Second-order mixing in networks

Shi Zhou

University College London

Dr. Shi Zhou *('Shee Joe')*

- Lecturer of Department of Computer Science, University College London (UCL)
- Holder of The Royal Academy of Engineering/EPSRC Research Fellowship
- s.zhou@cs.ucl.ac.uk

Outline

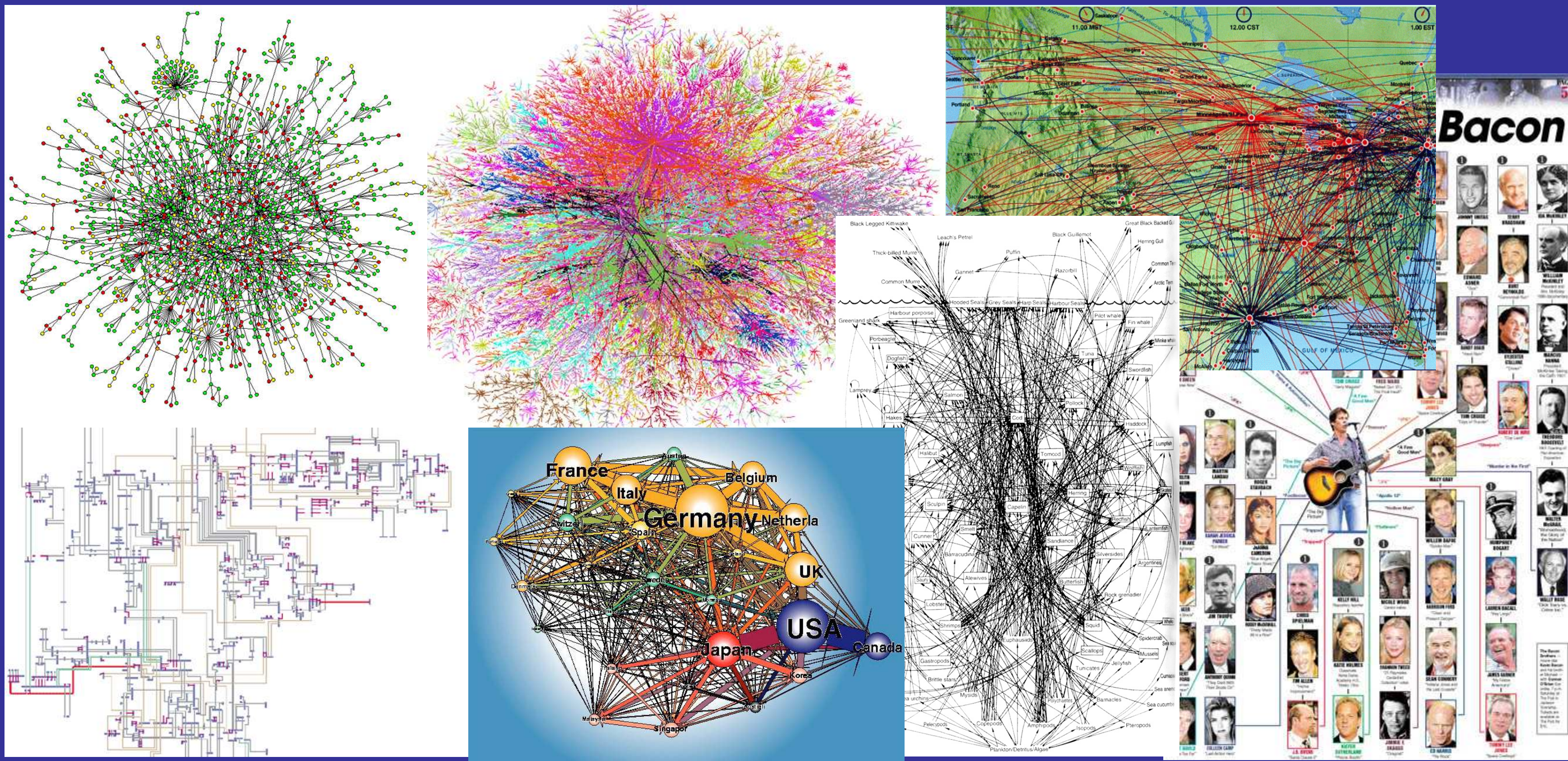
- Part 1.
  - Power-law degree distribution
  - (1st-order) assortative coefficient
  - Rich-club coefficient
    - Collaborated with Dr. Raul Mondràgon, Queen Mary University of London (QMUL).
- Part 2. Second-order mixing in networks
  - Ongoing work collaborated with
    - Prof. Ingemar Cox, University College London (UCL)
    - Prof. Lars K. Hansen, Danish Technical University (DTU).

Part 1

❑Power-law degree distribution, assortative mixing and rich-club

Complex networks

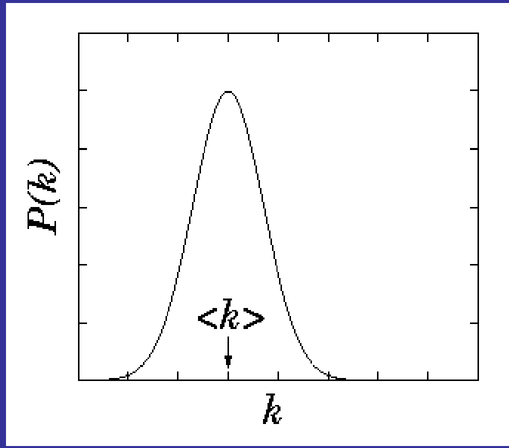
- Heterogeneous, irregular, evolving structure

