RANDOM NETWORKS

Analysis of Large Scale Social Networks

Bart Thijs

- We can study theoretical properties of real world graphs
- Compare properties with real observations
- Create realistic test graphs which enables development

WHY

Four random graph model

1. Erdös-Rényi Model
2. Watts-Strogatz Model
3. Barabási-Albert Model
4. Bianconi-Barabási Model

RANDOM NETWORK MODELS

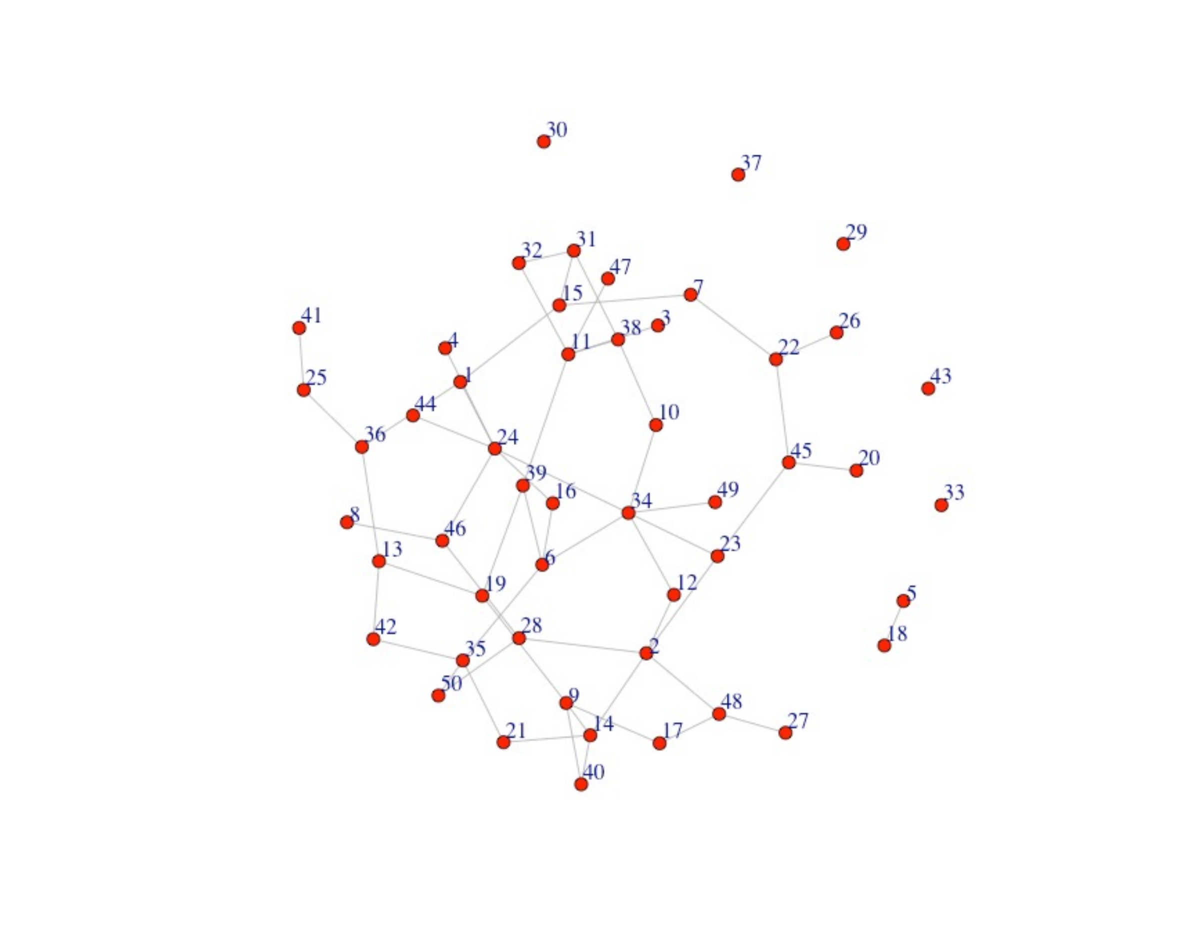
1. Erdös-Rényi Model
- Fixed number of vertices and edges.
- Any topology of graph is equally likely.

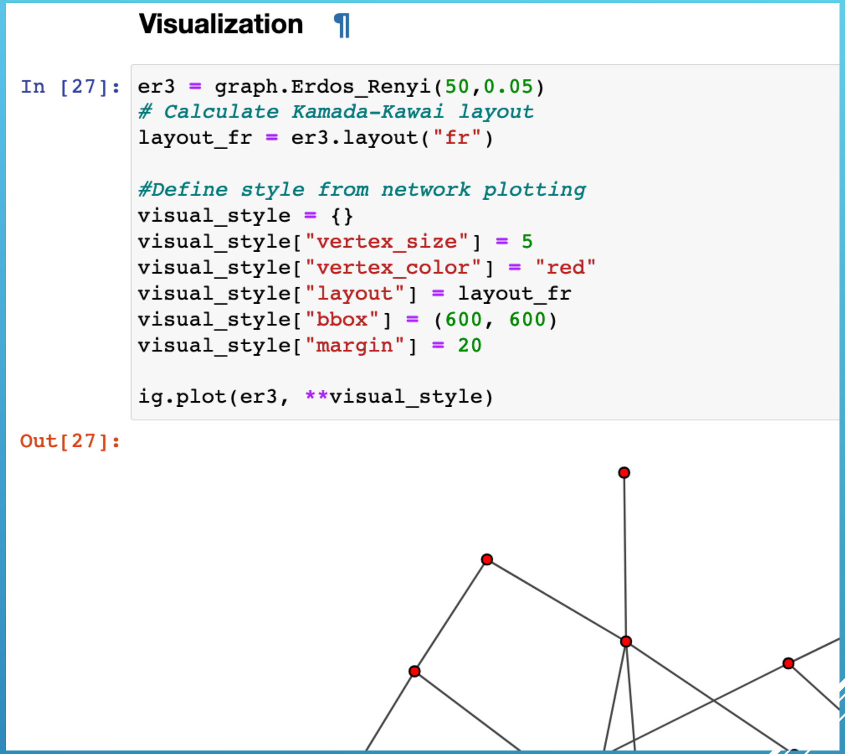
2. Watts-Strogatz Model
- Fixed number of vertices and edges in a regular ring lattice
- Rewiring of edges with given probability

3. Barabási-Albert Model
- The number of nodes increases over time
- Each new node comes with additional (*n*) edges
- The degree of existing nodes determines the probability to become target of additional edge

RANDOM NETWORK MODELS

Question 1: Which random model is used for this graph ?





