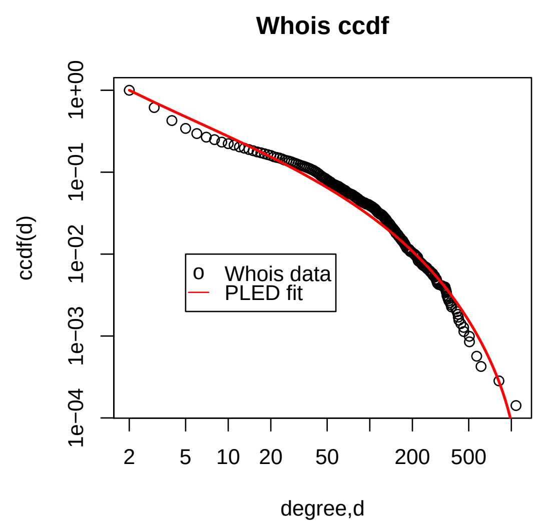

FIG. 6: WHOIS ccdf for $d \geq 2$. Data points are from the WHOIS tables. The solid line is the fit $\text{ccdf}(x) = A\sum_{j=x}^{\infty} x^{-b}\exp(-x/c)$, where $b = 1.63$ and $c = 350$. With this fit, $R = 0.985$, thus $R^2 = 0.970$.

[1] Mahadevan P, Krioukov D, Fomenkov M, Huffaker B, Dimitropoulos X, claffy kc, Vahdat A (2005) arxiv:cs.NI/0508033.

[2] Mahadevan P, Krioukov D, Fomenkov M, Huffaker B, Dimitropoulos X, claffy kc, Vahdat A (2006) *ACM SIGCOMM Comp Comm Rev* 36:17-26.

[3] Berger N, Borgs C, Chayes JT, D'Souza RM, Kleinberg RD (2004) *Lect Notes Comput Sci (ICALP 2004)*, 3142:208–221.

[4] Berger N, Borgs C, Chayes JT, D'Souza RM, Kleinberg RD (2005) *Combinatorics Probability Comput*, 14:697–721.