Network Science

Chapter 7: Degree Correlations

Albert-László Barabási

With

Emma K. Towlson and Sean P. Cornelius

www.BarabasiLab.com

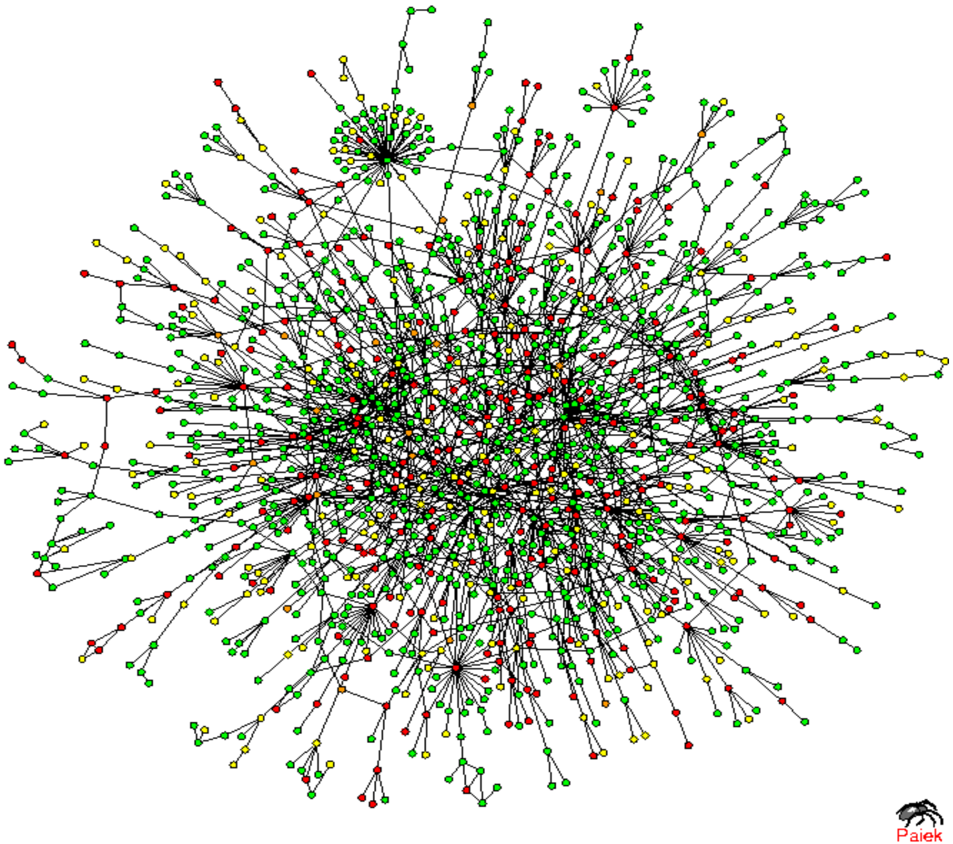
Questions

1. **What are degree correlations? Why do we want to study correlations?**
2. **What is the degree correlation matrix? What do we expect it to look like for random, assortative and disassortative networks? Why?**
3. **What is the degree correlation function? What do we expect to see for random, assortative and disassortative networks? Why?**
4. **What is the degree correlation coefficient *r*? What values do we expect for random, assortative and disassortative networks? Why?**
5. **What is structural disassortativity? What kind of network is affected and how can we detect it?**
6. **What is the impact of the degree correlations? Why do we study them? Why does the threshold for the phase transition in Fig. 7.15 change?**
7. **Summary**
8. **Differences between undirected and directed networks.**

TOPOLOGY OF THE PROTEIN NETWORK

Nodes: proteins
Links: physical interactions (binding)

Puzzling pattern:
Hubs tend to link to small degree nodes.
Why is this puzzling?



In a random network, the probability that a node with degree *k* links to a node with degree *k'* is:

$$p_{kk'} = \frac{kk'}{2L}$$

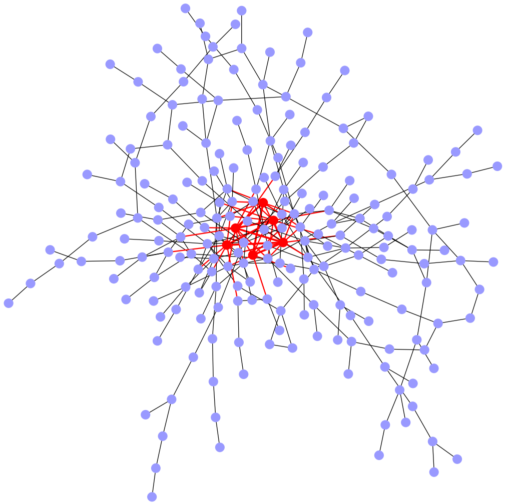
k≅50, k'=13, N=1,458, L=1746

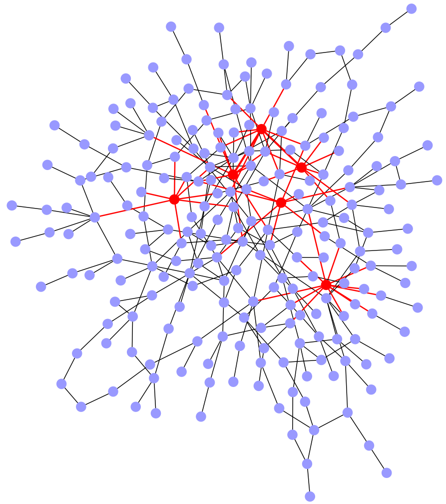
$$p_{50,13} = 0.15$$
$$p_{2,1} = 0.0004$$

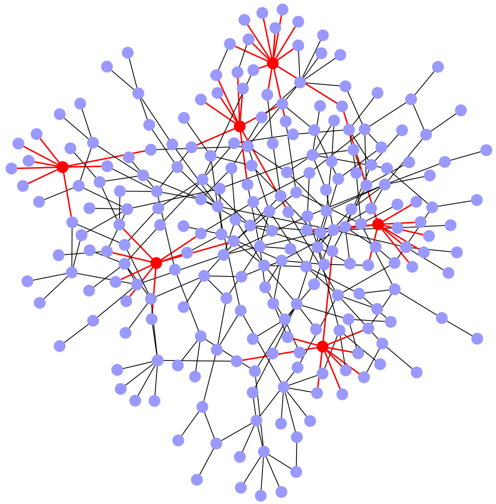
Yet, we see many links between degree 2 and 1 nodes, and no links between the hubs.

H. Jeong, S.P. Mason, A.-L. Barabasi, Z.N. Oltvai, Nature 411, 41-42 (2001)

DEGREE CORRELATIONS IN NETWORKS







Assortative:
hubs show a tendency to link to each other.

Neutral:
nodes connect to each other with the expected random probabilities.

Disassortative:
Hubs tend to avoid linking to each other.

Quantifying degree correlations (three approaches):

→ full statistical description (Maslov and Sneppen, Science 2001)
→ degree correlation function (Pastor Satorras and Vespignani, PRL 2001)
→ correlation coefficient (Newman, PRL 2002)

STATISTICAL DESCRIPTION

e_{jk}: probability to find a node with degree *j* and degree *k* at the two ends of a randomly selected edge

$$\sum_{j,k} e_{jk} = 1 \qquad \sum_{j} e_{jk} = q_k$$

q_k: the probability to have a degree *k* node at the end of a link.

