Modeling the Internet: stat. observables, dynamical approaches, parameter proliferation…………………..

A.Vespignani

Indiana University School of informatics

the biocomplexity institute
Indiana University

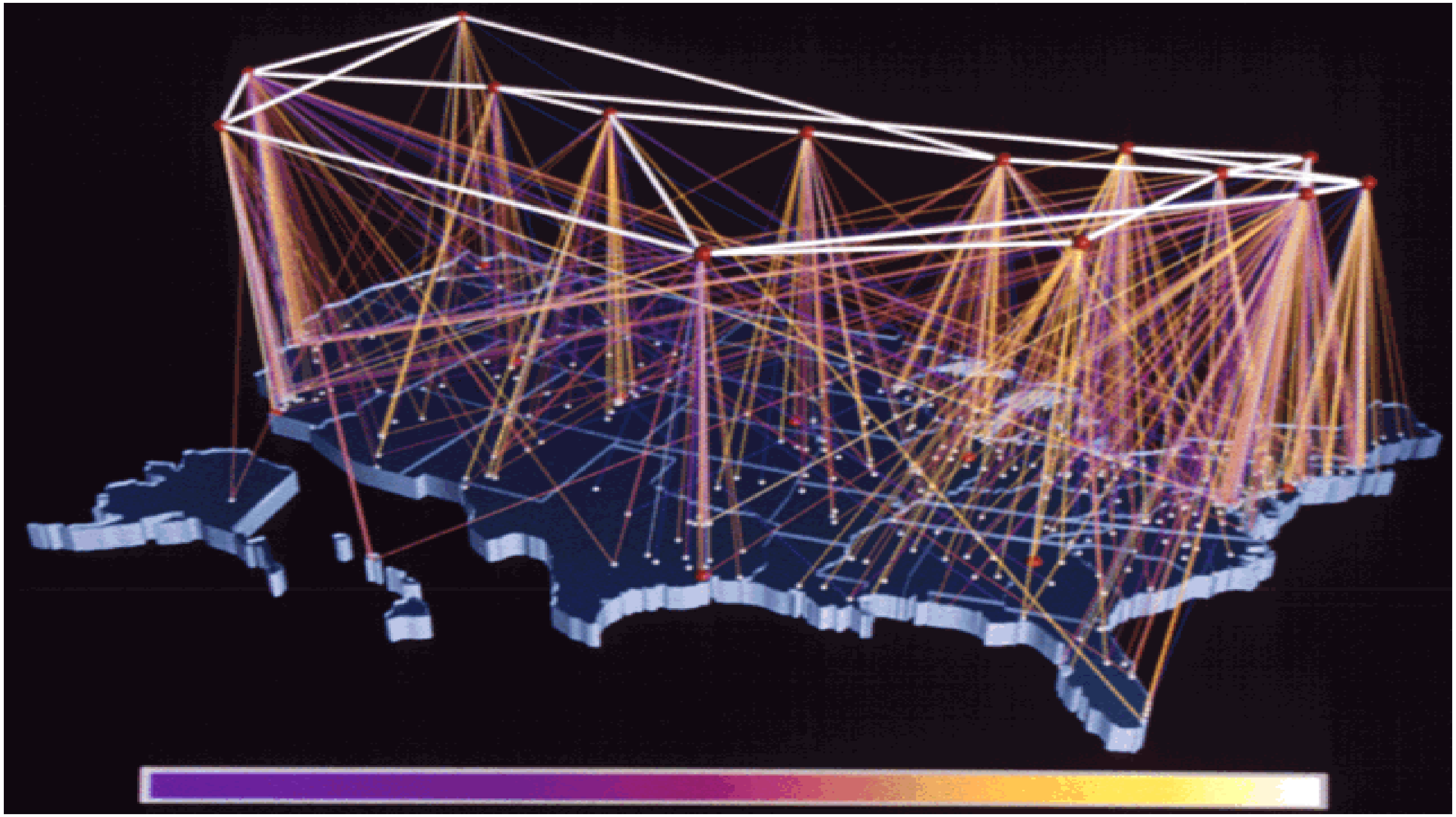
CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

INSTITUTE
FOR SCIENTIFIC INTERCHANGE
FOUNDATION

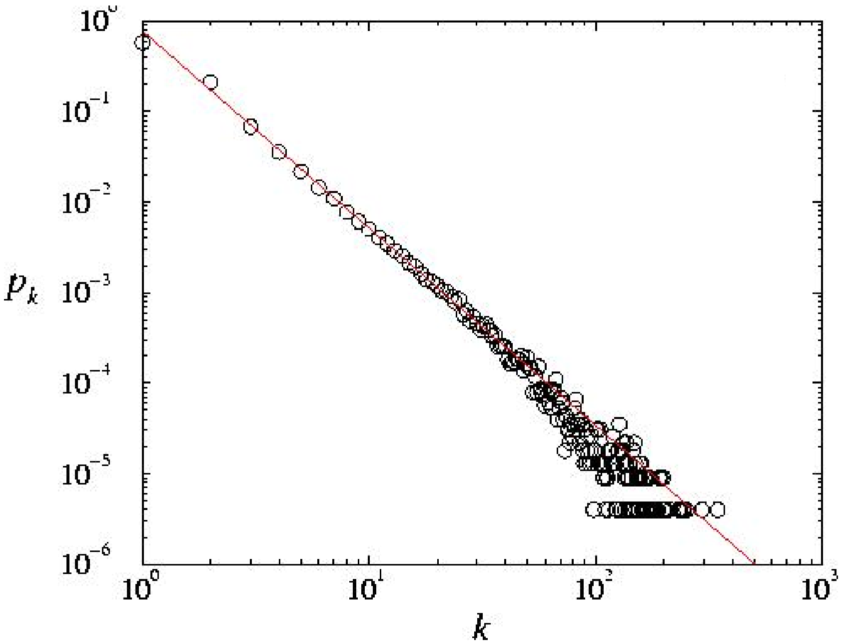
Collaborators

- Romualdo Pastor-Satorras
- Ignacio Alvarez-Hamelin
- Luca Dall’Asta
- Alain Barrat
- Vic Colizza
- Mark Meiss
- Filippo Menczer
- Mariangels Serrano
- Alexei Vazquez

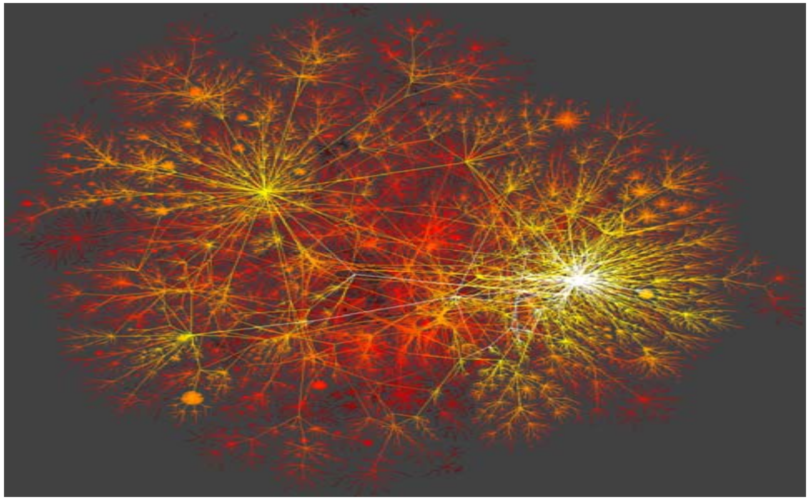
Once upon a time there was the physical Internet….



The beginning…..



Faloutsos et al. 1999

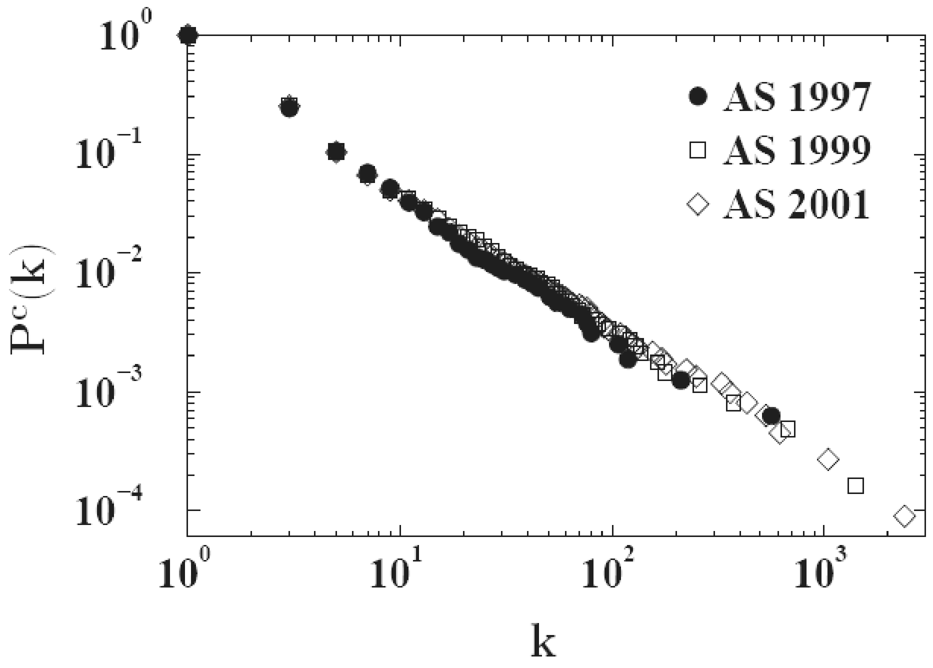


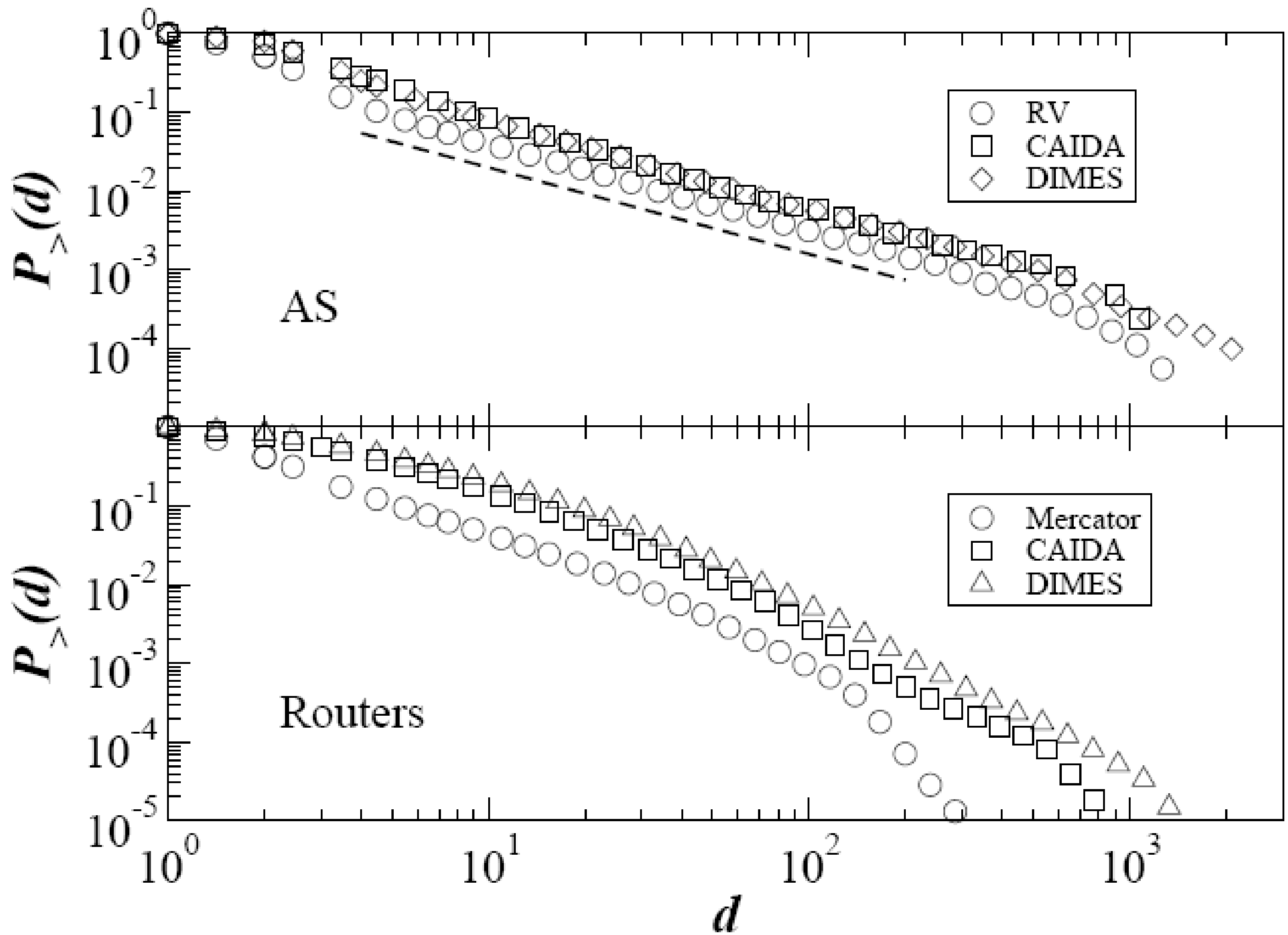
Degree distribution of the Internet graph
(AS and Router level)

Measurement infrastructures
Passive/active measurements
(CAIDA; NLANR; Lumeta…)

Internet graphs…..

