Self-similarity of complex networks & hidden metric spaces

M. ÁNGELES SERRANO
Departament de Química Física
Universitat de Barcelona

TERA-NET:
Toward Evolutive Routing Algorithms for scale-free/internet-like NETworks

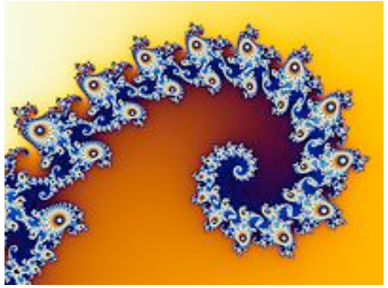
Bordeux, France, July 5 2010

Self-similarity

Scale invariance

Exact form of self-similarity

$$f(\lambda v) = \lambda^r f(v)$$

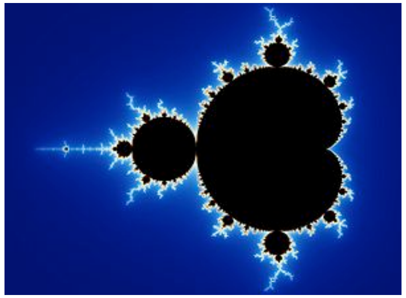


Self-similarity

The same properties at different length scales (exact or approximate or statistical)

Fractality

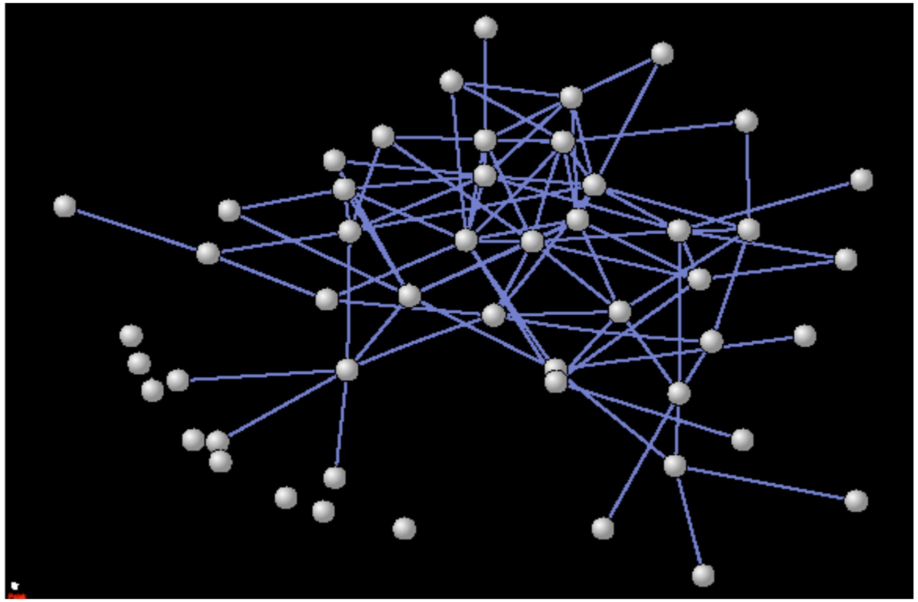
Self-similarity is a property of fractals, objects with $N = s^d$ Hausdorff dimension greater than its topological dimension (usually non-integral) – the measured length depends on the measuring scale



Self-similarity

Scale invariance Self-similarity Fractality

Self-similarity of complex networks?



Scale-free degree distributions

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

Power laws are scale-invariant

$$P(\lambda k) = \lambda^{-\gamma} P(k)$$

Box covering

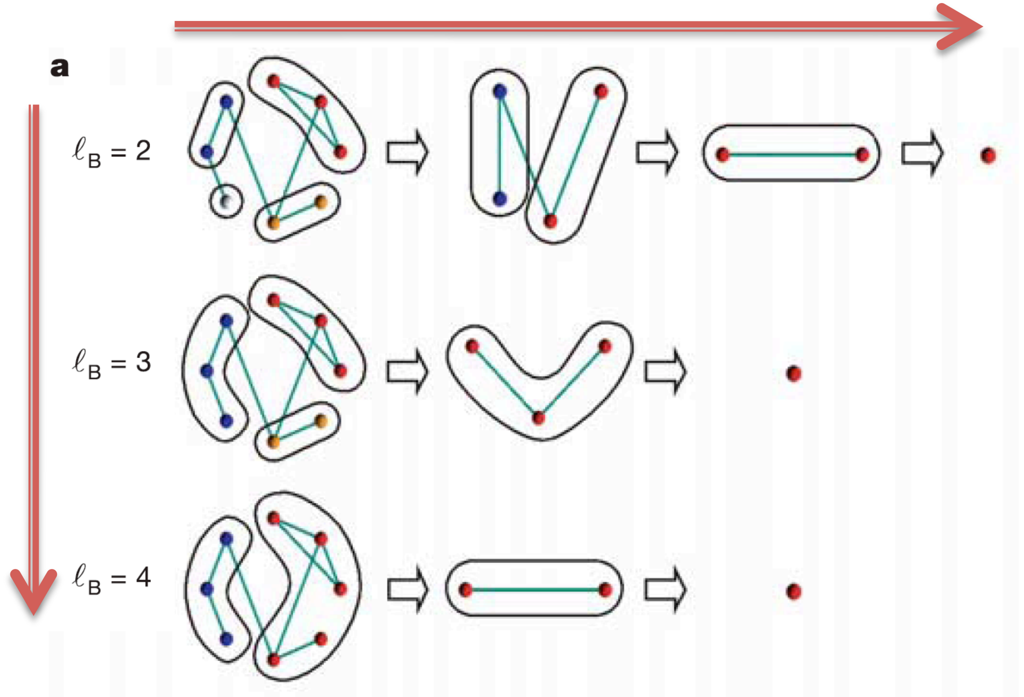
Topological self-similarity

Box-covering renormalization in complex networks

The network is tiled with boxes such as all nodes in a box are connected by a minimum distance smaller than a given ℓ_B

Each box is replaced by a single node and two renormalized nodes are connected if there is at least one link between the boxes

In fractal geometry, box-counting is the primary way to evaluate the fractal dimension of a fractal object



C. Song, S. Havlin, H. A. Makse, Nature 433, 392-395 (2005); Nature Physics 2, 275-281 (2006)

Box covering

Topological self-similarity

Box-covering renormalization in complex networks

Some systems (WWW, biological networks) are fractal and have degree distributions that remain invariant and have finite “fractal” dimension

$$N_B(\ell_B)/N \sim \ell_B^{-d_B} \quad \text{and} \quad k_B(\ell_B)/k_{hub} \sim \ell_B^{-d_k}$$

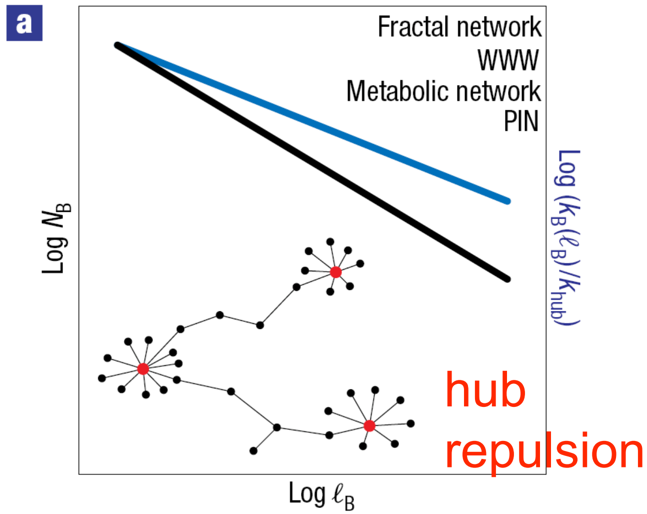
Some systems (the Internet) are non-fractal, with scaling laws that are replaced by exponentials

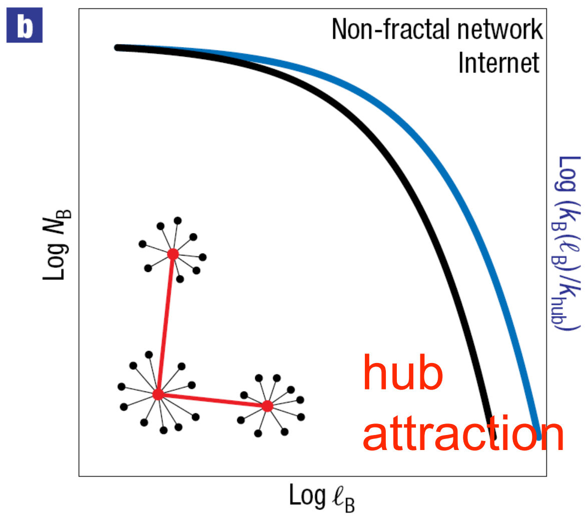
$$d_B \to \infty \text{ and } d_k \to \infty$$

$$N_B(l_B)/N \sim \exp(-(\ln n/2)\ell_B)$$

$$k_B(l_B)/k_{hub} \sim \exp(-(\ln s/2)\ell_B)$$

C. Song, S. Havlin, H. A. Makse,
Nature 433, 392-395 (2005);
Nature Physics 2, 275-281 (2006)





Box covering

Topological self-similarity
Box-covering renormalization in complex networks

Specific problems of the box-covering methodology

- For a given linear size of the boxes, the box covering partition is not univocal, results may depend on the specific partition of nodes into boxes
prescriptions: partition with the minimum number of boxes, NP-hard problem; sequentially centering the boxes around the nodes with the largest mass…

- Large fractal dimension may not be distinguishable from exponential behavior

- Self-similarity of the degree distribution under renormalization, but in general correlations do not scale

- Small range in which the scaling is valid (small world property)

Other proposals

- J S Kim, K-I Goh, B Kahng and D Kim, **Fractality and self-similarity in scale-free networks,** New J. Phys. **9** 177 (2007)

a slightly different box covering method
fractality and self-similarity are disparate notions in SF networks
the Internet is a non-fractal SF network, yet it exhibits self-similarity

- J I Alvarez-Hamelin, L Dall'Asta, A Barrat, and A Vespignani, **K-core decomposition of Internet graphs: hierarchies, self-similarity and measurement biases,** *Networks and Heterogeneous Media*, 3(2): 371-293 (2008)

the Internet shows a statistical self-similarity of the topological properties of k-cores

… and references therein ...

