On large random graphs of the "Internet type"

Hannu Reittu and Ilkka Norros
VTT Information Technology
{Hannu.Reittu, Ilkka.Norros}@vtt.fi

Introduction

’Complexity’ of the Internet

- connects various networks with no typical topology
- diversity in link capacity
- rapid growth and change of topology
- it is very large
- new ’killer applications’ may appear suddently.

$\Rightarrow$

Modelling is difficult, a desire for ’invariant’ properties.

Few invariants found

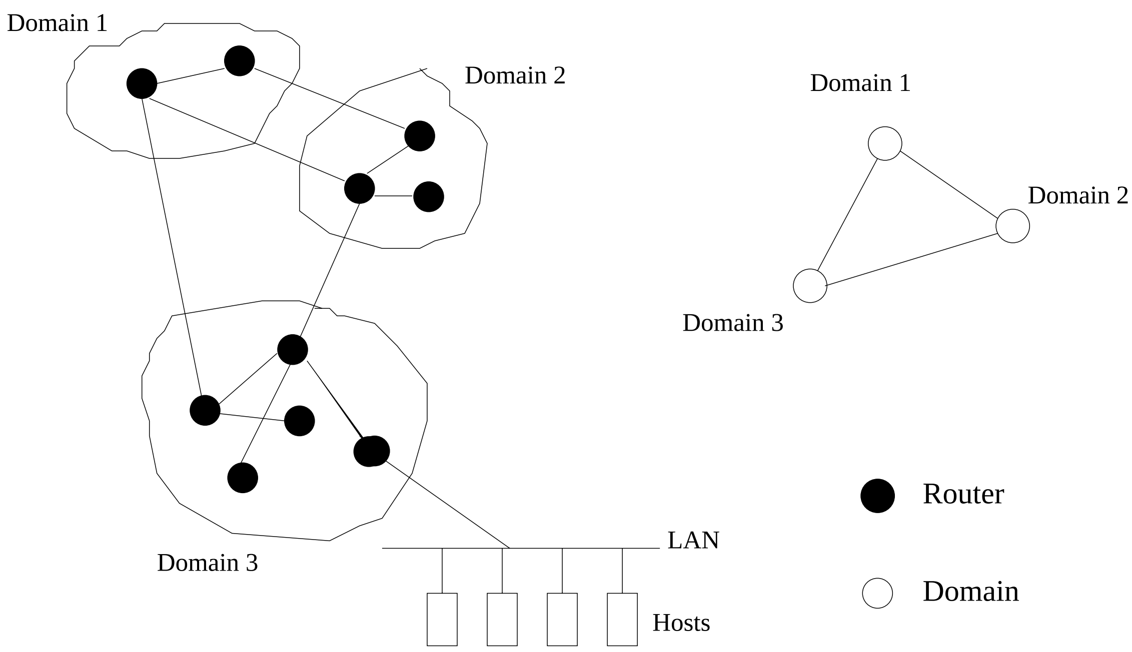
- diurnal changes of activity
- self-similarity of the data traffic
- Poisson arrivals of sessions
- log-normal session durations

A new candidate: a power-law topology of the Internet.

M. Faloutsos, P. Faloutsos, and Ch. Faloutsos. On power-law relationships of the Internet topology. *ACM SIGCOMM'99 (http://www.cs.ucr.edu/ ∼michalis/papers.html)*, pages 251–262, 1999.

Probably helps to understang routing dependent features of the Internet.

Suggested power-laws



Power-Law 1: degree exponent $\mathcal{O}$ The frequency f_d of a degree d is proportional to a constant power, $\mathcal{O}$: $f_d \sim d^{\mathcal{O}}$.

Power-Law 2: Rank exponent $\mathcal{R}$ The degree d_v of a node v is proportional to the rank of the node r_v to a constant power $\mathcal{R}$: $d_v \sim r_v^{\mathcal{R}}$. The rank of the node is found as its position in the list of all nodes ordered in the decreasing order of their degrees.

Power-Law 3 ('Approximation') hop-plot exponent $\mathcal{H}$ The total number of pairs of nodes $P(h)$ within h hops is proportional to the number of hops to a constant power $\mathcal{H}$: $P(h) \sim h^{\mathcal{H}}, h \ll \delta$, the diameter.

Power-Law 4: eigen exponent Υ The eigenvalues λ_i of a graph are proportional to the order i to a constant power Υ: $\lambda_i \sim i^{\Upsilon}$.

The model of Newman, Strogatz and Watts

M.E.J. Newman, S.H. Strogatz, and D.J. Watts. Random graphs with arbitary degree distribution and their applications. *http://arXiv.org/list/cond-mat/0007?200*, pages 1–18, 2000.