- Skewed
- Heterogeneity and high variability
- Very large fluctuations (variance>>average)
- Various fits : power-law+cut-off; Weibull etc.





RV
CAIDA
DIMES
AS
Mercator
CAIDA
DIMES
Routers
$P_>(d)$
$P_>(d)$
d

- Higher order statistical characterization….
- Model validation……
- Model construction…..

Multi-point correlations *P(k,k')*

- **0-dimensional projection (pearson coefficient)**

  M. Newman (2002)

- **One-dimensional projection (average nearest neighbor degree)**

  Pastor-Satorras & A.V. (2001)

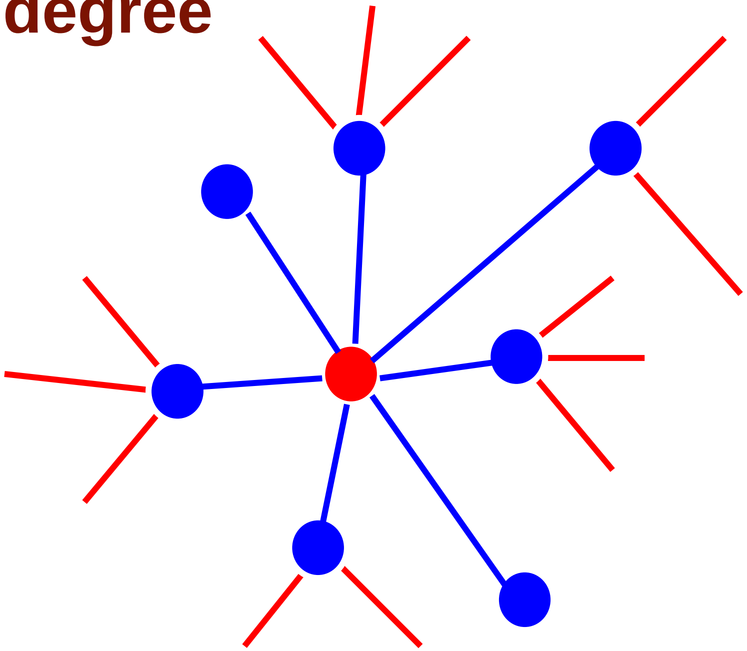
- **Three dimensional analysis**

  Maslov&Sneppen (2002)

Multi-point correlations *P(k,k')*

Average nearest neighbors degree

$$k_{nn}(i) = \frac{1}{k_i} \sum_j k_j$$



Correlation spectrum:

Average over degree classes $< k_{nn}(k) >$

$$< k_{nn}(k)> = \Sigma_{k'} k' \, p(k'|k)$$

Degree correlation function

- **Assortative** behaviour: growing $k_{nn}(k)$

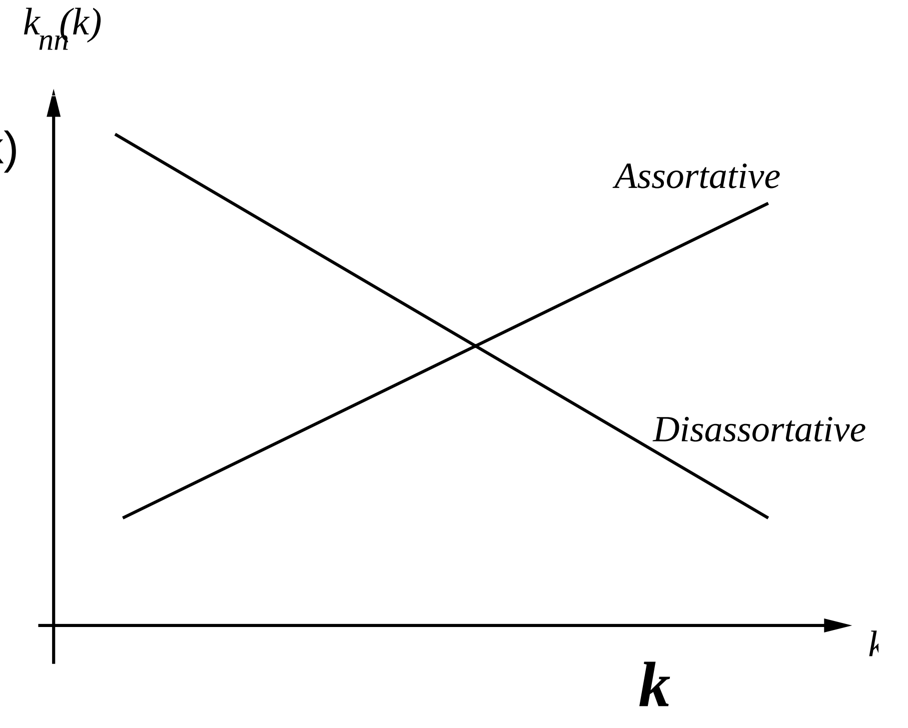
Example: social networks

Large sites are connected with large sites

- **Disassortative** behaviour: decreasing $k_{nn}(k)$

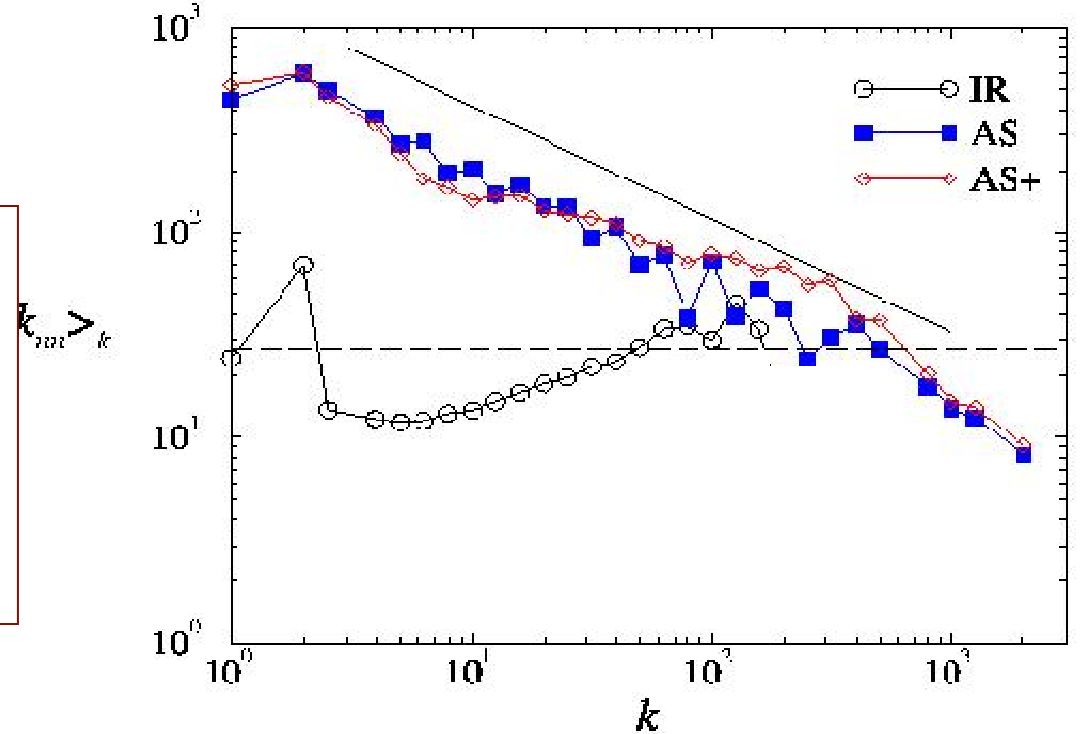
Example: internet

Large sites connected with small sites



$$< k_{nn}(k)> = \Sigma_{k'} k' \, p(k'|k)$$

Degree correlation function

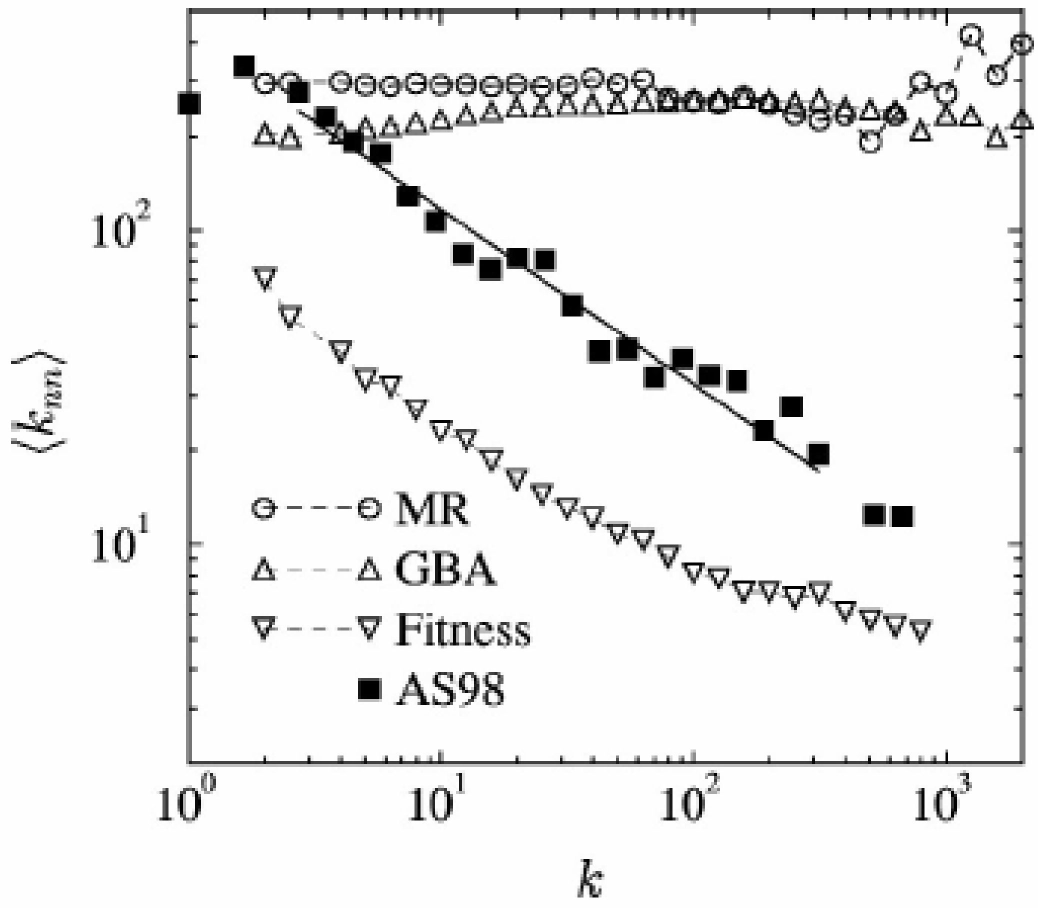


Highly degree ASs connect to low degree ASs

Low degree ASs connect to high degree ASs

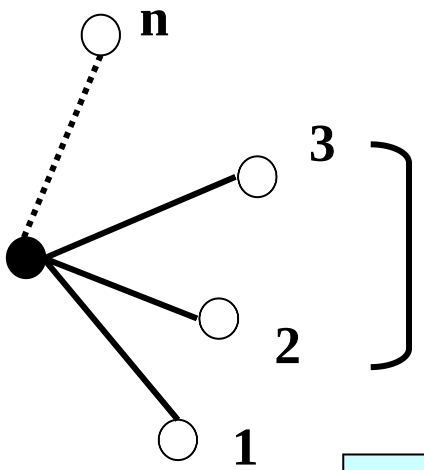
No hierarchy for the router map

Pastor Satorras, Vazquez & Vespignani,
PRL 87, 258701 (2001)