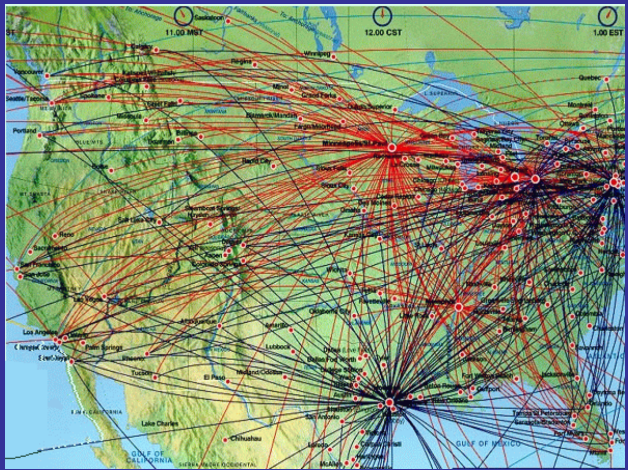


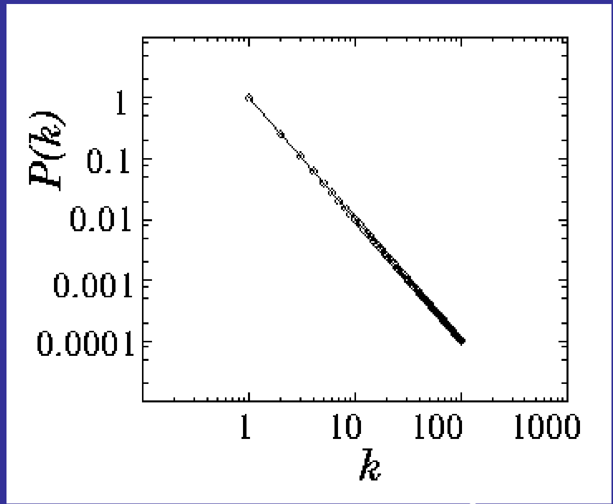
Degree

- **Degree, *k***
  - **Number of links a node has**
- **Average degree**
  - **$\langle k \rangle = 2L / N$**
- **Degree distribution, *P(k)***
  - **Probability of a node having degree *k***

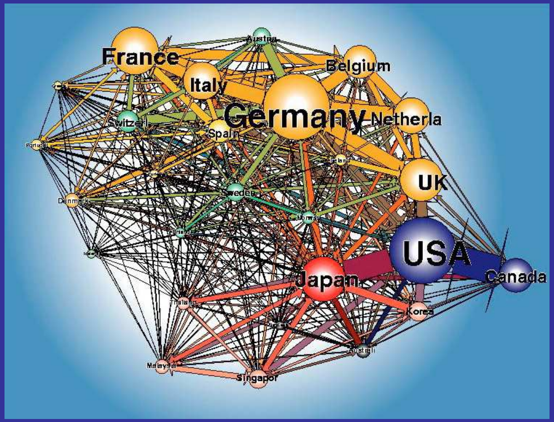
Poisson vs Power-law distributions



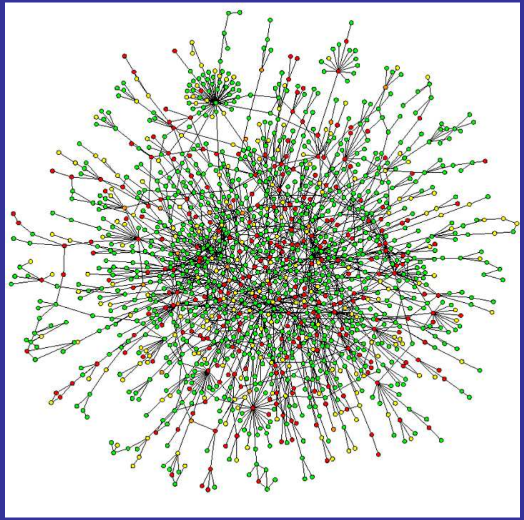




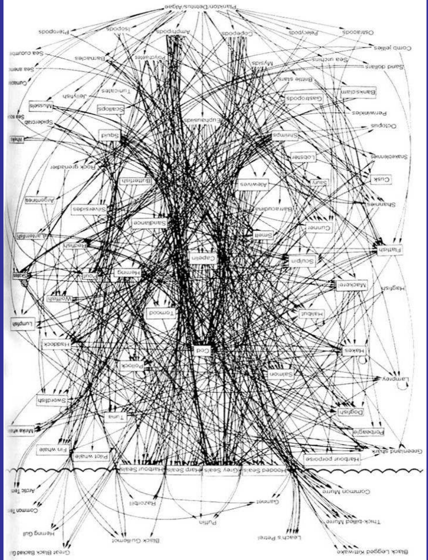
❑Power law (or scale-free) networks are everywhere



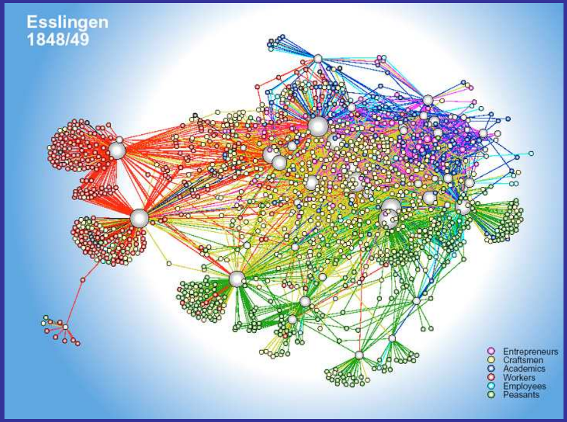
Trading networks



Yeast protein network



Food web



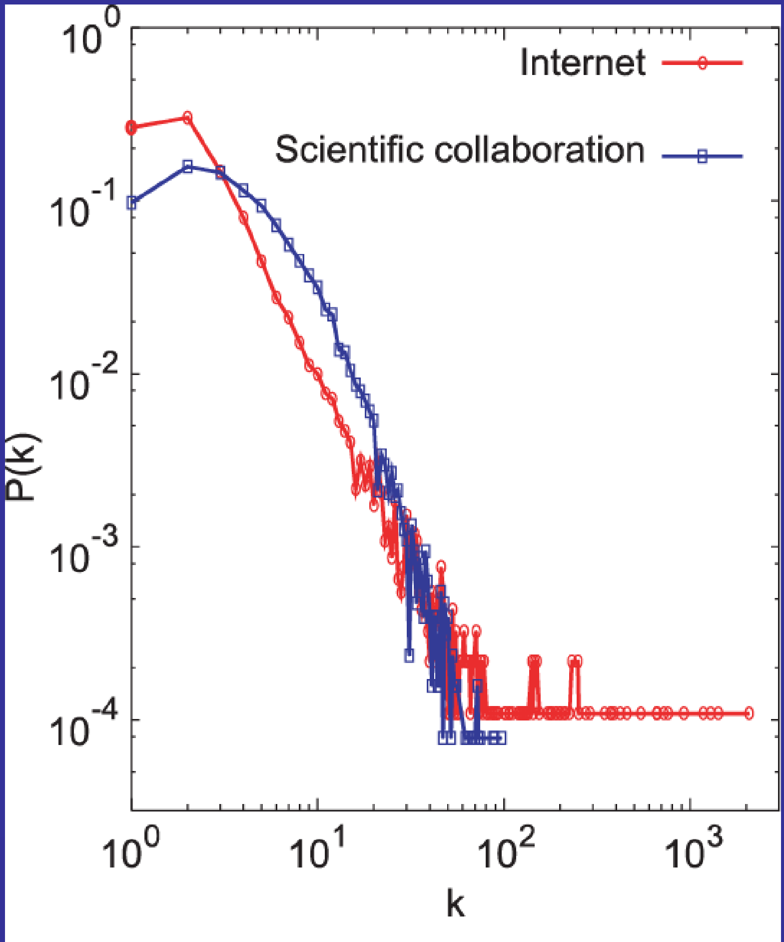
Business networks

Here we study two typical networks

- ❑Scientific collaborations network in the research area of condense matter physics
  - ➢ Nodes: 12,722 scientists
  - ➢ Links: 39,967 coauthor relationships