Where: $$q_k = \frac{kp_k}{\langle k \rangle}$$

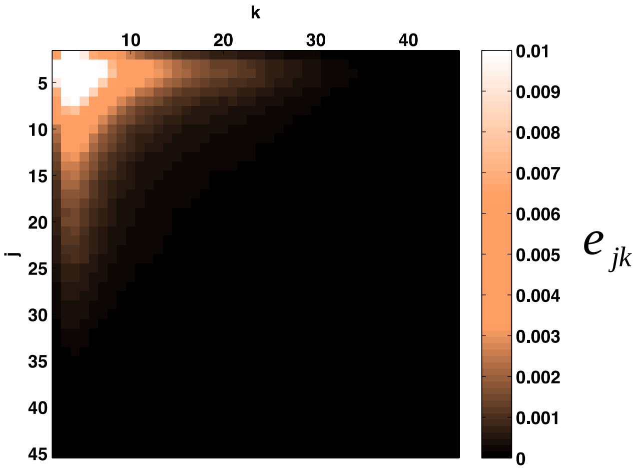
Probability to find a node at the end of a link is biased towards the more connected nodes, i.e. $q_k = Ckp_k$, where C is a normalization constant . After normalization we find C=1/<k>, or $q_k = kp_k/\langle k \rangle$

If the network has no degree correlations: $$e_{jk} = q_j q_k$$

Deviations from this prediction are a signature of *degree correlations.*

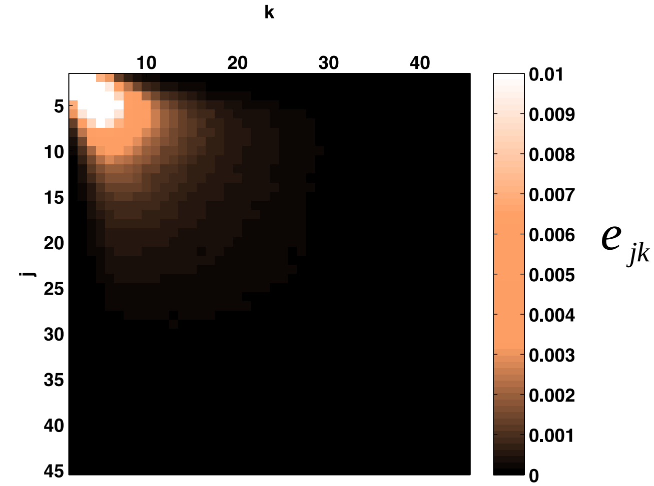
M. E. J. Newman, Phys. Rev. Lett. **89,** 208701 (2002)

EXAMPLE: e_{jk} FOR A SCALE-FREE NETWORK

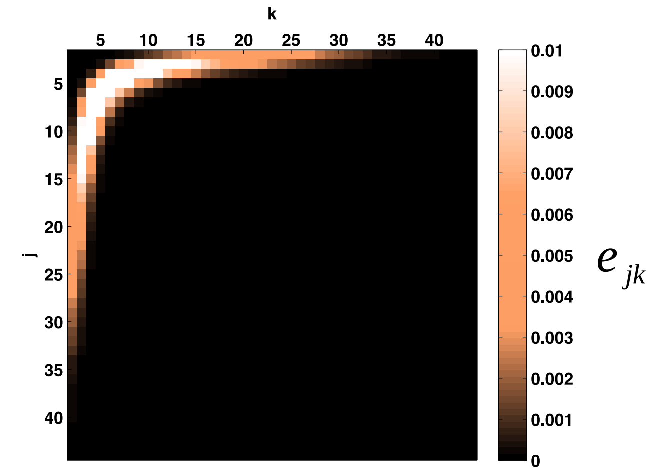


Neutral

Assortative:
More strength in the diagonal, hubs tend to link to each other.



Disassortative:
Hubs tend to connect to small nodes.



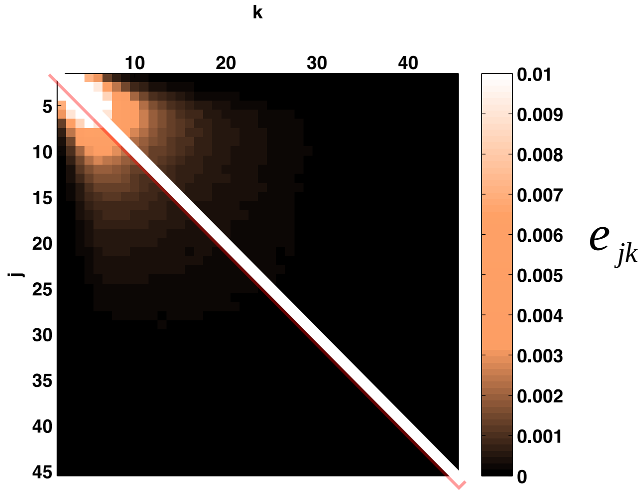
Each matrix is the average of 100 independent scale-free networks, generated using the static model with $N=10^4$, $\gamma=2.5$ and $\langle k \rangle=3$.

EXAMPLE: e_{jk} FOR A SCALE-FREE NETWORK

Perfectly assortative network:

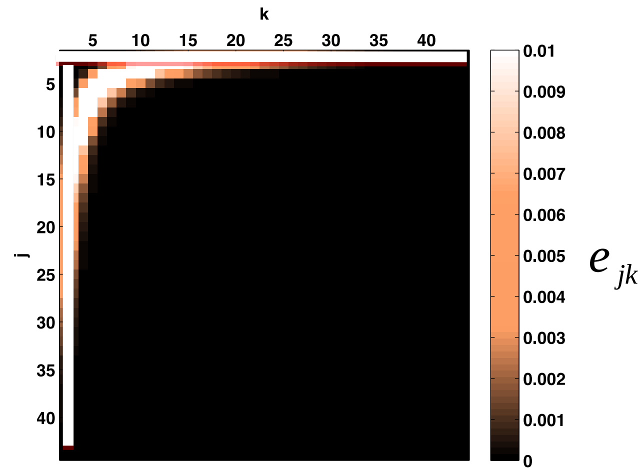
$$e_{jk}=q_k\delta_{jk}$$

Assortative:
More strength in the diagonal, hubs tend to link to each other.



Perfectly disassortative network:

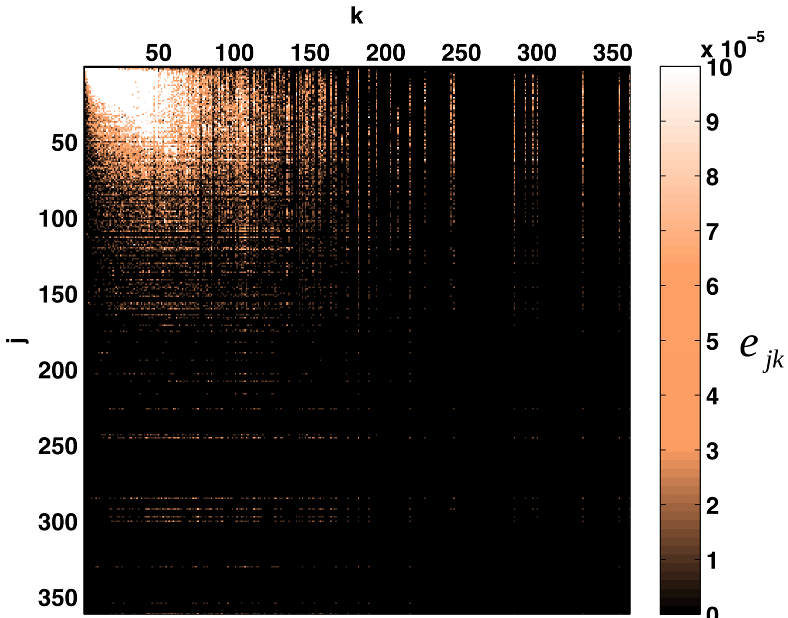
Disassortative:
Hubs tend to connect to small nodes.