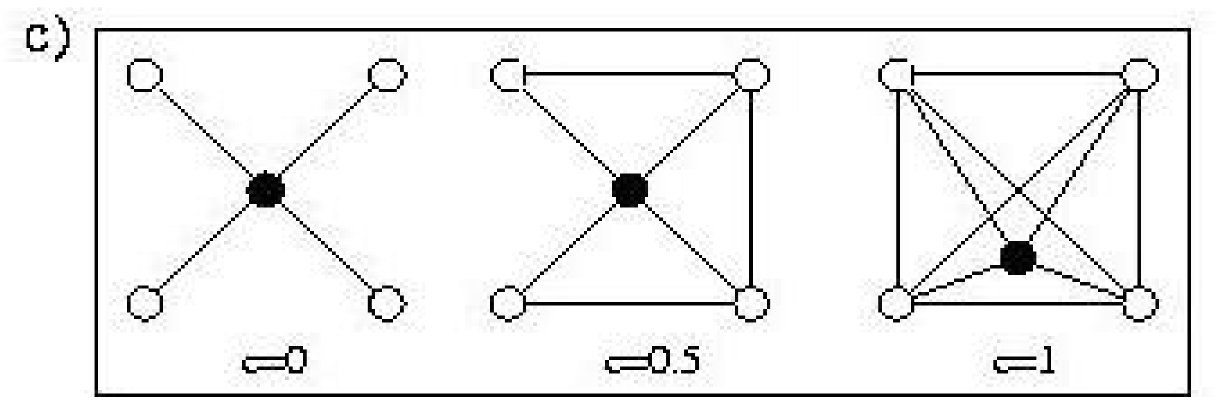
$\langle k_{nn} \rangle$
10^2
10^1
MR
GBA
Fitness
AS98
10^0
10^1
10^2
10^3
k

Clustering coefficient = **connected peers will likely know each other**



Higher probability to be connected

$$C = \frac{\text{\# of links between 1,2,…k neighbors}}{k(k-1)/2}$$



Clustering spectrum

$$C(k) = \frac{\sum_i \delta(k_i - k) c_i}{\sum_i \delta(k_i - k)}$$

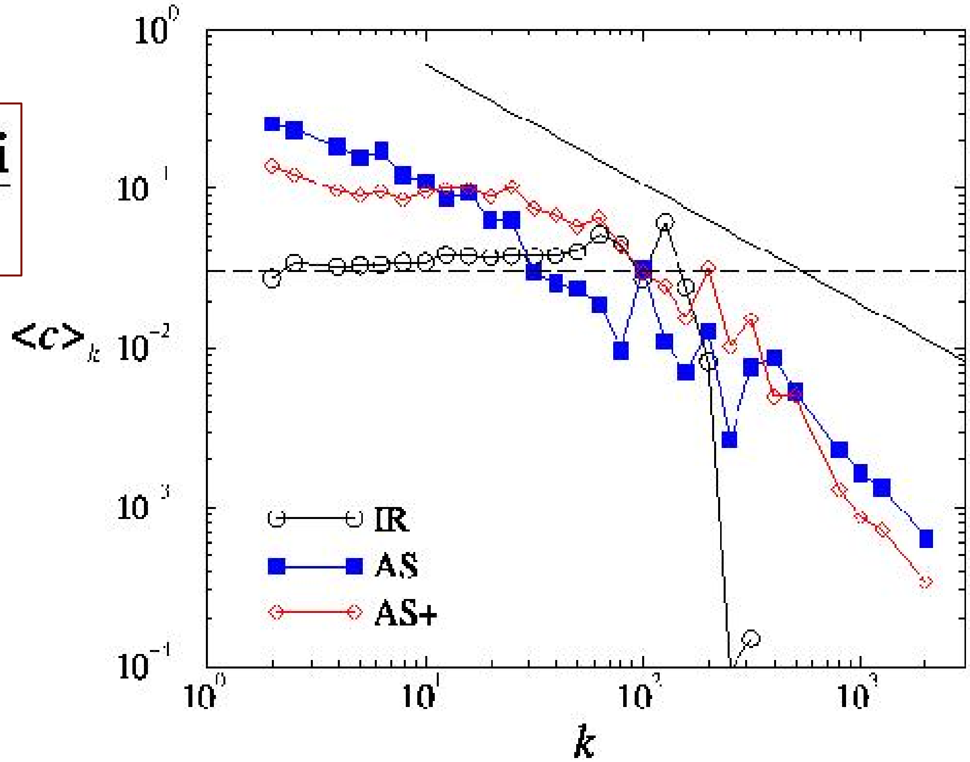
Clustering spectrum

This is a kind of three-points correlation function…..

Clustering Spectrum in the Internet

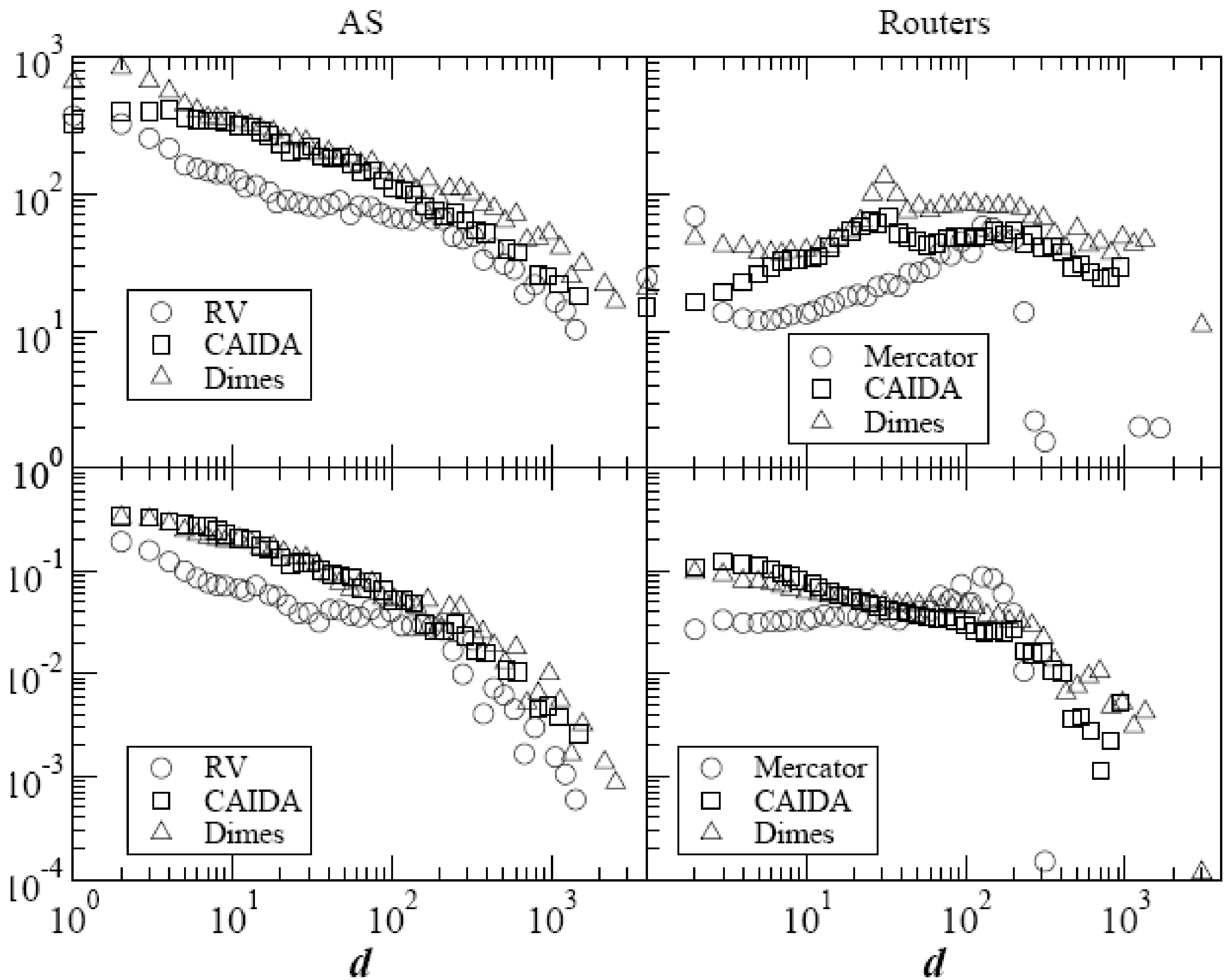
Clustering coefficient as a function of the vertex degree

$$\mathbf{C(k) = \frac{\Sigma_i\ \delta(k_i - k) c_i}{\Sigma_i\ \delta(k_i - k)}}$$



Highly degree ASs bridge not connected regions of the Internet
Low degree ASs have links with highly interconnected regions of the Internet

Vazquez et al.
Physical Review E 65, 066130 (2002).



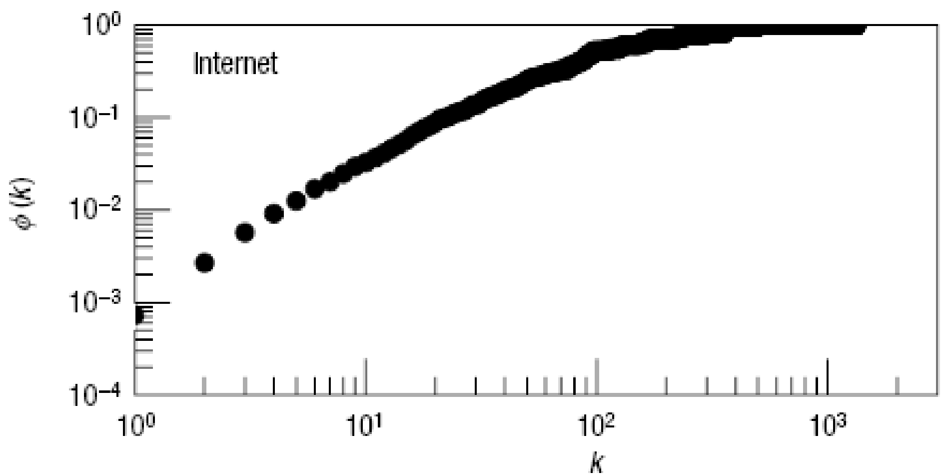
AS
Routers
RV
CAIDA
Dimes
Mercator
CAIDA
Dimes
RV
CAIDA
Dimes
Mercator
CAIDA
Dimes
d
d

Rich-Club coefficient

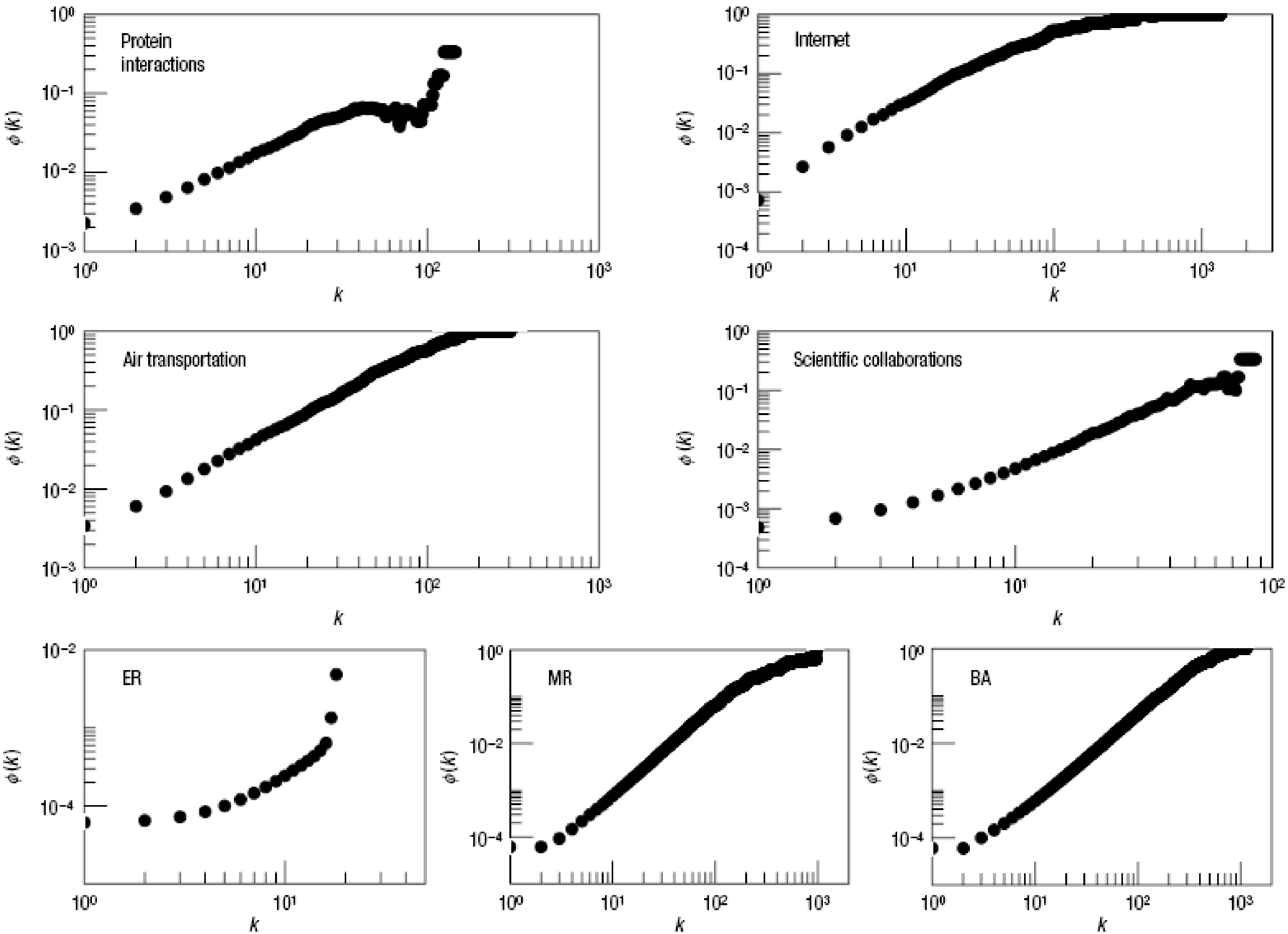
Fraction of edges shared by nodes of degree >*k* with respect to the Maximum allowed number.

