# Power laws as a pre-asymptotic regime of the PFP mode

Dmitri Krioukov (CAIDA/UCSD)
Paul Krapivsky (BU)
dima@caida.org

CAIDA WIT

May 10-12, 2006

# Motivation

- there are two classes of topology models:
  - static (equilibrium): matching observed data is “easy”
  - growth (non-equilibrium): matching observed data is “hard”
- the Positive-Feedback Preference (PFP) growth model yields a very good match with observed AS-level topology
  - comparisons are made based on the richest set of topology characteristics
  - the model matches them all almost perfectly
  - traceroute (skitter) data is used
- the model does not have analytic solution
  - preferential attachment with
  - super-linear preference rate with
  - multiple link additions
- let’s find an analytic solution and explain the model’s success

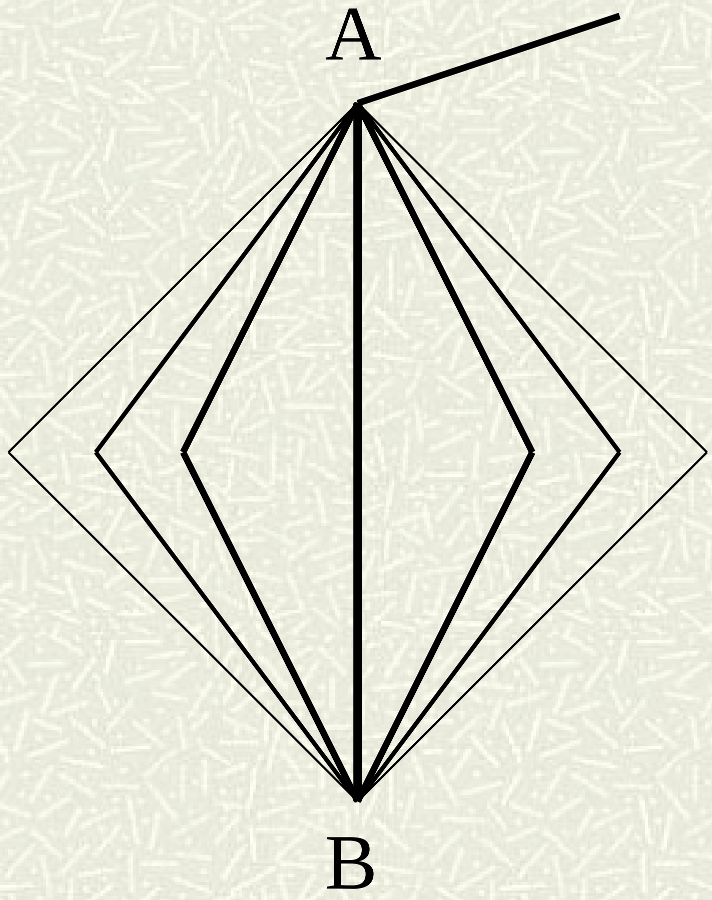
# Solving PFP: one-link additions (known case)

- preference rate is $\sim k^{\delta}$
- connectivity “phase transitions” at $\delta=1+1/p$, $p=1,2,3,\ldots$
  - $N_{k+1}(N) \sim N^{k+1-k\delta}$, if $1 \leq k < p$
  - $N_{k+1}(N)$ is finite otherwise
- extremal growth: $\delta \to \infty$: result is a star
- stars are open $1$-books

# Solving PFP: two-link additions

- link placement options
  - both links are connected to the new node
  - one link is connected to the new node, another – to the host node
- preference rate is as in the one-link case
- extremal growth
- both link placement option lead to the following degree distribution:
  - $N_{N-1}=N_{N-2}=N_1=1$
  - $N_2=N-3$

# Solving PFP: open *2*-book



# Solving PFP: three-link additions

- link placement option
  - one link is connected to a new node, another – to the host node, and the last one – to the host-peer node
- preference rate is as in the one-link case
- extremal growth
- result is the following degree distribution:
  - $N_{N-1}=N_{N-2}=N_{N-3}=N_2=N_1=1$
  - $N_3=N-5$

# Solving PFP: open *3*-book

- highest-degree nodes are *A*, *B*, and *C*
- *D* is the *2*-degree node, *E* is the *1*-degree node
- *F*'s are the *N-5 3*-degree nodes
- binding is triangle *ABC*
- pages are **3**-simplexes (tetrahedrons) *ABCF*'s
- bookmarks are triangles *ABD* and link *AE*

# Solving PFP: *m*-link additions

- link placement option
  - one link is connected to a new node, another – to the host node, and the last one – to the last host-peer node
- preference rate is as in the one-link case
- extremal growth
- result is the following degree distribution:
  - $N_{N-1}=N_{N-2}=\ldots=N_{N-m}=N_{m-1}=N_{m-2}=\ldots=N_1=1$
  - $N_m=N-2m+1$

# Solving PFP: open *m*-book

- space is $(m+1)$-dimensional
- pages are $1$-codimensional (***m***-simplices)
- binding is $2$-codimensional ($(m-1)$-simplices)
- bookmarks are $1,2,\ldots,(m-1)$-dimensional