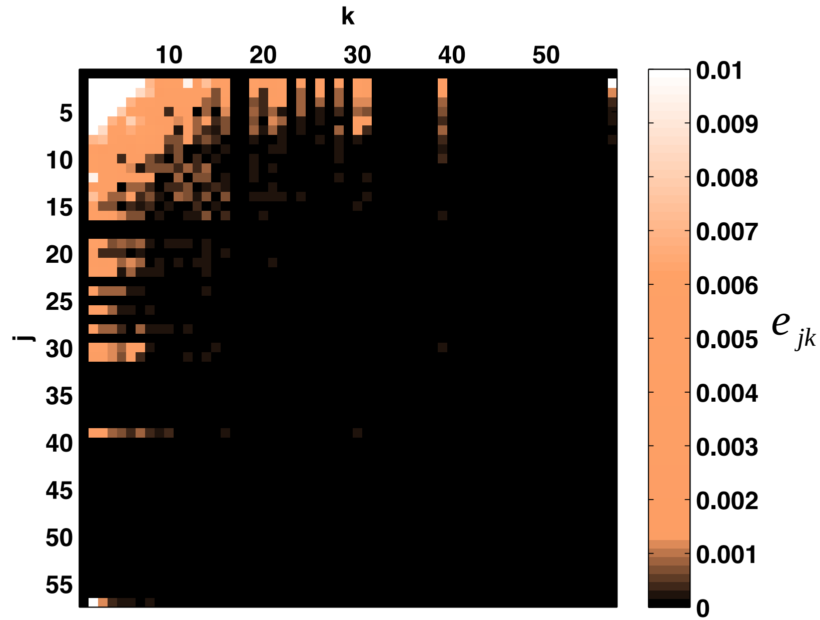
Each matrix is the average of 100 independent scale-free networks, generated using the static model with $N=10^4$, $\gamma=2.5$ and $<k>=3$.

REAL-WORLD EXAMPLES

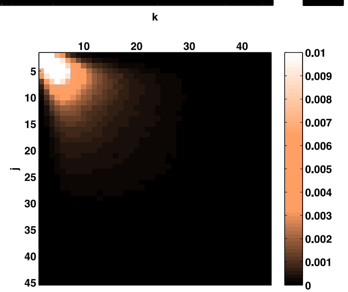
Astrophysics co-authorship network



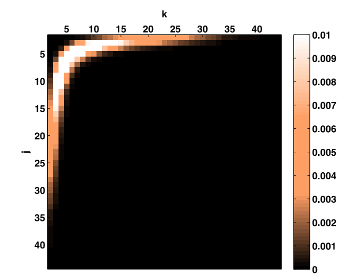
Yeast PPI



Assortative: More strength in the diagonal, hubs tend to link to each other.

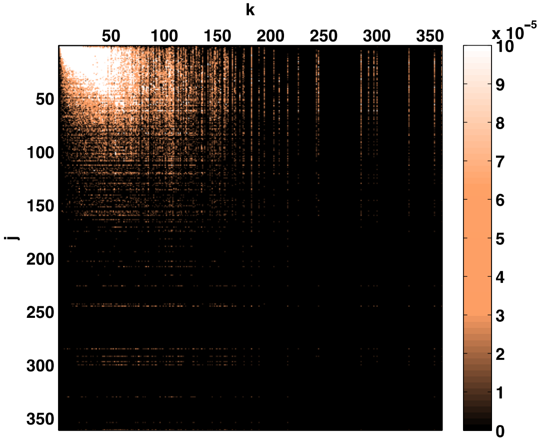


Disassortative: Hubs tend to connect to small nodes.



PROBLEM WITH THE FULL STATISTICAL DESCRIPTION

(1) Difficult to extract information from a visual inspection of a matrix.



(2) Based on e_{jk} and hence requires a large number of elements to inspect:

$$\frac{k_{\max}\left(k_{\max}-1\right)}{2}-1-k_{\max}$$

Nr. of independent elements

Undirected network:
k_{max} x k_{max} matrix

$$\sum_{j,k} e_{jk}=1 \qquad \sum_{j=1,k_{\max}} e_{jk}=q_k$$

Constraints

We need to find a way to reduce the information contained in e_{jk}

M. E. J. Newman, Phys. Rev. Lett. **89,** 208701 (2002)

Measuring Degree Correlations

Average next neighbor degree

$k_{nn}(k)$: average degree of the first neighbors of nodes with degree k.

$$k_{nn}(k) \equiv \sum_{k'} k' P(k' \mid k)$$

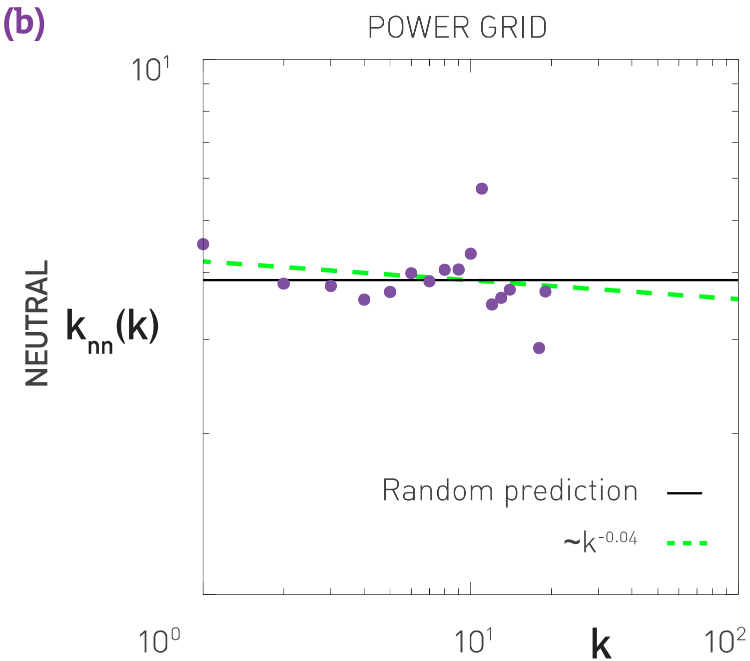
- **Neutral Network**

  For a neutral network (7.3)-(7.5) predict

$$P(k' \mid k) = \frac{e_{kk'}}{\sum_{k'} e_{kk'}} = \frac{e_{kk'}}{q_k} = \frac{q_{k'} q_k}{q_k} = q_{k'}.$$

  This allows us to express $k_{nn}(k)$ as

$$k_{nn}(k) = \sum_{k'} k' q_{k'} = \sum_{k'} k' \frac{k' p(k')}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}.$$