$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k}-1)}$$

Increasing interconnectivity for increasing *k*



Rich-club phenomenon??



Protein interactions
Internet
Air transportation
Scientific collaborations
ER
MR
BA
$\phi(k)$
k
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^1
10^2
10^3

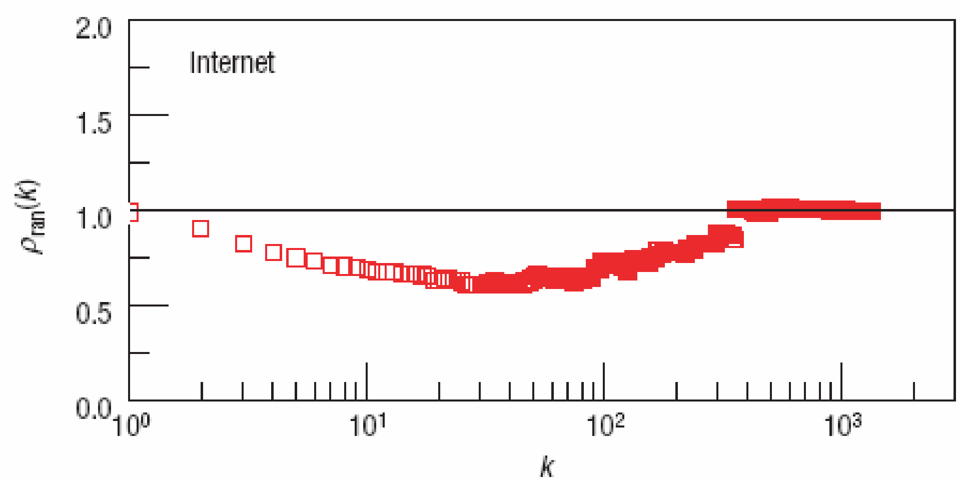
Normalized rich-club coefficient

It is possible to show that for a completely uncorrelated network

$$\phi_{\mathrm{unc}}(k) \underset{k,k_{\max}\to\infty}{\sim} \frac{k^2}{\langle k\rangle N}$$

$$\rho_{\mathrm{ran}}(k) = \phi(k)/\phi_{\mathrm{ran}}(k)$$

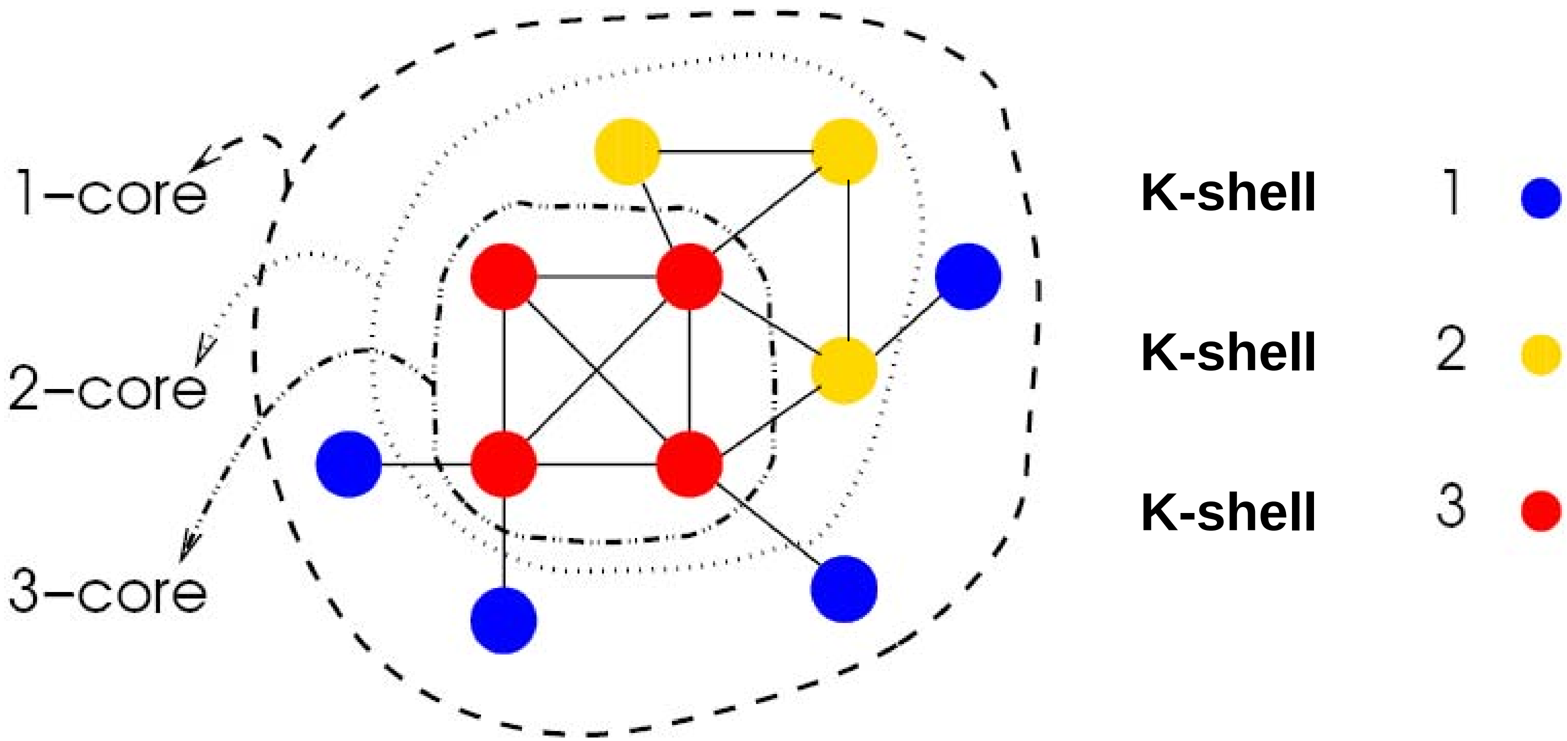
Coefficient of the maximally randomized equivalent graph