Visualization ¶

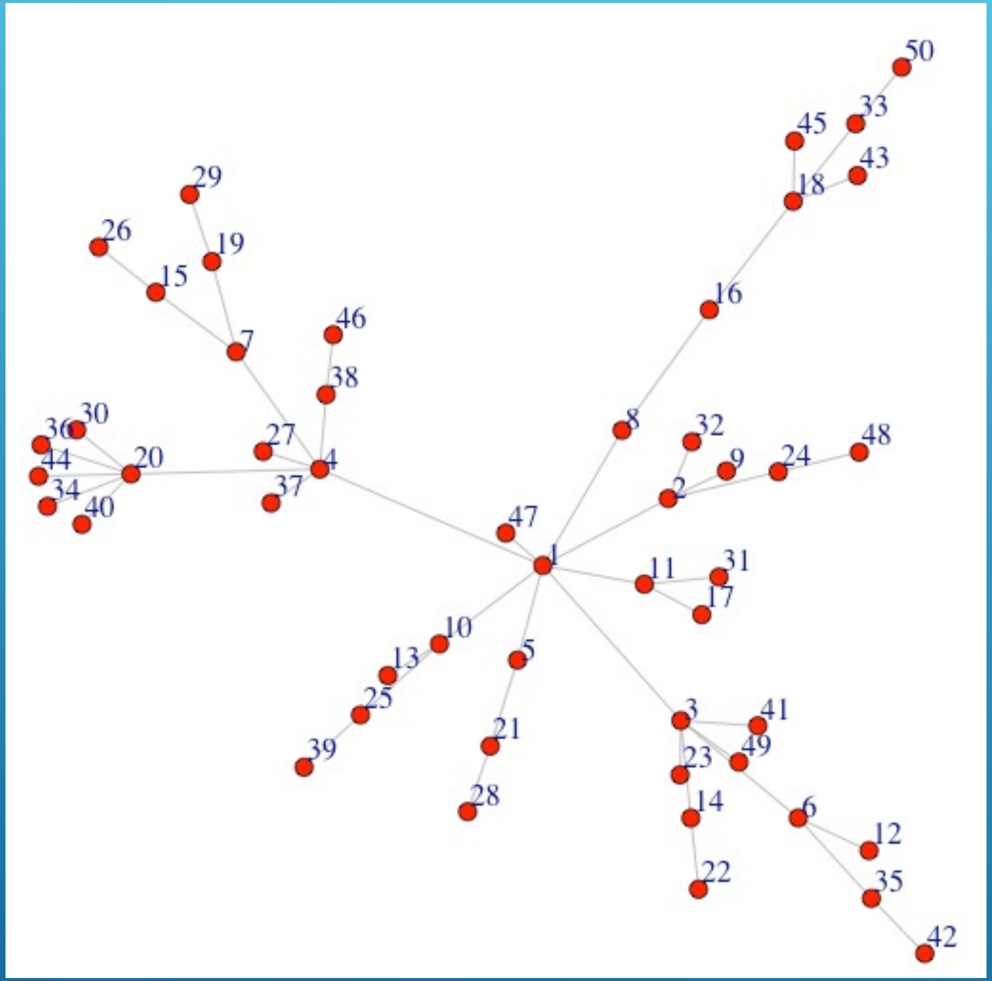
In [27]:
```
er3 = graph.Erdos_Renyi(50,0.05)
# Calculate Kamada-Kawai layout
layout_fr = er3.layout("fr")

#Define style from network plotting
visual_style = {}
visual_style["vertex_size"] = 5
visual_style["vertex_color"] = "red"
visual_style["layout"] = layout_fr
visual_style["bbox"] = (600, 600)
visual_style["margin"] = 20

ig.plot(er3, **visual_style)
```

Out[27]:

Erdös-Renyi network

http://www.networksciencebook.com/chapter/3#number-of-links

Question 2: Which random model is used for this graph ?

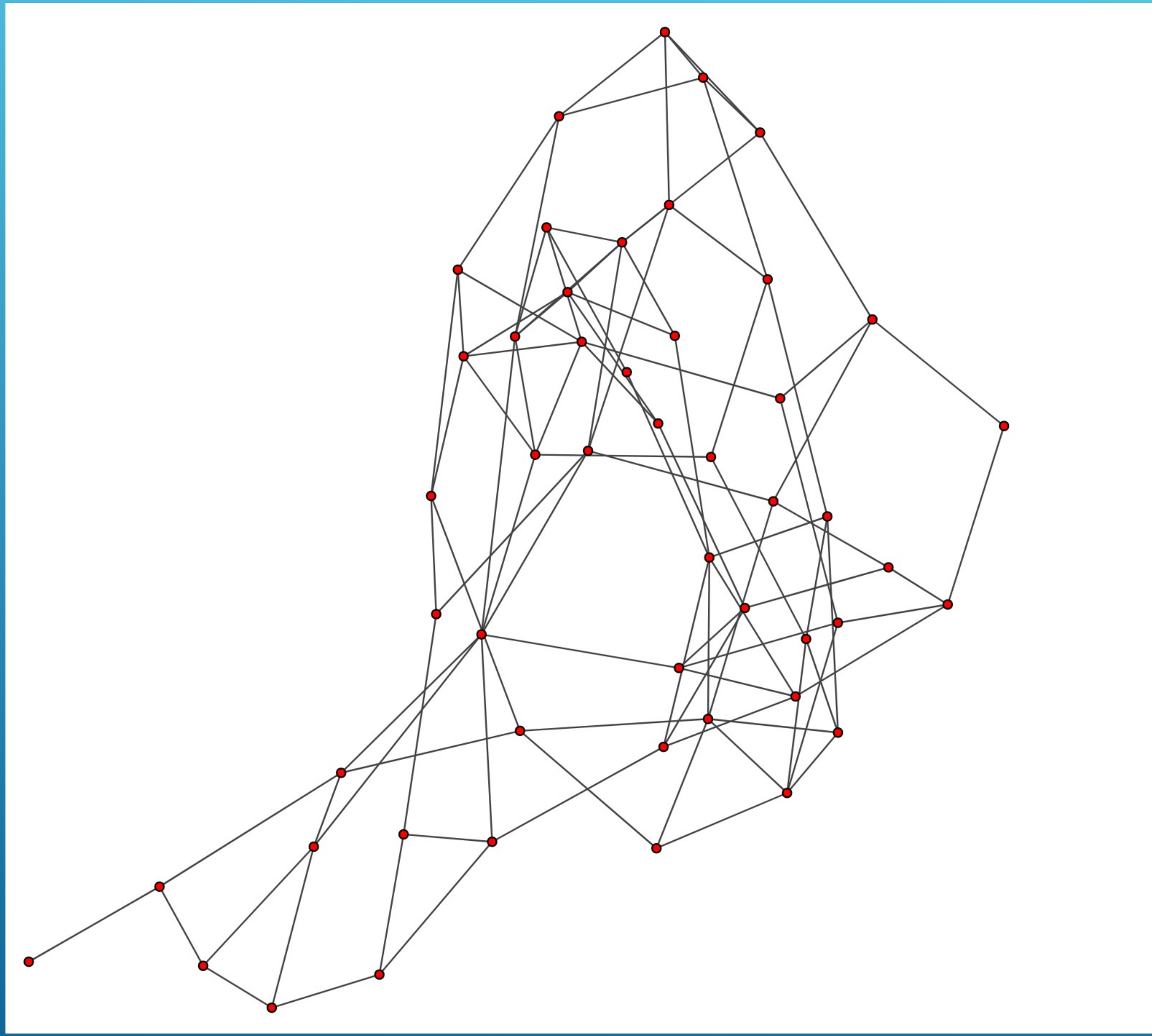



```
In [60]: ba3 = graph.Barabasi(50,1)
         # Calculate Kamada-Kawai layout
         layout_fr = ba3.layout("fr")

         #Define style from network plotting
         visual_style = {}
         visual_style["vertex_size"] = 5
         visual_style["vertex_color"] = "red"
         visual_style["layout"] = layout_fr
         visual_style["bbox"] = (600, 600)
         visual_style["margin"] = 20

         ig.plot(ba3, **visual_style)

Out[60]:
```

