

CHAPTER 7

DEGREE CORRELATIONS

Introduction

Assortativity and disassortativity 1

Measuring degree correlations 2

Structural cutoffs 3

Degree correlations in real networks 4

Generating correlated networks

The impact of degree correlations

Summary

ADVANCED TOPICS 7.A

Degree correlation coefficient

ADVANCED TOPICS 7.B

Structural Cutoffs

Bibliography



INTRODUCTION

Angelina Jolie and Brad Pitt, Ben Affleck and Jennifer Garner, Harrison Ford and Calista Flockhart, Michael Douglas and Catherine Zeta-Jones, Tom Cruise and Katie Holmes, Richard Gere and Cindy Crawford. An odd list, yet instantly recognizable to those immersed in the headline-driven world of celebrity couples. They are Hollywood stars that are or were married in the past. Their weddings (and breakups) sold millions of magazines and drawn countless hours of media coverage. Thanks to them we take for granted that celebrities marry each other. We rarely pause to ask: is this normal? What is the expected chance that a celebrity marries another celebrity?

Assuming that a celebrity could date anyone from a pool of about a billion (10^9) eligible individuals, the chances that their mate would be another celebrity from a generous list of 1,000 other celebrities is only 10^{-6} . Therefore, if dating is driven by random encounters, celebrities would never marry each other. Yet, they do, with some puzzling implications.

Even if you do not care about the dating habits of celebrities, we must pause and explore what this phenomenon tells us about the structure of the social network. Hollywood celebrities, political leaders, and CEOs of major corporations tend to know an exceptionally large number of individuals and are known by even more. They are hubs. Hence celebrity dating is a manifestation of an interesting property of social network: hubs tend to have ties to other hubs.

As obvious this may sound, this property is not present in all networks. Consider for example the protein-interaction network of yeast, shown in Fig. 7.2. Each node corresponds to a protein and a link between two proteins indicates a binding interaction. A quick inspection of the network reveals its scale-free nature: numerous one- and two-degree proteins coexist with a few highly connected hubs. These hubs, however, tend avoid linking to *each other*. They link instead to many small-degree nodes, generating a hub-and-spoke pattern. This is particularly obvious for the two hubs highlighted in Fig. 7.2: they almost exclusively interact with small-degree proteins while avoiding linking to each other.



Figure 7.1
Hubs Dating Hubs

Celebrity couples, offering a vivid demonstration that in social networks hubs tend to know, date and marry each other (Images from <http://www.whosdatedwho.com>).

A brief calculation illustrates how unusual this pattern is. Let us assume that each node chooses randomly the nodes it connects to. Therefore the probability that two nodes with degree k and k' link to each other is

$$p_{k,k'} = \frac{kk'}{2L}. \quad (7.1)$$

[Eq. 7.1](#) tells us that hubs, by the virtue of the many links they have, are much more likely to connect to each other than to small degree nodes. Yet, the hubs highlighted in [Fig. 7.2](#) almost exclusively connect to degree one nodes. By itself this is not unexpected: [Fig. 7.1](#) also predicts that a hub with $k = 56$ connections should link to $N_1 P_{1,56} \approx 12$ nodes with degree 1. The problem is that this hub connects to 46 degree one neighbors, i.e. four times the expected number.

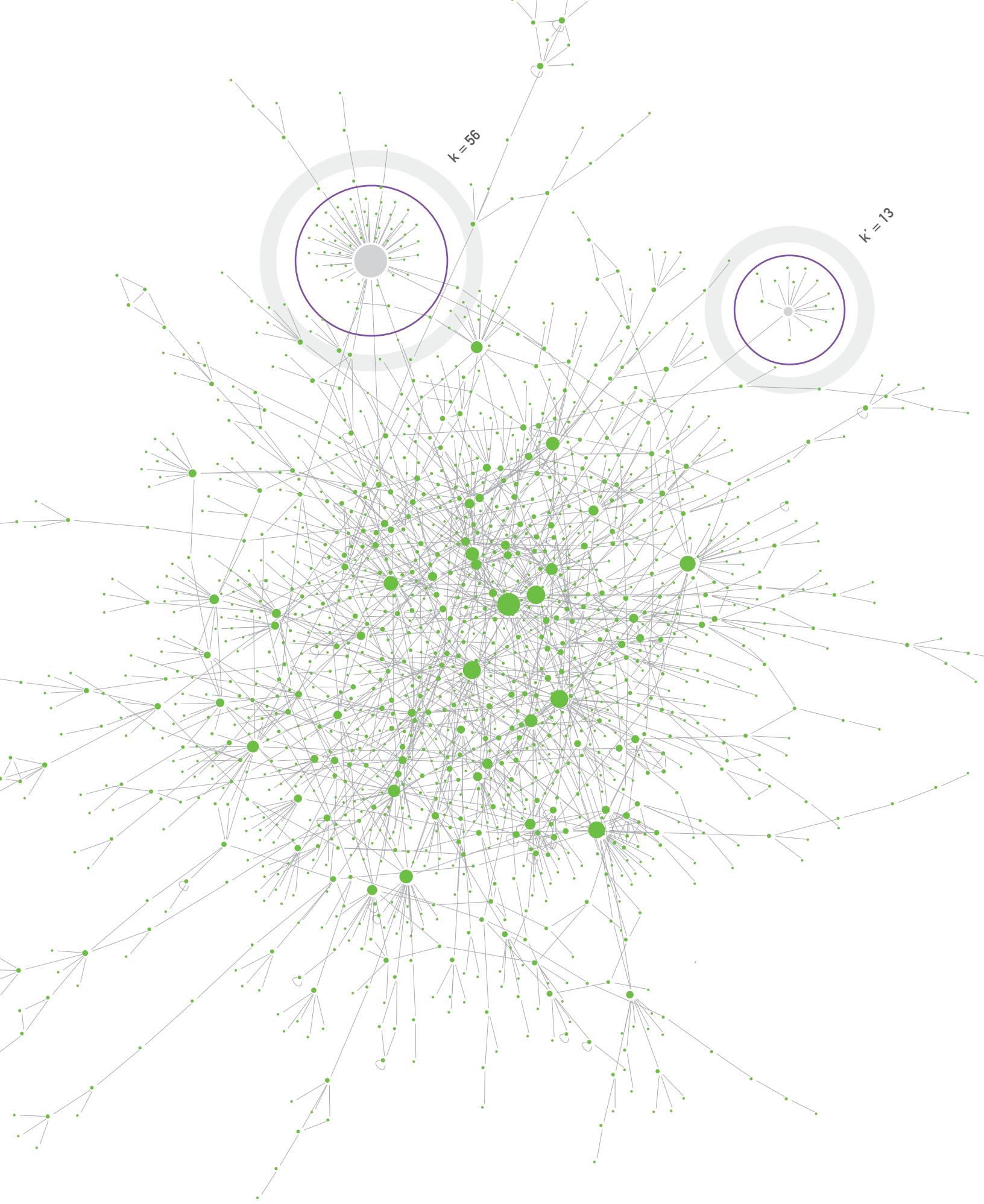
Furthermore, the likelihood that two largest hubs with degrees $k=56$ and $k'=13$ have a direct link between them [Fig. 7.2](#), is $p_{k,k} = 0.15$, which is 400 times larger than $p_{1,2} = 0.0004$, the likelihood that a degree-two node links to a degree-one node. Yet, there are no direct links between the hubs in [Fig. 7.2](#), but we observe numerous direct links between small degree nodes.

In summary, while in social networks hubs tend to “date” each other, in the protein interaction network the opposite is true: the hubs avoid linking to other hubs. While it is dangerous to extrapolate generic principles from two examples, the purpose of this chapter is to show that these patterns are manifestations of a general property of real networks: they exhibit a phenomena called *degree correlations*. We discuss how to measure such degree correlations and explore their impact on the network topology.

[Figure 7.2 \(folowing page\)](#)
Hubs Avoiding Hubs

The protein interaction map of yeast. Each node corresponds to a protein and two proteins are linked if there is experimental evidence that they can bind to each other in the cell. The two largest hubs, with degrees $k = 56$ (left) and $k' = 13$ (right) are highlighted in the figure. They both connect to many small degree nodes and avoid linking to each other.

The network has $N = 1,870$ proteins connected by $L = 2,277$ links, representing one of the earliest protein interaction maps [1, 2]. Only the largest component is shown. Note that the protein interaction network of yeast, discussed in [TABLE 4.1](#), represents a later, more detailed map. Hence, it contains more nodes and links than the network shown in this figure. Redrawn after [3].



ASSORTATIVITY AND DISASSORTATIVITY

Just by the virtue of the many links they have, hubs are expected to link to each other. Yet, as we have seen in the previous section, in some networks they do, in others they don't. This is illustrated in Fig. 7.3, that shows three networks with identical degree sequence but different topology:

- **Neutral Network**

Fig. 7.3b shows a network whose wiring is truly a random. The network of Fig. 7.3b is *neutral*, meaning that the number of links between the hubs coincides with what we expect by chance, as predicted by Eq. 7.1. For clarity we highlighted in red the five largest nodes and the direct links between them, observing a few red links as the likelihood that two nodes link to each other increases with their degree.

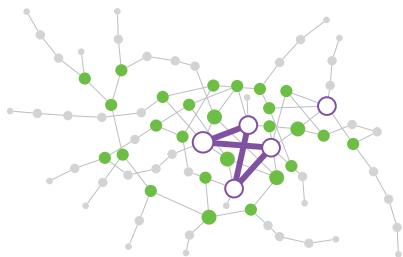
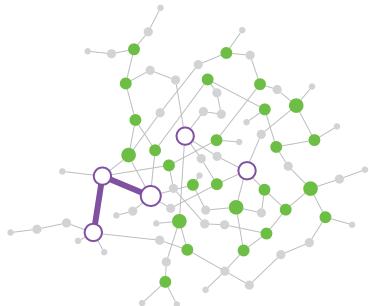
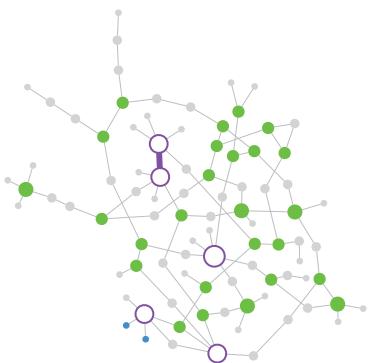
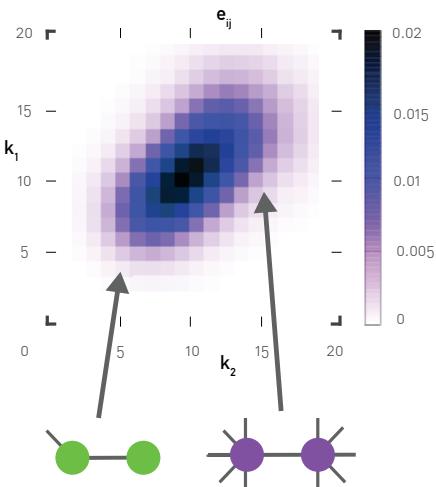
- **Assortative Network**

The network of Fig. 7.3a has precisely the same degree sequence as the one in Fig. 7.3b. Yet, there is a noticeable difference between the two networks: the hubs in Fig. 7.3a tend to link to each other, while avoiding linking to small-degree nodes. At the same time the small-degree nodes tend to connect to other small-degree nodes. Networks displaying such trends are *assortative*. An extreme manifestation of this pattern is a perfectly assortative network, in which degree- k nodes connect only to other degree- k nodes Fig. 7.4.