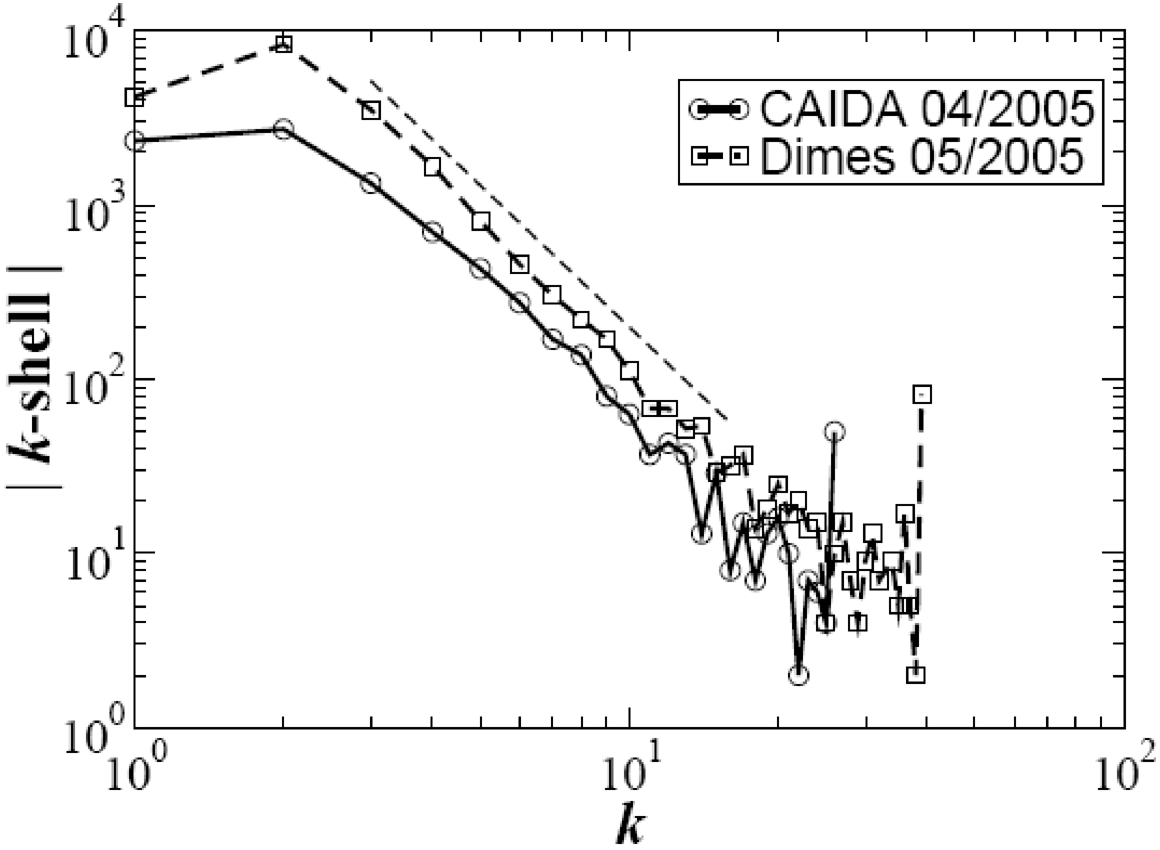
NO rich club phenomenon

V.Colizza et al. nature physics | VOL 2 | FEBRUARY 2006

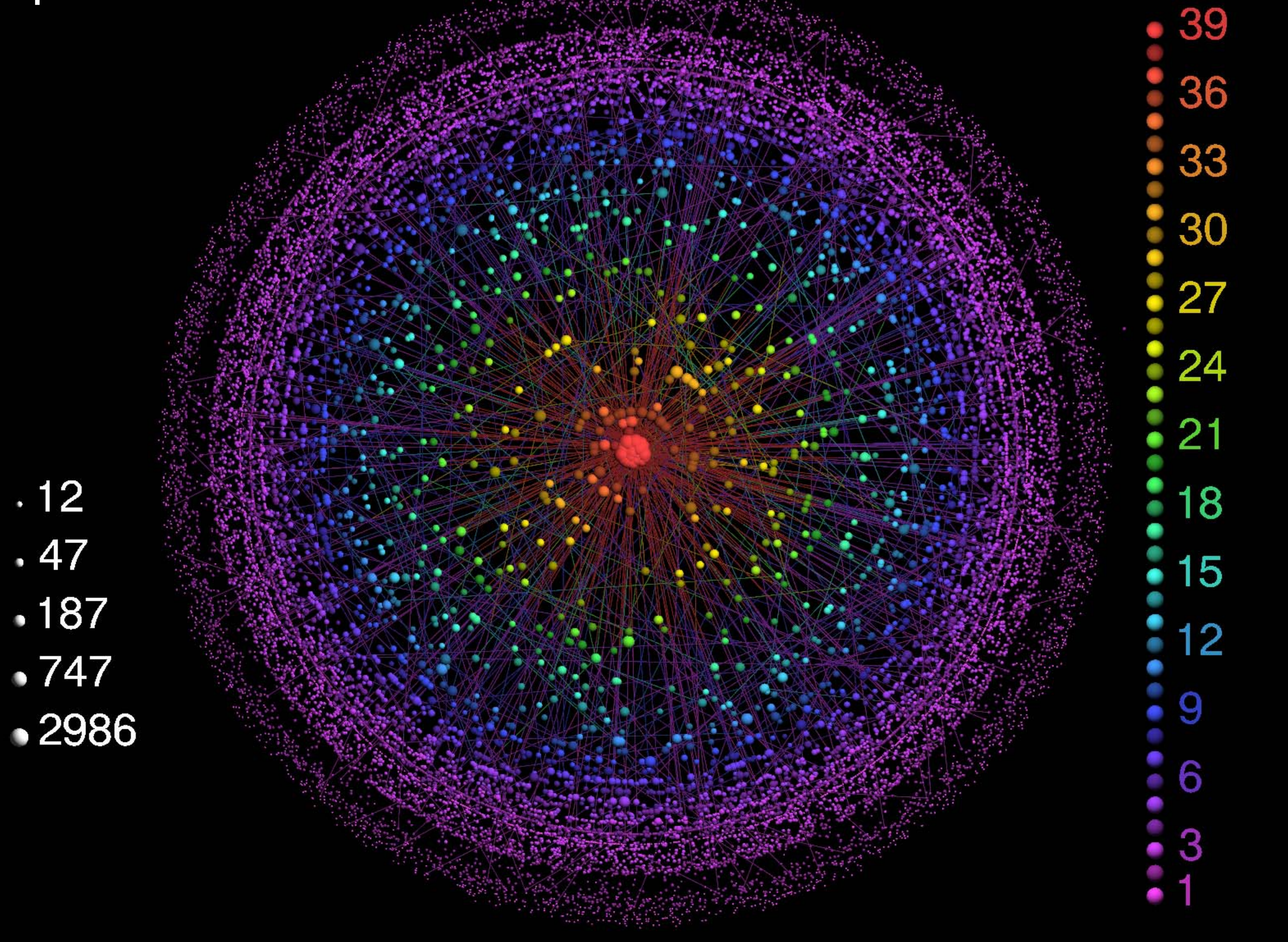
K-core decomposition



K-core structure…

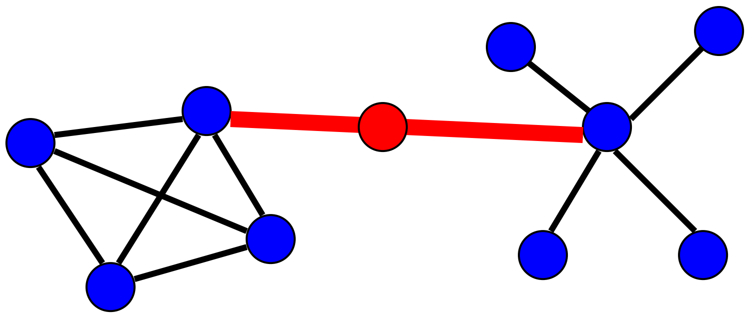


http://xavier.informatics.indiana.edu/lanet-vi

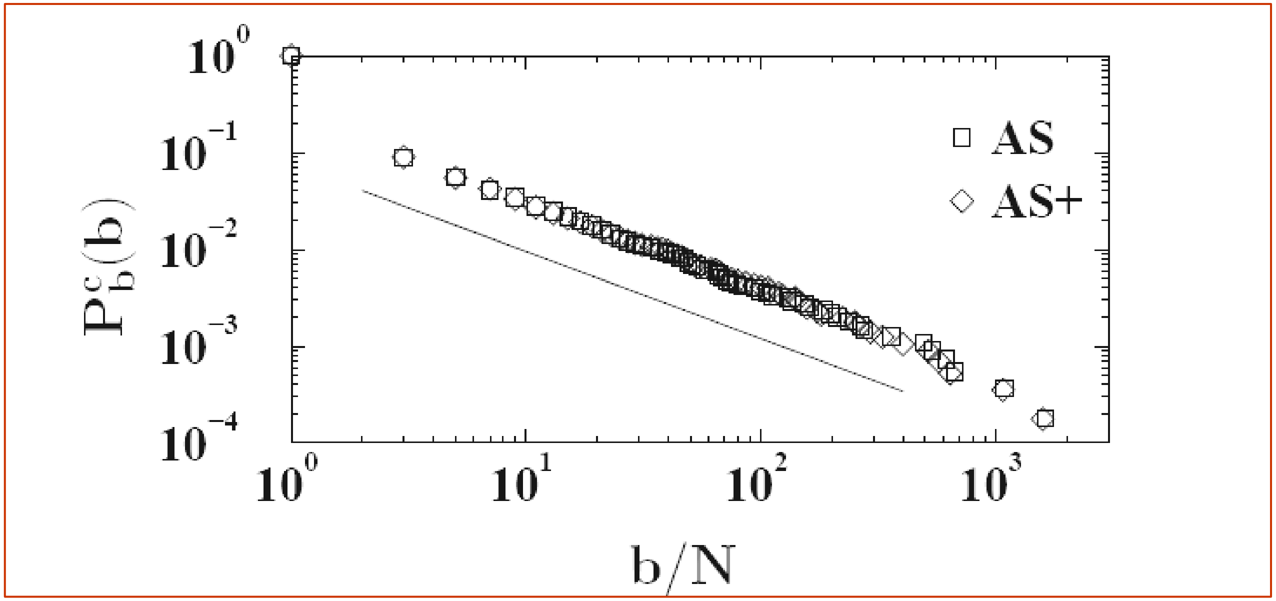


Non-local measure of centrality

Betweenness centrality = # of shortest paths traversing a vertex or edge (flow of information) if each individuals send a message to all other individuals



Beteweenness Probability distribution



Heavy-tailed and highly heterogeneous

Classical topology generators

- Waxman generator
- Structural generators
  - Transit-stub
  - Tiers

→ Exponentially Bounded Degree distributions

Scale-free topology generators

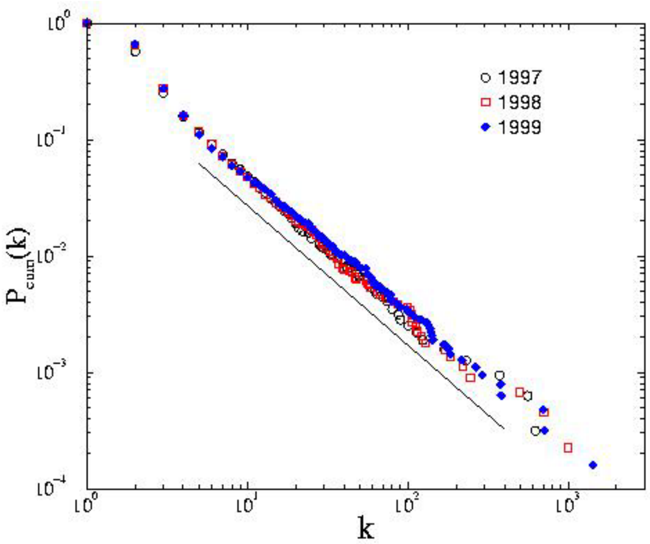
INET (Jin, Chen, Jamin)

BRITE (Medina & Matta)

Modeling of the Network structure with ad-hoc algorithms tailored on the properties we consider more relevant

What about the degree distribution ?

Heavy tails ?



Static construction

Molloy-reed
Position model
Hidden variables
Etc.

Generalized random graphs with pre-assigned degree distribution

Shift of focus:

Static constructionDynamical evolution

Direct problem

Evolution rules → Emerging topology

Inverse problem

Given topology → Evolution rules

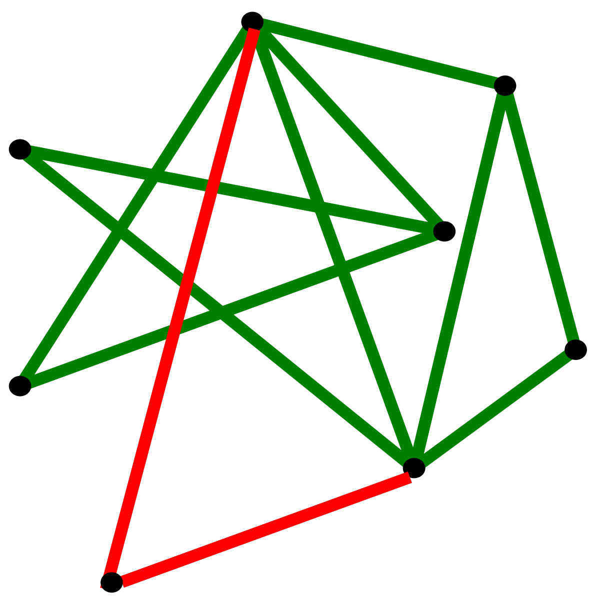
The rich-get-richer mechanism

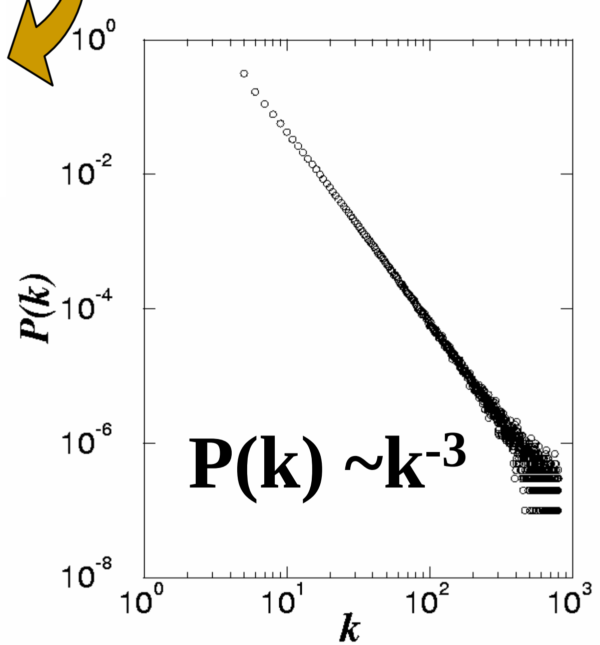
(Barabasi& Albert 1999)

Growth: The network starts with a small core of m_0 connected vertices. Every time step we add a new vertex, with m edges ($m < m_0$) connected to old vertices in the system.

Preferential attachment: The new edges are connected to the old s-th vertex with a probability proportional to its degree k_s.

$$\Pi[k_s(t)] = \frac{k_s(t)}{\sum_j k_j(t)}$$





Continuous approximations

Average degree value $k_s(t)$ that the node born at time s has a time t

$\{p(k,s,t)=\delta(k-k_s(t))\}$

Evolution equation

$$\frac{\partial k_s(t)}{\partial t} = m\Pi[k_s(t)]$$

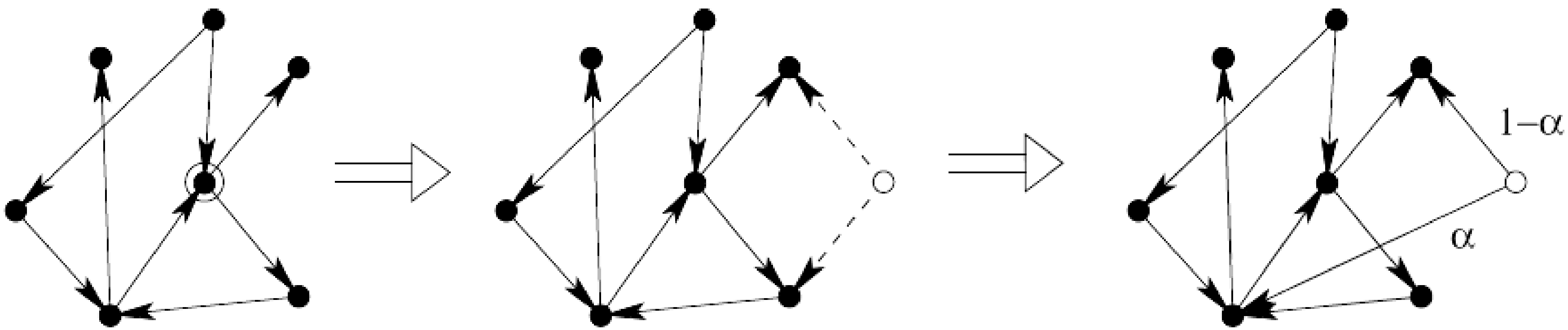
Degree distribution

$$P(k,t) = \frac{1}{t+m_0}\int_0^t \delta(k-k_s(t))\,\mathrm{d}s \equiv -\frac{1}{t+m_0}\left(\frac{\partial k_s(t)}{\partial s}\right)^{-1}\Bigg|_{s=s(k,t)}$$

- **BA is a conceptual model….**
- **It has not been thought to specifically model the Internet**
- **More details/realism/ingredients needed**

COPY MODEL

Illustration of the rules of the copying model. A prototype vertex (black dot surrounded by a circle) is selected and a new vertex (hollow dot) is created with virtual edges pointing to the neighbors of the prototype. With probability $1-\alpha$ the virtual edges are kept; with probability α they are rewired to a randomly chosen vertex.