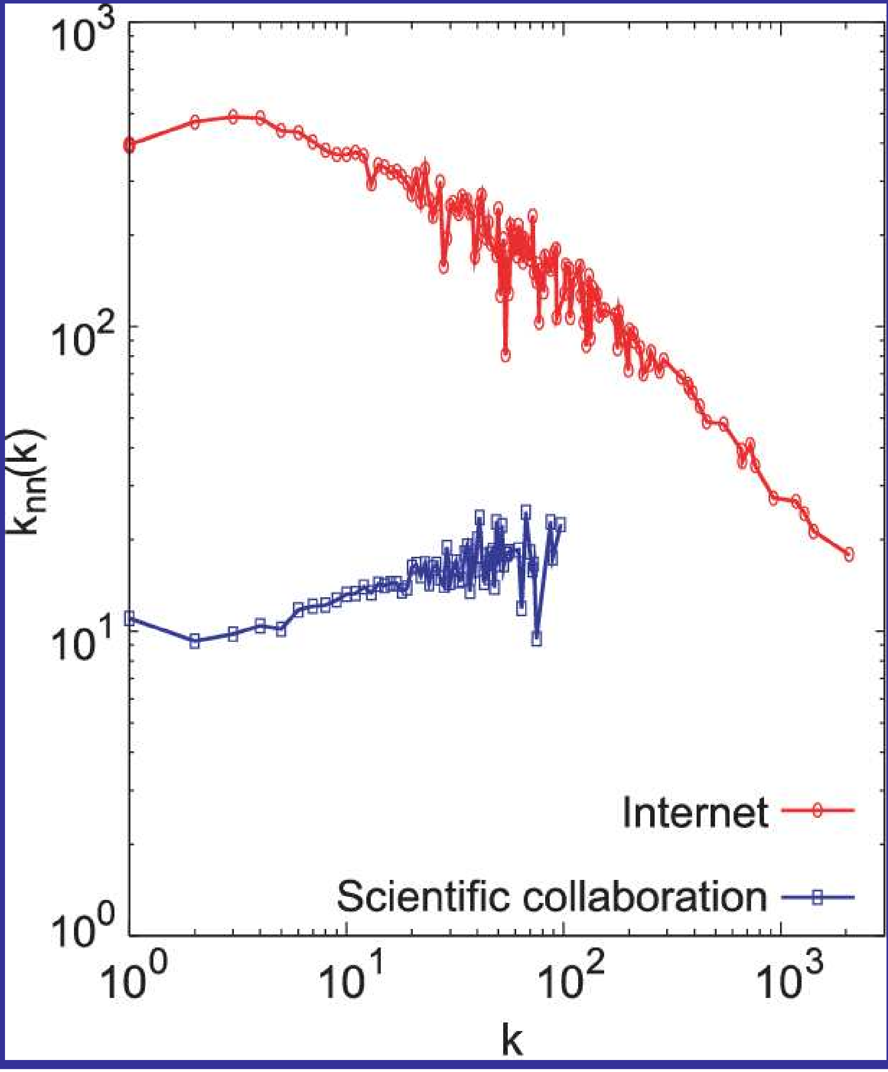
- ❑The Internet at autonomous systems (AS) level
  - ➢ Nodes: 11,174 Internet service providers (ISP)
  - ➢ Links: 23,409 BGP peering relationship

Degree properties



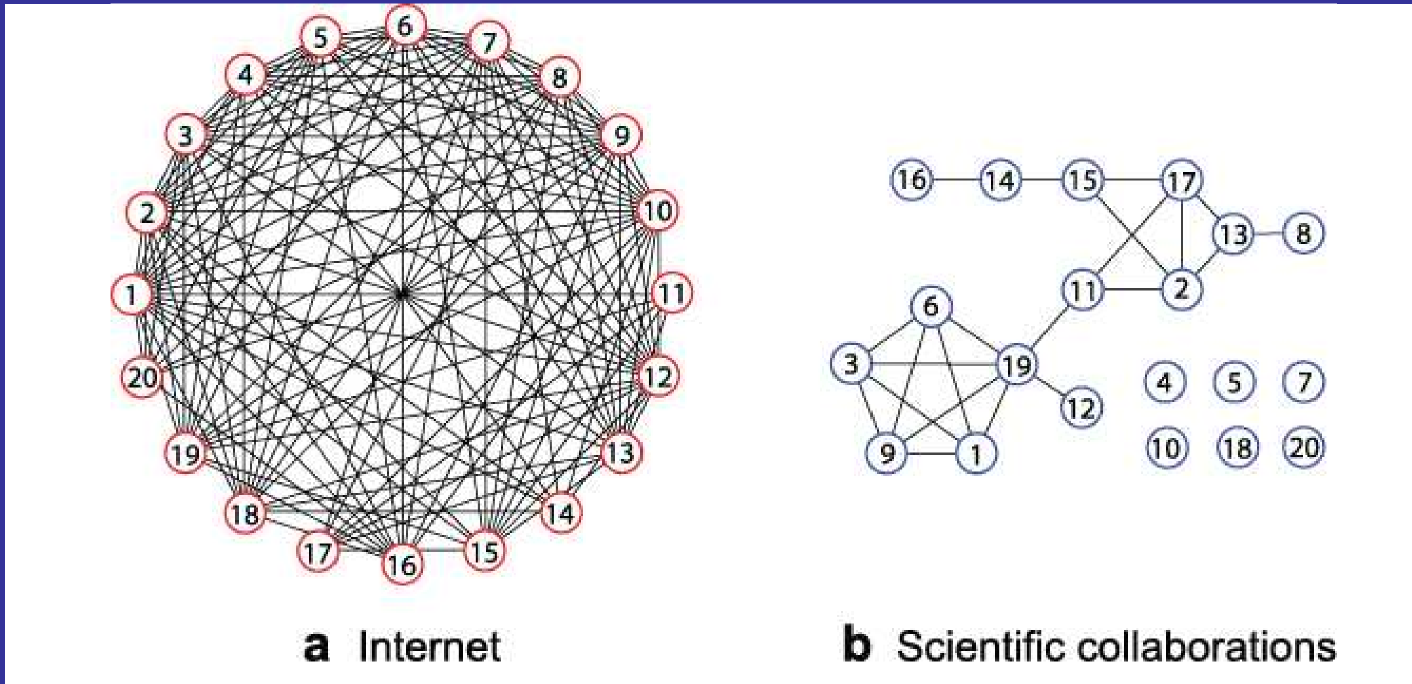
- Average degree is about 6
- Sparsely connected
  - $^{L}/_{N(N-1)/2}$ <0.04%
- Degree distribution
  - Non-strict power-law
- Maximal degree
  - 97 for scientist network
  - 2389 for Internet

Mixing pattern: who connects with whom?