- **Disassortative Network**

We observe the opposite trend in Fig. 7.3c, where the hubs completely avoid each other, linking mainly to small-degree nodes. Consequently the network displays a hub and-spoke character, making it *disassortative*.

In general a network displays degree correlations if the number of links between the high and low-degree nodes is systematically different from what is expected by chance. In other words, in correlated networks the number of links between nodes of degrees k and k' deviates from Eq. 7.1.

a ASSORTATIVE**b** NEUTRAL**c** DISASSORTATIVE**d****Figure 7.3****Degree correlation matrix**

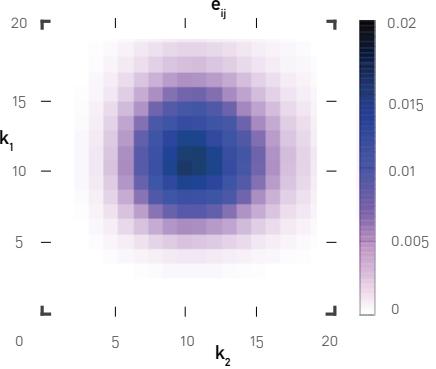
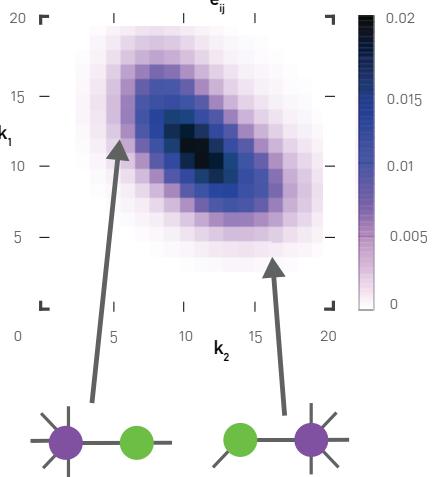
a, b, c Three networks that have precisely the same degree distribution (Poisson p_k), but display different degree correlations. We show only the largest component and we highlight the five nodes with the highest degree in red, together with the direct links between them.

d, e, f The degree correlation matrix e_{ij} for (d) an assortative, (e) a neutral and (f) a disassortative network with Poisson degree distribution and $N=1,000$, and $\langle k \rangle=10$. The colors correspond to the probability that there is a link between nodes with degrees k_1 and k_2 .

a, d For assortative networks e_{ij} takes higher values along the main diagonal. This indicates that nodes of similar degree tend to link to each other: small-degree nodes to small-degree nodes and hubs to hubs. The network in (a) illustrates this by having numerous links between its hubs.

b, e In neutral networks nodes link to other nodes randomly. Hence, the density of links is symmetric around the average degree, indicating the lack of correlations in the linking pattern.

c, f In disassortative networks e_{ij} is higher along the secondary diagonal, indicating that hubs tend to connect to small-degree nodes, and small-degree nodes to hubs. This is illustrated by the hub and spoke character of the network in (c).

e**f**

The complete information about potential degree correlations is contained in the degree correlation matrix, e_{ij} , which is the probability of finding a node with degrees i and j at the two ends of a randomly selected link. As e_{ij} is a probability, it obeys the normalization condition

$$\sum_{i,j} e_{ij} = 1. \quad (7.2)$$

In SECT. 5.8 we derived the probability q_k that there is a degree- k node at the end of the randomly selected link Eq. 5.29.

$$q_k = \frac{kp_k}{\langle k \rangle} \quad (7.3)$$

We can connect q_k to e_{ij} via

$$\sum_j e_{ij} = q_i. \quad (7.4)$$

In neutral networks, we expect

$$e_{ij} = q_i q_j. \quad (7.5)$$

A network displays degree correlations if e_{ij} deviates from the random expectation captured by Eq. 7.5, Eqs. 7.2 - 7.5 are valid for networks with an arbitrary degree distribution, hence they apply to both random and scale-free networks. Given that e_{ij} contains the complete information about potential degree correlations, we start with its visual inspection. Figs. 7.3 d, e, f shows e_{ij} for an assortative, a neutral and a disassortative network. In a neutral network small and high-degree nodes connect to each other randomly, hence e_{ij} lacks any trend Fig. 7.3e. In contrast, assortative networks show high correlations along the main diagonal, indicating that nodes predominantly connect to other nodes with comparable degree. Therefore low-degree nodes tend to link to other low-degree nodes and hubs to hubs Fig. 7.3d. In disassortative networks e_{ij} displays the opposite trend: it has high correlations along the secondary diagonal, indicating that high-degree nodes tend to connect to low-degree nodes Fig. 7.3f.

In summary information about degree correlations is carried by the degree correlation matrix e_{ij} . Yet, the study of degree correlations through the inspection of e_{ij} has numerous disadvantages:

- It is difficult to extract information from the visual inspection of a matrix.
- Unable to infer the magnitude of the correlations, it is difficult to compare networks with different correlations.
- e_{jk} contains approximately k^2_{max} independent variables, representing a huge amount of information that is difficult to model in analytical calculations and simulations.

We therefore need to develop a more compact way to detect the presence and the magnitude of degree correlations.

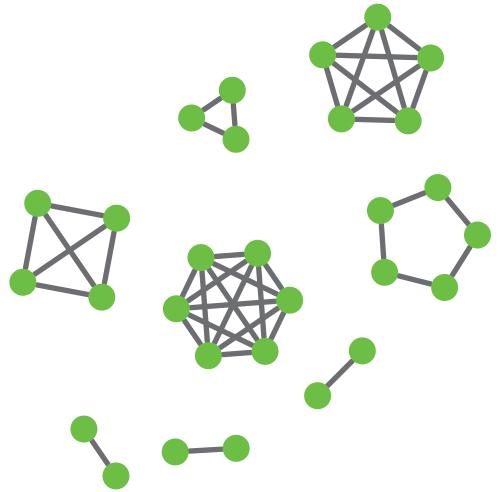


Figure 7.4
A perfectly associative network

Maximal assortativity is obtained when each degree- k node links only to other degree- k nodes. For such a perfectly assortative network $e_{jk} = \delta_{jk} q_k$, where δ_{jk} is the Kronecker delta. In this case the non-diagonal elements of the e_{jk} matrix are zero. The figure shows such a perfectly assortative network, consisting of complete k -clusters.

MEASURING DEGREE CORRELATIONS

While e_{ij} contains the complete information about the potential degree correlations characterizing a network, it is difficult to interpret its content. The purpose of this section is to introduce the degree correlation function, which offers a simpler way to measure degree correlations.

Degree correlations capture the relationship between the degrees of nodes that link to each other. One way to quantify their magnitude is to measure for each node i the average degree of its neighbors Fig. 7.5.

$$k_{nn}(k_i) = \frac{1}{k_i} \sum_{j=1}^N A_{ij} k_j. \quad (7.6)$$

If we wish to calculate Eq. 7.6 for all nodes with the same degree k , we define the degree correlation function as [4, 5]

$$k_{nn}(k) \equiv \sum_{k'} k' P(k' | k) \quad (7.7)$$

where $P(k' | k)$ is the conditional probability that following a link of a k -degree node we reach a degree- k' node. To quantify degree correlations we inspect the dependence of $k_{nn}(k)$ on k . For *neutral networks*, using Eqs. 7.3-7.5, we have

$$P(k' | k) = \frac{e_{kk'}}{\sum_{k'} e_{kk'}} = \frac{e_{kk'}}{q_k} = \frac{q_k q_{k'}}{q_k} = q_{k'}. \quad (7.8)$$

Hence $k_{nn}(k)$ can be expressed as

$$k_{nn}(k) = \sum_{k'} k' q_{k'} = \sum_{k'} k' \frac{k' p(k')}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}. \quad (7.9)$$

Therefore, in a neutral network the average degree of a node's neighbors is independent of the node's degree k and depends only on $\langle k \rangle$ and $\langle k^2 \rangle$. So plotting $k_{nn}(k)$ in function of k is expected to result in a horizontal line at $\langle k^2 \rangle / \langle k \rangle$, as observed in the case of the power grid in Fig. 7.6b. Eq. 7.9 also reflects an intriguing property of real networks: that our friends are more popular than we are, a phenomenon called the *friendship paradox* BOX 7.1.

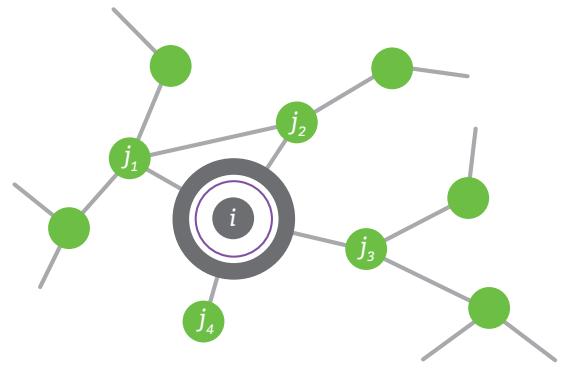


Figure 7.5
Nearest neighbor degree: K_{nn}

To determine $k_{nn}(k)$, we calculate the average degree of a node's neighbors. The figure illustrates the calculation of $k_{nn}(k)$ for node i shown in red. As the degree of the node i is $k_i = 4$, by averaging the degree of its neighbors j_1, j_2, j_3 and j_4 , we obtain $k_{nn}(4) = (4 + 3 + 3 + 1)/4 = 2.75$.

- **Assortative Network**

In this case hubs tend to connect to other hubs, hence the higher is the degree k of a node, the higher should be the average degree of its nearest neighbors. Consequently for assortative networks $k_{nn}(k)$ increases with k , as observed in collaboration networks in Fig. 7.6a.

- **Disassortative Network**

In this case hubs prefer to link to low-degree nodes. Consequently $k_{nn}(k)$ decreases with k , as observed for the protein-protein interaction network Fig. 7.6c.

The scaling observed in Fig. 7.6 prompts us to approximate the degree correlation function with [4]

$$k_{nn}(k) = ak^{\mu}. \quad (7.10)$$

If Eq. 7.10 holds, then the nature of degree correlations characterizing a network is determined by the sign of the *correlation exponent* μ :

- **For assortative Networks $\mu > 0$**

Indeed, a fit to $k_{nn}(k)$ for the science collaboration network provides $\mu = 0.37 \pm 0.11$ Fig. 7.6a.

- **For neutral networks we have $\mu = 0$**

As according to Eq. 7.9 $k_{nn}(k)$ is independent of k . For the power grid we obtain $\mu = 0.04 \pm 0.05$, which is indistinguishable from zero Fig. 7.6b.

- **For disassortative networks we expect $\mu < 0$**

Indeed, for the metabolic network we obtain $\mu = -0.76 \pm 0.04$ Fig. 7.6c.