R. Pastor-Satorras, A. Vázquez, A. Vespignani, Phys. Rev. E **65**, 066130 (2001)

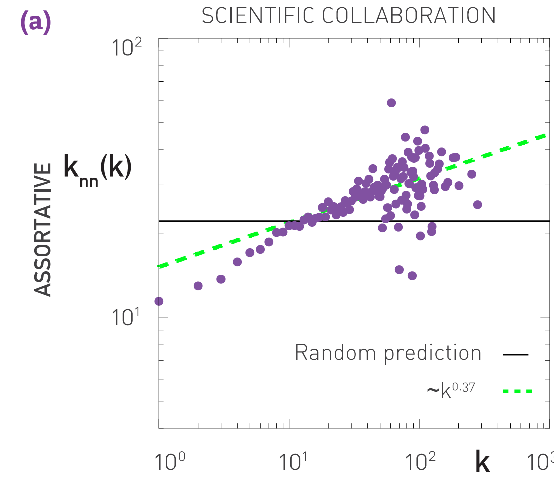
Average next neighbor degree

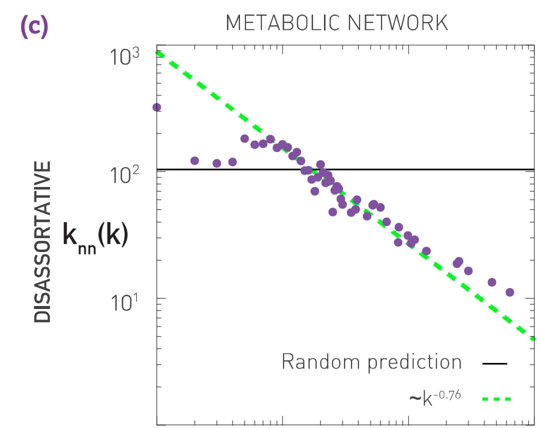
$k_{nn}(k)$: average degree of the first neighbors of nodes with degree k.

$$k_{nn}(k) \equiv \sum_{k'} k' P(k' \mid k)$$

- **Assortative Network**
  In assortative networks hubs tend to connect to other hubs, hence the higher is the degree k of a node, the higher is the average degree of its nearest neighbors. Consequently for assortative networks $k_{nn}(k)$ *increases with* k, as observed for scientific collaboration networks (Figure 7.6a).

- **Disassortative Network**
  In disassortative network hubs prefer to link to low-degree nodes. Consequently $k_{nn}(k)$ *decreases with* k, as observed for the metabolic network (Figure 7.6c).





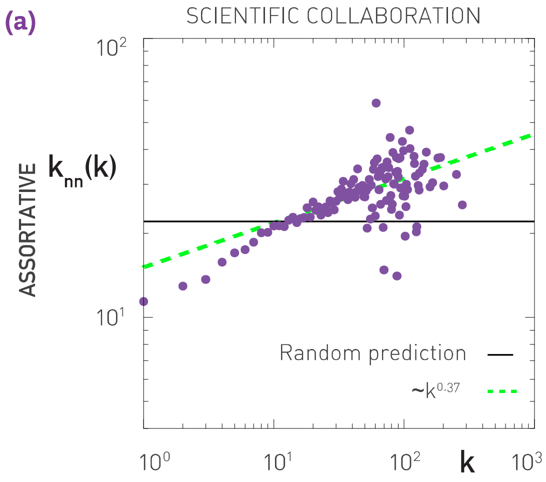
Average next neighbor degree

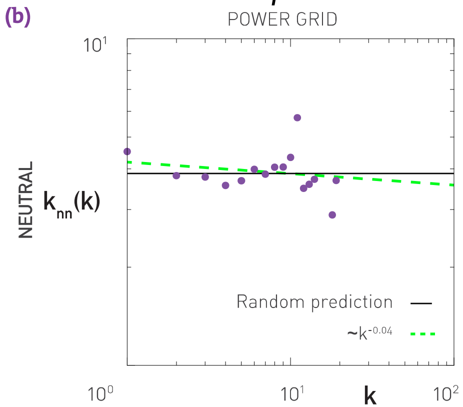
$$k_{nn}(k) = ak^{\mu}$$

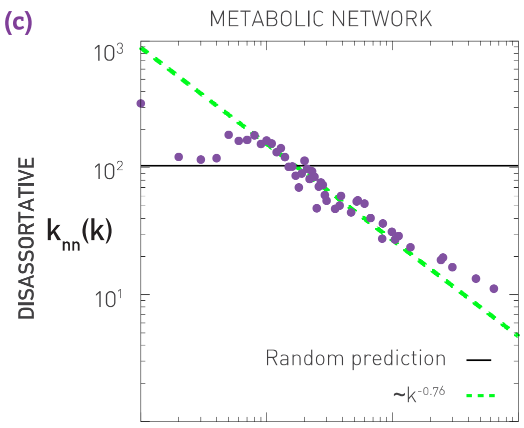
- **Assortative Networks: μ > 0**
  A fit to $k_{nn}(k)$ for the science collaboration network provides μ = 0.37 ± 0.11 (Figure 7.6a).

- **Neutral Networks: μ = 0**
  According to (7.9) $k_{nn}(k)$ is independent of k. Indeed, for the power grid we obtain μ = 0.04 ± 0.05, which is indistinguishable from zero (Figure 7.6b).

- **Disassortative Networks: μ < 0**
  For the metabolic network we obtain μ = – 0.76 ± 0.04 (Figure 7.6c).







Degree Correlation Coefficient

If there are degree correlations, e_{jk} will differ from $q_j q_k$. The magnitude of the correlation is captured by <jk>-<j><k> difference, which is:

$$\sum_{jk} jk(e_{jk} - q_j q_k)$$

<jk>-<j><k> is expected to be:

positive for *assortative* networks,
zero for *neutral* networks,
negative for *dissasortative* networks

To compare different networks, we should normalize it with its maximum value; the maximum is reached for a *perfectly assortative network*, i.e. $e_{jk}=q_k\delta_{jk}$

normalization: $$\sigma_r^2 = \max \sum_{jk} jk(e_{jk} - q_j q_k) = \sum_{jk} jk(q_k \delta_{jk} - q_j q_k)$$

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_r^2} \qquad -1 \le r \le 1$$

$r \le 0$ *disassortative*
$r = 0$ *neutral*
$r \ge 0$ *assortative*

M. E. J. Newman, Phys. Rev. Lett. **89,** 208701 (2002)

REAL NETWORKS

Social networks are *assortative*

Network	n	r
Physics coauthorship (a)	52 909	0.363
Biology coauthorship (a)	1 520 251	0.127
Mathematics coauthorship (b)	253 339	0.120
Film actor collaborations (c)	449 913	0.208
Company directors (d)	7 673	0.276
Internet (e)	10 697	−0.189
World-Wide Web (f)	269 504	−0.065
Protein interactions (g)	2 115	−0.156
Neural network (h)	307	−0.163
Marine food web (i)	134	−0.247
Freshwater food web (j)	92	−0.276
Random graph (u)		0
Callaway *et al.* (v)		$\delta/(1+2\delta)$
Barabási and Albert (w)		0