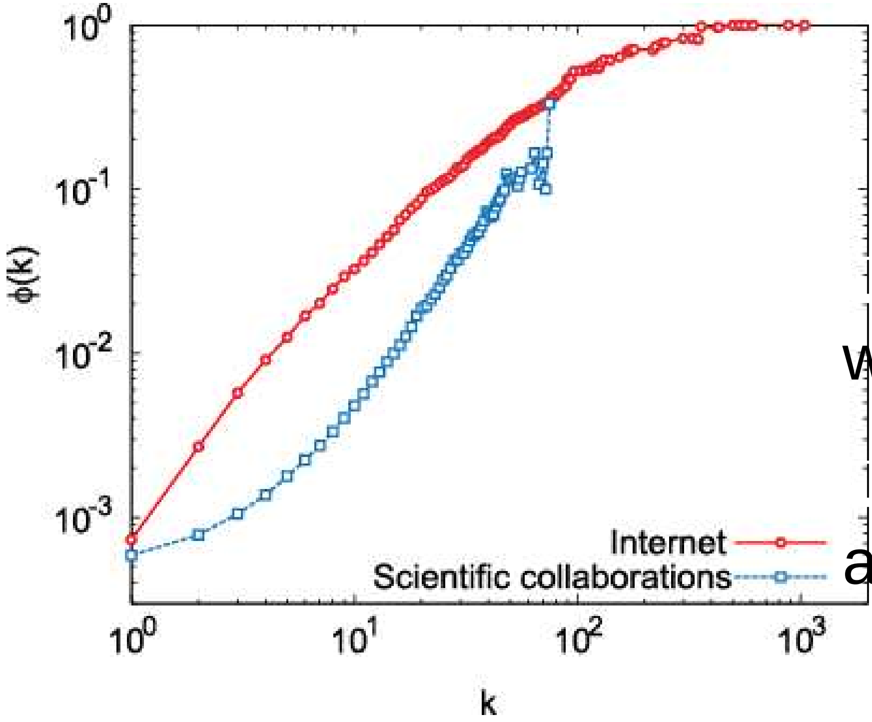
- Correlation between degrees of the two end nodes of a link
  - Assortative coefficient r
  - Neigbhours average degree vs degree
- Assortative mixing
  - Scientific collaboration
  - r = 0.161
- Disassortative mixing
  - Internet
  - r = -0.236

Rich-club



Connections between the top 20 best-connected nodes themselves

Rich-club coefficient



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

$N_{>k}$ is the number of nodes with degrees > k.

$E_{>k}$ is the number of links among the $N_{>k}$ nodes.

Summary

	Scientist network	Internet
Degree distribution	Power-law	Power-law, extra long tail
Mixing pattern	Assortative,	Disassortative
Rich nodes	Sparsely connected, no rich-club	Tightly interconnected, rich-club

Discussion 1

- Mixing pattern and rich-club are not trivially related
  - Mixing pattern is between TWO nodes
  - Rich-club is among a GROUP of nodes.
  - Each rich node has a large number of links, a small number of which are sufficient to provide the connectivity to other rich nodes whose number is small anyway.
- These two properties together provide a much fuller picture than degree distribution alone.

Discussion 2

- ❑Why is the Internet so ‘small’ in terms of routing efficiency?
  - ➢ Average shortest path between two nodes is only 3.12
- ❑This is because
  - ➢ Disassortative mixing: poorly-connected, peripheral nodes connect with well-connected rich nodes.
  - ➢ Rich-club: rich nodes are tightly interconnected with each other forming a core club.
    - ❖ shortcuts
    - ❖ redundancy

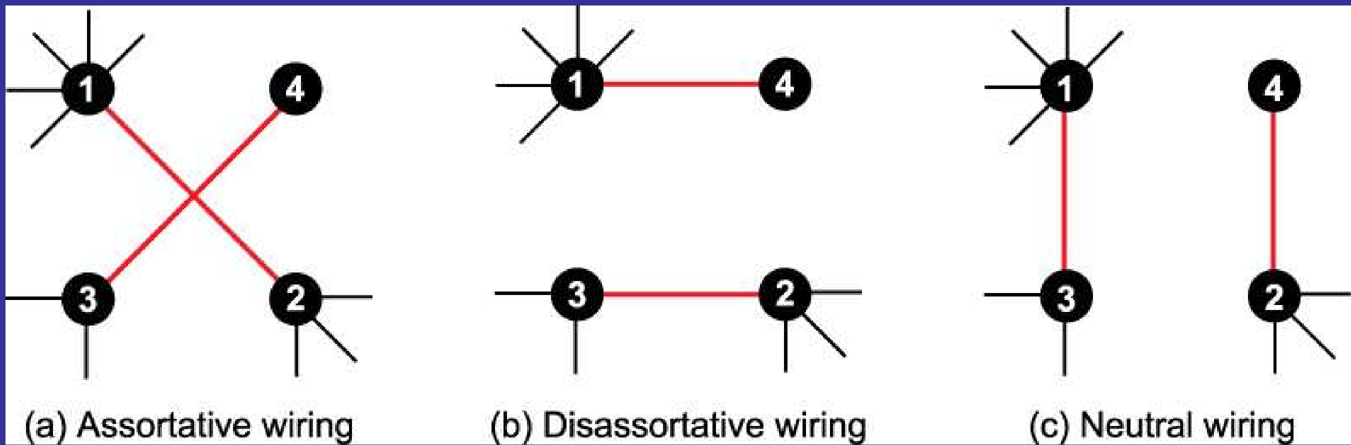
Jellyfish model

Discussion 3 (optional)

- How are the three properties related?
  - Degree distribution
  - Mixing pattern
  - Rich-club