Barabasi-Albert network

http://www.networksciencebook.com/chapter/5#barabasi-model

Visualization of Watts-Strogatz network



click to expand output; double click to hide output

n [17]:

```
ws3 = graph.Watts_Strogatz(1,50,2,0.2)
# Calculate Kamada-Kawai layout
layout_fr = ws3.layout("fr")

#Define style from network plotting
visual_style = {}
visual_style["vertex_size"] = 5
visual_style["vertex_color"] = "red"
visual_style["layout"] = layout_fr
visual_style["bbox"] = (600, 600)
visual_style["margin"] = 20

ig.plot(ws3, **visual_style)
```

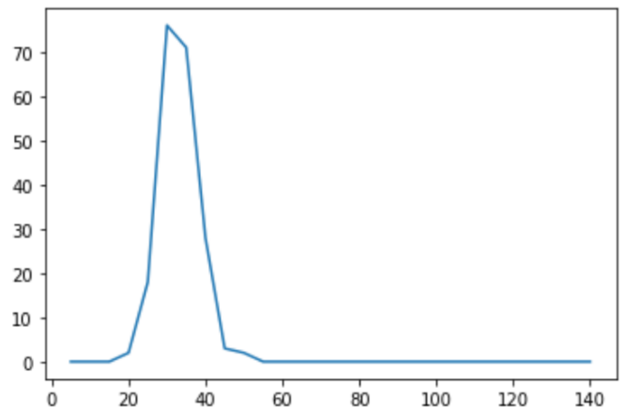
ut[17]:

http://www.networksciencebook.com/chapter/3#clustering-3-9

Question 3: Which two models can have similar degree distributions?

```
In [59]:
er_hist,er_bins=np.histogram(er2_deg,bins=np.linspace(0,140,29))
plt.plot(er_bins[1:],er_hist)
```

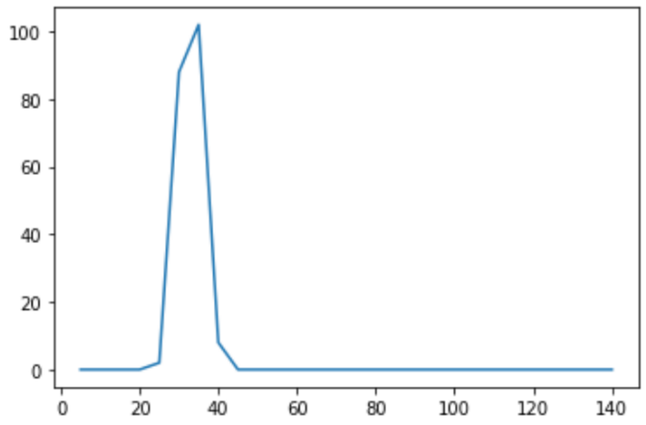
```
Out[59]: [<matplotlib.lines.Line2D at 0x120a8ff50>]
```



Erdös-Renyi network

```
In [24]: ws_hist,ws_bins=np.histogram(ws2_deg,bins=np.linspace(0,140,29))
         plt.plot(ws_bins[1:],ws_hist)
```

```
Out[24]: [<matplotlib.lines.Line2D at 0x11f951c50>]
```



Watts-Strogatz network

1. Erdös-Rényi Model

- Binomial with parameter p

$$P(\deg(v) = k) = \binom{n-1}{k} p^k (1-p)^{(n-1-k)}$$

- For large n, distribution becomes a Poisson distribution

$$P(\deg(v) = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

2. Watts-Strogatz Model

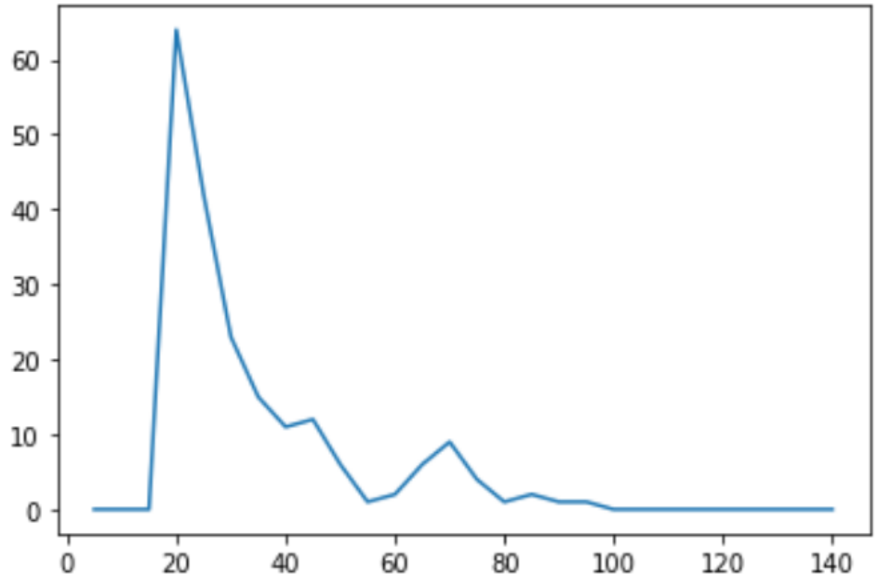
- Distribution becomes Poisson when rewiring probability close to 1

DEGREE DISTRIBUTION

Question 4: What is the degree distribution of a BA-model?