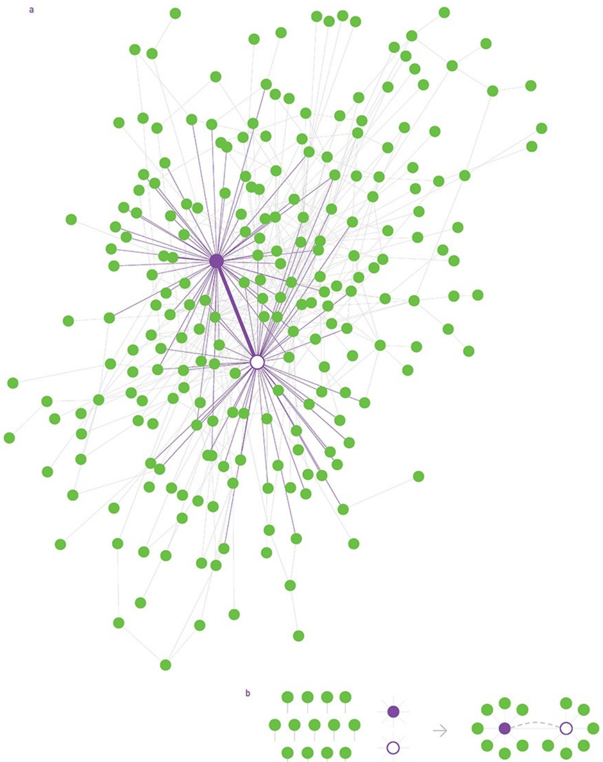
Biological, technological networks are *disassortative*

r>0: assortative network:
Hubs tend to connect to other hubs.

r<0: disassortative network:
Hubs tend to connect to small nodes.

Structural cutoffs

Example: Degree sequence introduces disassortativity



Scale-free network generated with the configuration model (N=300, L=450, γ=2.2).

The measured r=-0.19! → Dissasortative!

Purple hub: 55 neighbors.
White hub: 46 neighbors.

Calculation of the expected number of links between purple node (k=55) and white node (k=46) for uncorrelated networks:

k , p_k , k' , $p_{k'}$, $\langle k \rangle$

$$E_{55,46} = \langle k \rangle N \cdot e_{55,46} = 900 \cdot \frac{55\frac{1}{300} \cdot 46\frac{1}{300}}{3^2} \approx 2.8 > 1$$

In order for the network to be neutral, we need 2.8 links between these two hubs.

$$k_s(N) \sim (\langle k \rangle N)^{1/2}.$$

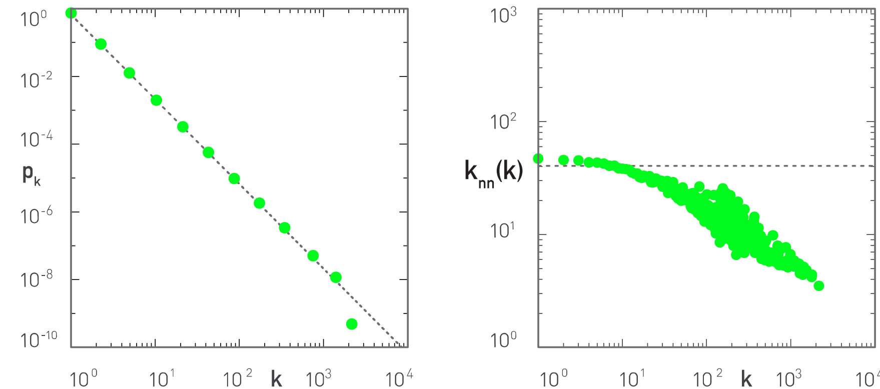
$$k_{max} \sim N^{\frac{1}{\gamma-1}}.$$

- **No Stuctural Cutoff**

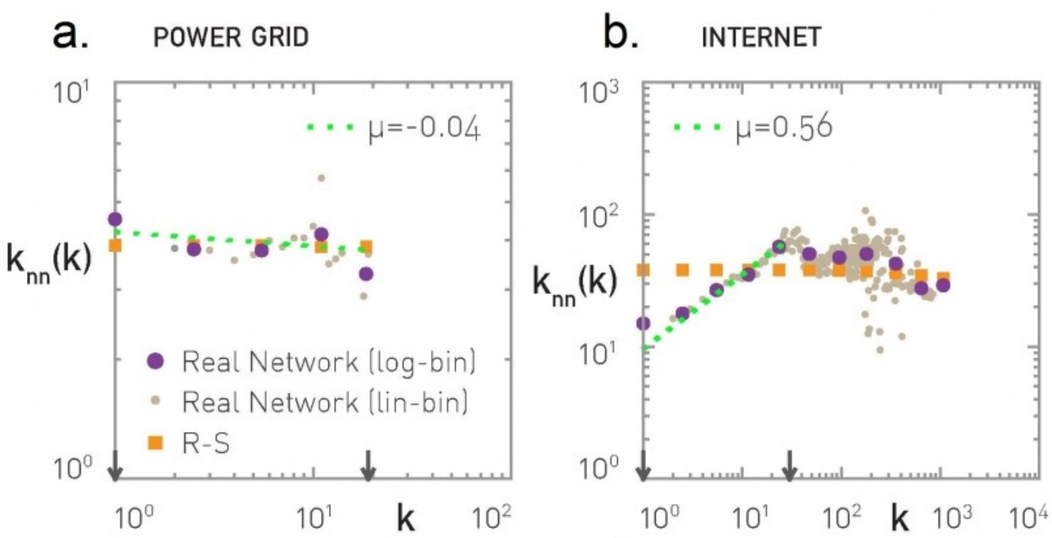
  For random networks and scale-free networks with $\gamma \geq 3$ the exponent of k_{max} is smaller than 1/2, hence k_{max} is always smaller than k_s. In other words the node size at which the structural cutoff turns on exceeds the size of the biggest hub. Consequently we have no nodes for which $E_{kk'} > 1$. For these networks we do not have a conflict between degree correlations and the simple network requirement.

- **Stuctural Disassortativity**

  For scale-fee networks with $\gamma < 3$ we have $1/(\gamma-1) > 1/2$, i.e. k_s can be smaller than k_{max}. Consequently nodes whose degree is between k_s and k_{max} can violate $E_{kk'} > 1$. In other words the network has fewer links between its hubs than (7.14) would predict. These networks will therefore become disassortative, a phenomenon we call *structural disassortativity*. This is illustrated in Figures 7.8a,b that show a simple scale-free network generated by the configuration model. The network shows disassortative scaling, despite the fact that we did not impose degree correlations during its construction.