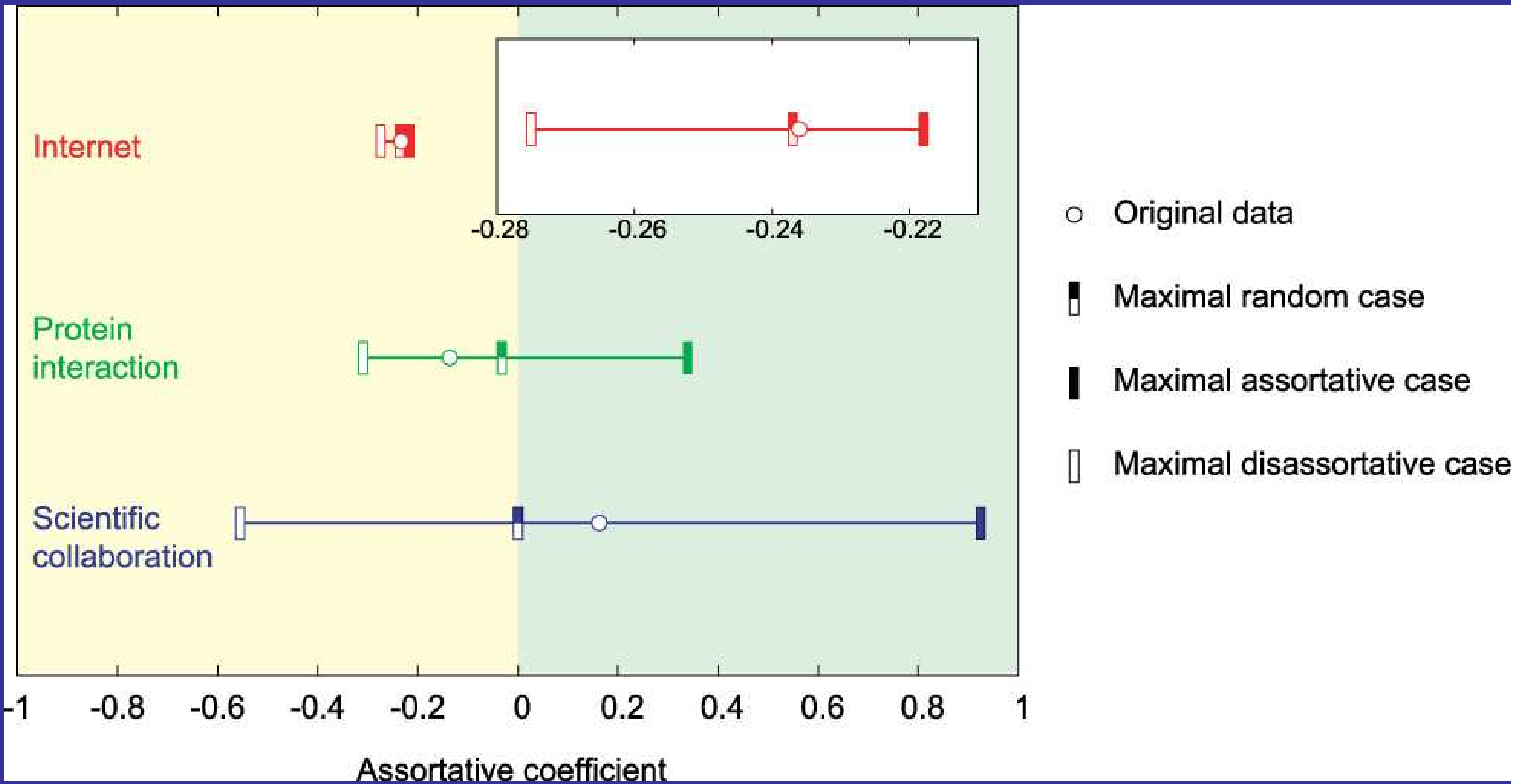
- We use the link rewiring algorithms to probe the inherent structural constraints in complex networks.

Link-rewiring algorithm

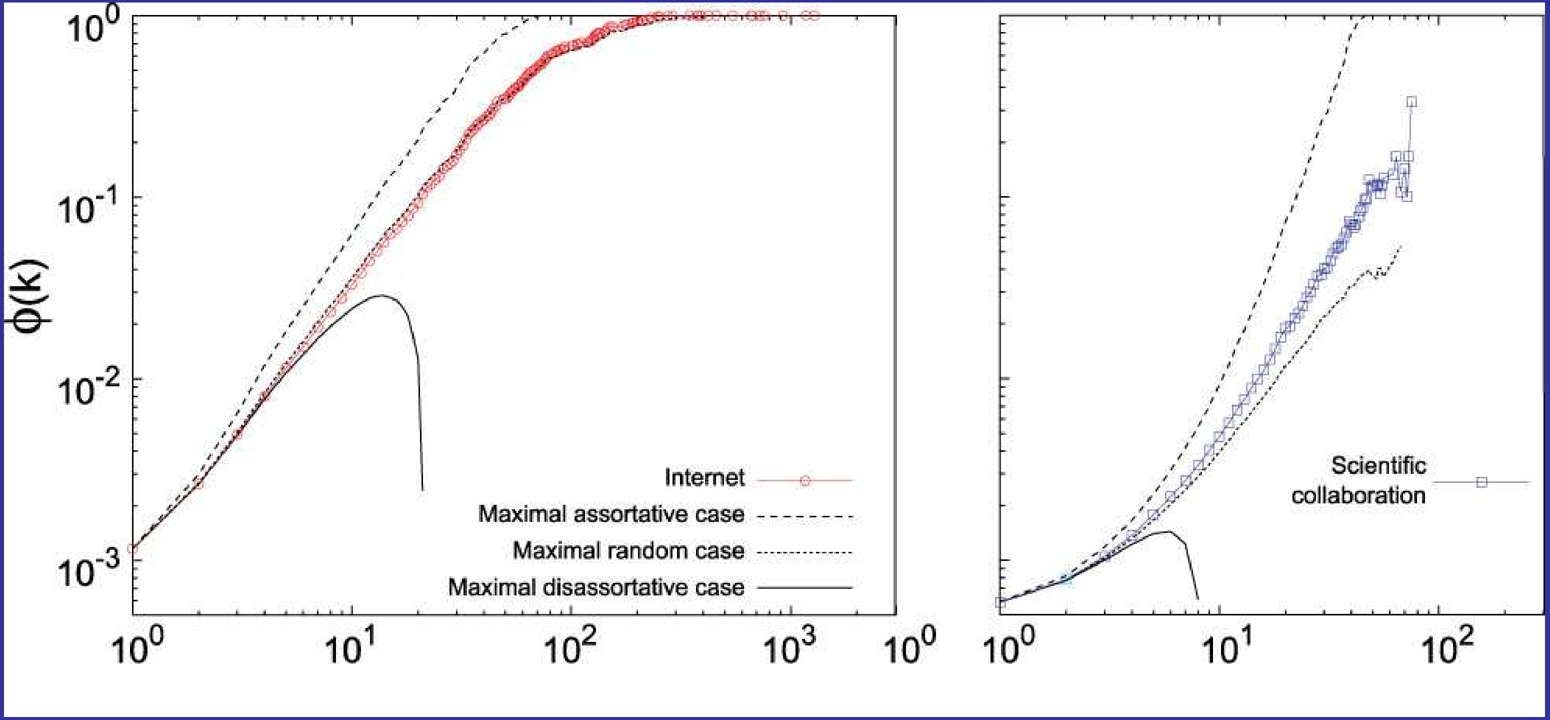


- Each node’s degree is preserved
- Therefore, surrogate networks is generated with exactly the same degree distribution.
  - Random case, maximal assortative case and maximal disassortative