Internet $\approx$ a random graph:

N number of nodes in the graph

$D_1, \ldots, D_N$ i.i.d. random variables, $\in \{1, 2, \ldots\}$.

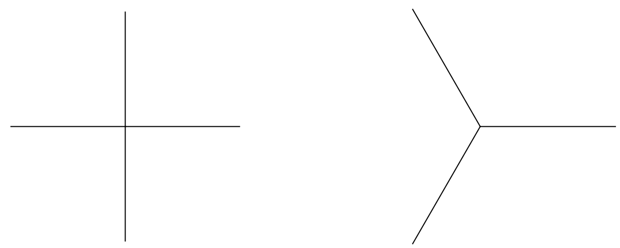
$$\mathbb{P}(D = d) \sim const \cdot d^{-\tau},$$

measurements suggest $\tau \in (2, 3)$, we take

$$\mathbb{P}(D \geq d) = d^{-\tau+1}, \quad d = 1, 2, \ldots.$$

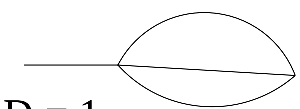
Large graphs: let $N \to \infty$

N = 2



$D_1 = 4$ $D_2 = 3$

One realization:

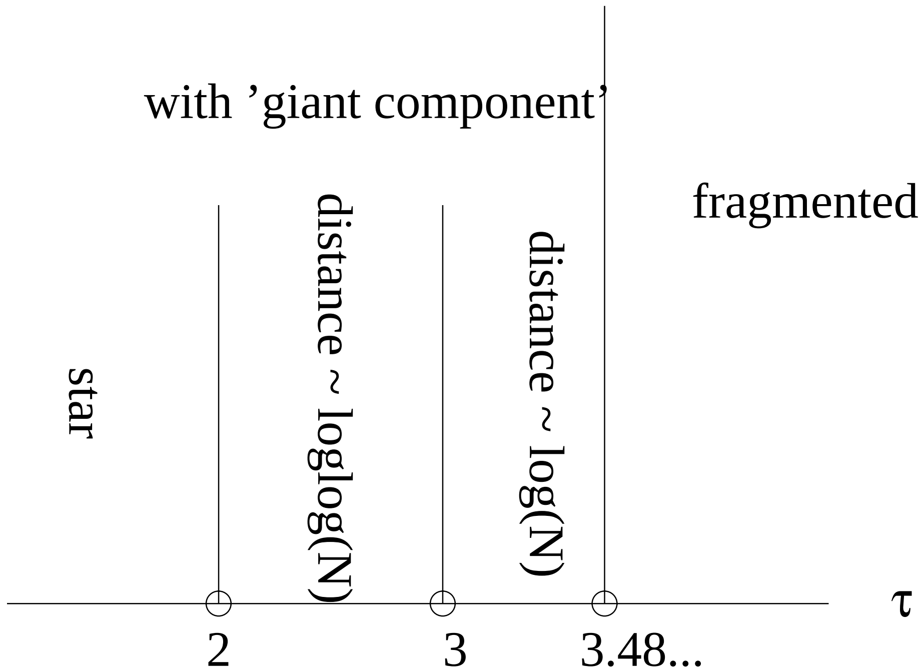


$D_0 = 1$

Distance in a power-law graph with $2 < \tau < 3$

τ-dependent graphs 'phases':

Distance $\mathcal{L}$?



$P(l) \sim m^l \Rightarrow, \mathcal{L} \sim \log(N), 2 < \tau < 3, \Rightarrow m = \infty$

Newman et al. "exclude large degrees" then m is finite...

Our claim: 'large' nodes are important.

Basic notions

$A = A^{(N)}$ happens *asymptotically almost surely (a.a.s.)*, if

$$\lim_{N \to \infty} \mathbb{P}\left(A^{(N)}\right) = 1.$$

Lemma 4.1 For any $\eta > 0$,

$$\sum_{i=1}^{N} D_i \in [(\mathbb{E}\{D\} - \eta)N, (\mathbb{E}\{D\} + \eta)N] \quad \text{a.a.s.}$$

Lemma 4.2 Let $\phi, \psi : \mathbb{N} \to \mathbb{R}$ be functions such that $\phi \to \infty$, $\psi \to \infty$, and that there exists a limit

$$\lim_{N \to \infty} \frac{\psi(N)}{\phi(N)} = a \in [0, \infty].$$

Then

$$\lim_{N \to \infty} \left(1 - \frac{1}{\phi(N)}\right)^{\psi(N)} = e^{-a} \in [0, 1].$$

Helps to reveal '0 or 1-laws'.