In summary, the degree correlation function helps us capture the presence or absence of correlations in real networks. The $k_{nn}(k)$ function also plays an important role in analytical calculations, allowing us to calculate the impact of degree correlations on various network characteristics SECT. 7.6. Note that it is often convenient to extract a single number to capture the magnitude of correlations present in a network. This can be achieved either through the correlation exponent μ defined in Eq. 7.10, or using the degree correlation coefficient discussed in BOX 7.2.

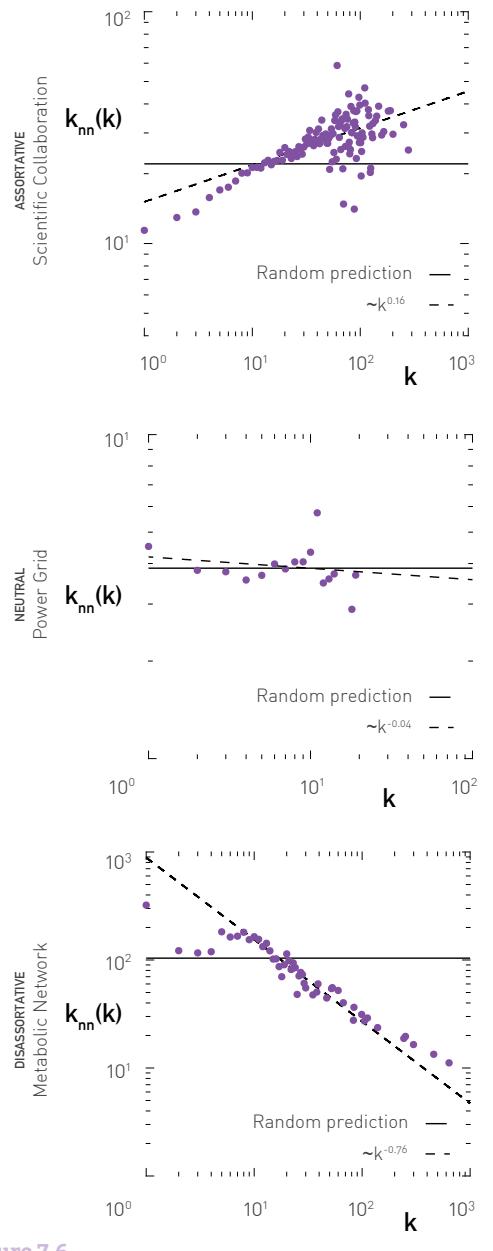


Figure 7.6

Degree correlation function

The degree correlation function $k_{nn}(k)$ for three real networks. The panels show $k_{nn}(k)$ on a log-log plot to test the validity of Eq. 7.10.

- (a) Collaboration network of astrophysicists. The increasing $k_{nn}(k)$ with k indicates that the network is assortative.
- (b) Power grid. The horizontal $k_{nn}(k)$ indicates the lack of degree correlations, as predicted by Eq. 7.9 for neutral networks.
- (c) Metabolic network. The decreasing $k_{nn}(k)$ documents the network's disassortative nature.

On each panel the horizontal dotted line corresponds to the prediction Eq. 7.9 and the oblique dashed line is a fit to Eq. 7.10. The slope in (a) is $\mu = 0.37$, in (b) is $\mu = 0.04$ while the slope in (c) is $\mu = -0.76$.

BOX 7.1

FRIENDSHIP PARADOX

While most people believe that they have more friends than their friends [7], the friendship paradox, discovered by sociologist Scott L. Feld, states the opposite: on average your friends are more popular than you are [6].

The roots of the friendship paradox is Eq. 7.9, telling us that the average degree of a node's neighbors is not simply $\langle k \rangle$, but depends on $\langle k^2 \rangle$ as well. Consider for example a random (Erdős-Rényi) network, for which $\langle k^2 \rangle = \langle k \rangle(1 + \langle k \rangle)$. According to Eq. 7.9

$$k_{nn}(k) = 1 + \langle k \rangle. \quad (7.11)$$

Therefore the average degree of a node's neighbors is always higher than the average degree of the network $\langle k \rangle$. The gap between $\langle k \rangle$ and our friends' degree can be particularly large in scale-free networks, for which $\langle k^2 \rangle/\langle k \rangle$ is significantly larger than $\langle k \rangle$ Fig. 4.7. Consider for example the email network, for which $\langle k^2 \rangle/\langle k \rangle = 390.45$, or the actor network, for which $\langle k^2 \rangle/\langle k \rangle = 565.70$. Hence in these networks the average degree of the friends of a randomly selected node can be hundreds of times higher than the expected degree of the node itself, which is $\langle k \rangle$.

To understand the origin of the friendship paradox, we must realize that for a randomly chosen node, the degree distribution of the nodes at the other end of each link do not follow p_k , but are biased towards higher-degree nodes, as indicated by Eq. 7.3. In other words, we are more likely to be friends with hubs than with small-degree nodes, simply because hubs have more friends than the small-nodes. Hence our friends do not reflect the whole population - they are biased towards the hubs.

STRUCTURAL CUTOFFS

Throughout this book we assumed that the networks we explore are simple, meaning that there is at most one link between any two nodes [CHAPTER 2](#). For example, in the email network we place a single link between two individuals that are in email contact, despite the fact that they may have exchanged multiple messages; in the actor network we connect two actors with a single link if they acted together, independent of the number of movies they jointly made. All datasets discussed in [TABLE 4.1](#) are simple networks. In simple networks there is a puzzling conflict between the scale-free property and degree correlations [10, 11]. Consider for example the scale-free network of [Fig. 7.7a](#), whose two largest hubs have degrees $k = 55$ and $k' = 46$, connected by a link. In a network with degree correlations $e_{kk'}$ the expected number of links between k and k' is

$$E_{kk'} = e_{kk'} \langle k \rangle N \quad (7.14)$$

For a neutral network $e_{kk'}$ is given by [Eq. 7.5](#), which, using [Eq. 7.3](#), predicts

$$E_{kk'} = \frac{k_k p_k k' p_{k'}}{\langle k \rangle} N = \frac{55 \cdot 46}{\frac{300 \cdot 300}{3}} 300 = 2.8. \quad (7.15)$$

Therefore, given the size of these two hubs, they should be connected to each other by two to three links to comply with the network's neutral nature. Yet, in a simple network we are allowed only one link between them, raising a conflict between degree correlations and the scale-free property. Such conflict emerges in a simple network each time the degrees violate the $E_{kk'} \leq 1$ condition. The goal of this section is to understand the origin and the consequences of this conflict.

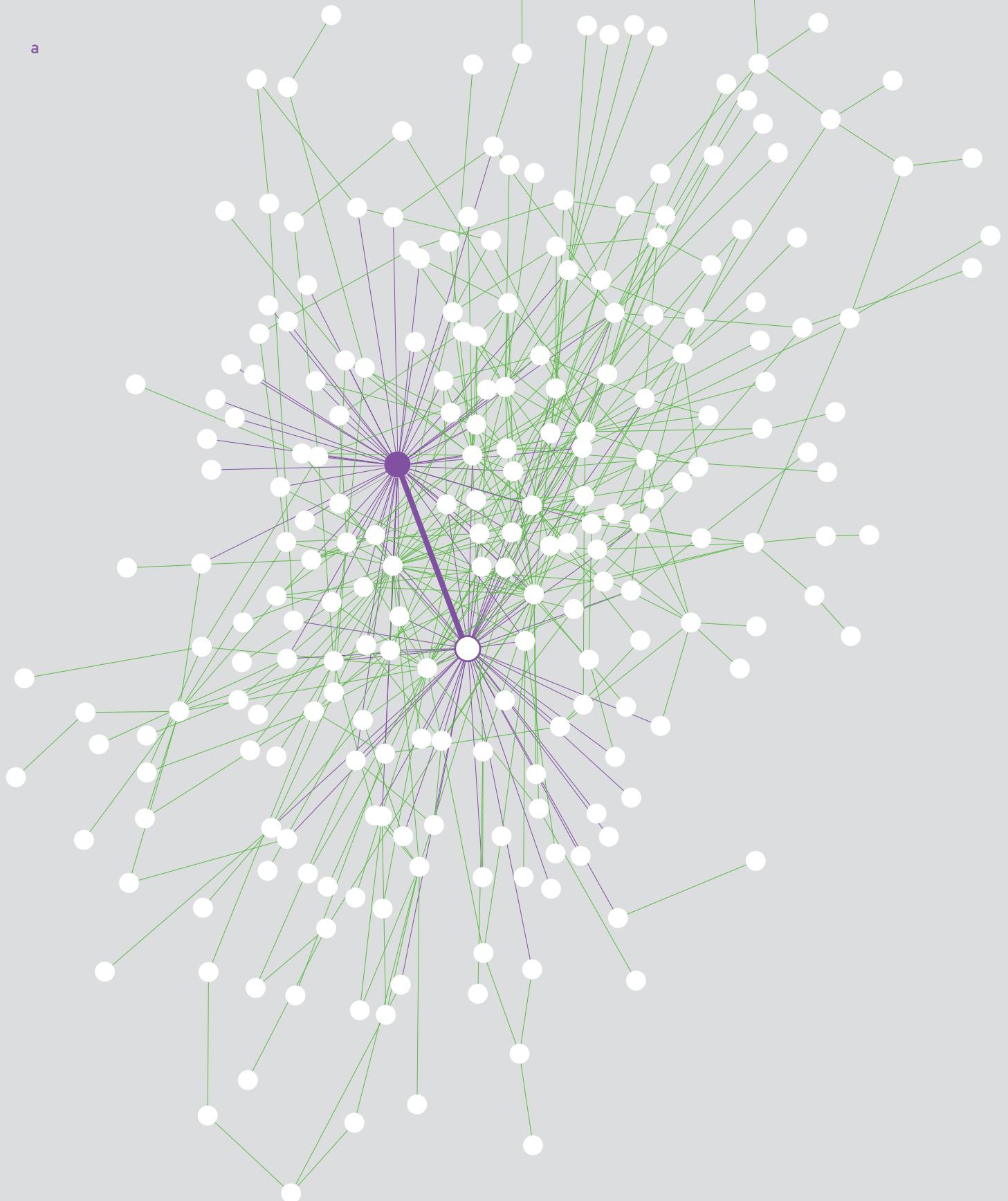
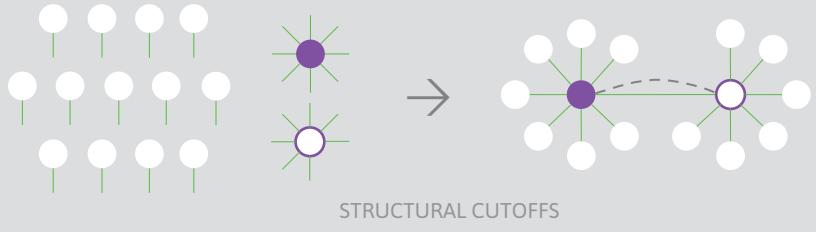
For small k and k' [Eq. 7.15](#) predicts that $E_{kk'}$ is also small, i.e. we expect less than one link between the two nodes. Only for nodes whose degree exceeds some threshold k_s will [Eq. 7.15](#) predict multiple links. As we show in [ADVANCED TOPICS 7.B](#), this k_s , that we

[Figure 7.7 \(following page\)](#)

Structural disassortativity

(a) A scale-free network with $N=300$, $L=450$, according to [Eq. 7.15](#), and $\gamma=2.2$, generated by the configuration model, while forbidding self-loops and multiple links between two nodes, making the network simple. The blue and the red nodes are the two largest nodes in the network and are connected by the red link. As [Eq. 7.15](#) predicts, to maintain the network's neutral nature, we would need two to three links between these two nodes. The fact that we do not allow multiple links (simple network representation) makes the network disassortative, a phenomena we call structural disassortativity.