In [27]:
```
ba_hist,ba_bins=np.histogram(ba2_deg,bins=np.linspace(0,140,29))
plt.plot(ba_bins[1:],ba_hist)
```

Out[27]: [<matplotlib.lines.Line2D at 0x11fa90790>]

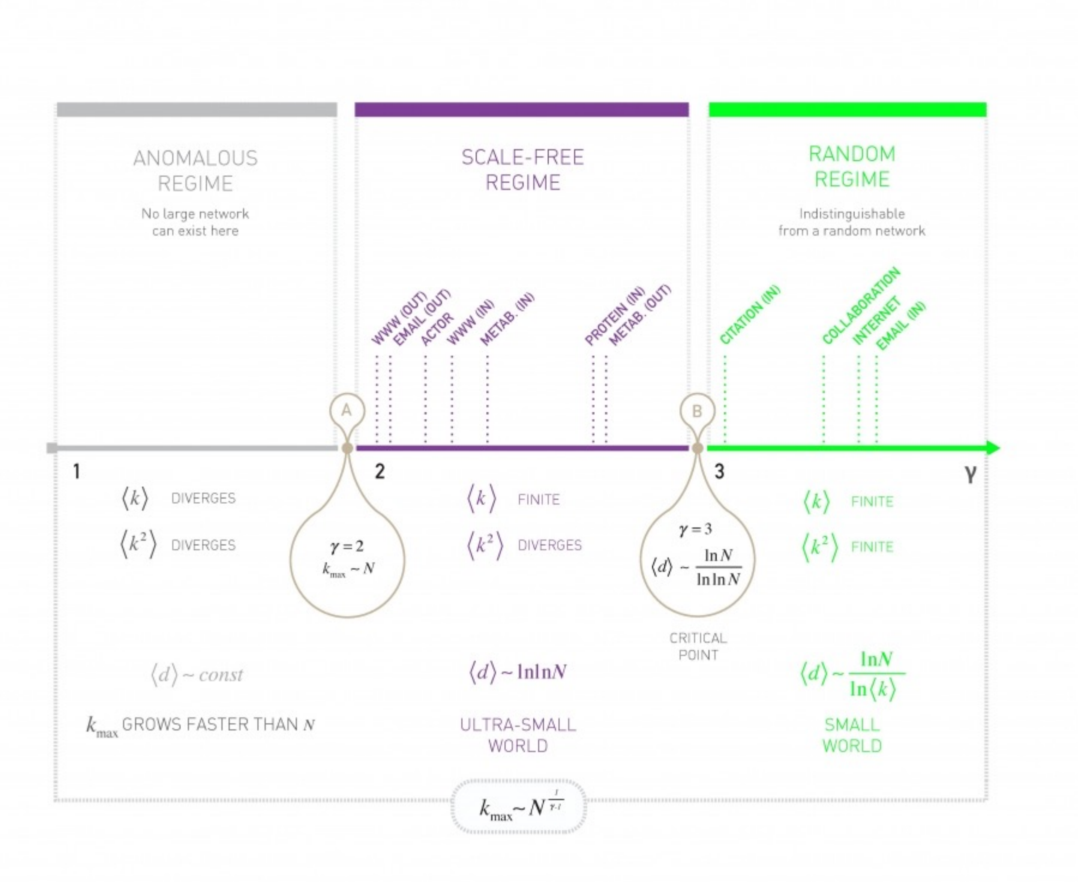


3. Barabasi-Albert Model

- Scale free, power law

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

For many real-world networks: $\lambda \approx 3$



See: http://www.networksciencebook.com/chapter/4#degree-exponent

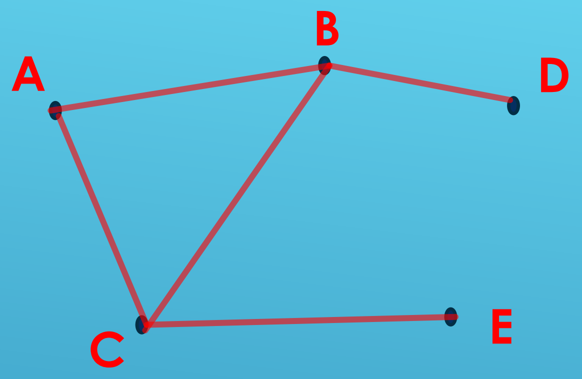
DEGREE DISTRIBUTION

Go to http://pollev.com/bartthijs

Question 5: Which model shows the highest clustering?

- Global clustering (Transitivity)

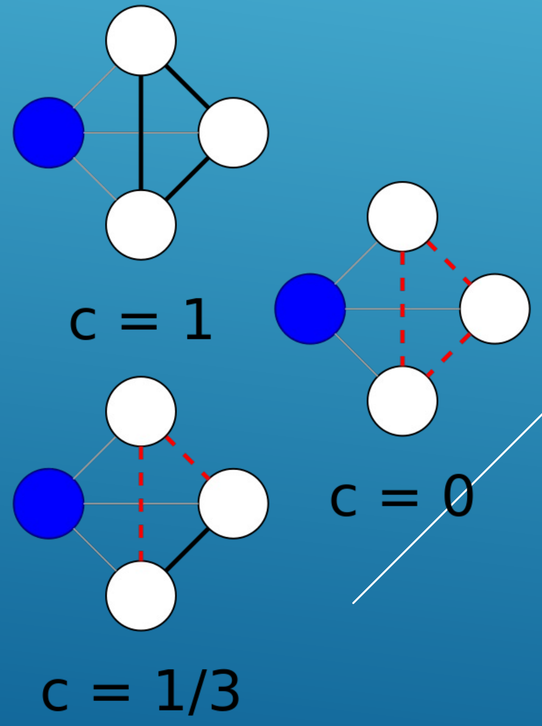
$$C = \frac{3 * number\ of\ trianles}{number\ of\ connected\ triplets}$$