Mixing pattern vs degree distribution



Rich-club vs degree distribution



- P(k) does not constrain rich-club.
- For the Internet, a minor change of the value of *r* is associated with a huge change of rich-club coef.

Observations

❑Networks having the same degree distribution can be vastly different in other properties.

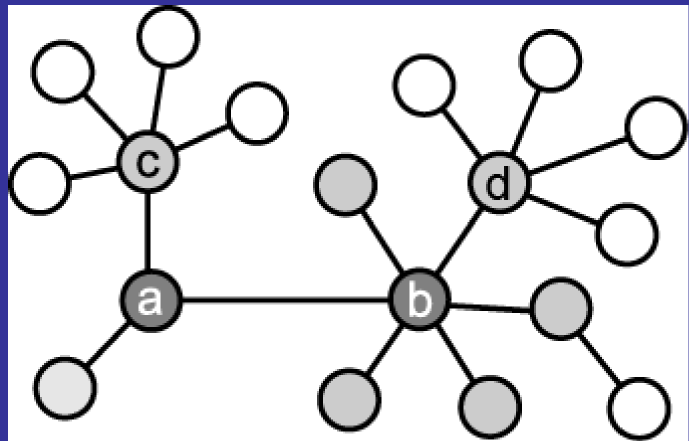
❑Mixing pattern and rich-club phenomenon are not trivially related.

- They are two different statistical projections of the joint degree distribution P(k, k’).
- Together they provide a fuller picture.

Part 2

Second-order mixing in networks

Neighbour's degrees

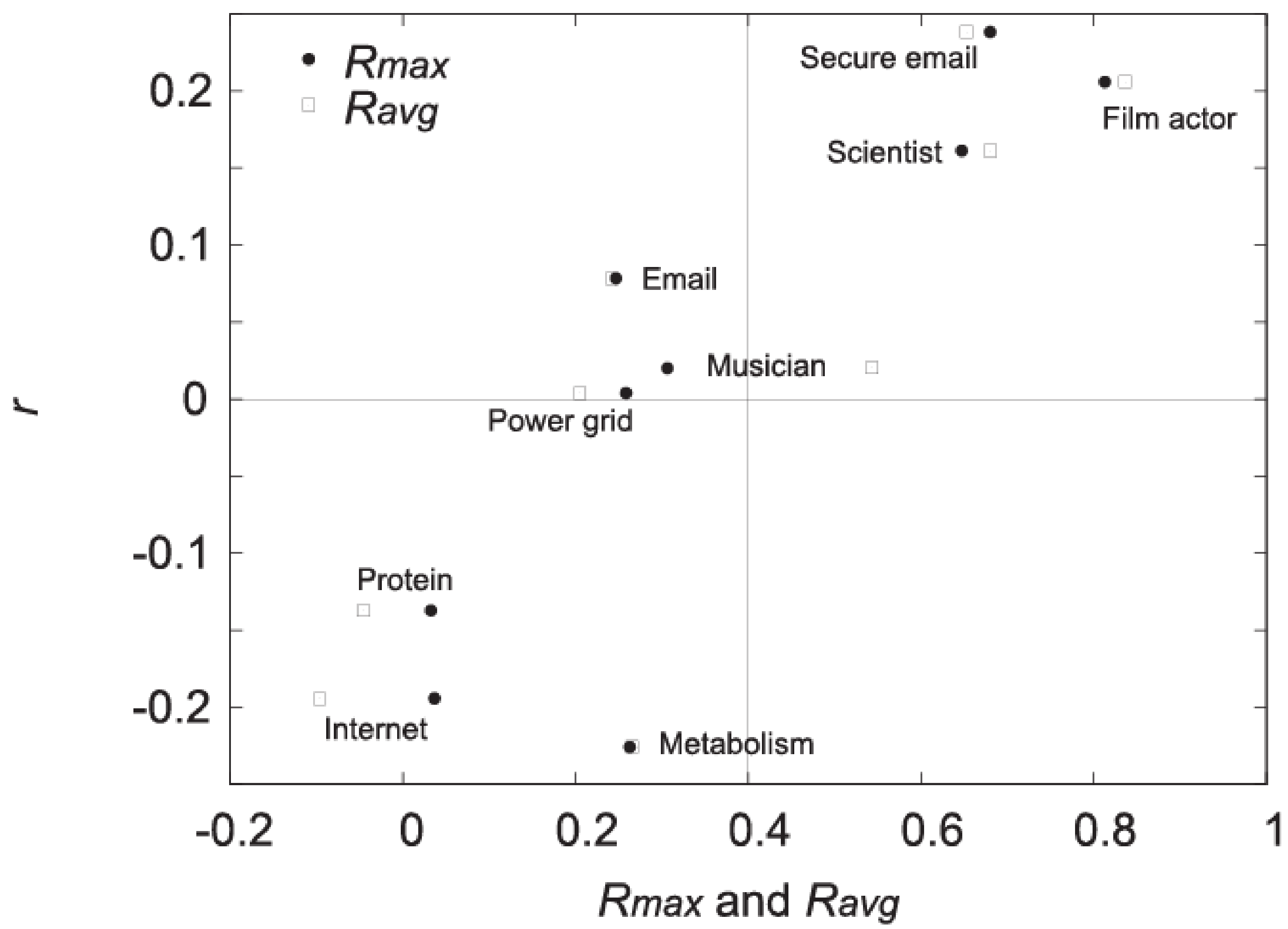


- Considering the link between nodes ***a*** and ***b***
- 1st order mixing: correlation between degrees of the two end nodes ***a*** and ***b***.
- 2nd order mixing: correlation between degrees of neighbours of nodes ***a*** and ***b***