(b) To illustrate the origins of structural correlations, we start from a fixed degree sequence, shown as stubs on the left, and we randomly connect the stubs (configuration model). In this case, the expected number of links between the nodes with degree 8 and 7 is $8 \times 7 / 28 \approx 2$. Yet, if we do not allow multilinks, there can only be one link, making the network structurally disassortative.

a**b**

$$k_s(N) \sim (\langle k \rangle N)^{1/2}. \quad (7.16)$$

In other words, nodes whose degree exceeds Eq. 7.16 are expected to have $E_{kk} > 1$, a conflict that as we show below gives rise to degree correlations.

To fully understand the consequences of the described conflict, we must first ask if a network has nodes whose degrees exceeds Eq. 7.16. For this we compare the structural cutoff, k_s , with the natural cutoff, k_{max} , which is the expected largest degree in a network with degree distribution p_k . According to Eq. 7.14, for a scale-free network $k_{max} \sim N^{\frac{1}{\gamma-1}}$. The relative magnitude of k_{max} vs. k_s , gives raise to two regimes:

- For scale-free networks with $\gamma \geq 3$ and random networks, k_s is always larger than k_{max} , hence we lack nodes for which $E_{kk} > 1$.
- For scale-free networks with $\gamma < 3$, k_s is smaller than k_{max} , hence all nodes between k_s and k_{max} violate $E_{kk} > 1$. Consequently, the network has fewer links between its hubs than expected based on Eq. 7.15. As a result, these networks will be disassortative, a phenomenon we call *structural disassortativity*. This is illustrated in Figs. 7.8a, b that show a simple scale-free network generated by the configuration model. The network shows disassortative tendencies, despite the fact that we did not impose degree correlations.

We have two avenues to generate networks that are free of structural disassortativity:

- (i) We relax the simple network requirement, allowing multiple links between the nodes. The conflict disappears and the network will be neutral Figs. 7.8c, d.
- (ii) If we insist of having a simple scale-free network that is neutral or associative, we must remove all hubs with degrees larger than k_s . This is illustrated in Fig. 7.8 e, f: the obtained network, missing nodes with $k \geq 100$, is neutral.

How can we convince ourselves that the correlations observed in a particular network are a consequence of structural dissasortativity, or are generated by some unknown process? Degree-preserving randomization Fig. 4.14 helps us distinguish these two possibilities:

- (i) Degree preserving randomization with simple links (R-S): We apply degree-preserving randomization to the original network, while making sure that we do not allow for more than one link between any pair of nodes. On the algorithmic side this means that each rewiring that results in multiple links between two nodes is discarded. If the real $k_{nn}(k)$ and the randomized $k_{nn}^{R-S}(k)$ are indistinguishable, then the correlations observed in a real system are all structural,

BOX 7.2

DEGREE CORRELATION COEFFICIENT

If we wish to characterize degree correlations using a single number, we can also use the degree correlation coefficient, introduced by Mark Newman and defined as [8,9]

$$r = \sum_{jk} \frac{jk(e_{jk} - q_j q_k)}{\sigma_r^2} \quad (7.12)$$

with

$$\sigma_r^2 = \sum_k k^2 q_k - \left[\sum_k k q_k \right]^2 \quad (7.13)$$

Hence r is the Pearson correlation coefficient between the degrees at the two end of the same link. It varies between $-1 \leq r \leq 1$: for $r < 0$ the network is assortative, for $r = 0$ the network is neutral and for $r > 0$ the network is disassortative. For example, for the collaboration network we obtain $r = 0.13$, in line with its assortative nature; for the protein interaction network $r = -0.04$, supporting its disassortative nature and for the power grid $r = 0$. Note that the degree correlation coefficient r assumes that $k_{nn}(k)$ is a linear function of k with slope r . In contrast the correlation exponent μ assumes that $k_{nn}(k)$ follows the power law Eq. 7.10. Naturally, both scaling laws cannot be valid simultaneously. The analytical models of SECT. 7.7 offer some guidance, supporting the validity of Eq. 7.10. As we show in ADVANCED TOPICS 7.B, while r correlates with μ , we need to be cautious when we use it to measure degree correlations.

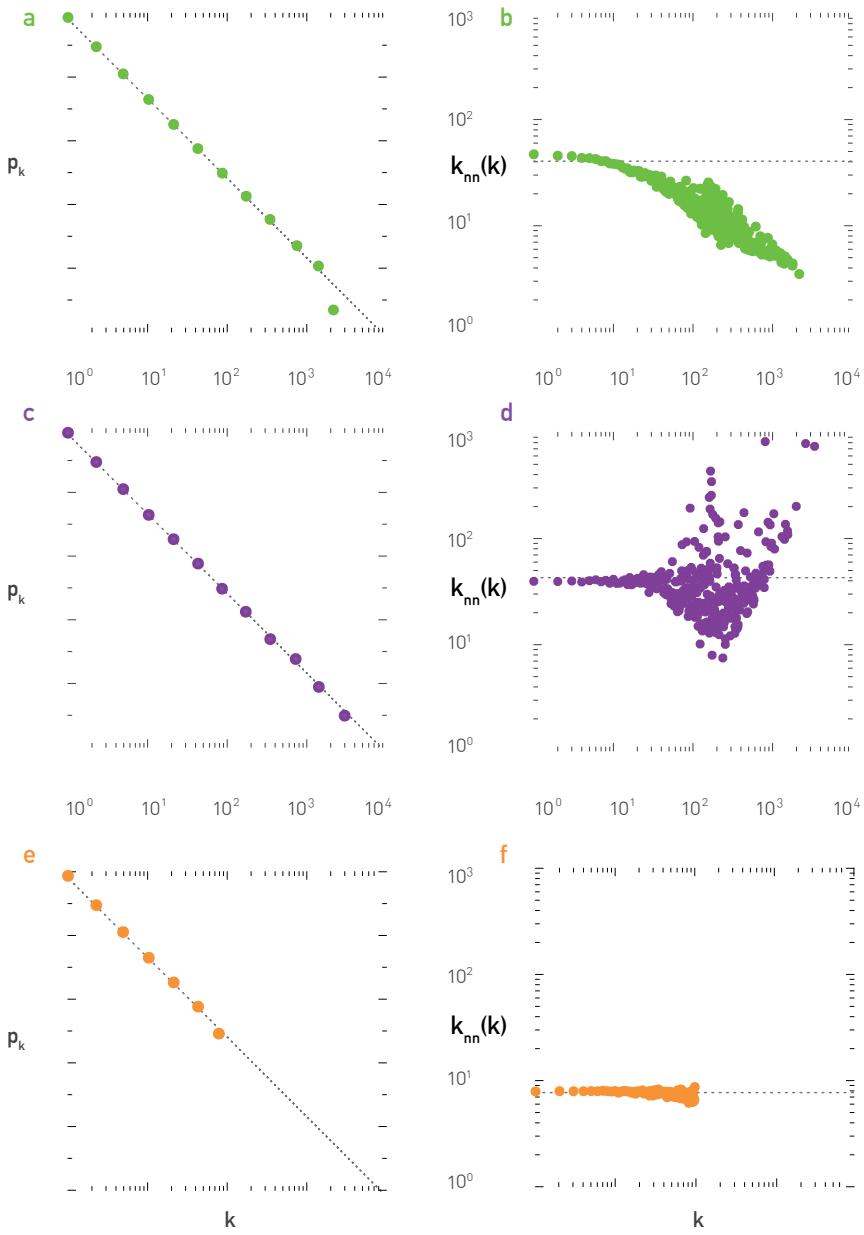


Figure 7.8
Natural and structural cutoffs

The figure illustrates the tension between the scale-free property and degree correlations. It shows the degree distribution (left panels) and the degree correlation function $k_{nn}(k)$ (right panels) of a scale-free network with $N = 10,000$ and $\gamma = 2.5$, generated by the configuration model.

(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 decays.

(c, d) We can eliminate structural disassortativity by allowing multiple links, i.e. relaxing the simple network requirement. As shown in (c,d), in this case we obtain a neutral scale-free network.

(e, f) If we artificially impose an upper cutoff by removing all nodes with $k \geq k_s$ predicted by Eq. 7.16, the network becomes neutral, as seen in (f).

fully explained by the degree distribution. If the randomized $k_{nn}^{R-S}(k)$ does not show degree correlations while $k_{nn}(k)$ does, there is some unknown process that generates the observed degree correlations.

- (ii) Degree preserving randomization with multiple links (*R-M*): For a self-consistency check it is useful to also perform degree-preserving randomization that allows for multiple links between the nodes. On the algorithmic side this means that we allow each random rewiring, even if they lead to multiple links. This process eliminates all degree correlations.

We have taken the three networks of in Fig. 7.6 and performed the randomizations discussed above. As Fig. 7.9a shows, the assortative nature of the scientific collaboration network disappears under both randomizations. This indicates that the observed assortative correlations are not linked to the scale-free nature of the underlying network. In contrast, for the metabolic network the observed disassortativity remains unchanged under *R-S* Fig. 7.9c. This indicates that the disassortativity of the metabolic network is structural, induced by its degree distribution.

In summary, the scale-free property can induce disassortativity in simple networks. To be specific, in neutral or assortative networks we expect multiple links between the hubs. If such multiple links are forbidden (simple graph), the network will display disassortative tendencies. This conflict vanishes for scale-free networks with $\gamma \geq 3$ and for random networks. It also vanishes if we allow for multiple links between the nodes.