# Solving PFP: getting rid of the extremal growth assumption

- consider
  - *2*-link additions
  - second link placements option (only one link is attached to the new node)
  - networks of large size $j$ (neglecting difference between $j$, $j-1$, $j-2$, and $j-3$)
- assuming we have an open book, estimate probability that the network will remain an open book after adding a new node
  - since we have 2 $\sim j$-degree nodes and $\sim j$ 2-degree nodes, the probability is
  - $P_{j\to j+1} \sim (j^\delta+j^\delta)/(j^\delta+j^\delta+j2^\delta) \times j^\delta/(j^\delta+j2^\delta)$
  - where the first factor accounts for the link attached to the new node, and the second factor – for the other link

# Solving PFP: getting rid of the extremal growth assumption

- probability that an $N$-size network is still an open book $P_N \sim \Pi_{j=1}^{N}(1+2^{\delta-1}/j^{\delta-1})^{-1}(1+2^{\delta}/j^{\delta-1})^{-1}$, so that
  - $\delta>2$: $P_N$ is finite
  - $\delta=2$: $P_N \sim N^{-6}$
  - $\delta<2$: $P_N \sim \exp(-aN^{2-\delta})$, $a=(2^{\delta}+2^{\delta-1})/(2-\delta)$
- for $\delta<2$, the network is thus not an open book, but the difference is small and still analyzable
- accounting for probabilities of attaching to low-degree nodes, we get
  - $N_3(N)\sim aN^{2-\delta}$, (note that $P_N \sim \exp(N_3(N))$ for $\delta\leq 2$)
  - $N_{k+2}(N)\sim N^{k+1-k\delta}$ for $\delta<1+1/k$
- conjecture for arbitrary $m$
  - $N_{k+m}(N)\sim N^{k+1-k\delta}$

# PFP's solved, but power-laws are not explained

- PFP *can* produce power-laws, because the Internet size is pre-asymptotically small
  - for $N=10^4$, $\delta=1.15$, the percentage of $3$-degree nodes is $N_3(N)/N \sim aN^{1-\delta}=0.98$
  - $N_3(N)/N \sim 0.1$ for $N \sim 10^{10}$, $N_3(N)/N \sim 0.01$ for $N \sim 10^{17}$
  - in derivation of $N_3(N)$, we neglected the loss terms. more accurate result is, for $\delta=1.15$, $N_3(N) \sim aN^{0.85}-bN^{0.7}+cN^{0.55}+\ldots$
  - fluctuations are also neglected
- PFP *does* produce power-laws, because it's designed to match the rich club connectivity

# Rich Club Connectivity vs. Joint Degree Distribution

- rich club connectivity (RCC) $\varphi(r/N)$ is the ratio of the number of links in the subgraph induced by $r$ highest degree nodes to the maximum possible number of such links $r(r-1)/2$
- joint degree distribution (JDD) $M_{kk'}$ is the number of links between $k$ and $k'$-degree nodes
- let $K$ be the maximum degree, $N_k^+ = \sum_{k'=k}^{K} N_{k'}$, then RCC is related to JDD via the total number of links between $k$- and higher-degree nodes
  $\varphi_k = \varphi(N_k^+/N)\, N_k^+(N_k^+-1)/2 - \varphi(N_{k+1}^+/N)\, N_{k+1}^+(N_{k+1}^+-1)/2$
  $= \sum_{k'=k}^{K} M_{kk'}$
- JDD defines RCC, up to reordering of the same-degree-nodes
- RCC constrains JDD

# How RCC constrains JDD

- average neighbor connectivity $k_{nn}(k) = M_k^{-1} \sum_{k'} k' M_{kk'}$, where $M_{kk'}$ is $(1+ \delta_{kk'}) M_{kk'}$ now, so that $M_k = kN_k = \sum_{k'} M_{kk'}$ is the total number of stubs attached to $k$-degree nodes
- consider range of possible values $\Delta(k)$ = max[$k_{nn}(k)$] – min[$k_{nn}(k)$] of $k_{nn}(k)$ with and without RCC constraints
- without constraints, $\Delta(k)=K-1$
- with constraints $\alpha_k=\varphi_k/M_k$, $\Delta_\varphi(k)=(K-2k+2)\alpha_k+k-2$
- relative decrease of freedom $(\Delta(k)-\Delta_\varphi(k))/\Delta(k) \sim (1-k/K) - (1-2k/K)\alpha_k$ $\sim (1-k/K)$ for disassortative networks where $\alpha_k \sim 0$ unless $k/K \sim 0$

# Summary

- the PFP models is so successful, because it matches RCC, RCC constrains JDD, and JDD defines AS-level topologies (i.e., they are *2K*-random)
- PFP does so, not in its asymptotic, but pre-asymptotic regime
- the asymptotic regime does not have any power-laws, it is a collection of open *m*-books (which may be slightly torn ☺)
- what if the ubiquitously observed power-laws are all pre-asymptotes?