Self-similarity of complex networks

Self-similarity of complex networks is not well defined yet in a proper geometrical sense

Lack of a metric structure except lengths of shortest paths

small world property $N \approx e^{\bar{\ell}/\ell_0}$

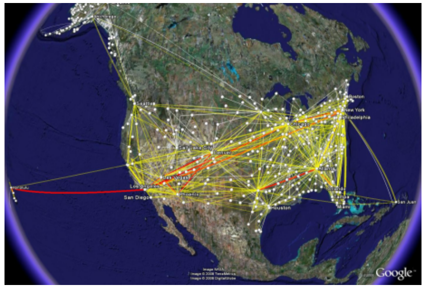
geometric length scale transformations?

self-similarity as the scale-invariance of the degree distribution (with the exception of the k-cores discussion)

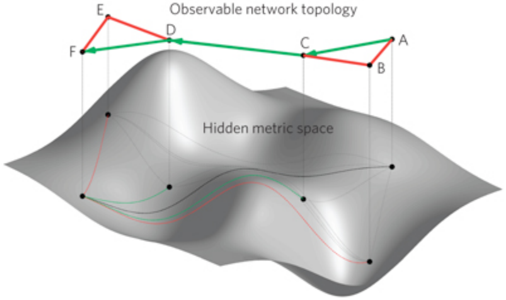
Underlying metric spaces

We propose networks embedded in metric spaces

- Geography as an obvious geometrical embedding: airport networks, urban networks…



- Hidden metric spaces: WWW (similarity between pages induced by content), social networks (closeness in social space)…



…how to identify hidden metric spaces and their meaning…

Underlying metric spaces

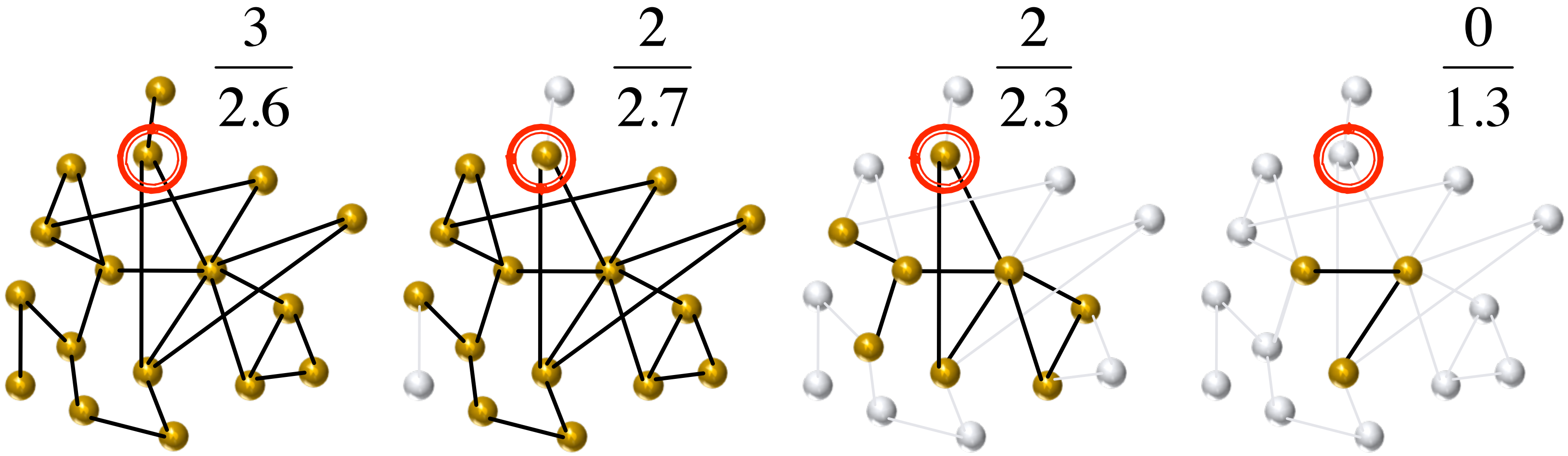
We go beyond and conjecture that hidden geometries underlying some real networks are a plausible explanation for their observed self-similar topologies

- Some real scale-free networks are self-similar (degree distribution, degree-degree correlations, and clustering) with respect to a simple degree-thresholding renormalization procedure (purely topological)
- A class of hidden variable models with underlying metric spaces are able to accurately reproduce the observed topology and self-similarity properties

Self-similarity of real complex networks

degree-thresholding renormalization procedure $k > k_T$

$$k_i \,/ < k_i(k_T) >$$



$k_T = 0$ $k_T = 1$ $k_T = 2$ $k_T = 3$

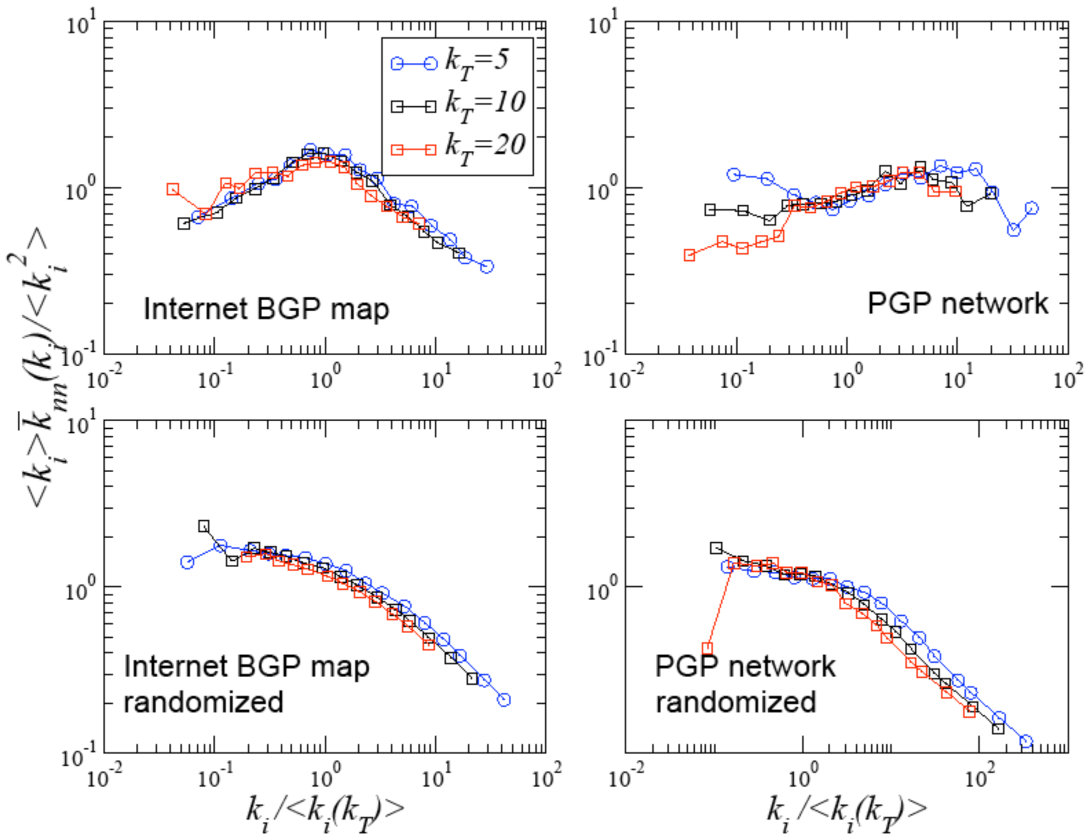
G **G(k_T)**

Self-similarity of real complex networks

Average nearest neighbors degree

BGP map of the Internet
at the AS level
SF with exponent 2.2
N=17446
<k>=4.68

PGP social web of trust
SF with exponent 2.5
N=57243
<k>=2.16



(also U.S. airports network, but not so challenging since it is embedded in geo)

A random model like the CM will produce self-similar networks regarding the degree distribution and degree-degree correlations, if the degree distribution of the complete graph is SF….