Dynamical evolution

Preferential attachment component

$$\frac{\partial k_{in,s}(t)}{\partial t} = m\left[\frac{\alpha}{t} + (1-\alpha)\frac{k_{in,s}(t)}{mt}\right]$$

Degree distribution

$$P(k_{in}) \sim k_{in}^{-(2-\alpha)/(1-\alpha}$$

More models

•Generalized BA model

Non-linear preferential attachment : $\Pi(k) \sim k^{\alpha}$

(Redner et al. 2000)

Initial attractiveness : $\Pi(k) \sim A + k^{\alpha}$

(Mendes & Dorogovstev 2000)

Rewiring

(Albert et al.2000)

•Highly clustered

(Eguiluz & Klemm 2002)

$$\Pi(k_i) \cong \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

•Fitness Model

(Bianconi et al. 2001)

•Multiplicative noise

(Huberman & Adamic 1999)

Heuristically Optimized Trade-offs (HOT)

Papadimitriou et al. (2002)

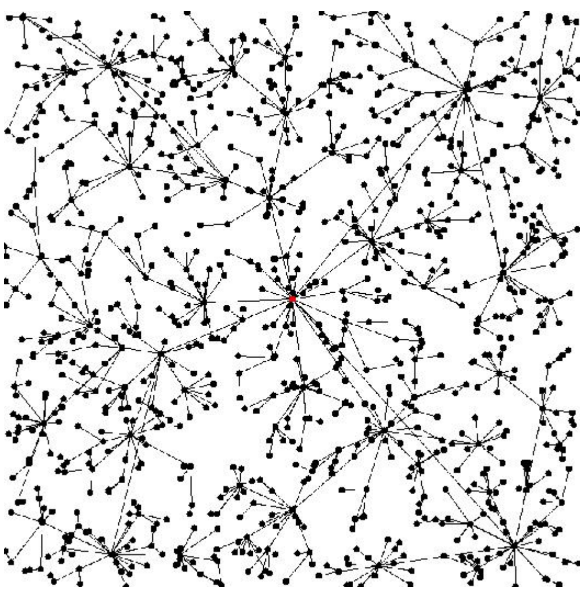
New vertex ***i*** connects to vertex ***j*** by minimizing the function

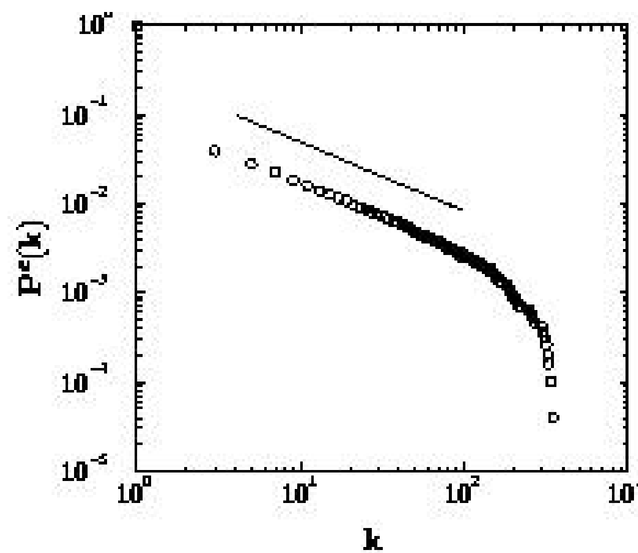
$$\mathbf{Y}(i,j) = \alpha\, \mathbf{d}(i,j) + \mathbf{V}(j)$$

d= euclidean distance

V(*j*)= measure of centrality

Optimization of conflicting objectives

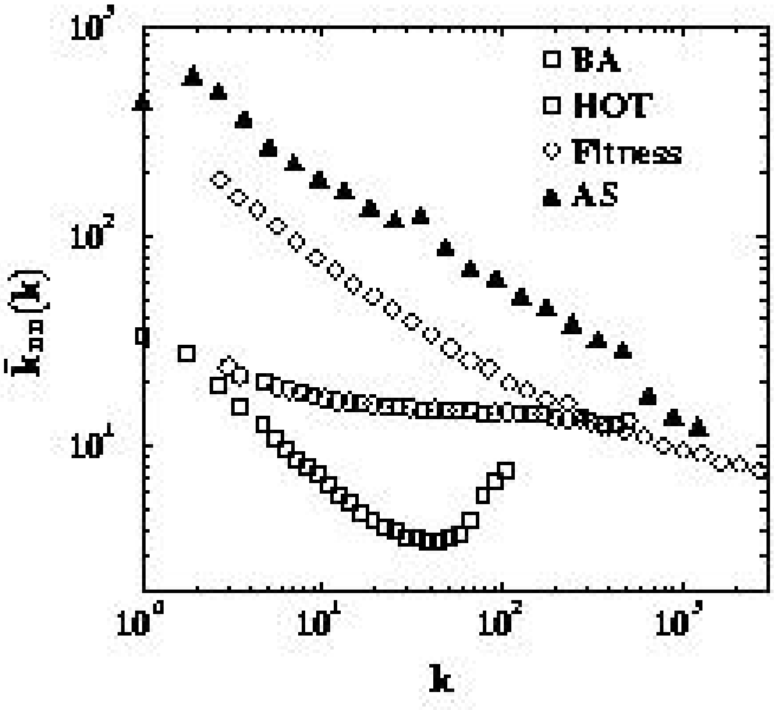




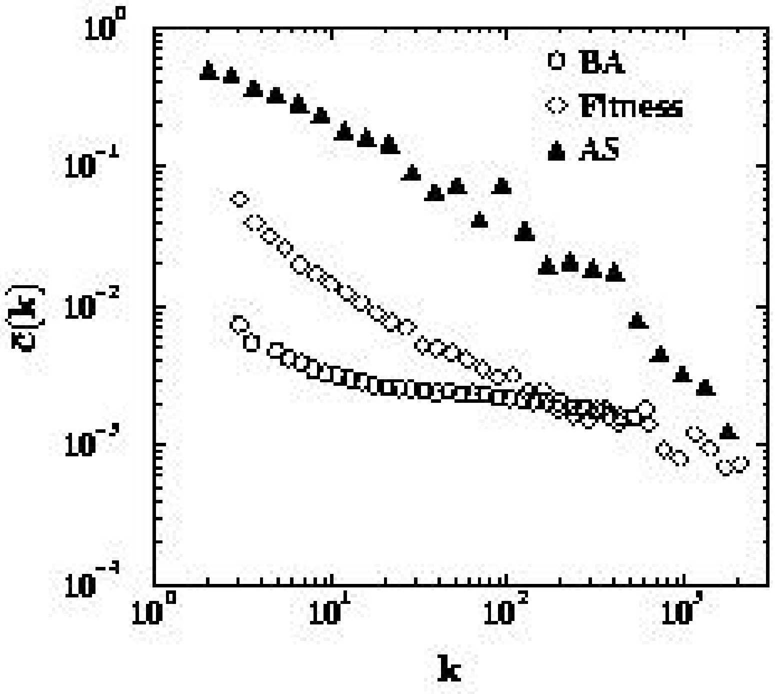
Model validation……

- Correlations
- Clustering
- Hierarchies (k-cores, modularity etc.)
- ………..