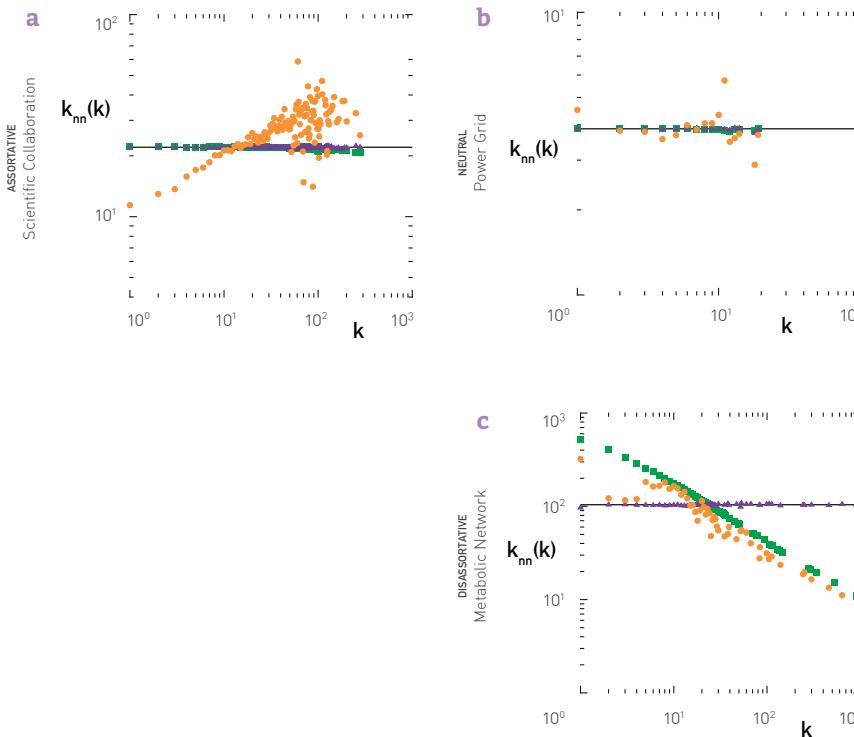


Figure 7.9
Randomization and degree correlations

To uncover the origin of the observed degree correlations, it is useful to compare $k_{nn}(k)$ with $k_{nn}^{R-S}(k)$ and $k_{nn}^{R-M}(k)$ obtained after degree-preserving randomization. We perform two different randomizations for this purpose.

Green symbols

Degree-preserving randomization with simple links (*r-s*), in which case at each step of the randomization process we check that we do not have more than one link between any node pairs.

Blue symbols

Degree-preserving randomization with multiple links (*R-M*), in which case we allow multiple links during the randomization processes.

We performed these two randomizations for the networks of Fig. 7.6. The *R-M* procedure always generates a neutral network, consequently $k_{nn}^{R-M}(k)$ is always horizontal. The true insight is provided when we compare $k_{nn}(k)$ with $k_{nn}^{R-S}(k)$, allowing us to decide if the observed correlations are structural:

(a) Scientific collaboration network

The increasing $k_{nn}(k)$ differs from the horizontal $k_{nn}^{R-S}(k)$, indicating that the network's assortativity is not structural (i.e. it is not a consequence of the degree distribution), but it is generated by some process that governs the network's evolution. This is not unexpected: structural effects can generate only disassortativity, not assortativity.

(b) Power grid

The horizontal $k_{nn}(k)$, $k_{nn}^{R-S}(k)$ and $k_{nn}^{R-M}(k)$ all support the lack of degree correlations (neutral network).

(c) Metabolic network

As both $k_{nn}(k)$ and $k_{nn}^{R-S}(k)$ decrease, we conclude that the network's disassortative nature is induced by its scale-free property. Hence the observed degree correlations are structural.

DEGREE CORRELATIONS IN REAL NETWORKS

To truly understand the prevalence of degree correlations, we need to inspect the correlations characterizing various real networks. Therefore, in Fig. 7.10 we show the $k_{nn}(k)$ function for the ten reference networks of TABLE 4.1. Let us discuss the observed behavior:

- **Power grid**

For the power grid $k_{nn}(k)$ is flat and indistinguishable from its randomized version, indicating a lack of degree correlations Fig. 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 Fig. 7.10b. The degree correlations vanish in the randomized networks. Hence the Internet is assortative, but structural cutoffs eliminate the effect for high k .

- **Social Networks**

The three networks capturing social phenomena, like the mobile phone network, science collaboration networks and actor network, all have an increasing $k_{nn}(k)$, indicating that they are assortative Figs. 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. For each of these networks the observed $k_{nn}(k)$, differs from the $k_{nn}^{R-S}(k)$, indicating that their assortative nature is not rooted in the degree distribution.

- **Email Network**

While the email network is often used as an example of a social network, its $k_{nn}(k)$ decreases with k , documenting a clear disassortative behavior Fig. 7.10f. The randomized $k_{nn}^{R-S}(k)$ 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 nega-

tive μ , suggesting that these networks are disassortative [Eq. 7.10](#). Yet, the scaling of $k_{\min}^{R-S}(k)$ is indistinguishable from $k_{nn}(k)$, indicating that we are observing structural disassortativity, rooted in the scale-free nature of these networks [Fig. 7.10g, h](#).

- **WWW**

The decaying $k_{nn}(k)$ implies disassortative correlations [Fig. 7.10i](#). The randomized $k_{\min}^{R-S}(k)$ 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 $k \leq 20$, $k_{nn}(k)$ shows a clear assortative trend; for $k > 20$, however, we observe equally clear disassortative scaling [Fig. 7.10j](#). Such mixed behavior can emerge in networks that display extreme assortativity [SECT. 7.6](#). This suggests that the citation network is strongly assortative up to k_s , but its scale-free nature reverses the trend for $k \gg k_s$.

In summary, [Fig. 7.10](#) indicates that to understand degree correlations, we must always compare $k_{nn}(k)$ to the degree randomized $k_{nn}^{R-S}(k)$. It also allows us to draw some interesting conclusions:

- (i) Of the ten reference networks the power grid appears to be the only that is truly neutral. Hence most real networks display degree correlations.
- (ii) All networks that display disassortative tendencies (email, protein, metabolic), do so thanks to their scale-free property. Hence, these are all structurally disassortative. Only the WWW shows disassortative correlations that are only partially explained by its degree distribution.
- (iii) The degree correlations characterizing associative networks are not explained by their degree distribution. Most social networks (mobile phone calls, scientific collaboration, actor network) are in this class and so is the Internet and the citation network.

A number of proposals exist to explain the origin of the observed assortativity. For example, the tendency of individuals to form communities [CHAPTER 9](#) has been shown to induce assortative scaling [12]. Similarly, the society has endless mechanisms, from professional committees to TV shows, to bring hubs together, enhancing the assortative nature of social and professional networks. Finally, homophily, a well documented social phenomena, [13], captures the fact that individuals have a tendency to associate with other individuals of similar background and characteristics. This tendency may also be responsible for the celebrity marriages discussed in [SECT. 7.0](#).

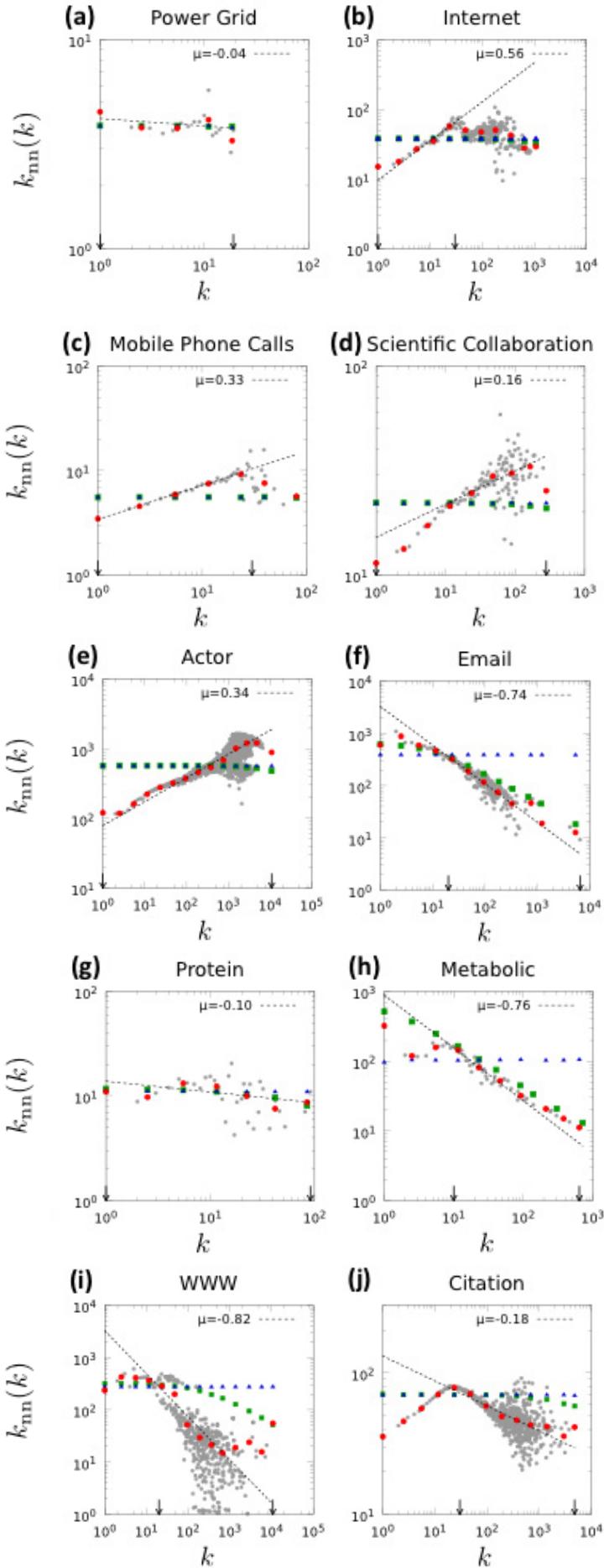


Figure 7.10
Randomization and degree correlations

The degree correlation function $k_{nn}(k)$ for the ten reference networks of [Table 4.1](#). The grey symbols show the $k_{nn}(k)$ function under linear binning; red symbols represent the same data using log-binning [SECT. 4.10](#). The dotted line corresponds to the best fit of the form [Eq. 7.10](#) and the small arrows at the bottom mark the fitting interval. Green squares represent $k_{nn}(k)$ obtained for 100 independent degree-preserving randomizations, making sure that we preserve the simple character of these networks; blue triangles correspond to $k_{nn}(k)$, i.e. randomization that does allow self-loops and multiple links between two nodes. Note that we made directed networks undirected when we measured $k_{nn}(k)$. To fully characterize the correlations emerging in directed networks we must use the directed correlation function [BOX 7.3](#).

BOX 7.3

CORRELATIONS IN DIRECTED NETWORKS

The degree correlation function $k_{nn}(k)$ in Eq. 7.7 is defined for undirected networks. To measure correlations in directed networks we must take into account that each node i is characterized by an incoming k_i^{in} and an outgoing k_i^{out} degree. Hence, we define four degree correlation functions, $k_{nn}^{\alpha, \beta}(k)$, where α and β refer to the in and out indices Figs. 7.11 a-d. In Fig. 7.11e we show $k_{nn}^{\alpha, \beta}(k)$ for citation networks, indicating a lack of in-out correlations, while a detectable assortative scaling for small k for the other three correlations.