Self-similarity of real complex networks

Degree-dependent clustering

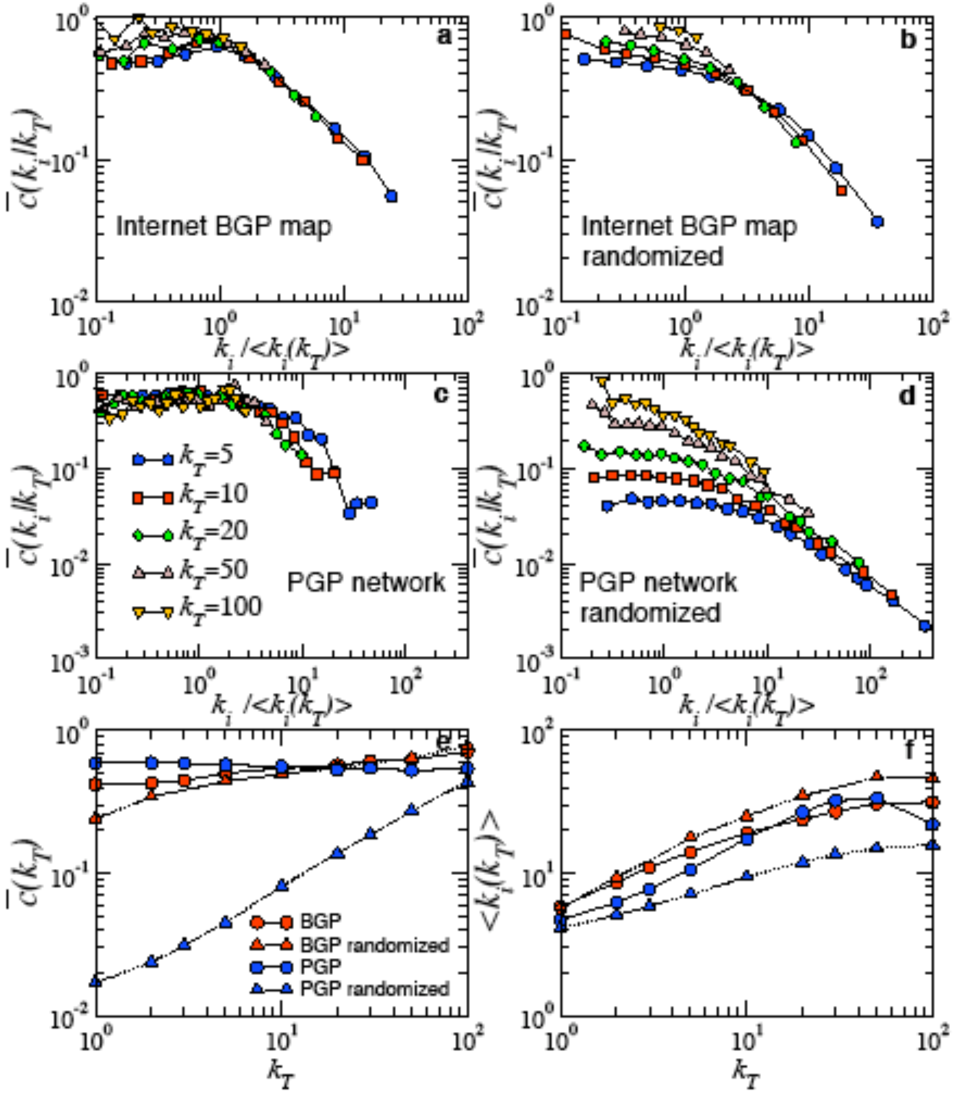
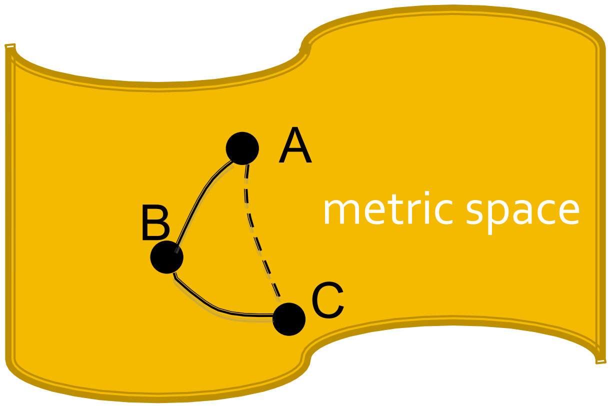


FIG. 2 (color online). (a)–(d) Degree-dependent clustering coefficient as a function of the rescaled internal degree for the Internet BGP map, the PGP web of trust, and their randomized versions. (e) Average clustering coefficient as a function of the threshold degree k_T for renormalized real networks and their randomized counterparts. (f) Internal average degree as a function of k_T for the same networks.

Models rerpoducing self-similarity

Why is clustering important?

CLUSTERING gives the clue for the connection between the observed topologies and the hidden geometries



TRIANGLE INEQUALITY for any triangle, the sum of the lengths of any two sides must be greater than the length of the remaining side

$$AC \leq AB + BC$$

the key point is to reproduce self-similar clustering

Models rerpoducing self-similarity

In order to explain the observed self-similar topologies, we propose a class of hidden variable models with underlying metric spaces, that in particular reproduce the self-similarity of clustering

Hidden variable model

- Set of nodes with hidden property $f(h)$
- Connection probability $p(h,h')$

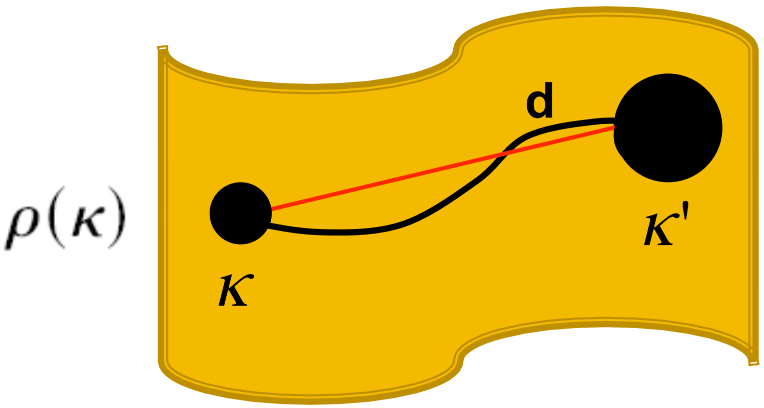
M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **68**, 036112 (2003).
G. Caldarelli, A. Capocci, P. D. L. Rios, and M. A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002).
B. Söderberg, Phys. Rev. E **66**, 066121 (2002).

Advantages:

Powerful and general
Analytic computations
No frustration

Models rerpoducing self-similarity

All nodes exist in an underlying metric space, so that distances can be defined between pairs



The connection probability is an integrable function of the form

$$r\left(\frac{d}{d_c}\right)$$

Nodes that are close to each other are more likely to be connected

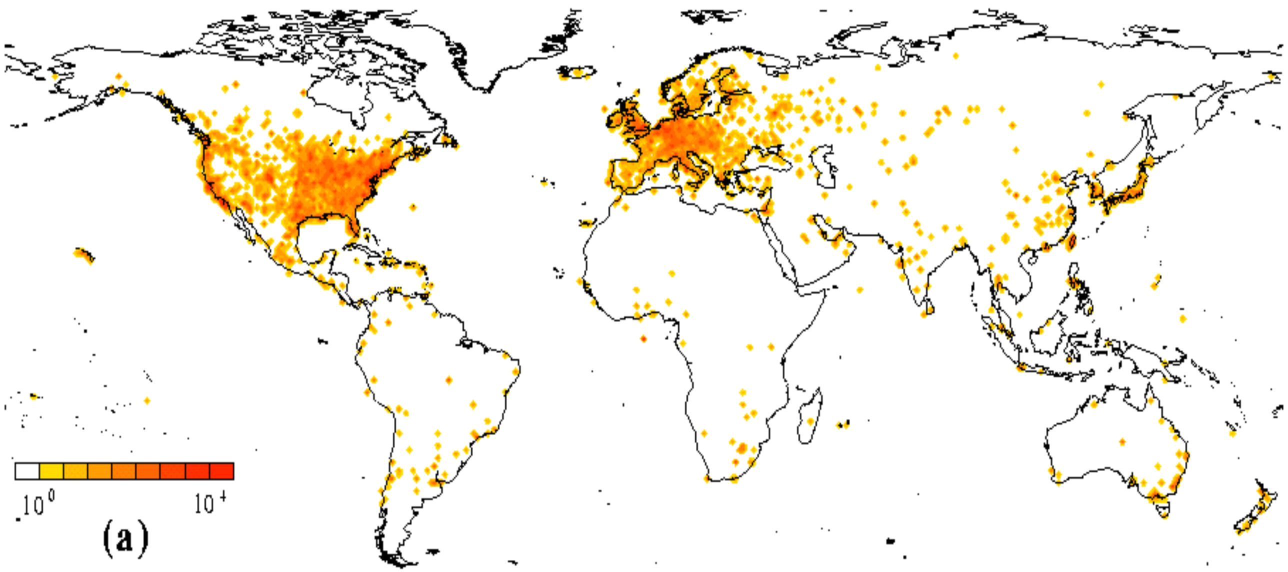
The characteristic distance depends on the expected degree

$$d_c(\kappa, \kappa') \propto (\kappa\kappa')^{1/D}$$

("gravity law")

(Precedent: user-driven growing model for the Internet)

Router density map, NETGEO tool CAIDA



OBSERVATION: The spatial distribution of users is that of routers and ASs, which is observed to be fractal with dimension *Df=1.5*

S.H. Yook, H. Jeong, and A.-L. Barabási, PNAS 99, 13382 (2002)

Box counting: $N(\ell) \equiv$ No. of boxes of size ℓ that contain routers

$\mathbf{N(\ell) \sim \ell^{-Df}}$

MODEL: Connections between ASs

Short distance links (connectivity costs)
vs
long distance links (structural necessity)