Lemma 4.3 Let U and V be two disjoint sets of nodes

$$F = \sum_{i \in U} D_i, \quad G = \sum_{j \in V} D_j.$$

If $FG/N \to \infty$ a.a.s., then a node from U is directly connected to a node in V a.a.s.

Fix $\ell : \mathbb{N} \to \mathbb{R}$ s.t.

$$\frac{\ell(N)}{\log \log N} = o(N), \quad \text{as } N \to \infty.$$

A "small number" is

$$\epsilon(N) = \frac{\ell(N)}{\log N}.$$

Note:

$$\epsilon(N) \to 0, \quad N^{\epsilon(N)} \to \infty \quad \text{as } N \to \infty. \tag{4.1}$$

Conjecture. *The random graph considered above has a giant component (with size proportional to N). Two randomly chosen nodes of the giant component are a.a.s. connected with at most $4k^*$ steps.*

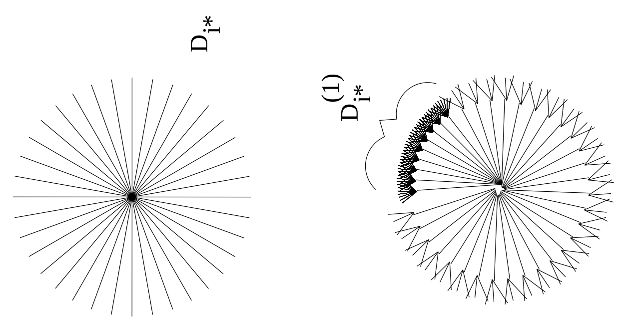
The sketch of proof

1. step

Find the first neighbours of the largest node.

$$i^* \doteq \mathrm{argmax}_{i\in\{1,\ldots,N\}} D_i.$$

The maximum degree D_{i^*} is about $N^{\frac{1}{\tau-1}}$. Note that $1/(\tau-1) > 1/2$.



Proposition 4.4 We have

$$D_{i^*} \in \left[N^{\alpha_1}, N^{\alpha_1+2\epsilon(N)}\right] \quad \text{a.a.s.},$$

where

$$\alpha_1 = \alpha_1(N) = \frac{1}{\tau - 1} - \epsilon(N).$$

Nodes

$$U_1 = \left\{ i : \; D_i \geq N^{\beta_1+\epsilon(N)} \right\}, \quad j = 1, 2, \ldots,$$

are a.a.s. among first neighbours and $\beta_1 = 1 - \alpha_1$

$$D_{i^*}^{(1)} = \sum_{i \in U_1} D_i.$$

2.step $D_{i^*}^{(1)}$ replaces D_{i^*}. We have analogously $\beta_2 = (\tau - 2)\beta_1 < \beta_1$. And so on:

k.step

$$\beta_1 \; = \; 1 - \alpha_1 = \frac{\tau - 2}{\tau - 1} + \epsilon(N),$$

$$\beta_2 = (\tau-2)\beta_1 + \epsilon(N),$$
$$\dots$$
$$\beta_k = (\tau-2)\beta_{k-1} + \epsilon(N) = \frac{(\tau-2)^k}{\tau-1} + \epsilon(N)\sum_{i=0}^{k-1}(\tau-2)^i.$$

and

$$U_j = \left\{ i:\ D_i \geq N^{\beta_j+\epsilon(N)} \right\}, \quad j = 1,2,\dots,$$

are within k-hops from i^*. $\beta_k \to \epsilon(N)/(3-\tau) > 0$.

Define the 'core' as:

$$C = \left\{ i:\ D_i \geq N^{2\epsilon(N)/(3-\tau)} \right\}. \tag{4.2}$$

Proposition 4.5 Let k^* denote the number

$$k^* = 1 + \left\lceil \frac{\log\log N + \log(3-\tau)}{-\log(\tau-2)} \right\rceil .$$

Then $C \subseteq U_{k^*}$. In particular, the nodes of C have at least $\frac{1}{2}N^{1-2\epsilon(N)/(3-\tau)}$ link stubs a.a.s.