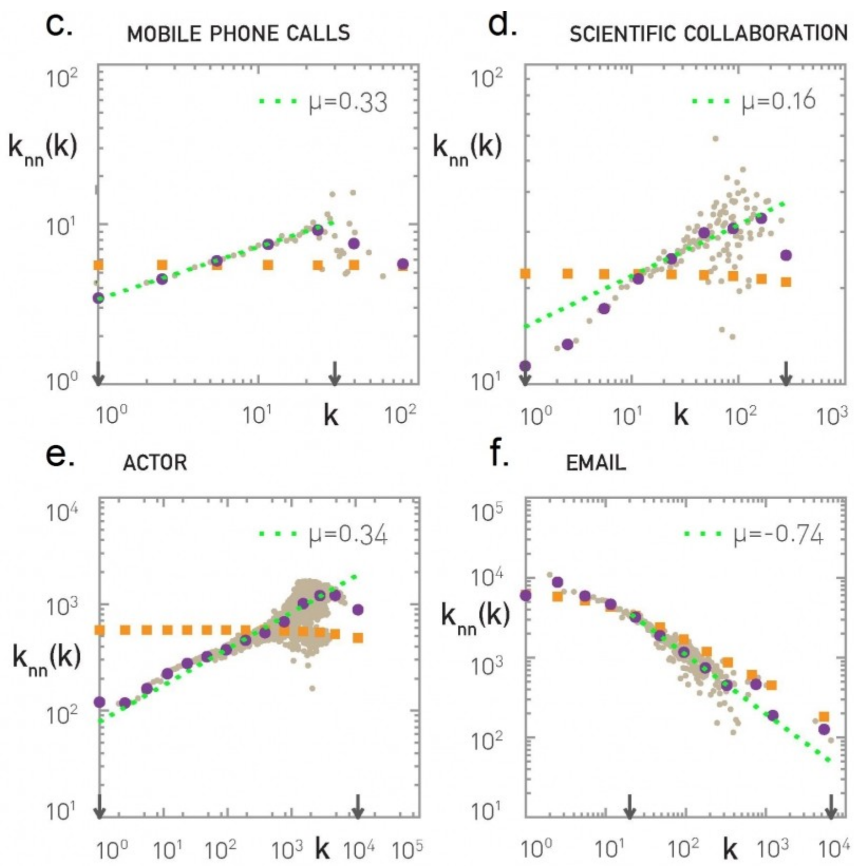


(a,b) If we generate a scale-free network with the power-law degree distribution shown in (a), and we forbid self-loops and multi-links, the network displays structural disassortativity, as indicated by $k_{nn}(k)$ in (b). In this case, we lack a sufficient number of links between the high-degree nodes to maintain the neutral nature of the network, hence for high k the $k_{nn}(k)$ function must decay.



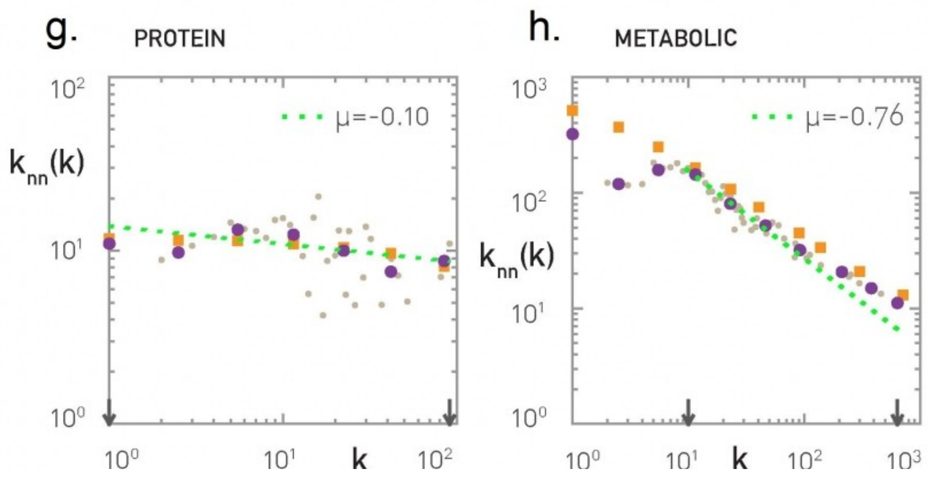
- **Power Grid**
  For the power grid $k_{nn}(k)$ is flat and indistinguishable from its randomized version, indicating a lack of degree correlations (Figure 7.10a). Hence the power grid is neutral.

- **Internet**
  For small degrees ($k \leq 30$) $k_{nn}(k)$ shows a clear assortative trend, an effect that levels off for high degrees (Figure 7.10b). The degree correlations vanish in the randomized version of the Internet map. Hence the Internet is assortative, but structural cutoffs eliminate the effect for high k.

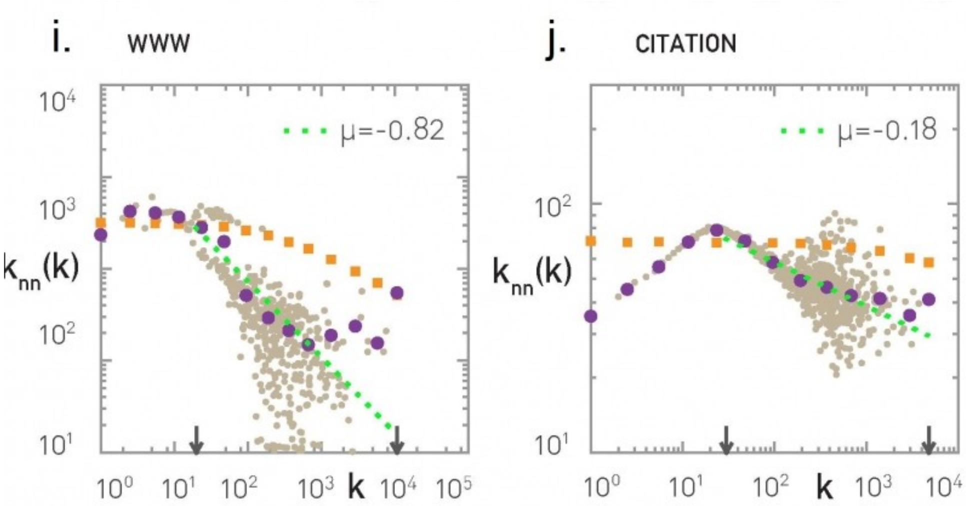


- **Social Networks**
  The three networks capturing social interactions, the mobile phone network, the science collaboration network and the actor network, all have an increasing $k_{nn}(k)$, indicating that they are assortative (Figures 7.10c-e). Hence in these networks hubs tend to link to other hubs and low-degree nodes tend to link to low-degree nodes. The fact that the observed $k_{nn}(k)$ differs from the $k_{nn}^{R\text{-}S}(k)$, indicates that the assortative nature of social networks is not due to their scale-free the degree distribution.

- **Email Network**
  While the email network is often seen as a social network, its $k_{nn}(k)$ decreases with k, documenting a clear disassortative behavior (Figure 7.10f). The randomized $k_{nn}^{R\text{-}S}(k)$ also decays, indicating that we are observing structural disassortativity, a consequence of the network's scale-free nature.



- **Biological Networks**
  The protein interaction and the metabolic network both have a negative μ, suggesting that these networks are disassortative. Yet, the scaling of $k_{min}^{R\text{-}S}(k)$ is indistinguishable from $k_{nn}(k)$, indicating that we are observing structural disassortativity, rooted in the scale-free nature of these networks (Figure 7.10 g,h).