ASs' sizes may affect link creation

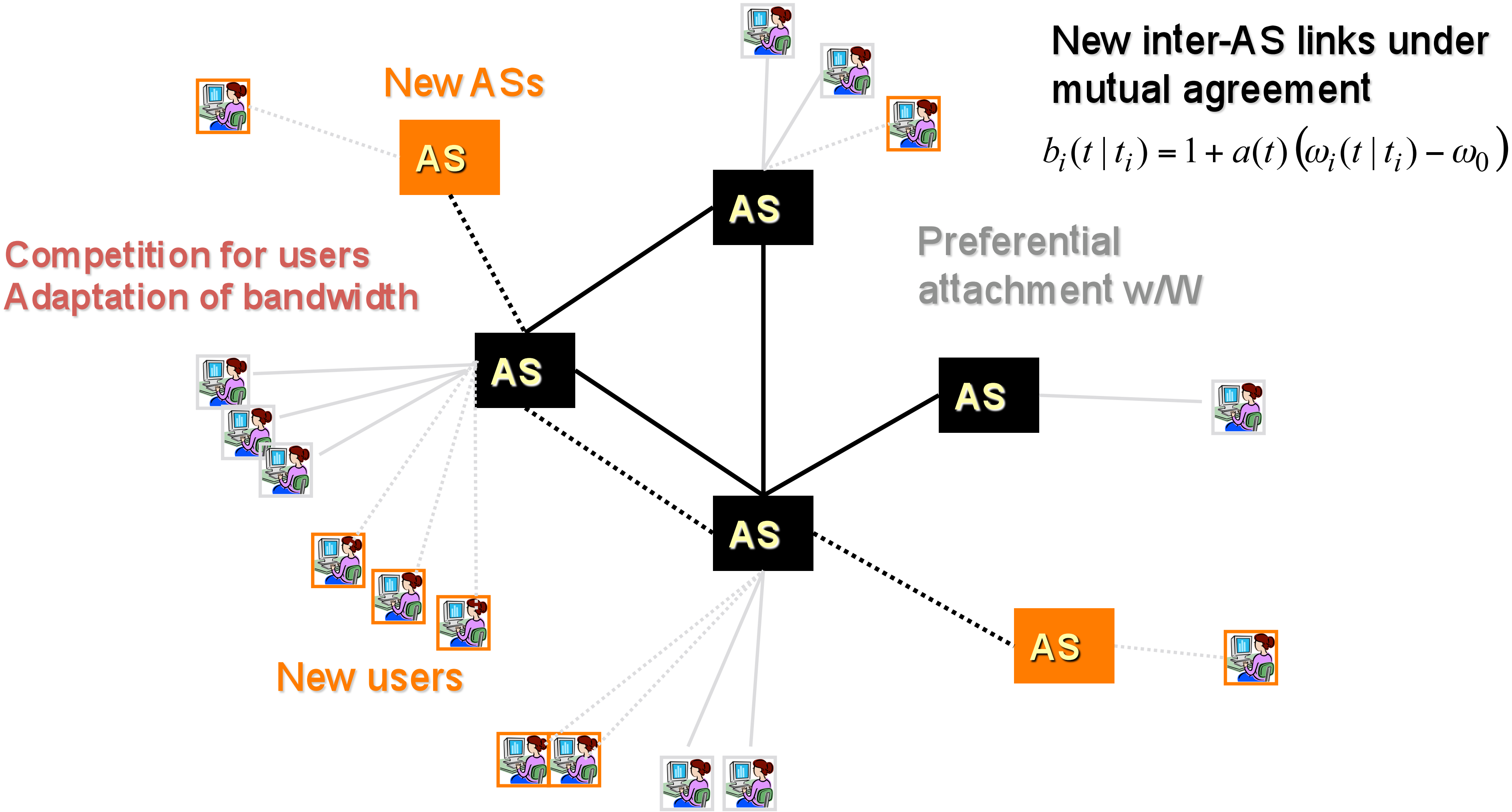
$$D(d_{ij}, w_i, w_j) = e^{-d_{ij}/d_c(w_i, w_j)}$$

$$d_c(w_i, w_j) = w_i w_j / \kappa\, W$$

κ - cost function

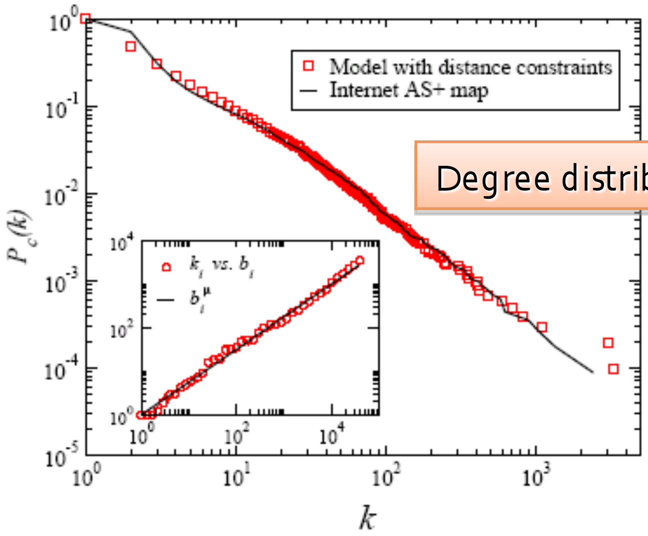
M.A. Serrano, M. Boguñá, and A. Díaz-Guilera, Phys. Rev. Lett. 94, 038701 (2005)

(Precedent: user-driven growing model for the Internet)

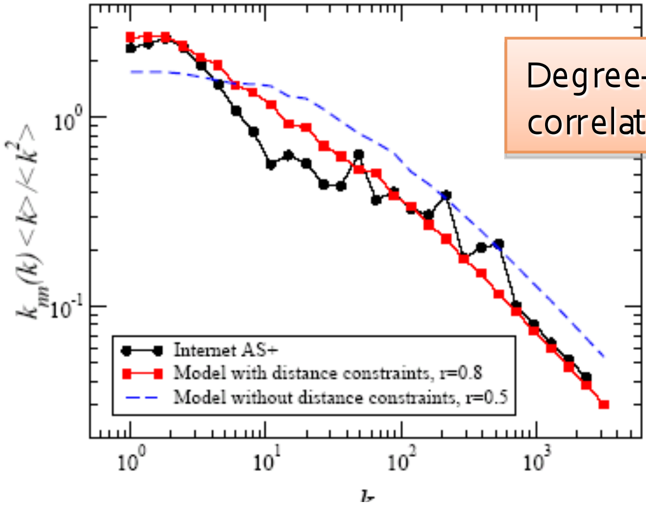


M.A. Serrano, M. Boguñá, and A. Díaz-Guilera, Phys. Rev. Lett. 94, 038701 (2005)

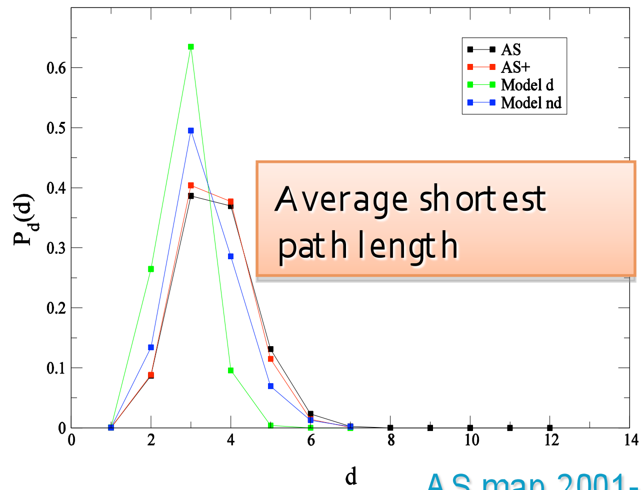
(Precedent: user-driven growing model for the Internet)