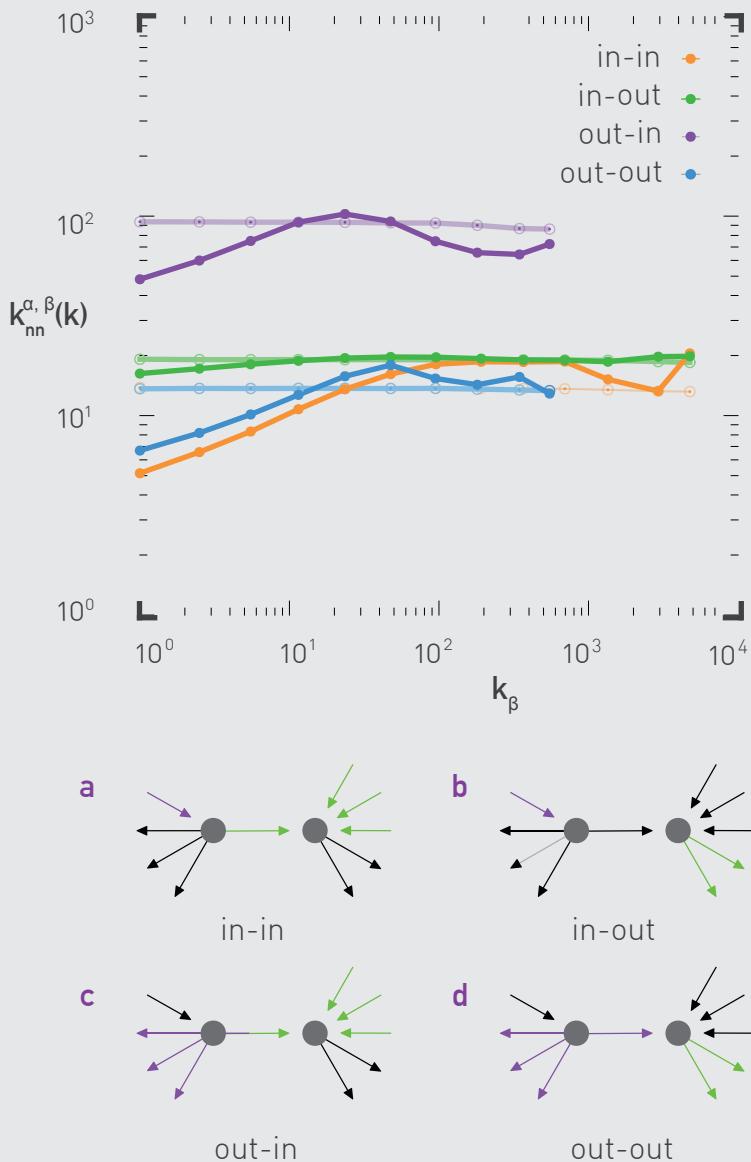


Figure 7.11
Correlation and directed network

Panels (a)-(d) illustrate the four possible correlations in directed networks. We show in red and green the (α, β) indices that define the appropriate correlation function [14]. For example, (a) describes the $k_{nn}^{in, in}(k)$ correlations between the in-degrees of two nodes connected by a link. (e) The $k_{nn}^{\alpha, \beta}(k)$ correlation function for citation networks, a directed network. For example $k_{nn}^{in, in}(k)$ is the average indegree of the in-neighbors of nodes with in-degree k_{in} . These functions show a clear assortative tendency for three of the four function up to degree $k \approx 100$. The lighter symbols capture the degree randomized $k_{nn}^{\alpha, \beta}(k)$ for each correlation function.

GENERATING CORRELATED NETWORKS

To study degree correlations and to explore their impact on various network characteristics, we need to build networks with tunable correlations. Given the conflicts between the scale-free property and degree correlations, this is not a trivial task. In this section we discuss the degree correlations characterizing some well-known network models, together with an algorithm capable of generating networks with tunable correlations.

DEGREE CORRELATIONS IN STATIC MODELS

Erdős-Rényi Model

The random network model is neutral by definition. As it lacks hubs, it does not develop structural correlations either. Hence for the Erdős-Rényi network $k_{nn}(k)$ is given by Eq. 7.9, predicting $\mu = 0$ for any $\langle k \rangle$ and N . *Configuration Model:* The configuration model SECT. 4.7 is also neutral, independent of our choice of the degree distribution p_k . This is because the model allows for both multi links and self-loops. Consequently, any conflicts caused by the hubs are relieved by multiple links between them. If, however, we force the network to be simple, then the generated network will develop structural disassortativity Fig. 7.8.

Hidden Parameter Model

In the model e_{jk} is the product of the hidden variables η_j and $\eta_{j'}$ which are chosen randomly, hence the network is technically uncorrelated SECT. 4.8. However, if we do not allow multiple links, for scale-free networks we again observe structural disassortativity. Analytical calculations indicate that in this case $k_{nn}(k) \sim k^{-1}$, i.e. we have $\mu = -1$ [10].

DEGREE CORRELATIONS IN EVOLVING NETWORKS

To understand the emergence (and absence) of degree correlations in growing networks, let us start with the initial attractiveness model discussed in SECT. 6.4. In the model preferential attachment follows $\Pi(k) \sim A + k$, where A is the initial attractiveness Eq. 6.23. The degree correlation function depends on A , the calculations predicting three scaling regimes [15]:

(i) If $\gamma < 3$ (i.e. $-m < A < 0$ according to Eq. 6.24), we have

$$k_{nn}(k) \approx m \frac{(m+A)^{1-\frac{A}{m}}}{2m+A} \left(\frac{2m}{2m+A} \right) N^{-\frac{A}{2m+A}} k^{\frac{A}{m}} \quad (7.17)$$

Hence the resulting network is disassortative, $k_{nn}(k)$ being characterized by the power-law decay [15, 16]

$$k_{nn}(k) \approx k^{\frac{|A|}{m}} \quad (7.18)$$

(ii) If $\gamma = 3$ ($A = 0$), the initial attractiveness model reduces to the Barabási-Albert model CHAPTER 5. In this case

$$k_{nn}(k) \approx \frac{m}{2} \ln N, \quad (7.19)$$

that is, $k_{nn}(k)$ is independent of k , hence the network is neutral.

(iii) If $\gamma > 3$ ($A > 0$), the calculations predict

$$K_{nn}(k) \approx (m+a) \ln \left(\frac{k}{m+a} \right). \quad (7.20)$$

As $k_{nn}(k)$ increases logarithmically with k , the resulting network displays a weak assortative tendency, but does not follow the scaling Eq. 7.10.

Bianconi-Barabási Model

With a uniform fitness distribution the Bianconi-Barabási model generates a disassortative network [5] Fig. 7.12. As the randomized version of the network is also disassortative, this is a structural disassortativity. Note, however, that the real $k_{nn}(k)$ and the randomized $k_{nn}^{R-S}(k)$ do not overlap, indicating that the Bianconi-Barabási model displays some disassortativity that is not fully explained by its scale-free nature.

TUNING DEGREE CORRELATIONS

Several algorithms exist to generate networks with desired degree correlations [8, 17, 18]. Here we discuss a simplified version of the algorithm proposed by Xalvi-Brunet and Sokolov that generates maximally correlated networks with a predefined degree sequence [19, 20, 21]. It consists of the following steps Fig. 7.13a:

- **Step 1: Link selection**

Choose at random two links. Label the four nodes at the end of these two links with a, b, c , and d such that their degrees k_a, k_b, k_c , and k_d are ordered as

$$k_a \geq k_b \geq k_c \geq k_d$$

- **Step 2: Rewiring**

Break the selected links and rewire them to form new pairs. Depending on the desired degree correlations the rewiring is done in two different ways:

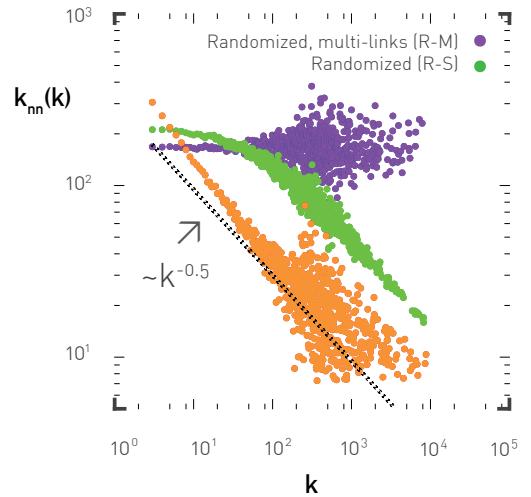


Figure 7.12
Correlations in the Bianconi-Barabási model

The degree correlation function of the Bianconi-Barabási model for $N = 10,000$, $m = 3$ and uniform fitness distribution SECT. 6.2. As the dotted line indicates, the network is disassortative, with $\mu = 0.5$. The green symbols show $k_{nn}^{R-S}(k)$, and the blue are for $k_{nn}^{R-M}(k)$. As $k_{nn}^{R-S}(k)$ also decreases, the bulk of the observed disassortativity is structural. But the difference between $k_{nn}^{R-S}(k)$ and correlations in the Bianconi-Barabási model suggests that structural effects cannot fully account for the observed degree correlation.

- **Step 2A: Assortative**

By pairing the two highest degrees (a with b) and the two lowest degrees (c with d), we are connecting nodes with comparable degrees, enhancing the network's assortative nature.

- **Step 2B, Disassortative**

By pairing the highest and the lowest degree nodes (a with d and b with c), we tend to connect nodes with rather different degrees, enhancing the network's disassortative nature.

By iterating these steps we gradually enhance the network's assortative (2A) or disassortative (2B) features. If we aim to generate a simple network (free of multi-links), after *Step 2* we check whether the particular rewiring leads to multi-links. If it does, we reject it, returning to *Step 1*.

The correlations characterizing the networks generated by this algorithm converge to the maximal or minimal value one can reach for the given degree sequence [Fig. 7.13b](#). We refer to these networks as maximally assortative or maximally disassortative. The model has no difficulty creating disassortative correlations [Figs. 7.13e, f](#). In the assortative limit simple networks displays a mixed $k_{nn}(k)$: assortative for small k and disassortative for high k [Figs. 7.13b](#). This is a consequence of structural cutoff: for scale-free networks the system is unable to sustain assortativity for high k . This behavior is reminiscent of the $k_{nn}(k)$ function observed for citation networks [Fig. 7.10](#).

The version of the Xalvi-Brunet & Sokolov algorithm discussed in [Fig. 7.13](#) generates maximally assortative or disassortative networks. We can tune the magnitude of the generated degree correlations if we use the original version of the proposed algorithm, discussed in [Fig. 7.14](#).

In summary, static models, like the configuration or hidden parameter models, are neutral if we allow multi-links, and develop structural disassortativity if we force them to generate simple networks. To generate networks with tunable correlations, we can use for example the Xalve-Brunet & Sokolov algorithm. An important result of this section is [Eq. 7.17](#), predicts the functional form of the degree correlation function for a growing network, offering analytical backing for the scaling hypothesis [Eq. 7.10](#).