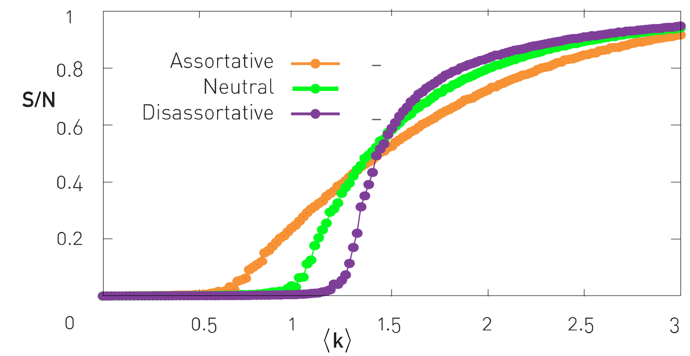
- **WWW**
  The decaying $k_{nn}(k)$ implies disassortative correlations (Figure 7.10i). The randomized $k^{R\text{-}S}_{min}(k)$ also decays, but not as rapidly as $k_{nn}(k)$. Hence the disassortative nature of the WWW is not fully explained by its degree distribution.

- **Citation Network**
  This network displays a puzzling behavior: for $k \leq 20$ the degree correlation function $k_{nn}(k)$ shows a clear assortative trend; for $k > 20$, however, we observe disassortative scaling (Figure 7.10j). Such mixed behavior can emerge in networks that display extreme assortativity (Figure 7.13b). This suggests that the citation network is strongly assortative, but its scale-free nature induces structural disassortativity, changing the slope of $k_{nn}(k)$ for $k \gg k_s$.

- **Assortative Networks**
  For assortative networks the phase transition point moves to a lower $\langle k \rangle$, hence a giant component emerges for $\langle k \rangle < 1$. The reason is that it is easier to start a giant component if the high-degree nodes seek out each other.

- **Disassortative Networks**
  The phase transition is delayed in disassortative networks, as in these the hubs tend to connect to small degree nodes. Consequently, disassortative networks have difficulty forming a giant component.

- **Giant Component**
  For large $\langle k \rangle$ the giant component is smaller in assortative networks than in neutral or disassortative networks. Indeed, assortativity forces the hubs to link to each other, hence they fail to attract to the giant component the numerous small degree nodes.

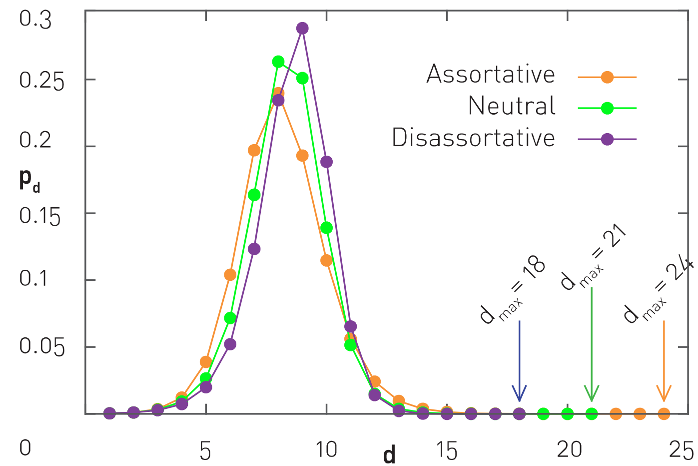


- Figure 7.16 shows the path-length distribution of a random network rewired to display different degree correlations. It indicates that in assortative networks the average path length is shorter than in neutral networks. The most dramatic difference is in the network diameter, d_{max}, which is significantly higher for assortative networks. Indeed, assortativity favors links between nodes with similar degree, resulting in long chains of $k = 2$ nodes, enhancing d_{max} (Figure 7.13c).

- Degree correlations influence a system's stability against stimuli and perturbations [26] as well as the synchronization of oscillators placed on a network [27, 28].

- Degree correlations have a fundamental impact on the vertex cover problem [29], a much-studied problem in graph theory that requires us to find the minimal set of nodes (cover) such that each link is connected to at least one node in the cover (BOX 7.4).

- Degree correlations impact our ability to control a network, altering the number of input signals one needs to achieve full control [30].

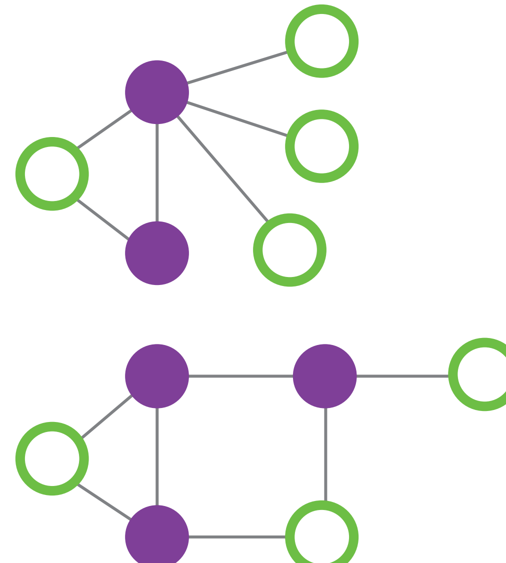


Distance distribution for a random network with size $N = 10,000$ and $\langle k \rangle = 3$. Correlations are induced using the Xalvi-Brunet & Sokolov algorithm with $p = 0.5$ (Figure 7.14). The plots show that as we move from disassortative to assortative networks, the average path length decreases, indicated by the gradual move of the peaks to the left. At the same time the diameter, d_{max}, grows. Each curve represents an average over 10 independent networks.

Imagine that you are the director of an open-air museum located in a large park. You wish to place guards on the crossroads to observe each path. Yet, to save cost you want to use as few guards as possible. How many guards do you need?

Let N be the number of crossroads and $m < N$ is the number of guards you can afford to hire. While there are $\binom{N}{m}$ ways of placing the m guards at N crossroads, most configurations leave some paths unsupervised [31].

The number of trials one needs to place the guards so that they cover all paths grows exponentially with N. Indeed, this is one of the six basic NP-complete problems, called the *vertex cover problem*. The vertex cover of a network is a set of nodes such that each link is connected to at least one node of the set **(Figure 7.17)**. NP-completeness means that there is no known algorithm which can identify a minimal vertex cover substantially faster than using as exhaustive search, i.e. checking each possible configuration individually. The number of nodes in the minimal a vertex cover depends on the network topology, being affected by the degree distribution and degree correlations [29].



Assortative mating reflects the tendency of individuals to date or marry individuals that are similar to them. For example, low-income individuals marry low-income individuals and college graduates marry college graduates. Network theory uses assortativity in the same spirit, capturing the degree-based similarities between nodes: In assortative networks hubs tend to connect to other hubs and small-degree nodes to other small-degree nodes. In a network environment we can also encounter the traditional assortativity, when nodes of similar properties link to each other (Figure 7.18).

Disassortative mixing, when individuals link to individuals wo are unlike them, is also common in some social and economic systems. Sexual networks are perhaps the best example, as most sexual relationships are between individuals of different gender. In economic settings trade typically takes place between individuals of different skills: the baker does not sell bread to other bakers, and the shoemaker rarely fixes other shoemaker's shoes.