Triangle = ABC
Connected triplet= A-B-C, B-A-C, A-C-B, C-B-D, A-B-D, A-C-E, B-C-E

- Local clustering

$$C = \frac{2\left|\left\{e_{jk}\colon v_j, v_k \in N_i, e_{jk} \in E\right\}\right|}{d_i(d_i - 1)}$$

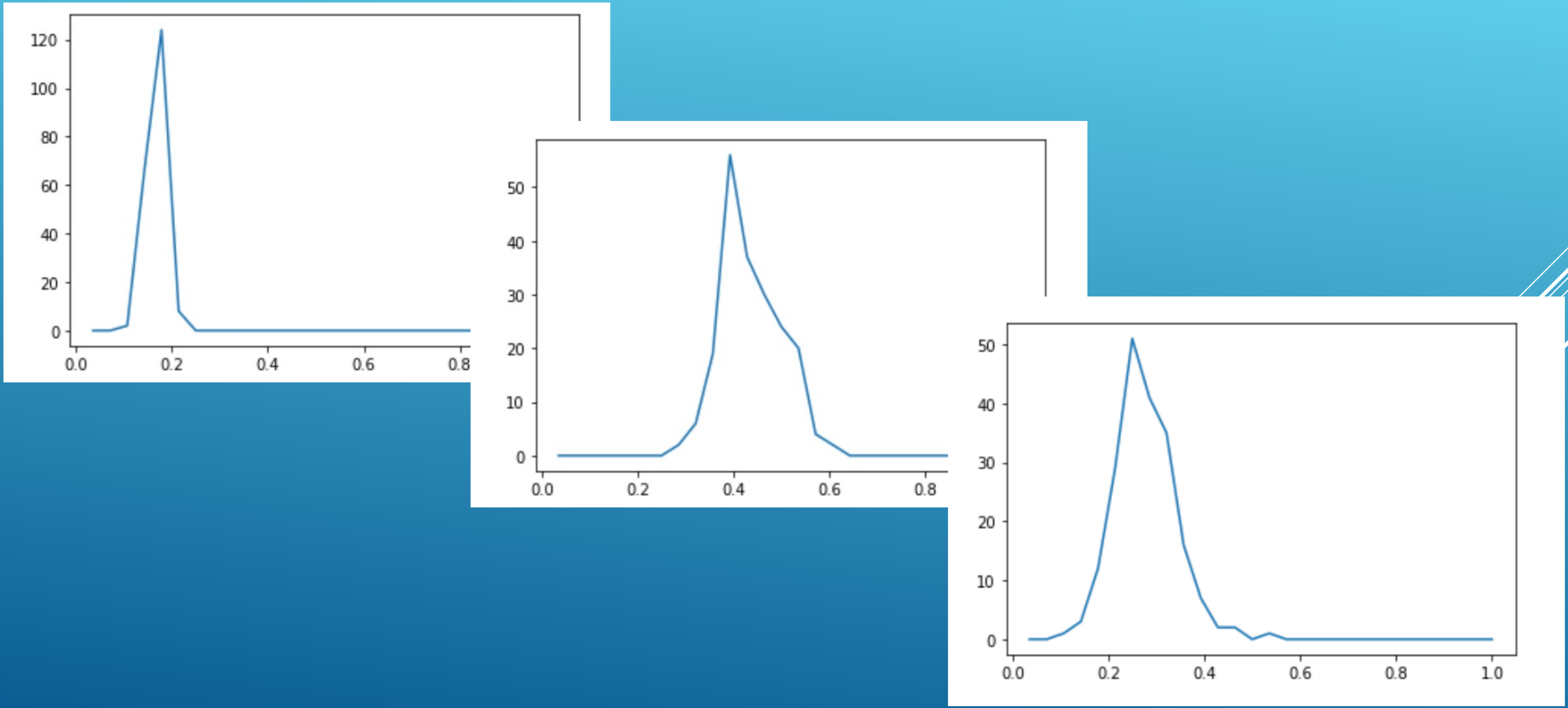


CLUSTERING COEFFICIENT

Question 5: Which model shows the highest clustering?

```
In [32]: er_t = er2.transitivity_undirected()
         print("Global Clustering Coefficient of Erdös-Renyi network: %5.3f \n" %(er_t ))

         ws_t = ws2.transitivity_undirected()
         print("Global Clustering Coefficient of Watts-Strogatz network: %5.3f \n" %(ws_t ))

         ba_t = ba2.transitivity_undirected()
         print("Global Clustering Coefficient of Barabasi-Albert network: %5.3f \n" %(ba_t ))

         Global Clustering Coefficient of Erdös-Renyi network: 0.151

         Global Clustering Coefficient of Watts-Strogatz network: 0.425

         Global Clustering Coefficient of Barabasi-Albert network: 0.260
```

Distribution of local clustering



Go to http://pollev.com/bartthijs

Question 6: Which model shows the shortest path lengths?

Question 6: Which model shows the shortest path lengths?

1. Erdös-Rényi Model
- This property can be present in the ER-model if connected

2. Watts-Strogatz Model
- Property was introduced by Watts & Strogatz
- The ring lattice shows both short path lengths and high clustering. Too much rewiring might disconnect graph

3. Barabasi-Albert Model
- Has the lowest shortest path length due to hubs

SMALL WORLD PROPERTY

Evolution of the Social Network of Scientific Collaboration

Physica A, 2002, 311, 590-614

Barabasi, Jeong, Néda, Ravasz, Schubert & Vicsek

RANDOM NETWORK PROPERTIES IN REAL WORLD NETWORKS

Data Set:

Co-authorship network of scientists in Mathematics and in Neuro Science.

Based on articles published between 1991 and 1998

71K Authors in M and 209K authors in NS

Evolution=> New authors enter the field and publish their first paper after 1991

SCIENTIFIC COLLABORATION

5 Questions on Toledo, Discuss in a small group of peers:

1. The properties of the degree distributions of both networks
2. The properties of the shortest path length
3. The properties of global and local clustering
4. The relation between the evolution of degree, separation and clustering coefficient.
5. What is the role of preferential attachement?