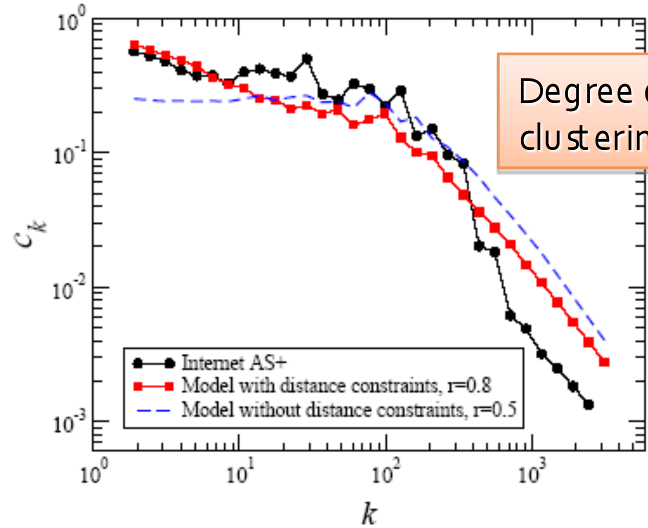
Degree distribution



Degree-degree correlations



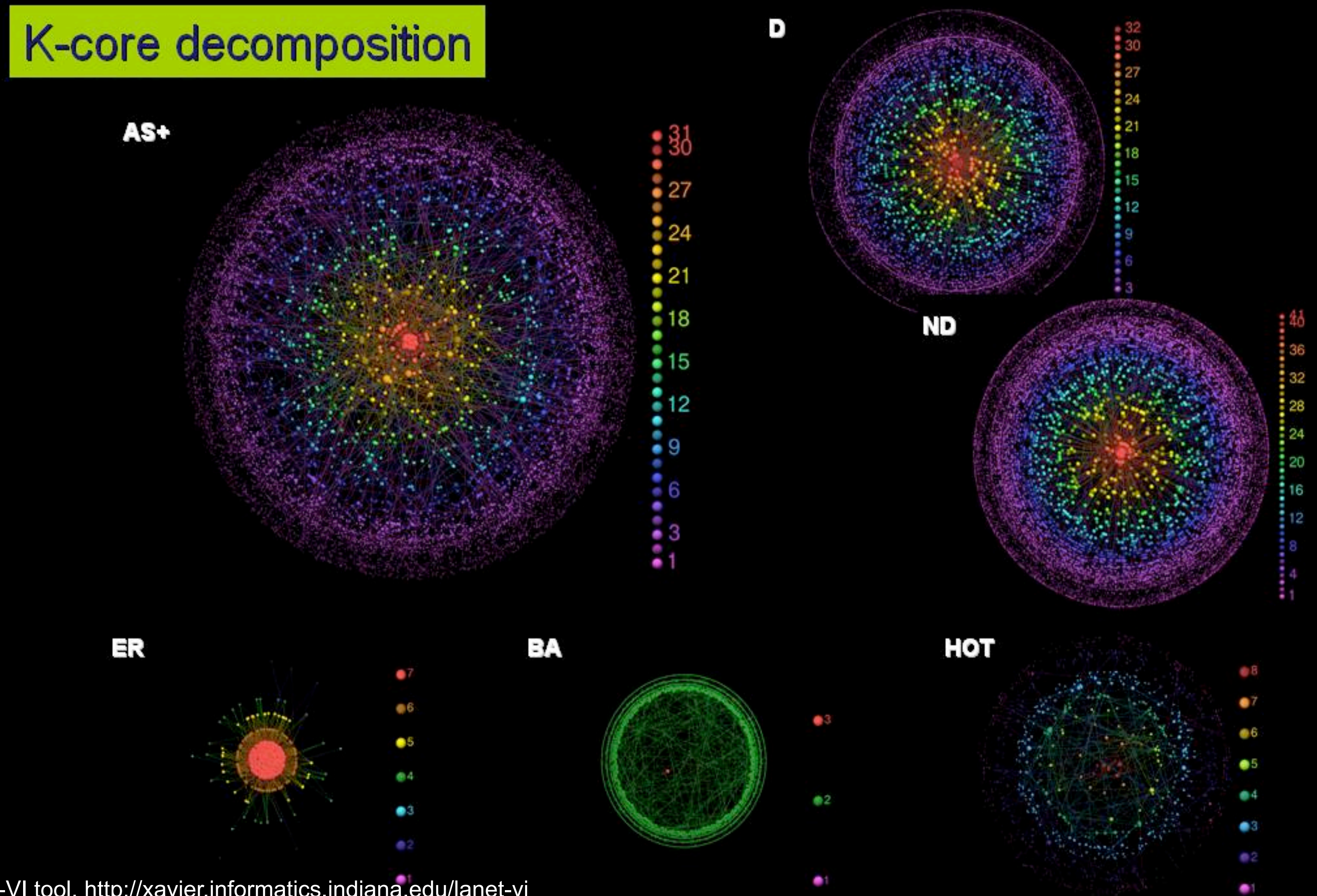
Average shortest path length



Degree dependent clustering

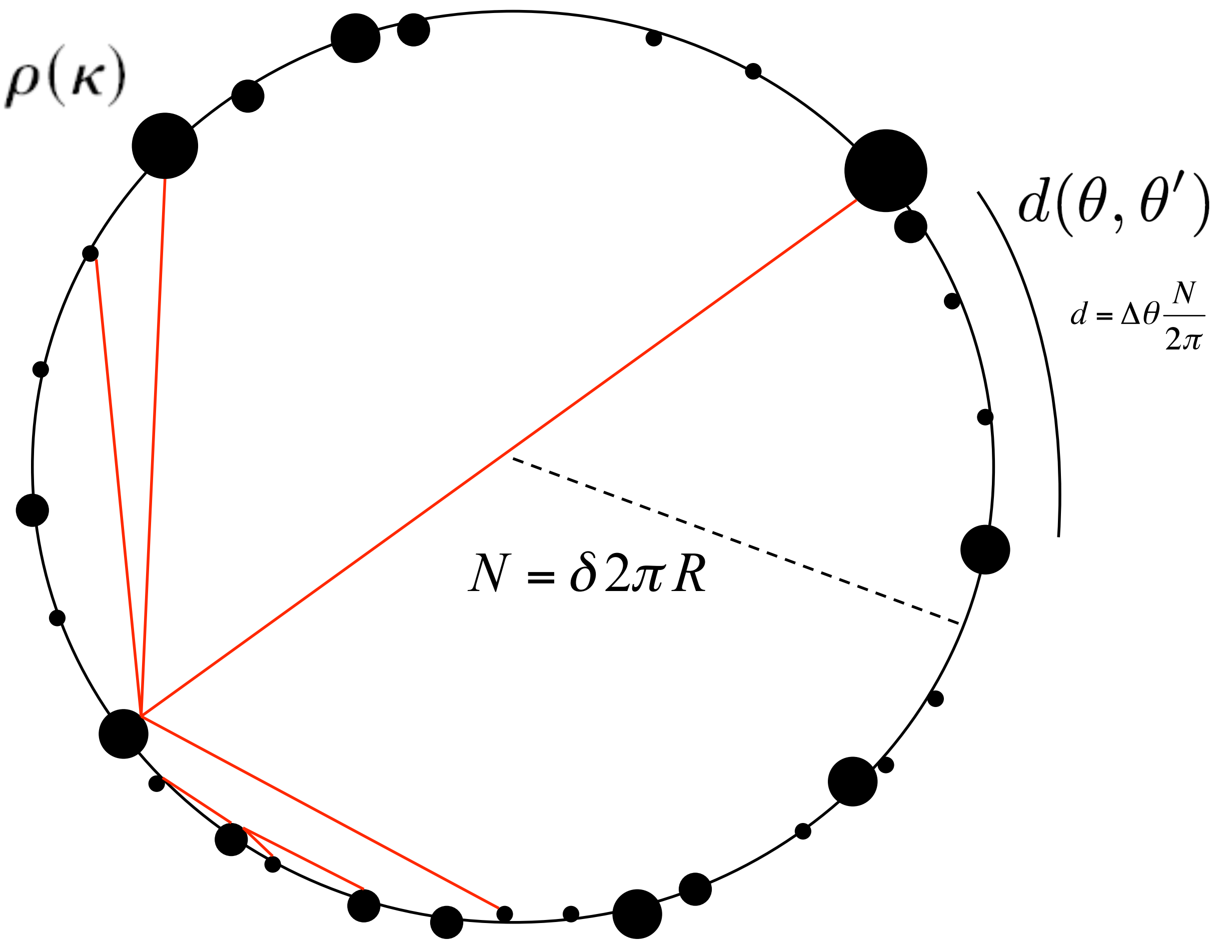
AS map 2001-ORVP
AS+ map 2001-ORVP+other sources

(Precedent: user-driven growing model for the Internet)



LANET-VI tool, http://xavier.informatics.indiana.edu/lanet-vi

The S1 model



Connection probabilities based on:

distances

product of degrees

$$\rho(\kappa) = (\gamma - 1)\kappa_0^{\gamma-1}\kappa^{-\gamma},$$

$$\kappa > \kappa_0 \equiv (\gamma - 2)\langle k\rangle/(\gamma - 1), \quad \gamma > 2,$$

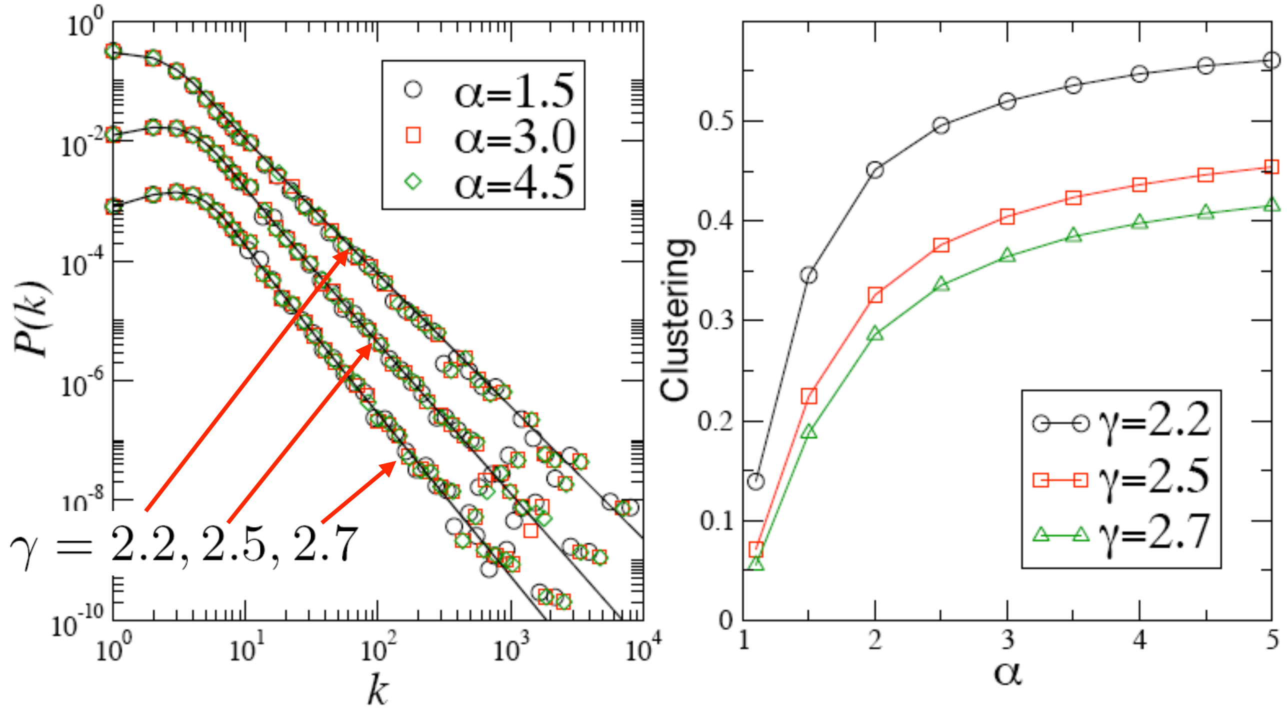
$$r(\theta, \kappa; \theta', \kappa') = \left(1 + \frac{d(\theta, \theta')}{\mu\kappa\kappa'}\right)^{-\alpha},$$

$$\mu = \left(\frac{\gamma - 2}{\gamma - 1}\right)^2 \frac{(\alpha - 1) < k >}{2\delta\kappa_0^2} \qquad \alpha > 1,$$

Underlying metric space

Heterogeneity

The S1 model – model behavior



$$\bar{k}(\kappa) = \kappa \qquad \rho(\kappa) \approx P(k)$$

$$P(k) = (\gamma - 1)\kappa_0^{\gamma-1}\frac{\Gamma(k+1-\gamma,\kappa_0)}{k!}, \quad P(k) \sim k^{-\gamma}$$

The S1 model - specifications

- $\rho(\kappa)$ controls the degree distribution, SF $\rho(\kappa) \approx P(k)$
- independently, α controls the level of clustering , strong clustering
- given α , the parameter $\mu = \left(\frac{\gamma-2}{\gamma-1}\right)^2 \frac{(\alpha-1) < k >}{2\delta\kappa_0^2}$ controls the average degree
- if $1 < \alpha < 2 \; or \; 2 < \gamma < 3$, small-world!!! but underlying metric space!

$$\boxed{\bar{d}(\kappa') = \int x p(x|\kappa') dx \xrightarrow{N,R\to\infty} \infty}$$