Degree Correlation Matrix e_{ij}

Neutral networks:

$$e_{ij} = q_i q_i = \frac{k_i p_{k_i} k_j p_{k_j}}{\langle k \rangle^2}$$

Degree Correlation Function

$$k_{nn}(k) = \sum_{k'} k' p(k' | k)$$

Neutral networks:

$$k_{nn}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

Scaling Hypothesis

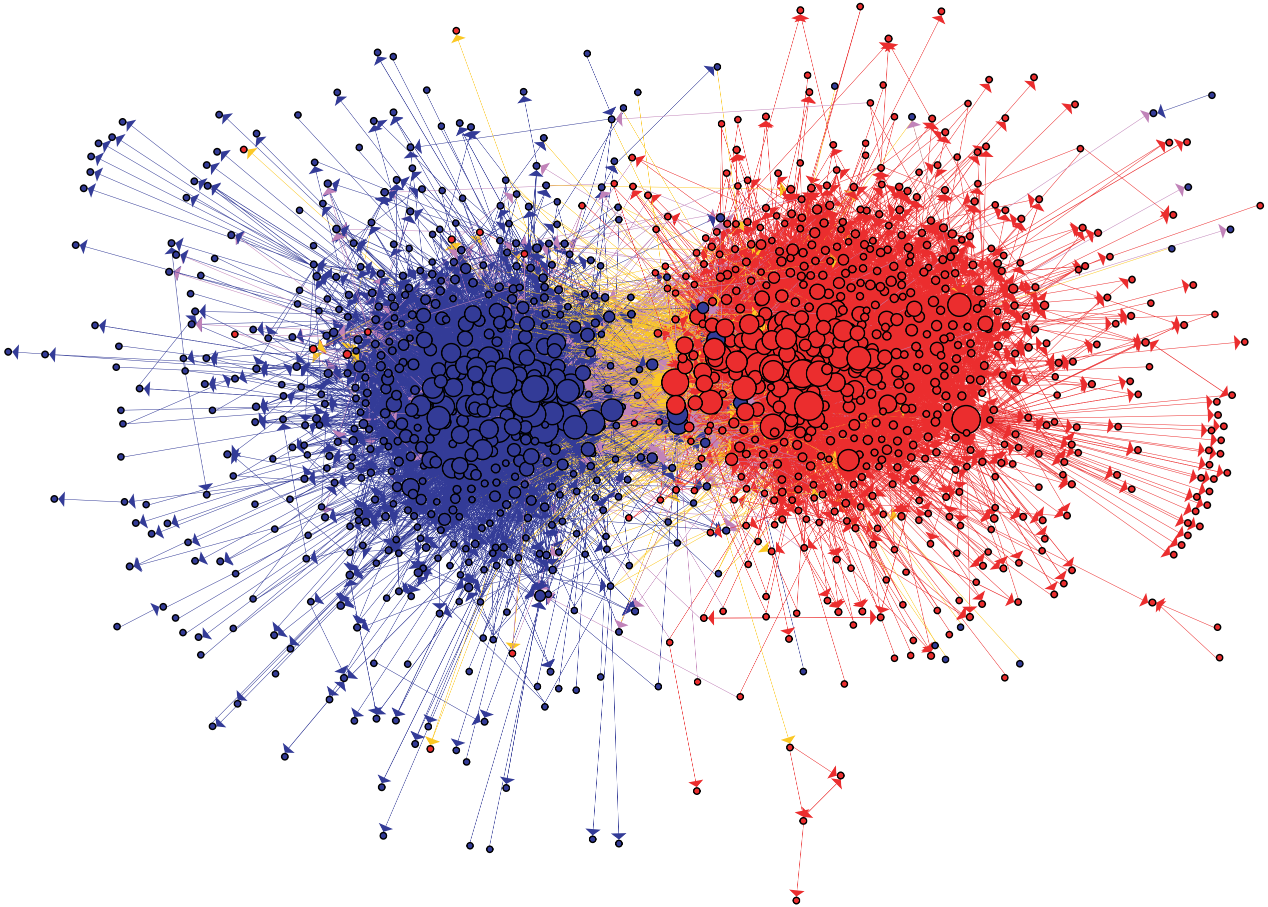
$$k_{nn}(k) \sim k^{\mu}$$

$\mu > 0$: *Assortative*
$\mu = 0$: *Neutral*
$\mu < 0$: *Disassortative*

Degree Correlation Coefficient

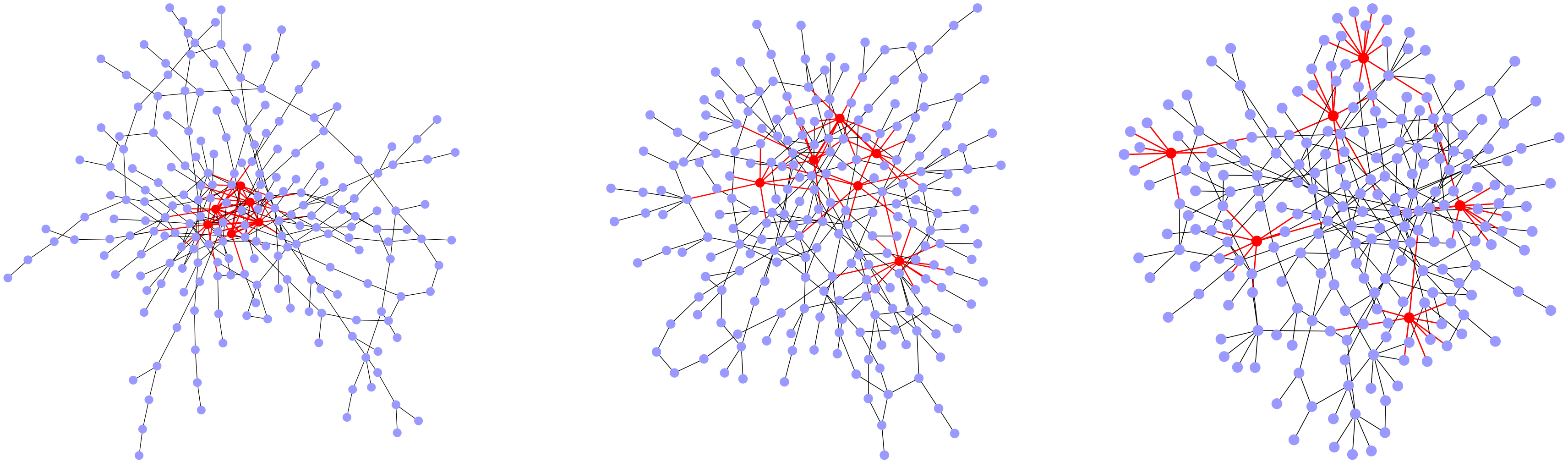
$$r = \sum_{jk} \frac{jk(e_{jk} - q_j q_k)}{\sigma_r^2}$$

$r > 0$: *Assortative*
$r = 0$: *Neutral*
$r < 0$: *Disassortative*



The network behind the US political blogosphere illustrates the presence of assortative mixing, as used in sociology, meaning that nodes of similar characteristics tend to link to each other. In the map each blue node corresponds to liberal blog and red nodes are conservative. Blue links connect liberal blogs, red links connect conservative blogs, yellow links go from liberal to conservative, and purple from conservative to liberal. As the image indicates, very few blogs link across the political divide, demonstrating the strong assortativity of the political blogosphere.

DEGREE CORRELATIONS IN NETWORKS



Assortative:
hubs show a tendency to link to each other.

Neutral:
nodes connect to each other with the expected random probabilities.

Disassortative:
Hubs tend to avoid linking to each other.

Quantifying degree correlations (three approaches):

→ full statistical description (Maslov and Sneppen, Science 2001)
→ degree correlation function (Pastor Satorras and Vespignani, PRL 2001)
→ correlation coefficient (Newman, PRL 2002)

Directed networks: knn