SCIENTIFIC COLLABORATION

Degree distribution

Both networks have scale-free properties:

$p_k \sim k^{-\gamma}$

For Mathematics:
$\gamma_M = 2.4$

For Neuro-Science:
$\gamma_M = 2.1$

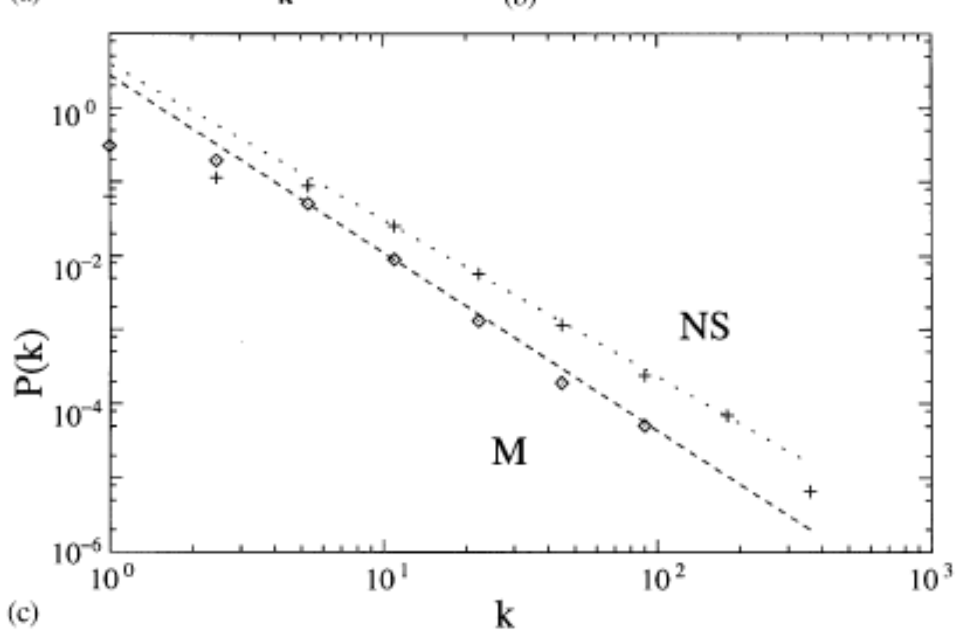


Fig. 2. Degree distribution for the (a) M and (b) NS database, showing the data based on the cumulative results up to years 1993 (×) and 1998 (●). (c) Degree distribution shown with logarithmic binning computed from the full dataset cumulative up to 1998. The lines correspond to the best fits, and have the slope 2.1 (NS, dotted) and 2.4 (M, dashed).

SCIENTIFIC COLLABORATION

Small World Properties

Surprisingly, the separation is decreasing over time.

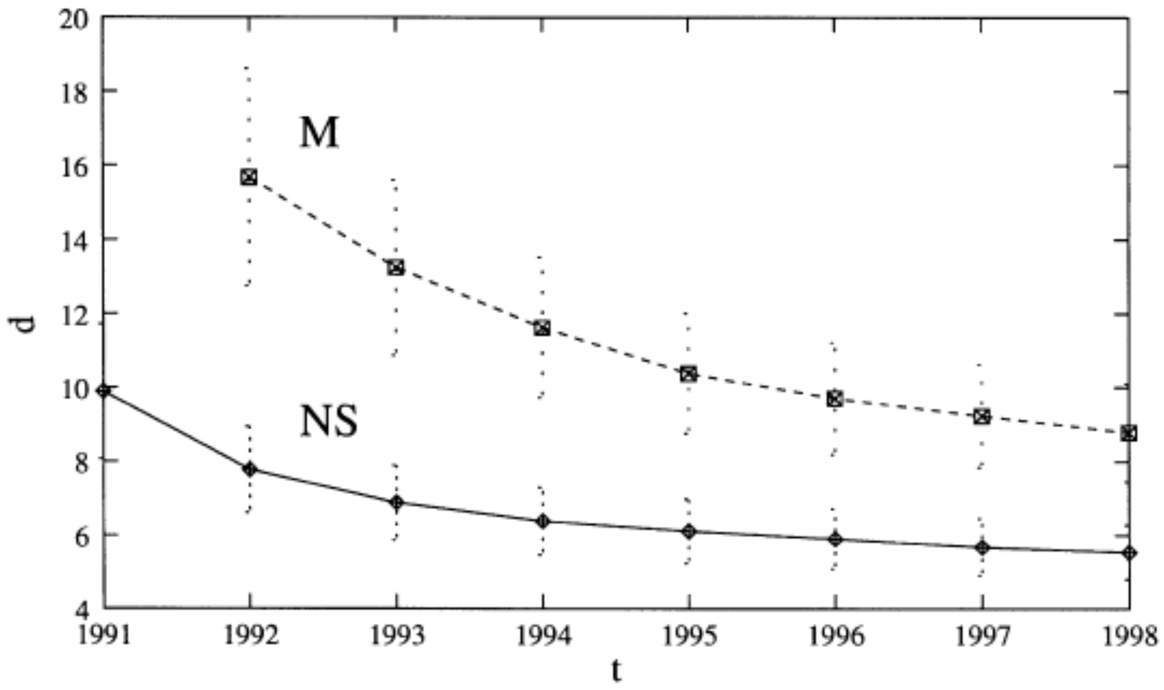


Fig. 3. Average separation in the M and NS databases. The separation is computed on the cumulative data up to the indicated year. The error bars indicate the standard deviation of the distances between all pairs of nodes.

SCIENTIFIC COLLABORATION

Clustering Coefficient

This coefficient is decaying over time.

The **smaller separation** of NS is in line with **the higher clustering coefficient** of NS.