1st and 2nd order assortative coefficients

Assortative coefficient	1st order (degree)	2nd order (Neighbours average degree)	2nd order (Neighbours max degree)
	r	R_{avg}	R_{max}
Scientist Collaborations	0.161	0.680	0.647
Internet	-0.195	-0.097	0.036

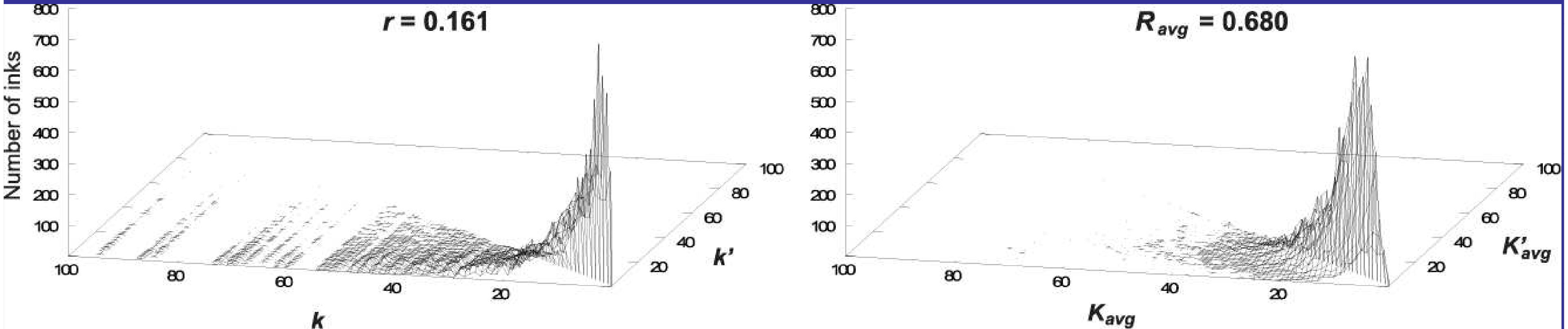
Assortative coefficients of networks



Statistic significance of the coefficients

- The Jackknife method
  - For all networks under study, the expected standard deviation of the coefficients are very small.
- Null hypothesis test
  - The coefficients obtained after random permutation (of one of the two value sequences) are close to zero with minor deviations.

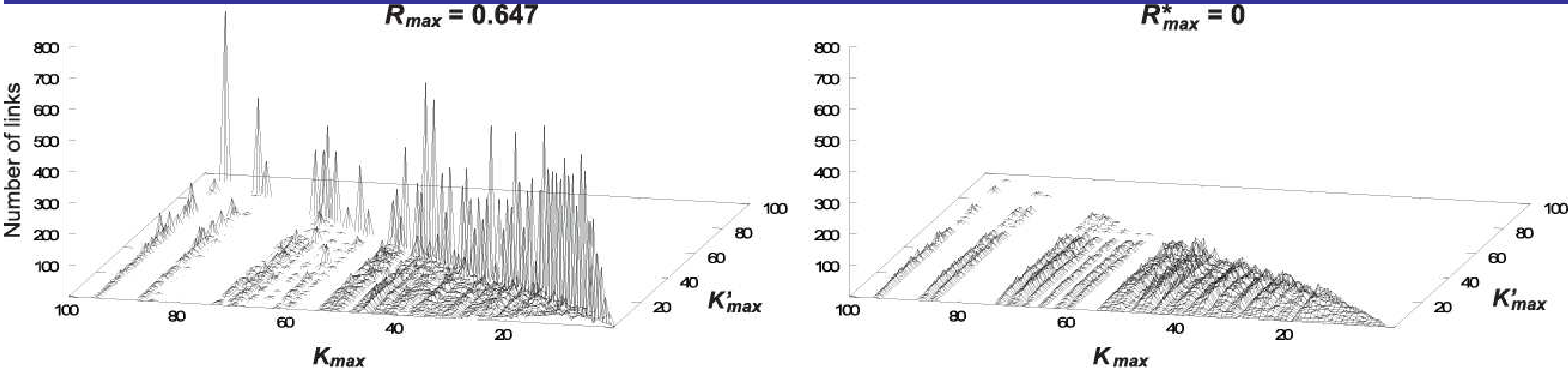
Mixing properties of the scientist network



- Frequency distribution of links as a function of
  1. degrees, ***k*** and ***k'***
  2. neighbours average degrees, ***Kavg*** and ***K'avg***

of the two end nodes of a link.

Mixing properties of the scientist network



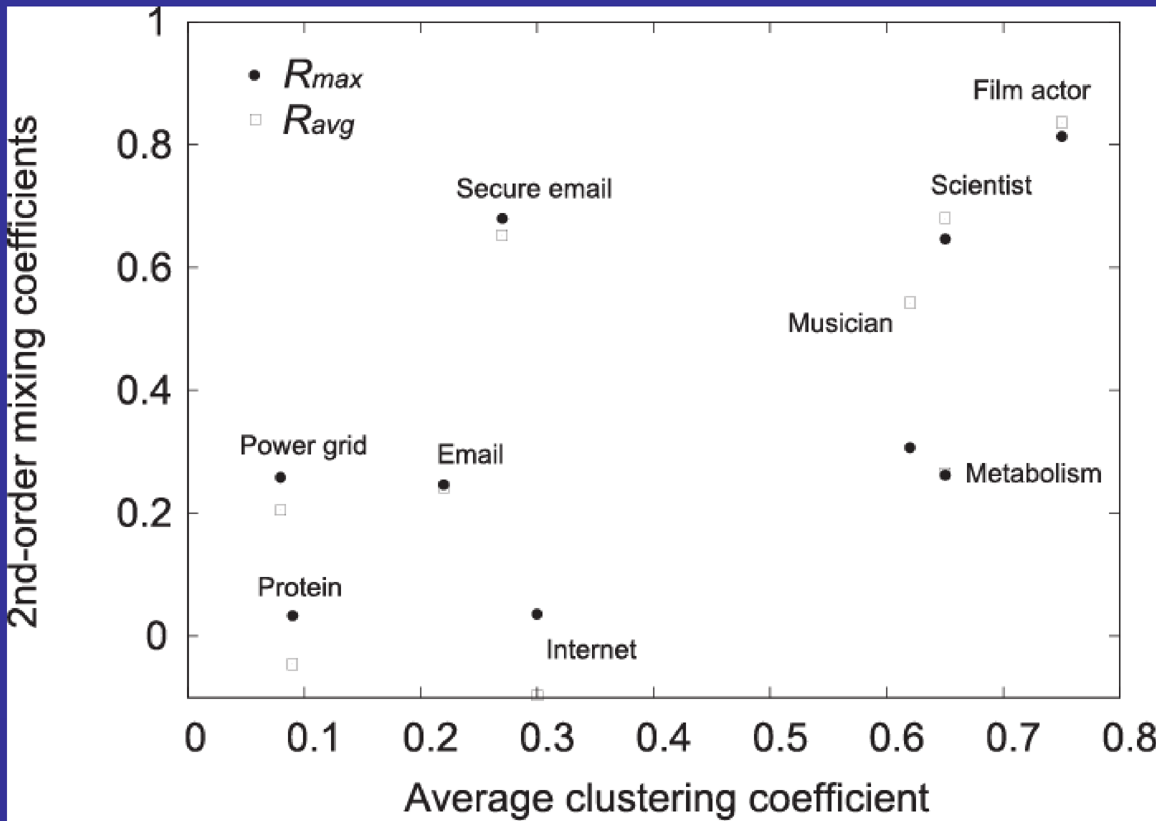
1. Link distribution as a function of neighbours max degrees, ***K**max* and ***K'**max*
2. That when the network is randomly rewired preserving the degree distribution.