The S1 model – comparing with real data

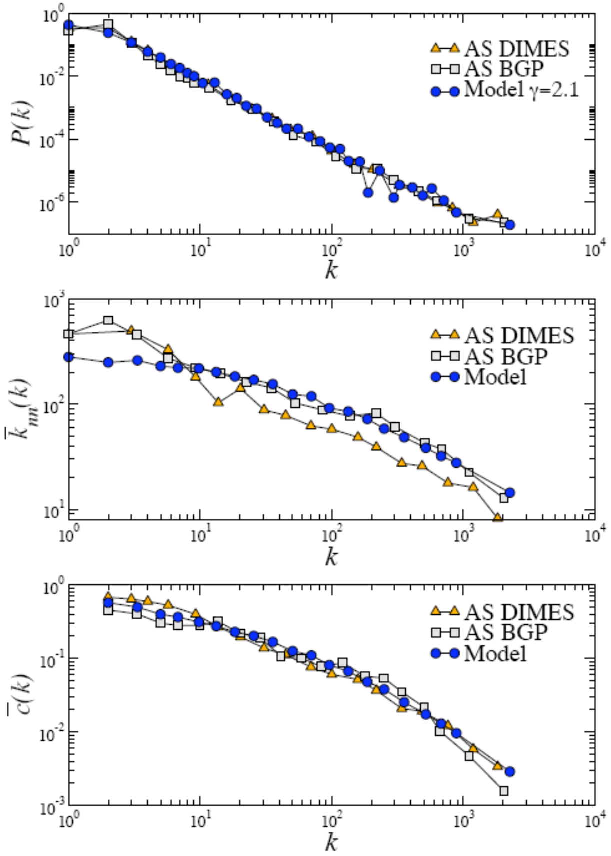


FIG. 1: Degree distribution $P(k)$, average nearest neighbours' degree $\bar{k}_{nn}(k)$, and degree-dependent clustering coefficient $\bar{c}(k)$ generated by our model with $\gamma = 2.1$ and $\alpha = 2$ compared to the same metrics for the real Internet map as seen by BGP data and the DIMES project.

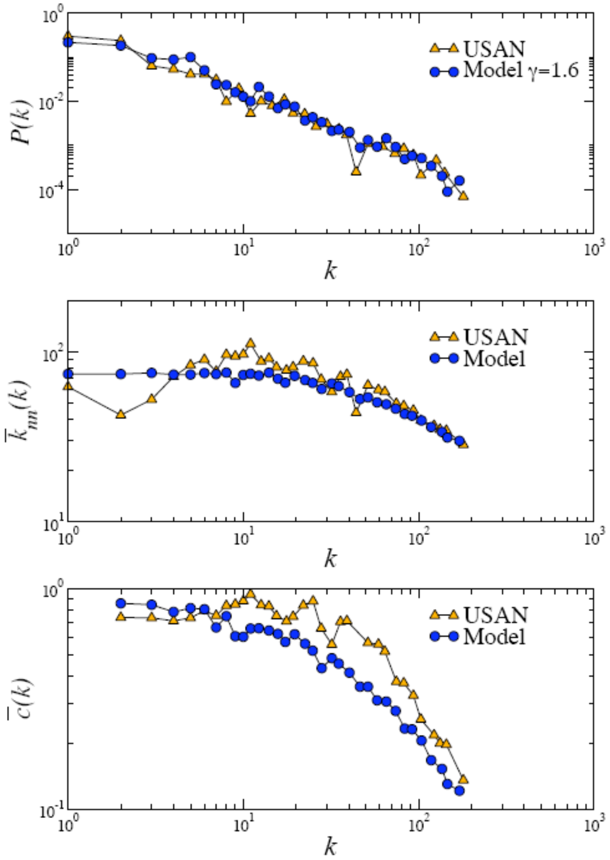
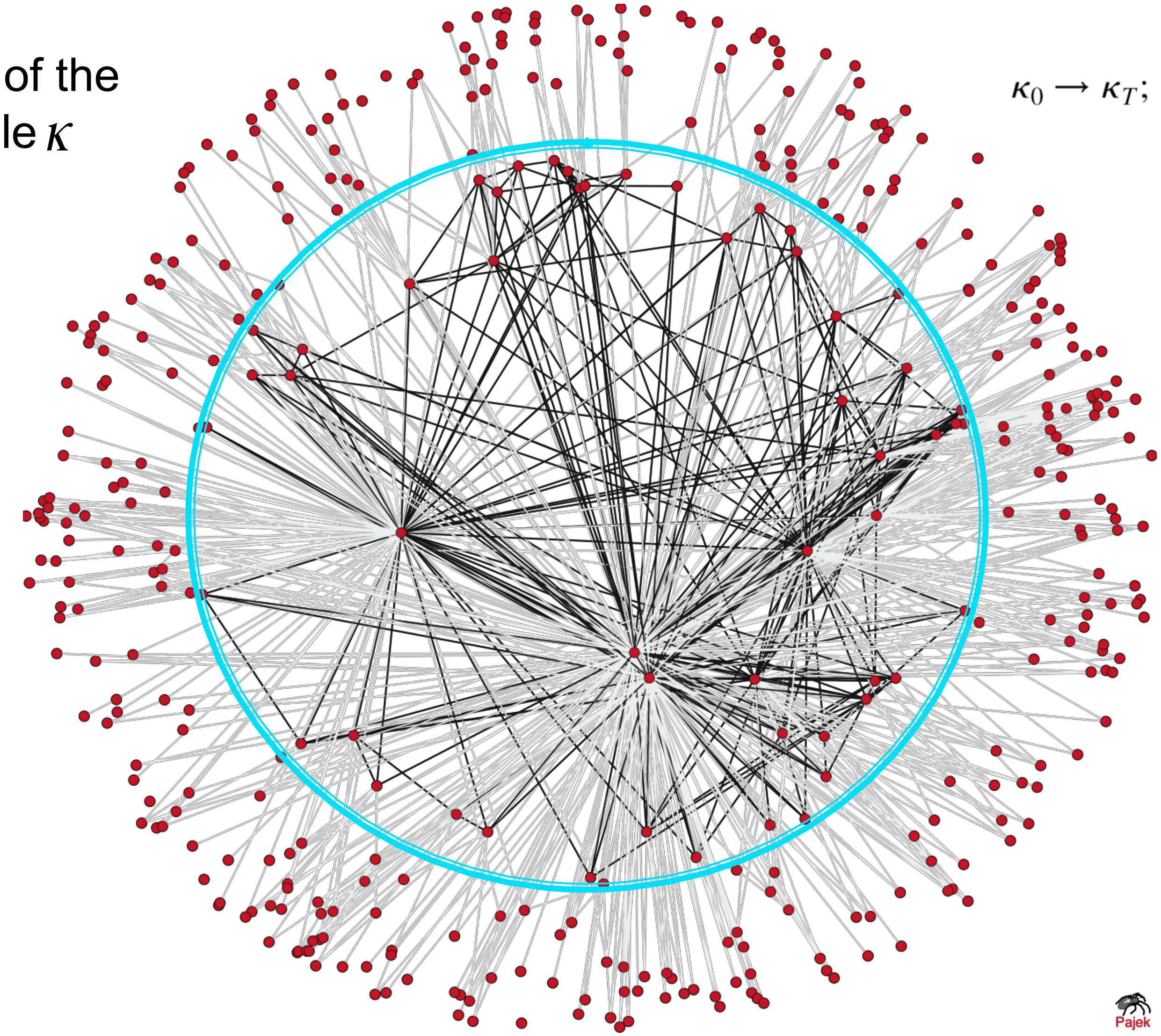


FIG. 2: Degree distribution $P(k)$, average nearest neighbours' degree $\bar{k}_{nn}(k)$, and degree-dependent clustering coefficient $\bar{c}(k)$ generated by our model with $\gamma = 1.6$, $\alpha = 5$ and a cut-off at $k_c = 180$ compared to the same metrics for the real US airport network.

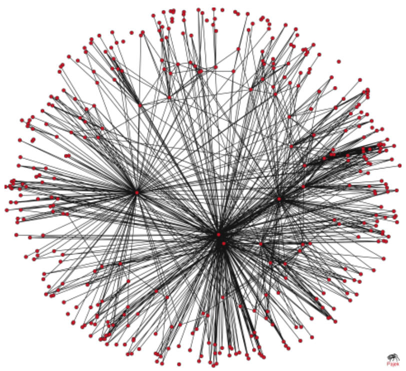
Self-similarity of the S1 model

Thresholding of the hidden variable κ

$$\kappa_0 \longrightarrow \kappa_T; \qquad \delta \longrightarrow \delta\left[\frac{\kappa_T}{\kappa_0}\right]^{1-\gamma}$$