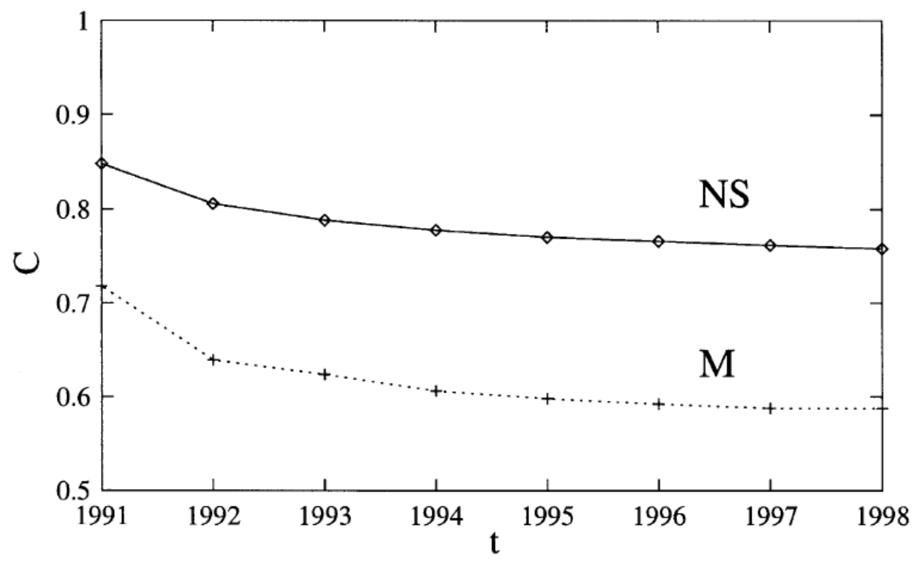
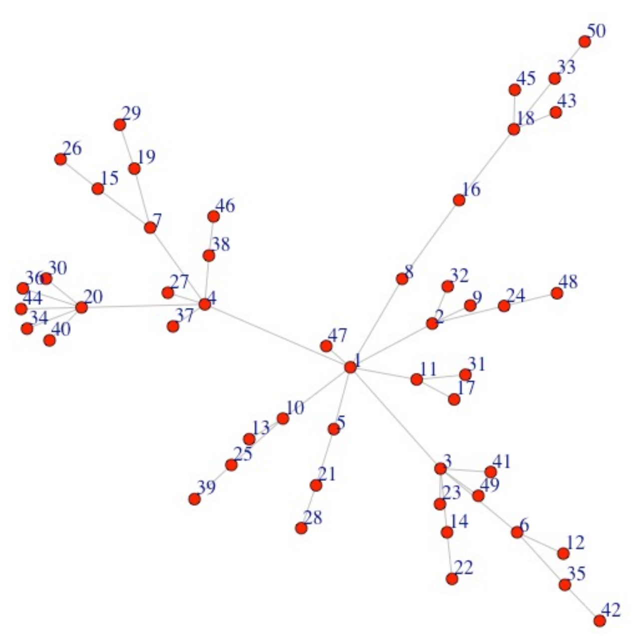


Fig. 4. Clustering coefficient of the M and NS database, determined for the cumulative data up to the year indicated on the *t*-axis.

SCIENTIFIC COLLABORATION

How is it possible that this indicator is decaying while the separation is decreasing as well?



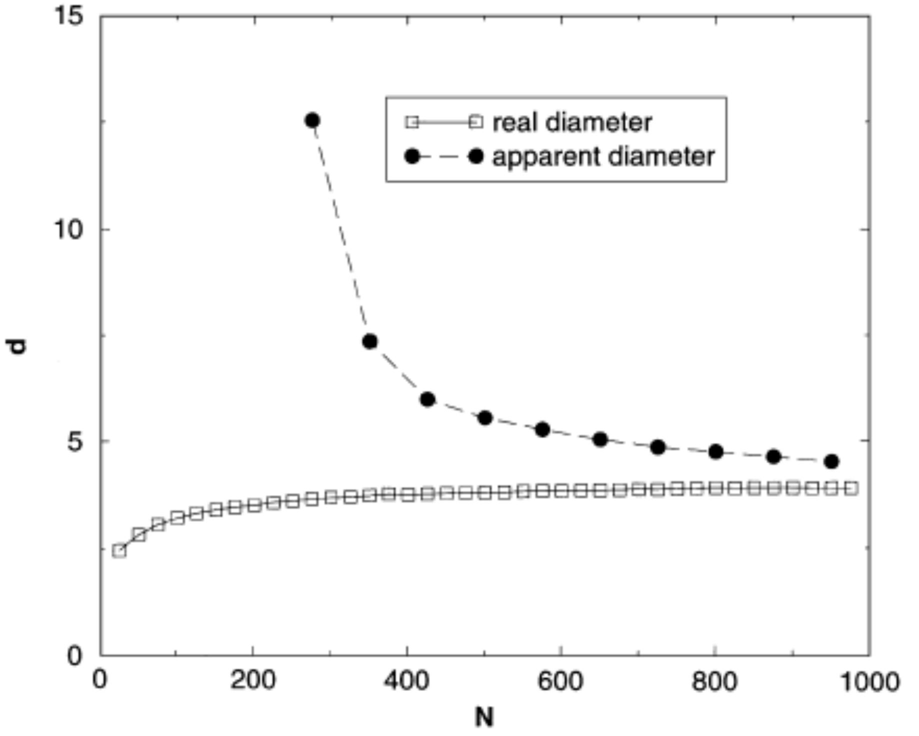


Fig. 12. Computer simulated dynamics for the real and apparently measured diameter value ($N_{max} = 1000$, $a = 0.001$, $\beta = 1$, $b = 2$ and $N_s = 200$).

SCIENTIFIC COLLABORATION

Clustering Coefficient

This coefficient is decaying with growing network.

But, it reaches a minimum when edges between existing nodes are possible. Beyond the minimum clustering coefficient starts to increase as size grows (nodes and edges).