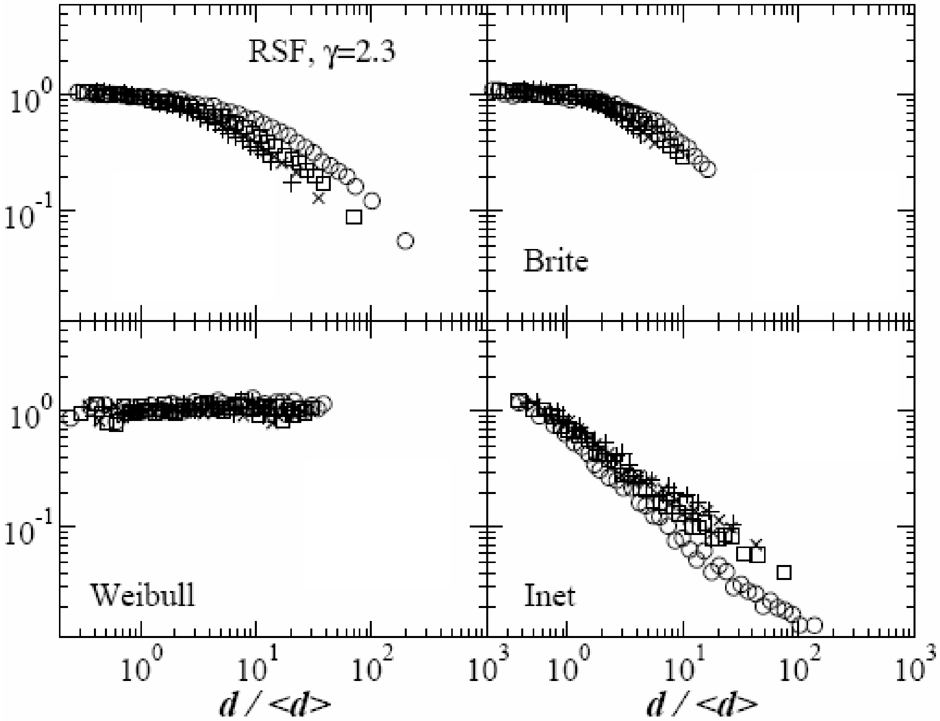
Correlation spectrum



Clustering spectrum

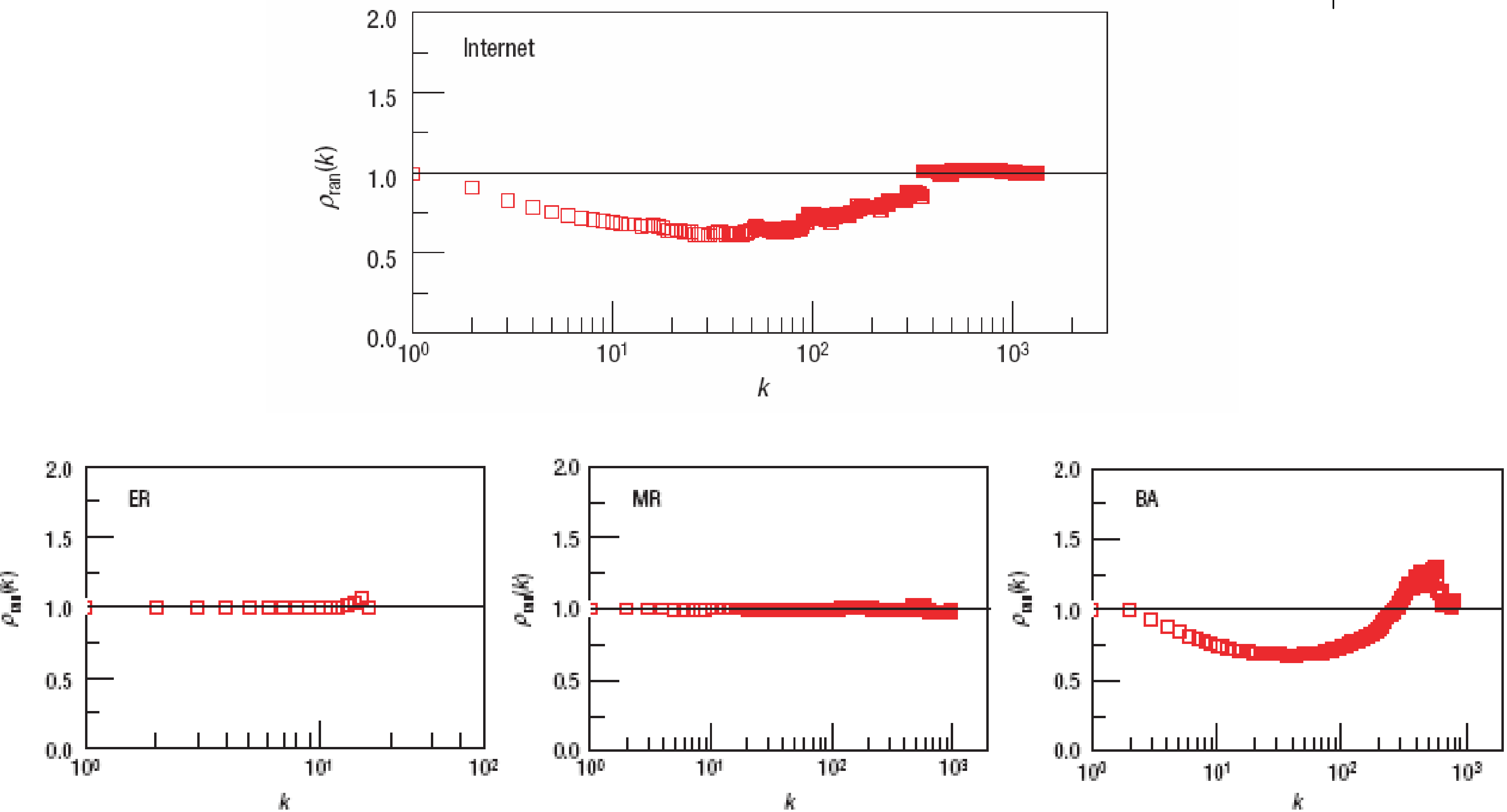


Clustering spectrum

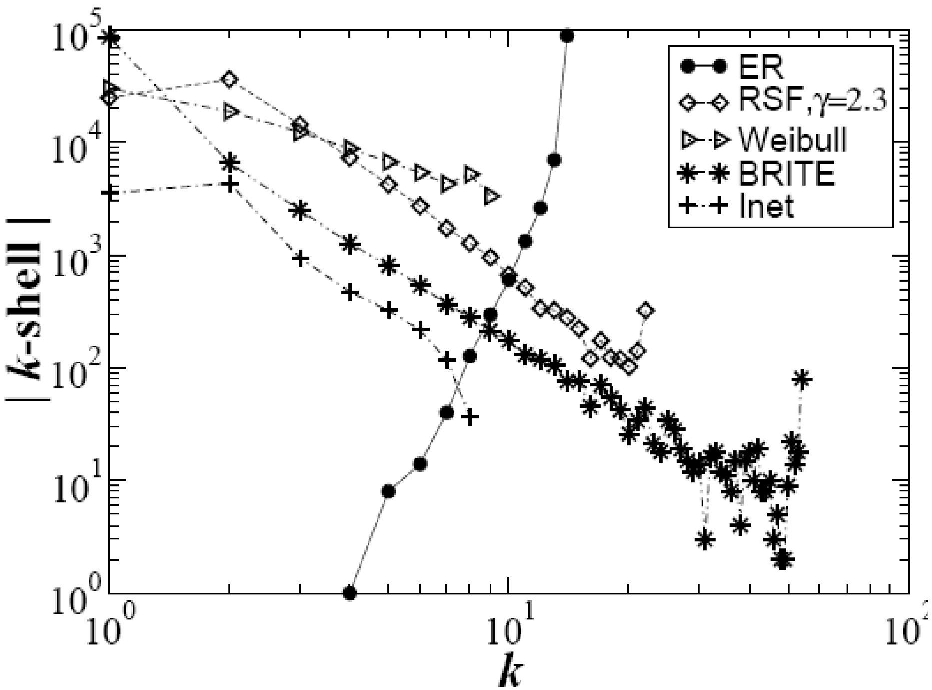


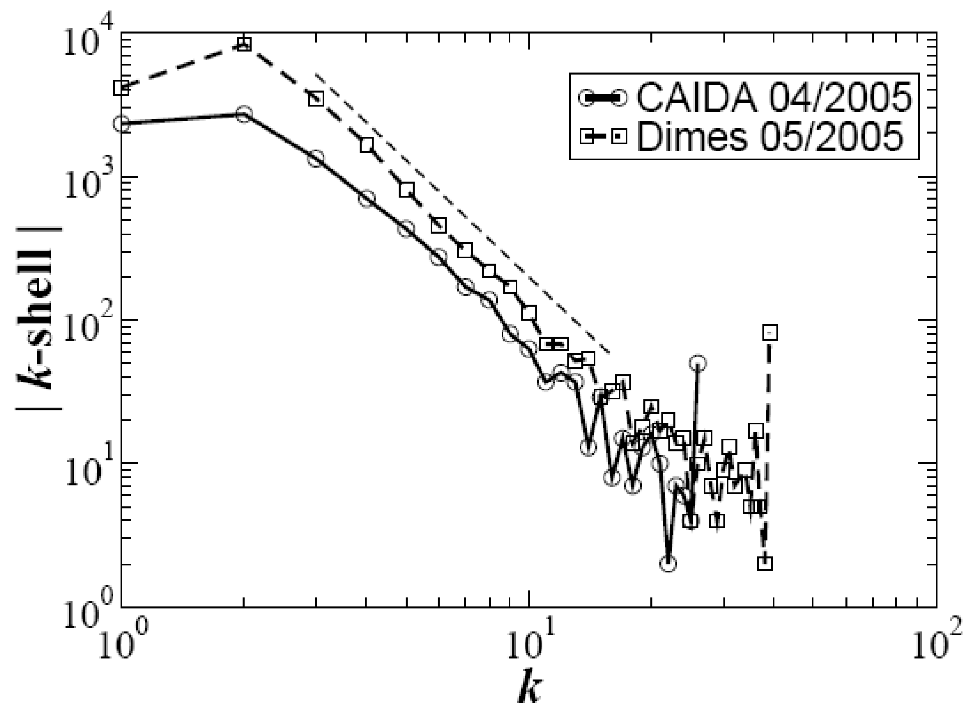
Correlation spectrum

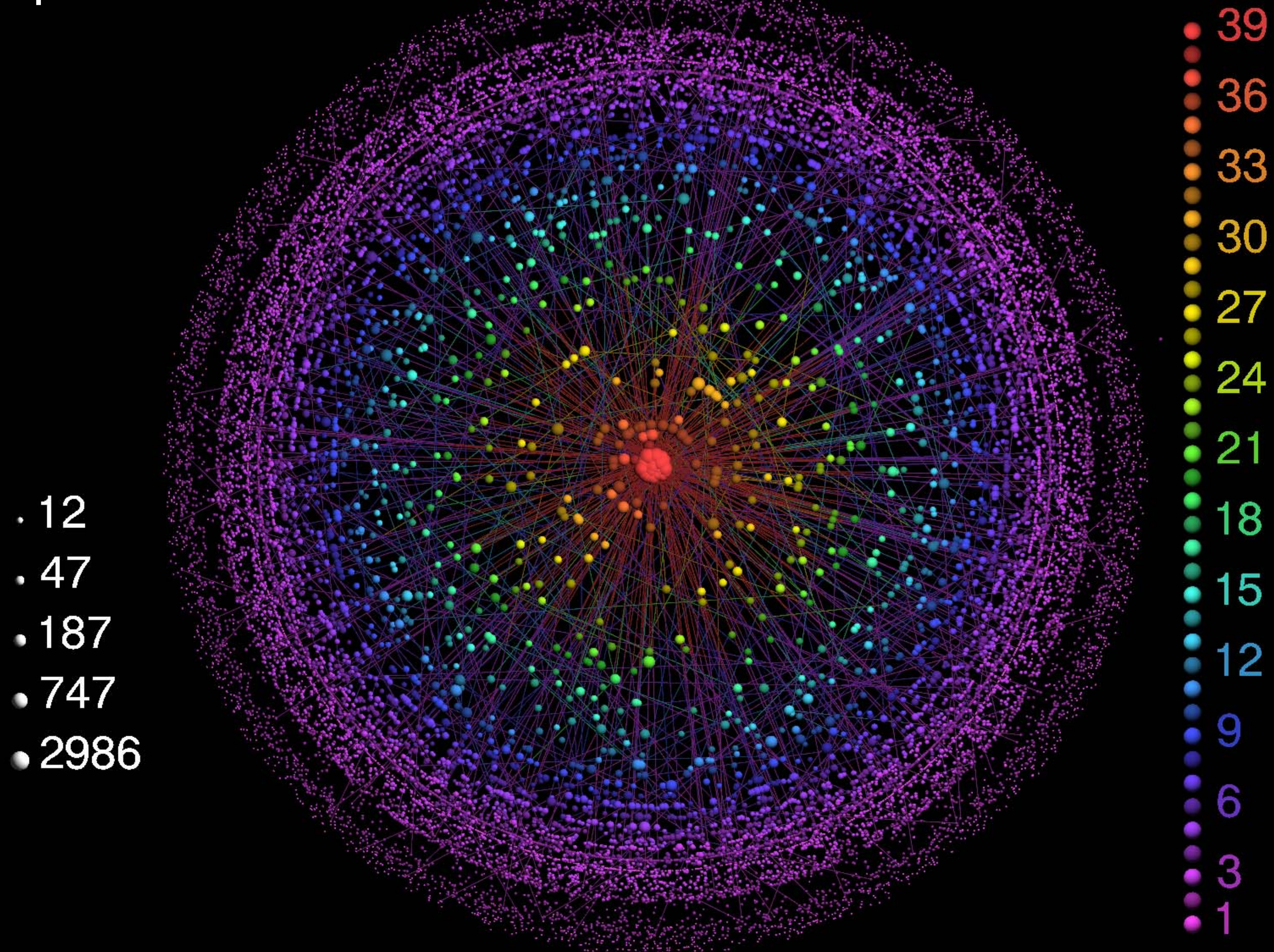
Rich-club coefficient



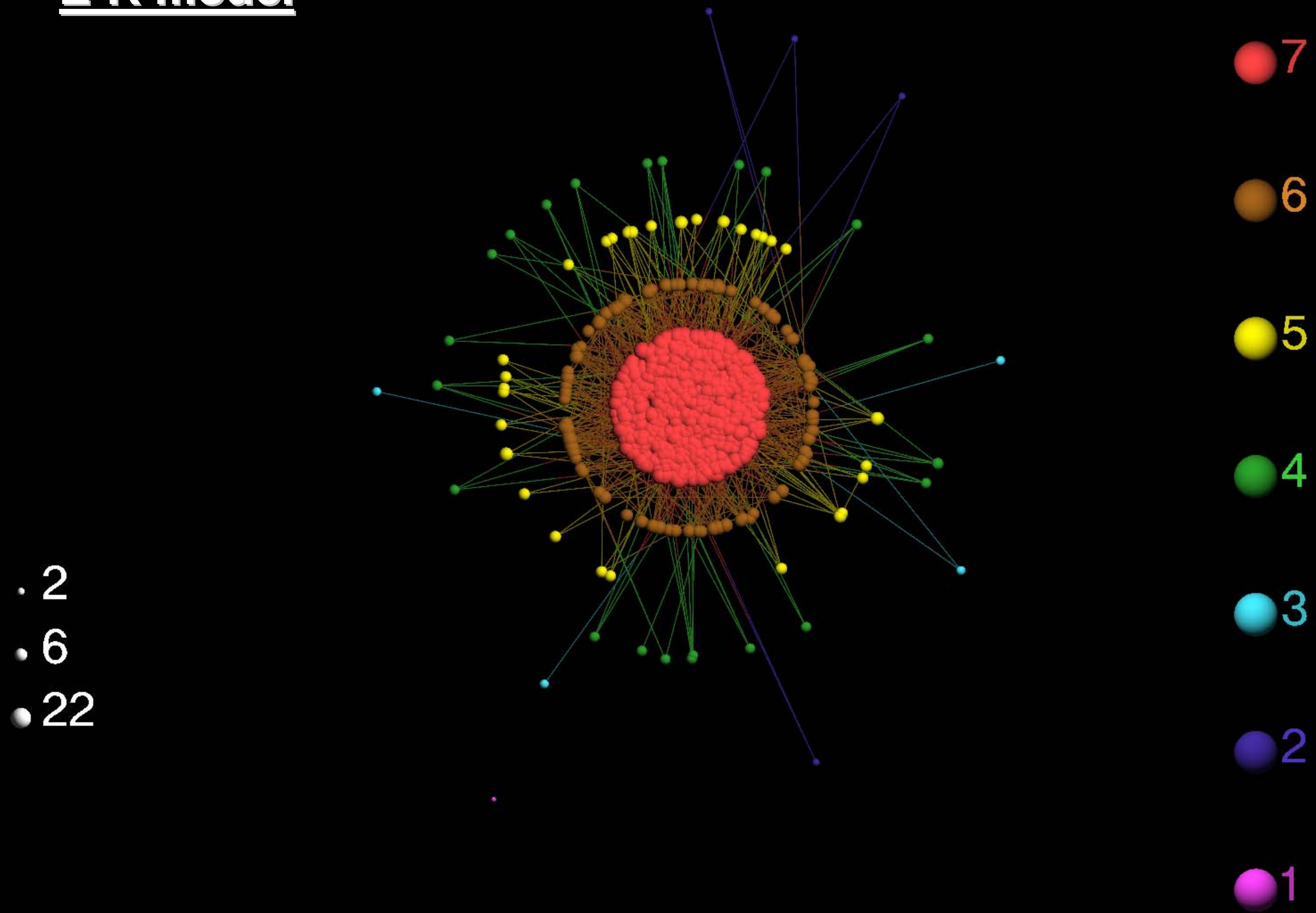
K-core structure



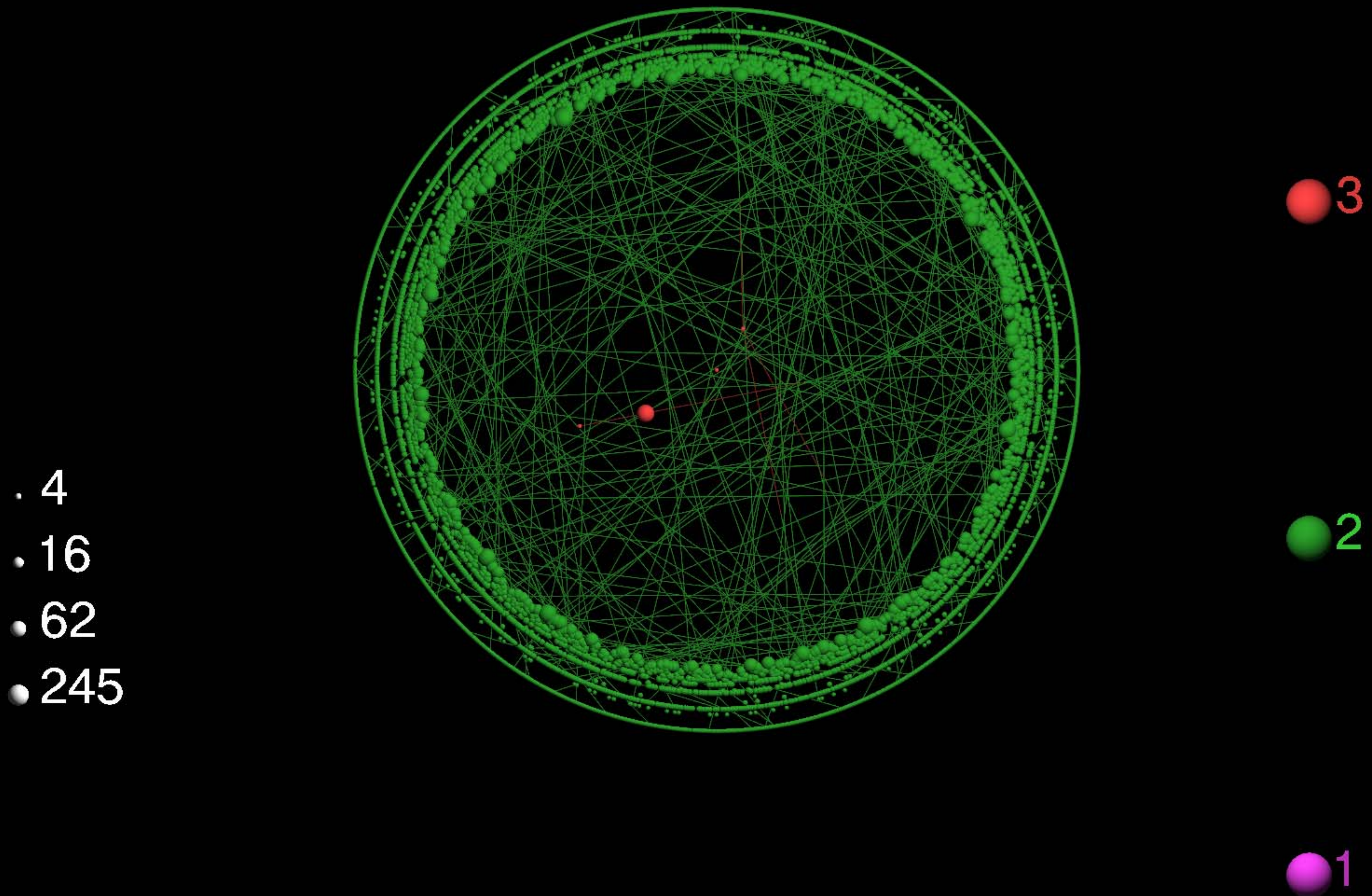



12
47
187
747
2986
39
36
33
30
27
24
21
18
15
12
9
6
3
1

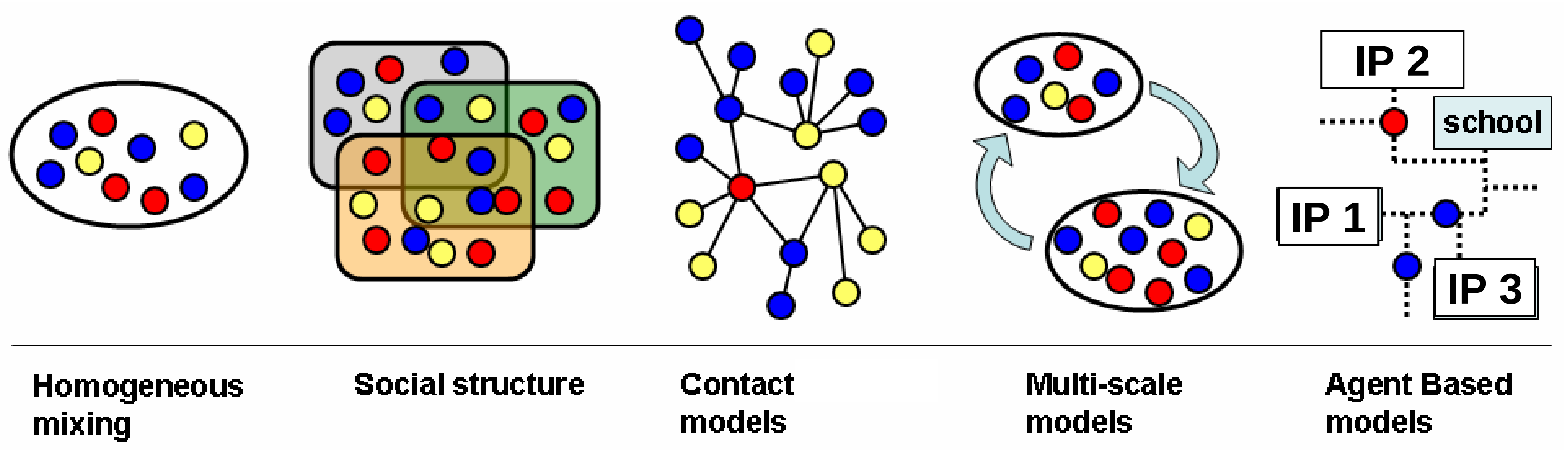
E-R model



B-A Model



Wide spectrum of complications and complex features to include…

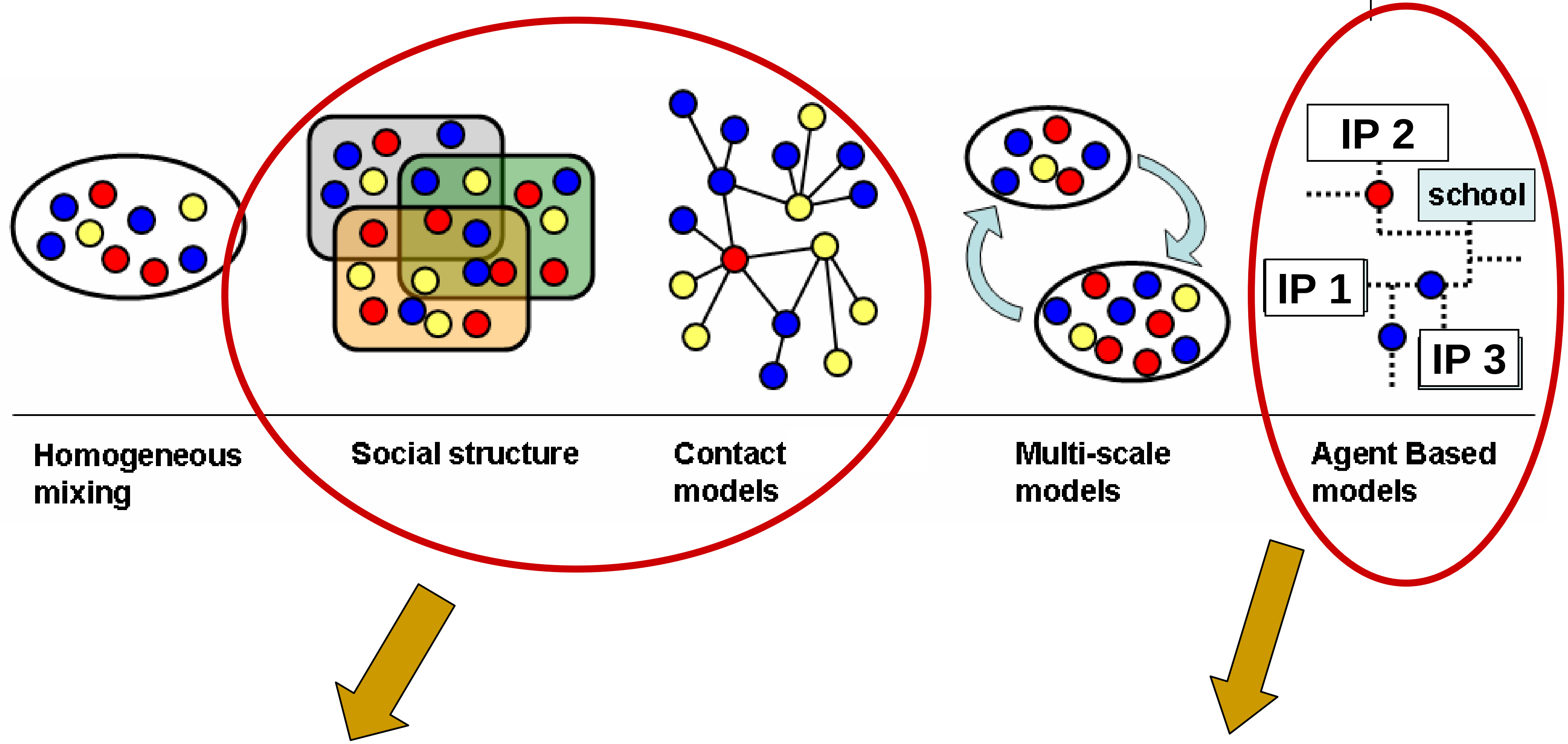


Simple → **Realistic**

Ability to explain (caveats) trends at a population level

Model realism looses in transparency.
Validation is harder.

Wide spectrum of complications and complex features to include…



Conceptual/theoretical

Data driven
Value if data are realistic and parameters are physical

Agent Based Modeling

- The good…
  - Data driven:

    Demographic, societal, census data from real experiments
  - Sensibility analysis / scenario evaluation

- The bad…
  - **Non-physical** parameters
    - (non-measurable/fitness/unmotivated parameters etc.)…

Physical Parameters ??

- Measurable quantity.
- Combination of measurable quantities.
- Parameters appearing from the symmetry and consistency of equations.

Hints..

- Minimum number of free (measurable) parameters….
- Falsifiable requisite for the model….

A few examples….

- BA model $$\Pi[k_s(t)] = \frac{k_s(t)}{\sum_j k_j(t)}$$
- Rewiring/copy model $$\left[\frac{\alpha}{t} + (1-\alpha)\frac{k_{in,s}(t)}{mt}\right]$$
- Fitness model $$\Pi(k_i) \cong \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$
- HOT $$Y(i,j) = a\, d(i,j) + V(j)$$

- Census/societal data
- Geographical data
- Traffic data
- In the lack of that ……..topology generators!!

(Using measurement data)

Effect of complex network topologies on physical processes

- Epidemic models

- Resilience & robustness

- Avalanche and failure cascades

- Search and diffusion…..