Self-similarity of the S1 model



$$\kappa_0 \longrightarrow \kappa_T;$$

$$\rho(\kappa) = (\gamma - 1)\frac{\kappa_0^{\gamma-1}}{\kappa^{\gamma}}$$

$$N = \delta 2\pi R$$

$$r(\theta, \kappa; \theta', \kappa') = \left(1 + \frac{d(\theta, \theta')}{\mu \kappa \kappa'}\right)^{-\alpha},$$

$$\mu = \left(\frac{\gamma - 2}{\gamma - 1}\right)^2 \frac{(\alpha - 1) < k >}{2\delta\kappa_0^2}$$

$$\rho_r(\kappa) = (\gamma - 1)\frac{\kappa_T^{\gamma-1}}{\kappa^{\gamma}}$$

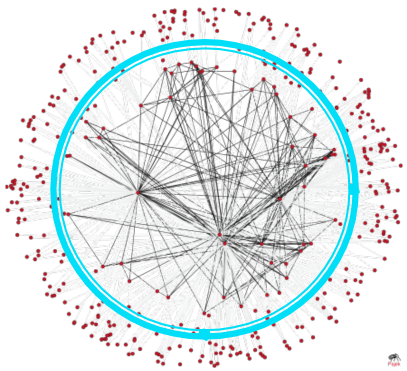
$$N_r = N\int_{\kappa_T}^{\infty} \rho(\kappa) d\kappa = N\left(\frac{\kappa_0}{\kappa_T}\right)^{\gamma-1}$$

$$\delta_r = \delta\left(\frac{\kappa_0}{\kappa_T}\right)^{\gamma-1}$$

$$r(\theta, \kappa; \theta', \kappa') = \left(1 + \frac{d(\theta, \theta')}{\mu \kappa \kappa'}\right)^{-\alpha},$$

$$\mu = \left(\frac{\gamma - 2}{\gamma - 1}\right)^2 \frac{(\alpha - 1) < k >_r}{2\delta_r\kappa_T^2}$$

$$\boxed{< k >_r = < k > \left(\frac{\kappa_T}{\kappa_0}\right)^{3-\gamma}}$$



Self-similarity of the S1 model

$G(\kappa_T)$ are replicas of G

after

$$< k >_r = < k > \left(\frac{\kappa_T}{\kappa_0} \right)^{3-\gamma}$$

In particular, clustering spectrum and clustering…

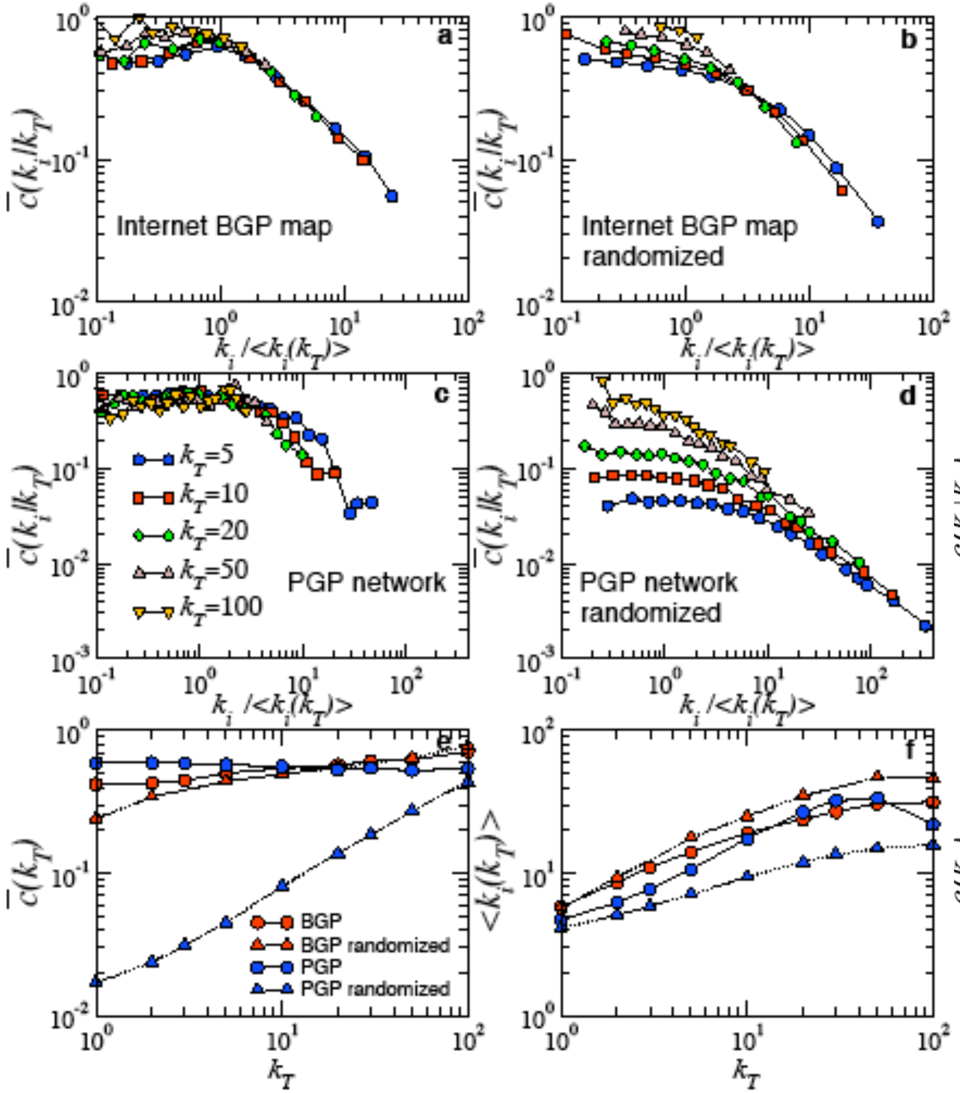
$$\bar{c}(\kappa|\kappa_T) = f(\kappa/\kappa_T), \quad \bar{c}(k_i|k_T) \approx f(k_i/k_T^{3-\gamma}) = \boxed{\tilde{f}[k_i/\langle k_i(k_T)\rangle],}$$

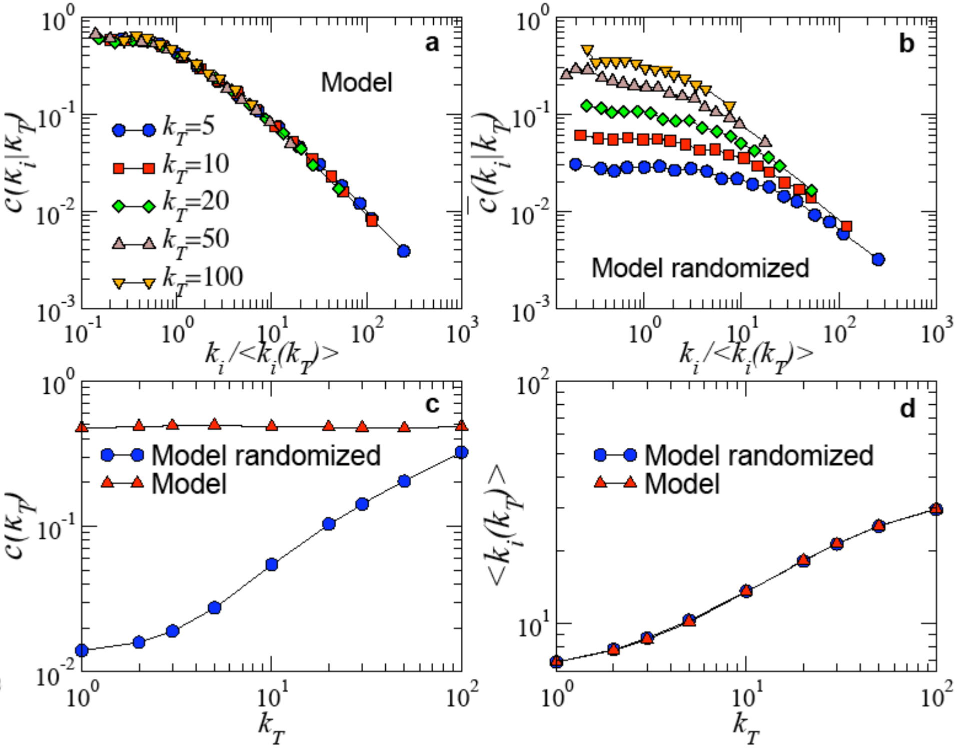
$$\bar{c}(\kappa_T) = \int_{\kappa_T} d\kappa \rho(\kappa|\kappa_T)\bar{c}(\kappa|\kappa_T),$$