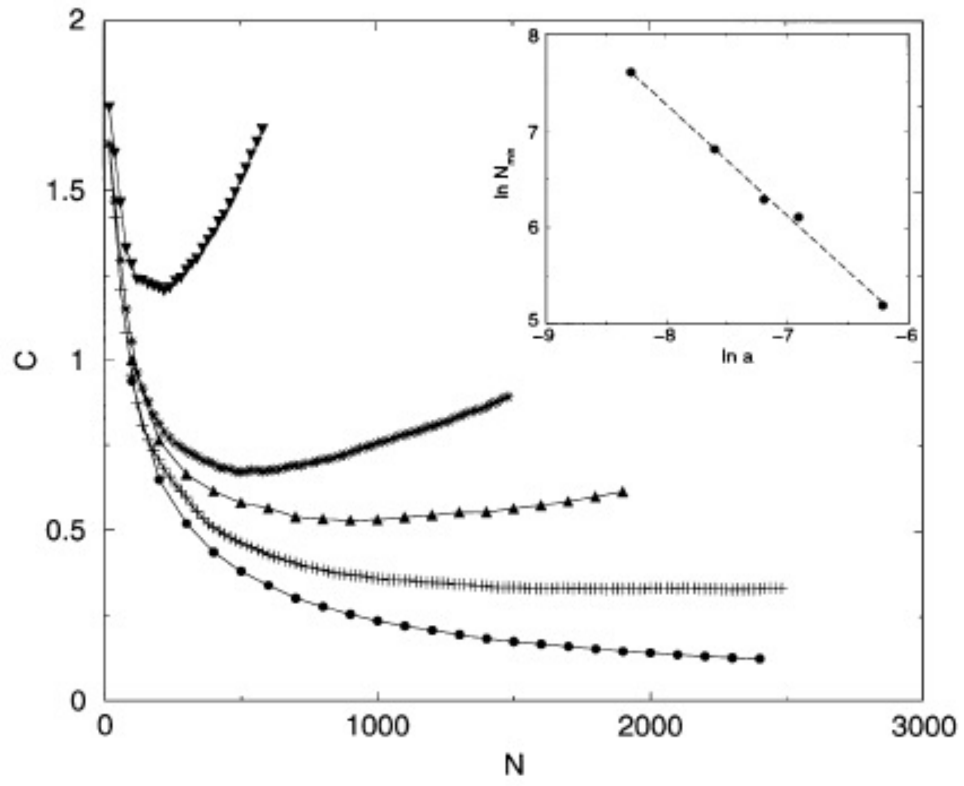


Fig. 13. Clustering coefficient for different values of the a parameter as a function of the system size N. ($N_{max} = 1000$, $\beta = 1$ and $b = 2$, values of a are 0 (●), 0.00025 (+), 0.0005 (▲), 0.00075 (∗), 0.002 (▼).)

SCIENTIFIC COLLABORATION

Connected Component

Both data sets have one giant connected component when considering the complete network 1991-1998.

Is this something we can expect?

SCIENTIFIC COLLABORATION

Preferrential Attachment

In a scale free network, the probability that the degree changes is dependent of the degree

$$\Pi(k) \sim k^{v}$$

v = 0.8 for M and *v=0.75* for NS

SCIENTIFIC COLLABORATION

Preferrential Attachment

The probability that two existing nodes get connected in the collaboration network depends on the degree of both authors.

$$\Pi(k_1, k_2) = \frac{N(k_1, k_2)}{D(k_1, k_2)}$$

SCIENTIFIC COLLABORATION

Preferrential Attachment

3 models explaining Preferrential Attachment

- Global model: All nodes have same information and calculate the same (similar) cost function
- Local Model 1: Nodes randomly pick edges to follow and select the target node of their connection.
- Local Model 2: Nodes pick random node to connect to with probability *p* or connects to one of its neighbours with probability *(1-p)*

SCIENTIFIC COLLABORATION

1. Erdös-Rényi Model
2. Watts-Strogatz Model
3. Barabási-Albert Model

4. Bianconi-Barabási Model

- Growth: new node *j* added to network
- Each new node comes with additional (*n*) edges
- Preferrential attachment
- Each new node has a fixed fitness so that:

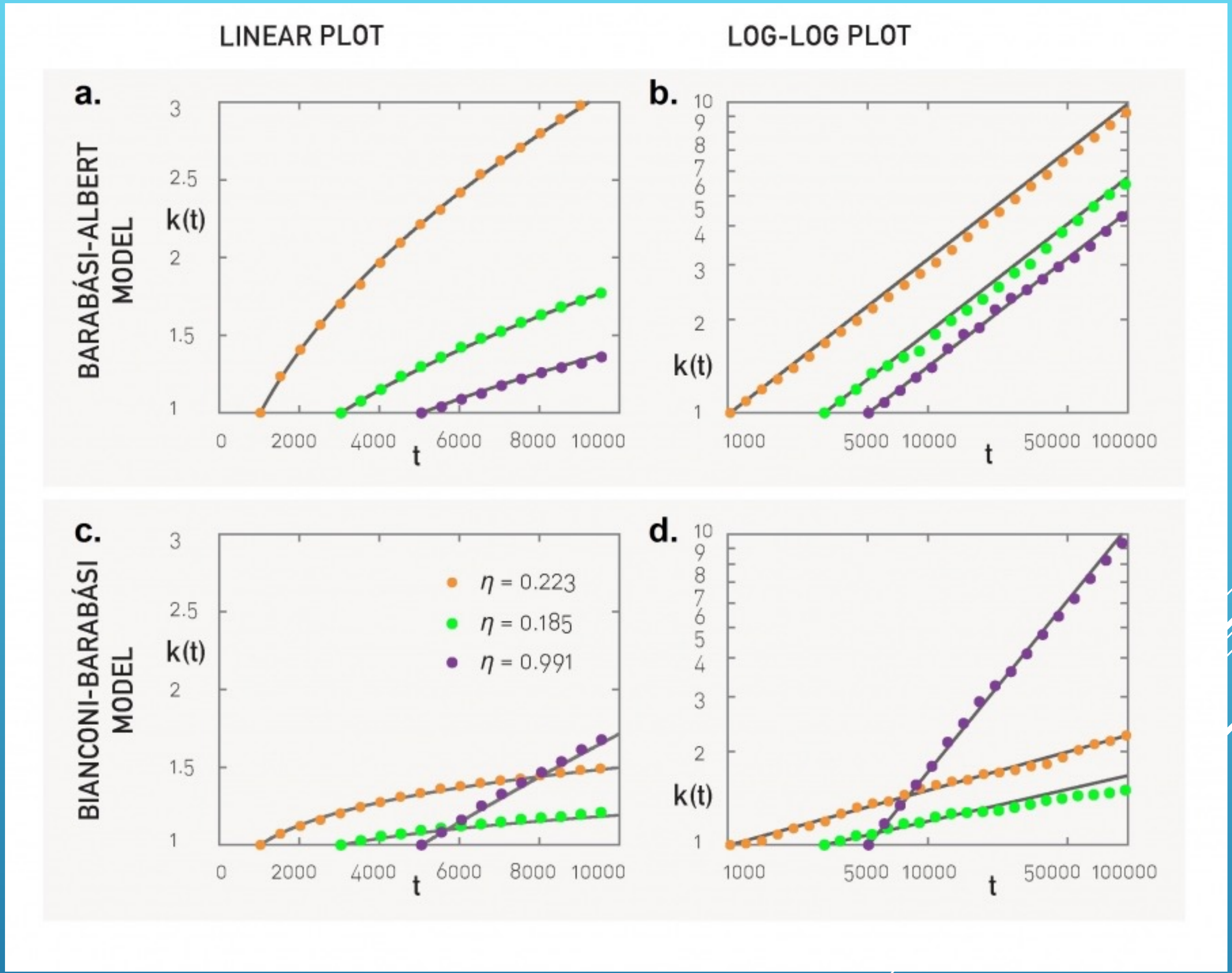
$$\prod_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

http://networksciencebook.com/chapter/6#bianconi-model

RANDOM NETWORK MODELS

Bianconi-Barabási Model

- Competition:
  New nodes can become more 'important' than older nodes
- Multiplication of fitness and preferential attachment



RANDOM NETWORK MODELS