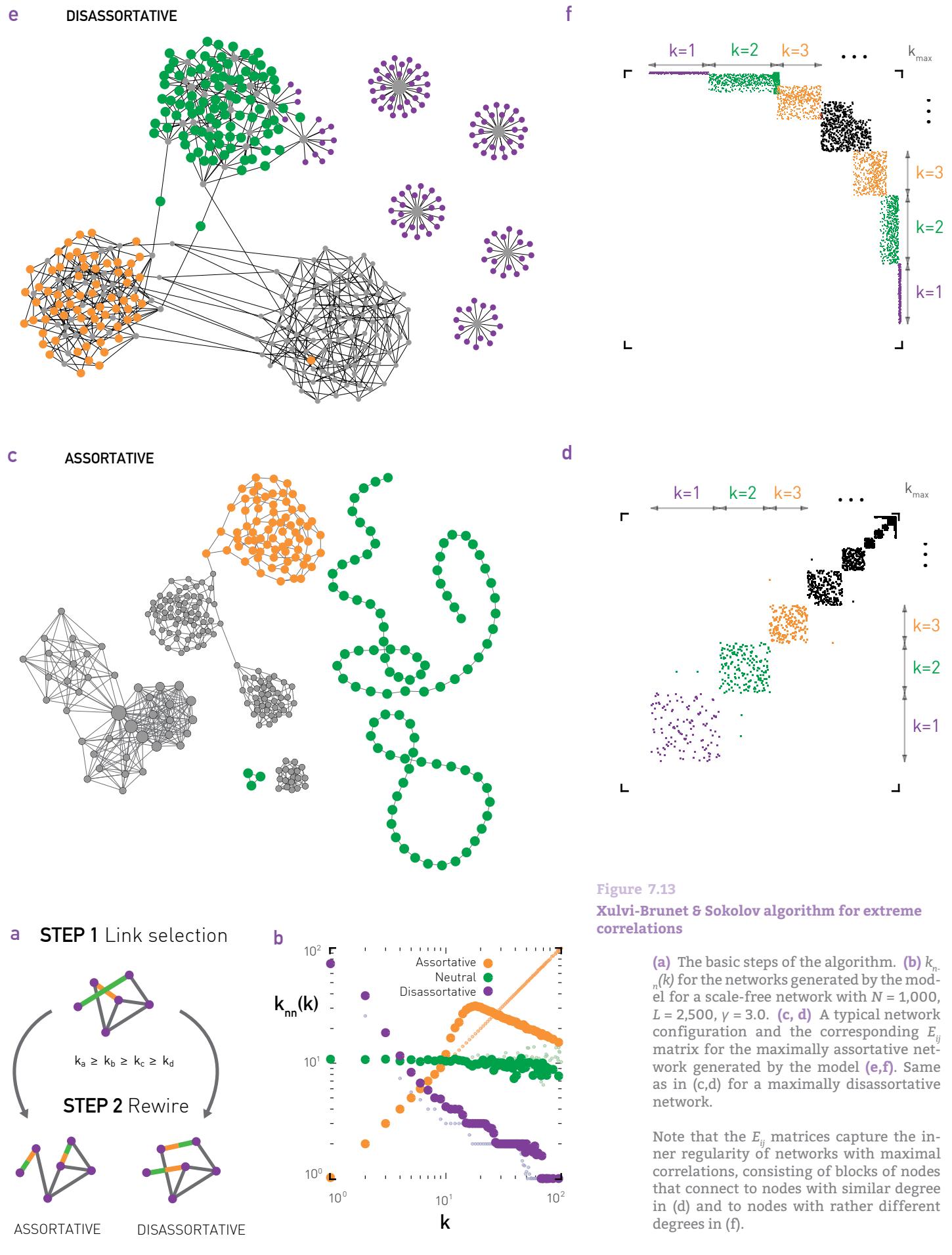
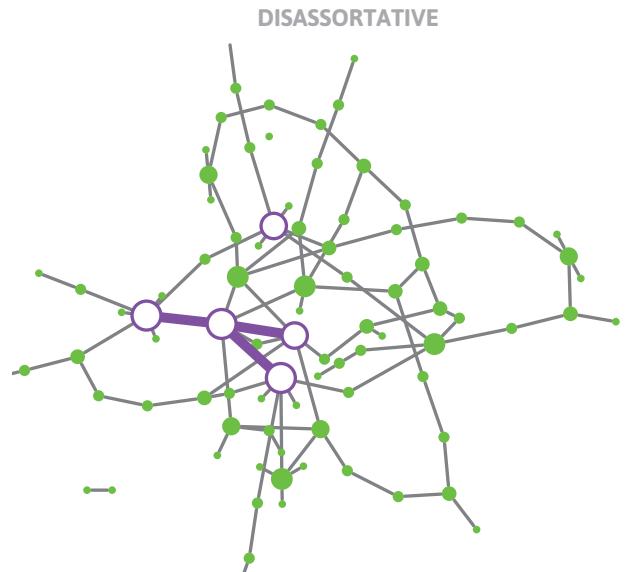
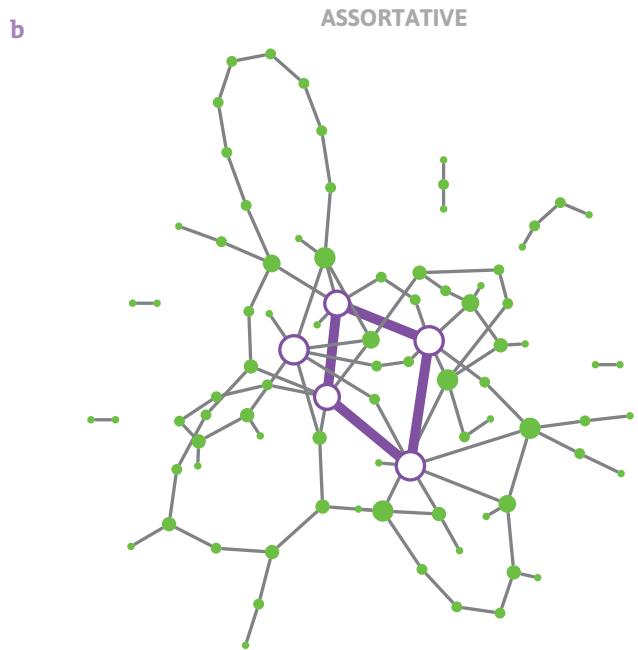
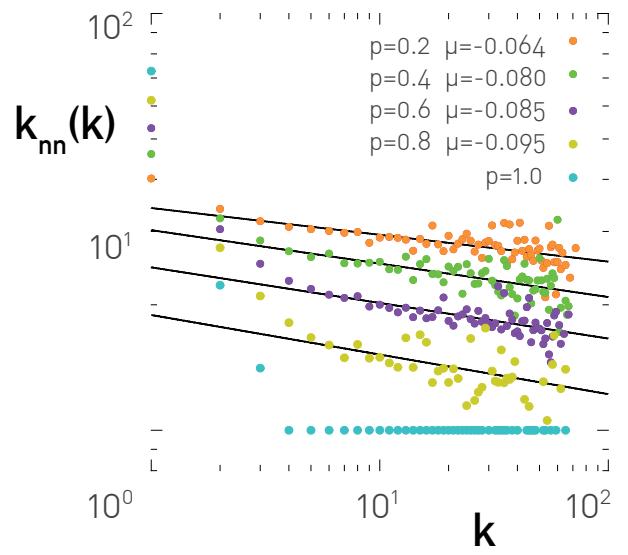
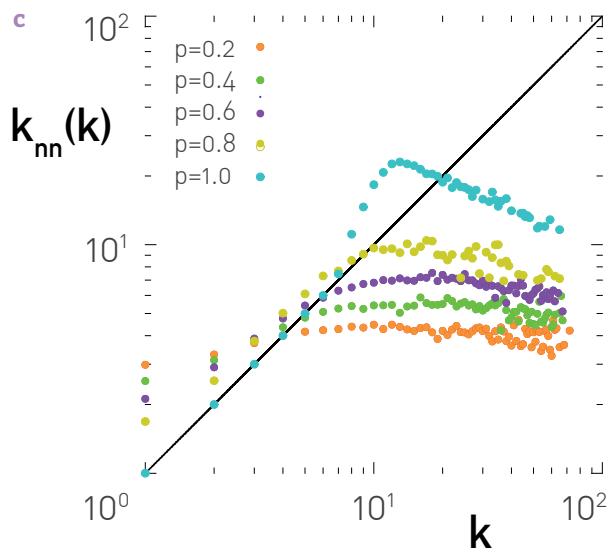
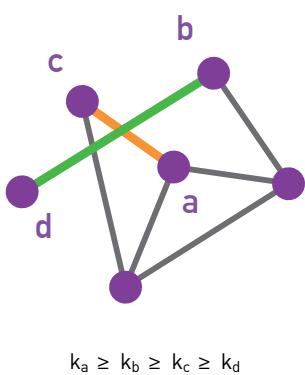


Figure 7.13
Xulvi-Brunet & Sokolov algorithm for extreme correlations

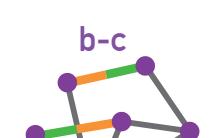
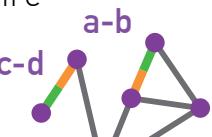


STEP 1 Link selection



STEP 2 Rewire

ASSORTATIVE



DISASSORTATIVE

p

$1 - p$

RANDOM REWIRE

(a) The original Xalvi-Brunet & Sokolov algorithm allows us to tune the magnitude of the observed degree correlations. For this we execute the deterministic rewiring step with probability p , and with probability $1 - p$ we randomly pair the a, b, c, d nodes with each other. For $p = 1$ we are back to the model of Fig. 7.13, generating maximal degree correlations; for $p < 1$ the induced noise tunes the magnitude of the effect.

(b) Typical network configurations generated for $p = 0.5$.

(c) The $k_{nn}(k)$ functions for various p values. The simulations are shown for a network with $N = 10,000$, $\langle k \rangle = 1$, and $\gamma = 3.0$.

Note that the fit of Eq. 7.10 is nonconclusive, as the exponents depend on the fitting region, especially in the assortative case.

THE IMPACT OF DEGREE CORRELATIONS

As we have seen in [SECT. 7.5](#), most real networks are characterized by some degree correlations. Social networks are assortative; biological networks display structural disassortativity. The presence of these correlations raise an important question: why do we care? In other words, do degree correlations alter the properties of a network? And which network properties do they influence? The purpose of this section is to briefly address these questions.

As we have seen in [SECT. 3.6](#), an important property of a random network is the emergence of a phase transition at $\langle k \rangle = 1$, marking the appearance of the giant component. [Fig. 7.15](#) shows the relative size of the giant component for networks with different degree correlations, indicating that [8, 19, 20]:

- **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 create a giant component if the high-degree nodes tend to link to other high-degree ones.

- **For disassortative networks**

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

- For large $\langle k \rangle$ the giant component is smaller in assortative networks than in neutral or disassortative networks. Indeed, the high-degree nodes form a core group of high mean degree. As assortativity forces these hubs to mostly link to each other, they fail to attract to the giant component the numerous small degree nodes.

These changes in the size and the structure of the giant component have implications on the spread of diseases [22, 23, 24], a topic discussed in [CHAP. 10](#). Indeed, as we have seen in [SECT. 7.4](#), social networks tend to be

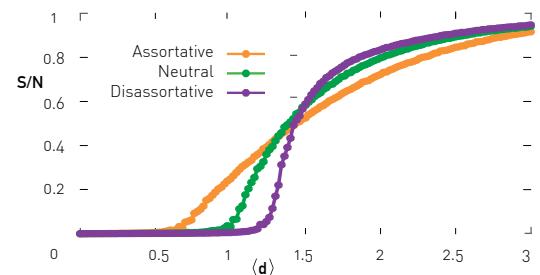


Figure 7.15
Degree correlations and the phase transition point

Relative size of the giant component for an Erdős-Rényi network of size $N=10,000$ (green curve), which is rewired using the Xalvi-Brunet & Sokolov algorithm with $p = 0.5$, to induce degree correlations (red and blue curve). Each point represents an average of 10 independent runs. The figure indicates that as we move from assortative to disassortative networks, the phase transition point is delayed and the size of the giant component increases for large $\langle k \rangle$.

assortative. The high degree nodes therefore form a giant component that acts as the “reservoir” for the disease, sustaining an epidemic even when on average the network is not sufficiently dense for the virus to persist.