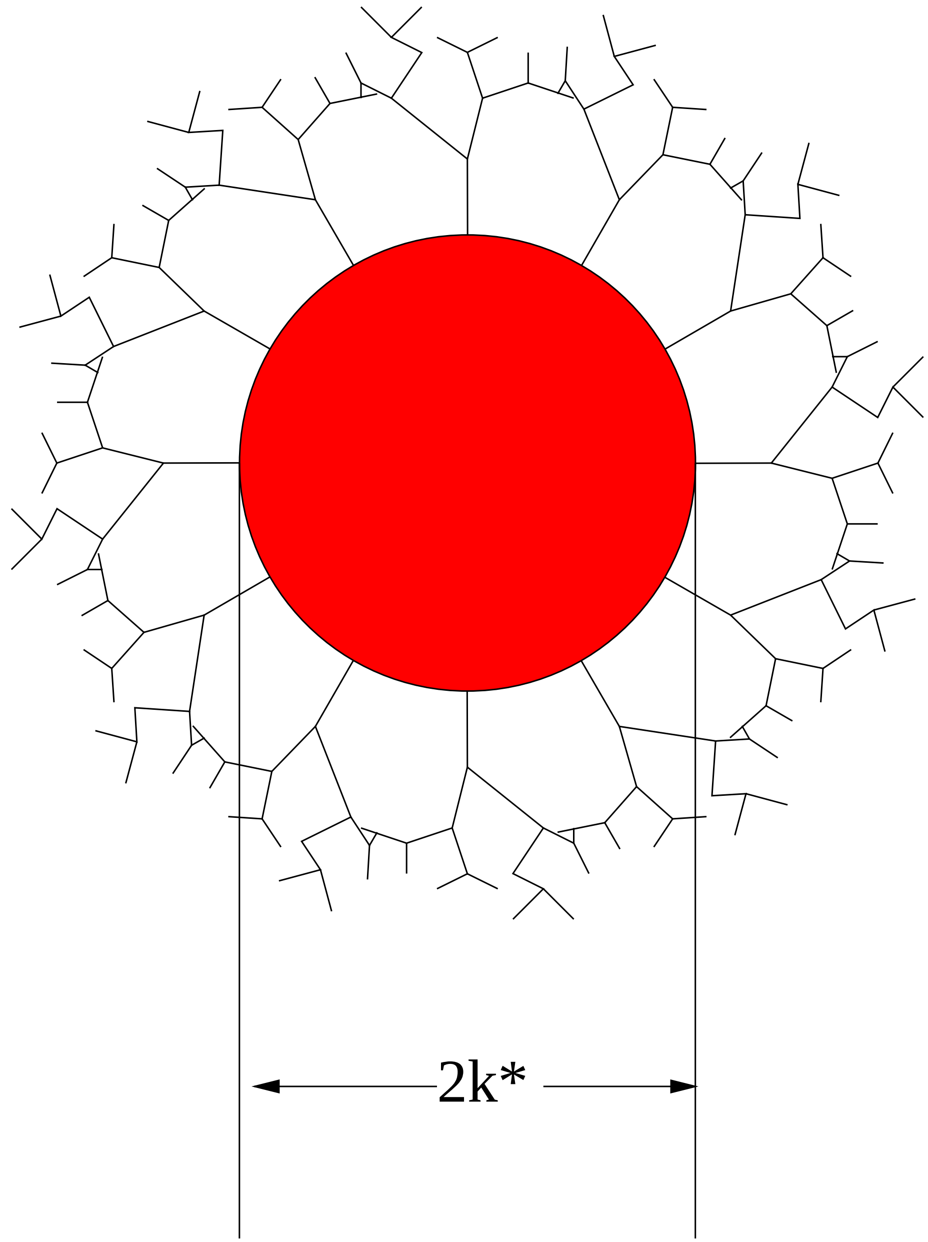
final steps

Take two random nodes, connect them to the core with less than k^* steps. Possible if: a random node i has its neibourghood $V_i(j(N))$ within $j(N)$ hops $j(N) \to \infty$, $j(N) \leq k^*(N)$ and environment grows at least exponentially with number of hops.

$\mid V_i(j(N)) \mid \geq \mu^{j(N)}$ a.a.s. for some $\mu > 1$

$$\frac{\mid V_i(k^*(N)) \mid N^{1-2\epsilon(N)/(3-\tau)}}{N} \geq \mu^{k^*(N)} e^{2l(N)/(3-\tau)} \to \infty$$

To be proved: the whole path exists a.a.s. and number of cycles is insignificant a.a.s.


2k*

Conclusions

- shown that a “core network” arises spontaneously from the distributional assumptions alone

- one-step connections between layers of the core, and many auxiliary results needed for that, are shown (we hope!) rigorously

- the remaining steps to prove the Conjecture in full are indicated

- properties of this kind of graphs have already been used in studies on denial of service attacs [] and multicast []; we hope that our results will turn out useful in that kind of studies