Discussion (1)

- Is the 2nd order assortative mixing due to increased neighbourhood?
  - No. In all cases, the 3rd order coefficient is smaller.
- Is it due to a few hub nodes?
  - No. Removing the best connected nodes does not result in smaller values of Rmax or Ravg.
- Is it due to power-law degree distribution?
  - No. As link rewiring result shows, degree distribution has little constraint on 2nd order mixing.

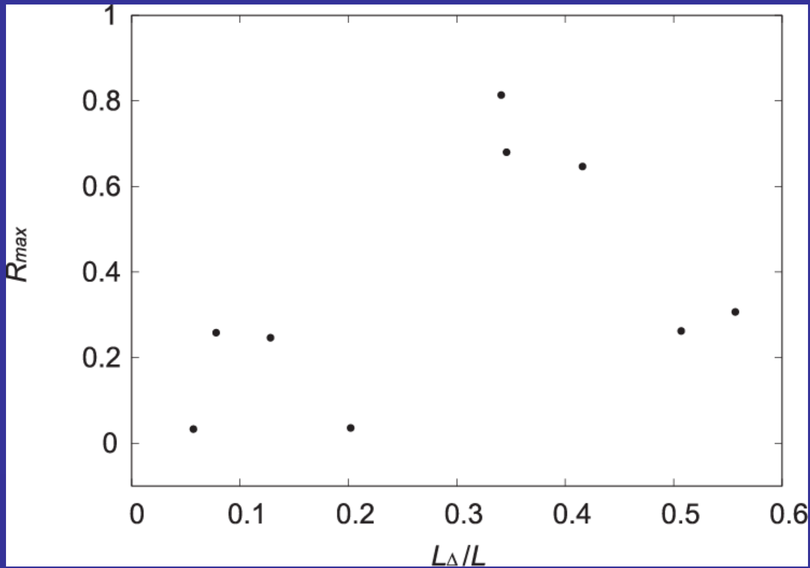
Discussion (2)

- ❑ Is it due to clustering?
  - ➢No.



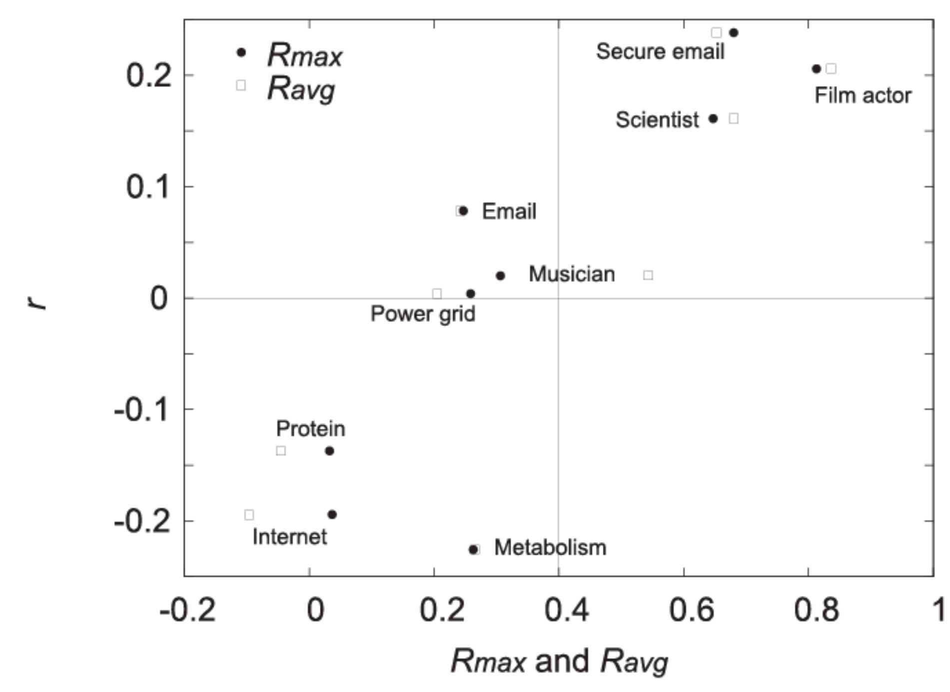
Discussion (3)

- How about triangles?
  - ➢There are many links where the best-connected neighbours of the two end nodes are one and the same, forming a triangle.
  - ➢ But there is no correlation between the amount of such links and the coefficient R_{max}.
  - ➢ There are also many links where the best-connected neighbours are not the same.
  - ➢ And there are many links …



Discussion (4)

- Is it due to the nature of bipartite networks?
  - No.
  - The secure email network is not a bipartite network, but it shows very strong 2nd order assortative mixing.
  - The metabolism network is a bipartite network, but it shows weaker 2nd order assortative mixing.



Discussion - summary

- Each of the above may play a certain role
- But none of them provides an adequate explanation.
- A new property?
  - New clue for networks modelling?

Implication of 2nd order assortative mixing

- It is not just ***how many*** people you know,
  - Degree
  - 1st order mixing
- But also ***who*** you know.
  - Neighbours average or max degrees
  - 2nd order mixing
- Collaboration is influenced less by our own prominence, but more by the prominence of who we know?

Reference

- The rich-club phenomenon in the Internet topology
  - IEEE Comm. Lett. 8(3), p180-182, 2004.
- Structural constraints in complex networks
  - New J. of Physics, 9(173), p1-11, 2007
- Second-order mixing in networks
  - http://arxiv.org/abs/0903.0687
  - An updated version to appear soon.

Thank You !