Independent of κ_T

Self-similarity of the S1 model

clustering

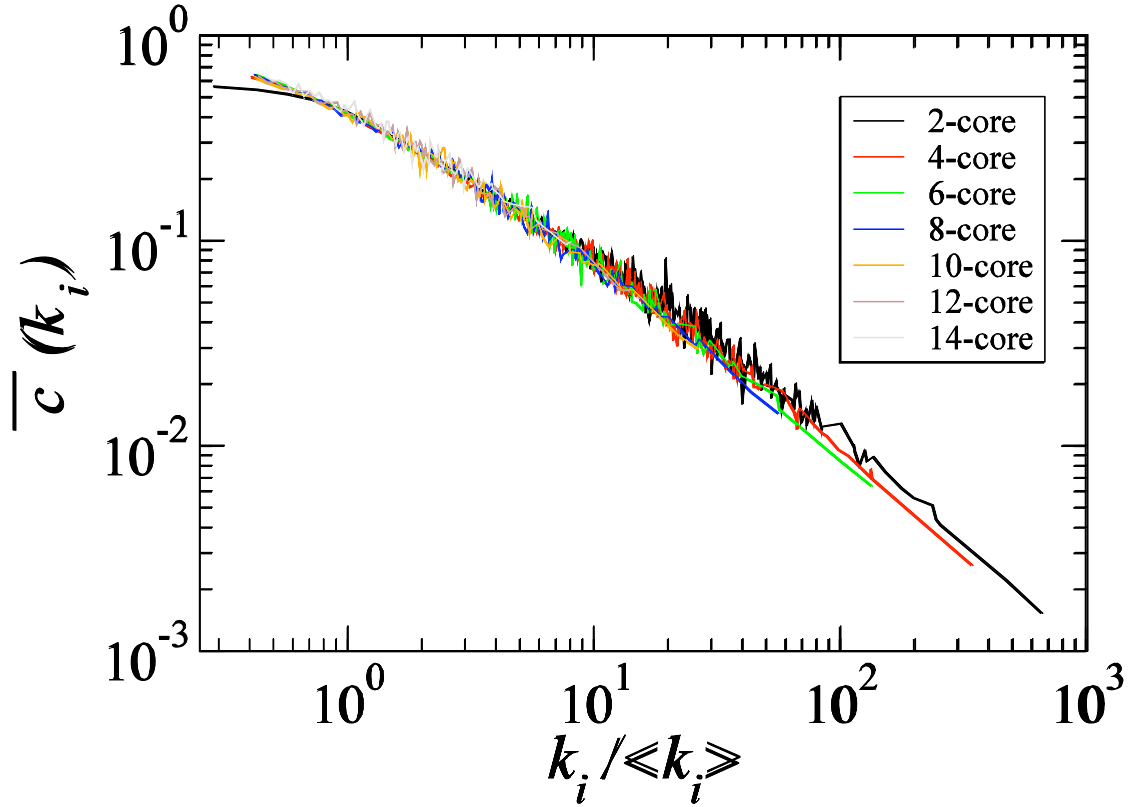




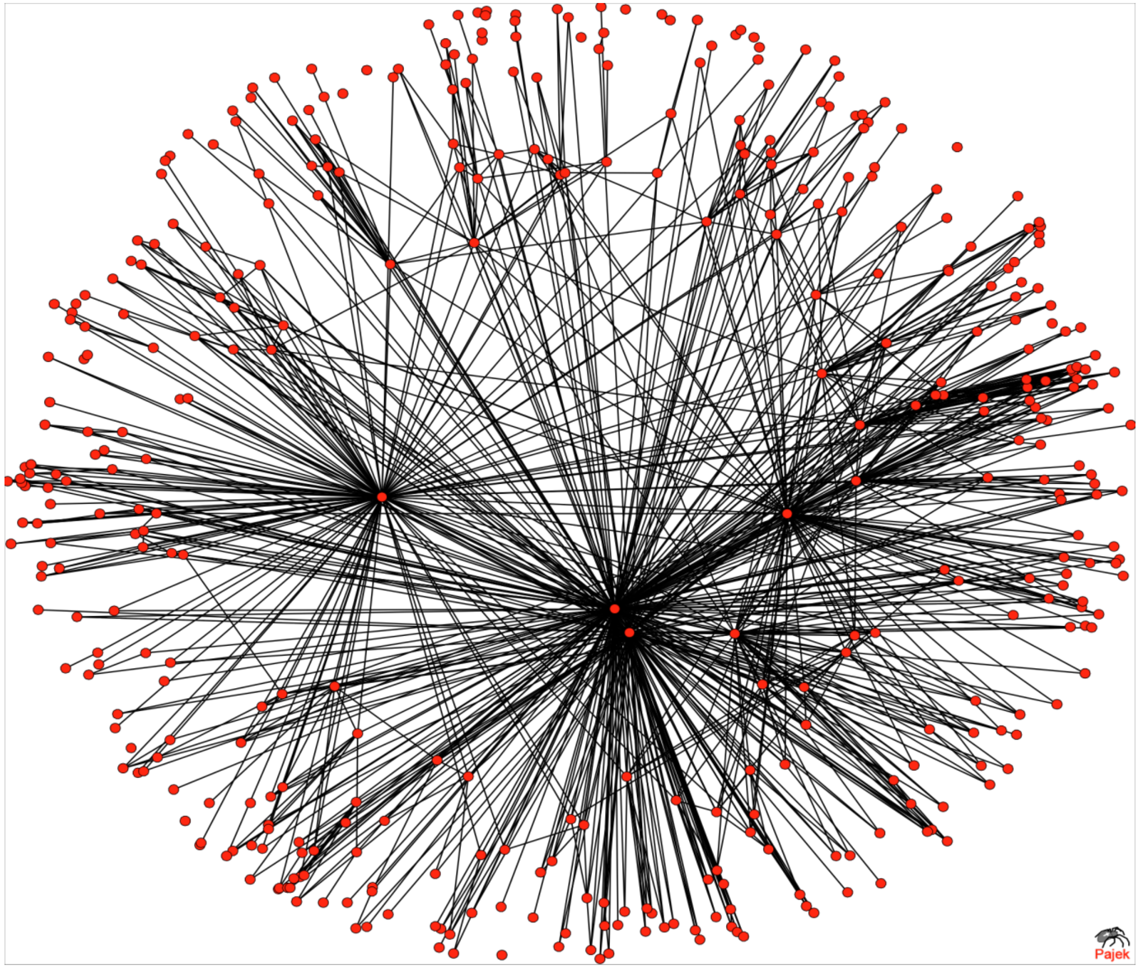
Self-similarity of the S1 model

0	100000
1	95085
2	80610
3	58580
4	36264
5	20149
6	10569
7	5839
8	3376
9	2066
10	1393
11	983
12	716
13	545
14	408
15	295
16	230
17	166
18	120
19	101
20	0

K-cores



Isomorphism with model in hyperbolic space



Curvature and temperature of complex networks,
Dmitri Krioukov, Fragkiskos Papadopoulos, Amin Vahdat, and Marián Boguñá,
Phys. Rev. E 80, 035101 (2009), H2 model

$$r\left(\frac{d}{\mu\kappa\kappa'}\right)$$

S1 distance

$$= r\left(\frac{\Delta\theta N}{2\pi\mu\kappa\kappa'}\right)$$

Relation between radius and hidden degree

$$r = R - \frac{2}{\zeta}\ln\left[\frac{\kappa}{\kappa_0}\right]$$

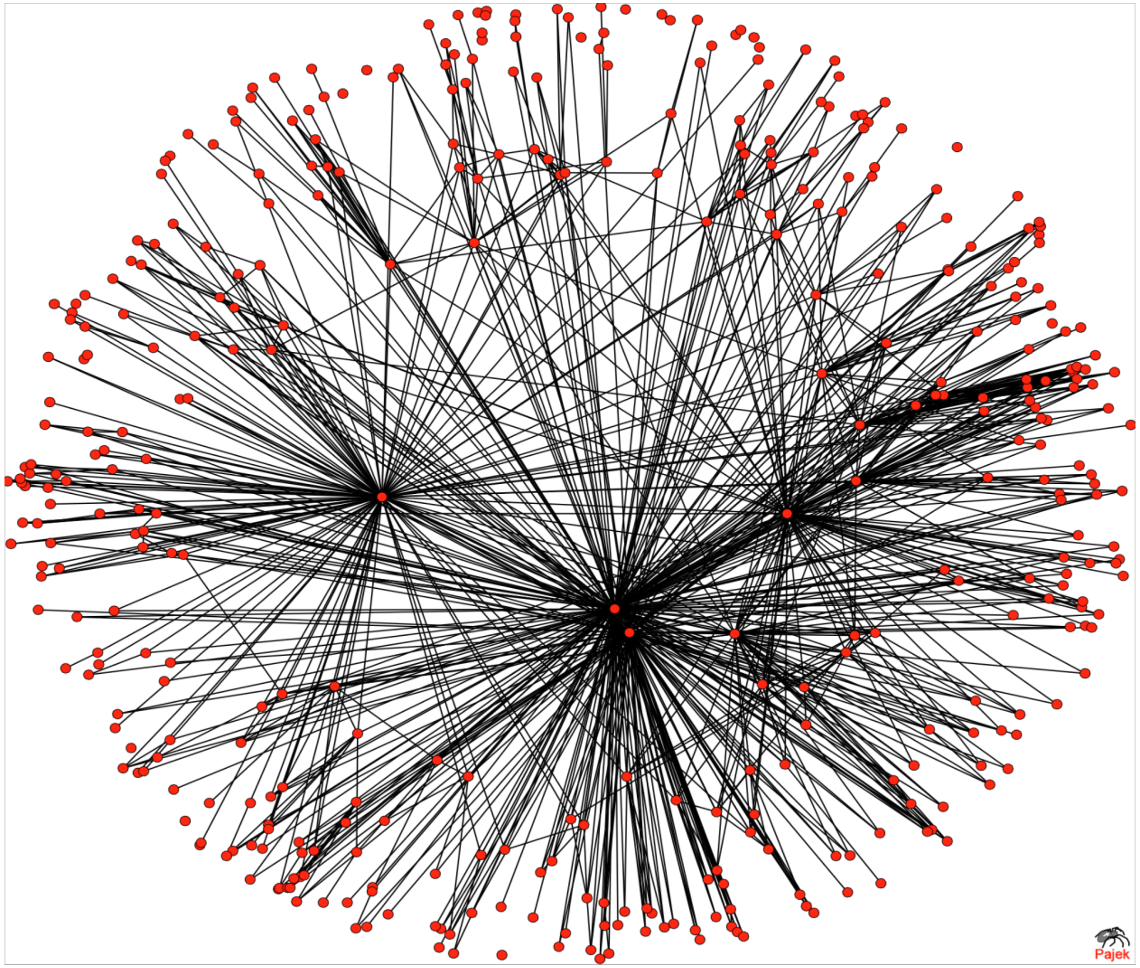
$$= \hat{r}\left(e^{\frac{\zeta}{2}(r+r'+\frac{2}{\zeta}\ln\frac{\Delta\theta}{2}-R)}\right)$$

$$\hat{r}\left(e^{\frac{\zeta}{2}(x-R)}\right)$$

$$\hat{r}(z)\begin{cases}\text{cte} & z \ll 1\\ 0 & z \gg 1\end{cases}$$

$p(x) = \Theta(R-x)$
nodes at hyperbolic distances smaller than R become connected

Isomorphism with model in hyperbolic space



Curvature and temperature of complex networks,
Dmitri Krioukov, Fragkiskos Papadopoulos, Amin Vahdat, and Marián Boguñá,
Phys. Rev. E 80, 035101 (2009), H2 model

S1 Newtonian
(gravity law)

vs

H2 Einstenian
or relativistic
(purely geometric)

Self-similarity of complex networks

In summary

hidden geometries underlying some complex networks appear to provide a simple a natural explanation of their self-similarity and observed topological properties

contrary to previous claims, the Internet is self-similar under appropriate renormalization

Future work

- Theoretical and practical implications of the self-similarity property
- Related concepts: renormalization and fractality

Thanks

M. A. Serrano, D. Krioukov, M. Boguñá

Phys. Rev. Lett. 100, 078701 (2008)

Work supported by DELIS FET Open 001907 and the SER-Bern 02.0234, by NSF CNS-0434996 and CNS-0722070, and by DGES FIS2004-05923-CO2-02, FIS2007-66485-C02-01, and FIS2007-66485-C02-02, Generalitat de Catalunya Grant No. SGR00889.

Dmitri Krioukov
Cooperative Association for Internet Data Analysis
CAIDA, UCSD, USA

Marián Boguñá
Dept. Física Fonamental
Universitat de Barcelona, Spain

marian.serrano@ub.edu