The altered giant component has implications for network robustness as well [25]. As we discuss in **CHAPTER 8**, a network can be fragmented by the removal of its hubs. In assortative networks hub removal makes less damage because the hubs cluster together, forming a core group, hence many of them are redundant. The removal of the hubs is more damaging in disassortative networks, as in these the hubs connect to many small-degree nodes, which fall off the network once a hub is deleted.

Let us mention a few additional consequences of degree correlations:

- Fig. 7.16 shows the path-length distribution for 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. Yet 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, hence it results in long chains of $k = 2$ nodes, enhancing d_{\max} Fig. 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 vertex cover problems [29], requiring us to find the minimal set of nodes such that each link is connected to at least one node in the vertex cover **BOX 7.4**.
- Finally, degree correlations have an impact on our ability to control a network, altering the number of input signals one needs to achieve full control [30].

In summary, degree correlations are not only of academic interest, but they alter numerous network characteristics and have a strong impact on various processes that take place on a network.

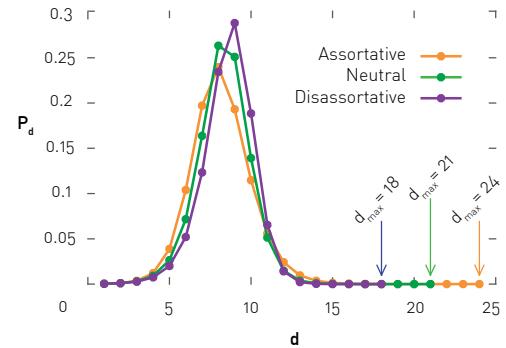


Figure 7.16
Degree correlations and path lengths

Shortest path distribution for a network with Poisson degree distribution of size $N = 10,000$ and $\langle k \rangle = 3$. Correlations are added using the Xalvi-Brunet & Sokolov algorithm with $p = 0.5$. Each curve presents is an average of 10 independent networks. The plots indicate that as we move from disassortative to assortative networks, the average path length decreases, but the diameter grows.

BOX 7.4

VERTEX COVER AND MUSEUM GUARDS

Imagine you are director of an open-air museum situated in a large park with numerous paths. You wish to place guards on crossroads to observe each path, but to save cost you want to use as few guards as possible. Let N be the number of crossroads and $m < N$ is the number of guards you can afford to hire. There are (Nm) ways of placing the m guards in the N positions, but most configurations will leave some paths unobserved [31].

The number of trials one needs to find a perfect solution grows exponentially with N . Indeed, this is one of the six basic NP-complete problems, called the vertex cover problem. By definition, 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. The NP-completeness means that there is no known algorithm which can identify a vertex cover substantially faster than using an exhaustive search, i.e. checking each possible configuration individually. Obviously, the number of nodes needed to obtain a vertex cover depends on the network topology, being affected by the degree distribution and potential degree correlations [29].

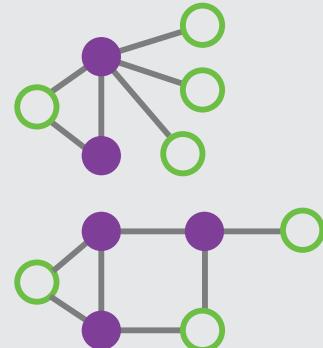


Figure 7.17
The minimum cover

Formally, a vertex cover of a network G is a set C of nodes such that each link of G connects to at least one node in C . A minimum vertex cover is a vertex cover of smallest possible size. The figure above shows examples of minimal vertex covers in two graphs, where the set C is shown in red. One can check that if we turn any of the red nodes into white nodes, we will have at least one link that does not connect to a red node.

SUMMARY

There are at least three important reasons why we care about degree correlations:

- Degree correlations are present in most real networks [SECT. 7.4](#).
- In the previous chapters we showed how much we can learn about a network by inspecting its degree distribution. Degree correlations force us to go beyond the degree distribution, demonstrating that there are quantifiable patterns that govern the way nodes link to each other that are not captured by p_k alone.
- Once present, degree correlations change a network's behavior [SECT. 7.6](#).

Despite the considerable effort devoted to characterizing degree correlations, our understanding of the phenomena is not yet complete. For example, while in [SECT. 7.6](#) we showed how to tune degree correlations, the problem is far from being fully resolved. Indeed, the full degree correlations characterizing a network is contained in the e_{ij} matrix. Generating networks with an arbitrary e_{ij} remains a difficult task.

The results of this chapter allow us to formulate the next network law:

Structural Correlations

Simple scale-free networks are disassortative.

Let us inspect the validity of this law in the light of the three criteria established in [CHAPTER 3](#):

A. Quantitative Formulation

The quantitative basis of this law is provided in [SECT. 5.3](#) and [ADVANCED TOPICS 7.B](#), where we derived the magnitude of the structural cutoff and the emergence of disassortative correlations beyond k_s .

B. Universality

In SECT. 7.4 we showed that many real networks, from biological to email networks, display structural disassortativity.

C. Non-random Character

As we showed in SECT. 5.3, structural disassortativity cannot appear in the random network model, as the degree of the largest node in a random network is smaller than the structural cutoff k_s .

BOX 7.5

DEGREE CORRELATIONS: BRIEF HISTORY

Degree correlations were first reported in 2001 in the context of the Internet in a classic paper by Romualdo Pastor-Satorras, Alexei Vazquez, and Alessandro Vespignani [4, 5]. This work introduced the degree correlation function $k_{nn}(k)$ and the scaling Eq. 7.10. A year later Kim Sneppen and Sergey Maslov used the full $p(k_i, k_j)$, rooted in the e_{ij} matrix, to discover the presence of degree correlations in protein-interaction networks [32]. In 2003 Mark Newman introduced the degree correlation coefficient [8, 9], allowing him to realize that two kinds of correlations can emerge in real systems. He also introduced the terminology “assortativity” and “disassortativity” to characterize this diversity. These terms have their roots in social sciences where they are used to capture mating preferences [33].

Assortative mating

reflects the tendency of individuals to date or marry individuals that are similar to them. For example, low-income individuals tend to 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 small-degree nodes. In a network environment we can also encounter the traditional assortativity, when nodes of similar properties link to each other Fig. 7.18.

Disassortative mixing

when individuals link to individuals who are unlike them, is also observed in social systems. Sexual networks are perhaps the best example of this phenomena, as most sexual relationships are between individuals of different gender. Disassortative mixing is also common in economic settings. For example, 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.



Figure 7.18
Politics is rarely neutral

The network behind the political blogosphere in the US 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 node corresponds to a blog, colored blue if the blog is considered liberal and red if 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, the linkage patterns is not random: liberal blogs predominantly cite other liberal blogs and conservative blogs connect mainly to conservative-leaning blogs. Very few blogs link across the political divide. After [34].

BOX 7.6

TWO-POINT, THREE-POINT CORRELATIONS

In their most general form, the degree correlations present in a network are determined by the conditional probability $P(k^{(1)}, k^{(2)}, \dots, k^{(K)})$ that a node of degree k connects to nodes with degrees $k^{(1)}$, $k^{(2)}, \dots, k^{(K)}$.

Two-point correlations

The simplest of these is the two-point degree correlation discussed in this chapter, being the conditional probability $P(k')$ that a node with degree k is connected to a node with degree k' . For uncorrelated networks this conditional probability is independent of k , hence $P(k') = k' P(k') / \langle k \rangle$ [18]. As the empirical evaluation of $P(k')$ in real networks is a cumbersome task, it is more practical to analyze the degree correlation function $k_{nn}(k)$ defined in Eq. 7.7.

Three-point correlations

In principle there is no reason to stop at two-point correlations. Correlations involving three nodes are determined by the probability $P(k^{(1)}, k^{(2)} | k)$ that a node with degree k is connected to nodes with degrees $k^{(1)}$ and $k^{(2)}$. This conditional probability determines the clustering coefficient Eq. 7.20. Indeed, the average clustering coefficient $C(k)$ of nodes with degree k [22, 23] can be formally written as the probability that a node of degree k is connected to nodes with degrees $k^{(1)}$ and $k^{(2)}$, and that those two are joined by a link, averaged over all the possible values of $k^{(1)}$ and $k^{(2)}$,

$$C(k) = \sum_{k^{(1)}, k^{(2)}} P(k^{(1)}, k^{(2)} | k) p_{k^{(1)}, k^{(2)}}^k,$$

where $p_{k^{(1)}, k^{(2)}}^k$ is the probability that nodes $k^{(1)}$ and $k^{(2)}$ are connected, provided that they have a common neighbor with degree k [18]. For neutral networks the clustering coefficient is independent of k , following

$$C(k) = \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{\langle k \rangle^3 N}.$$

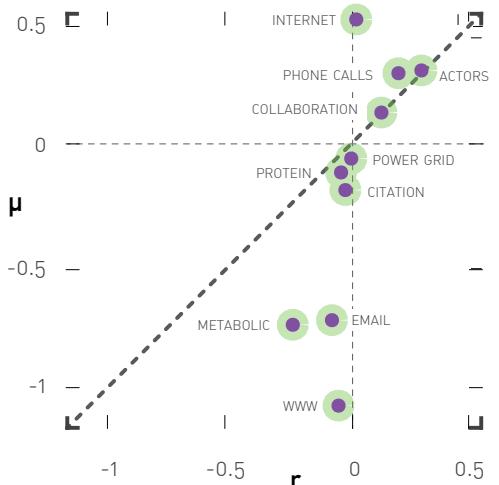


Figure 7.19

Correlation between r and μ

To illustrate the relationship between r and μ , we estimated μ by fitting the knn function to Eq. 7.10, whether or not the power law scaling was statistically significant Fig. 7.8.

ADVANCED TOPICS 7.A

DEGREE CORRELATION COEFFICIENT

In **BOX 7.2** we defined the degree correlation coefficient r as an alternative measure of degree correlations characterizing a network [8, 9]. The use of a single number to characterize degree correlations is extremely attractive, as it also offers an easy way to compare the correlations observed in networks of different nature and size. Yet, before we use r we must be aware of some of its limitations.

The hypothesis behind the correlation coefficient r is that the $k_{nn}(k)$ function can be approximated by the linear function

$$k_{nn}(k) \sim rk. \quad (7.21)$$

This is different from the scaling [Eq. 7.10](#), which assumes a power law dependence on k . [Eq. 7.21](#) raises several important issues:

- The linear dependence [Eq. 7.21](#) is not supported by empirical data, numerical simulations, or analytical calculations. Indeed, analytical calculation of the initial attractiveness model predict a power law [Eq. 7.18](#) or a logarithmic k -dependence [Eq. 7.20](#) for the degree correlation function. Therefore, r forces a linear fit to an inherently nonlinear function. This discrepancy is illustrated in [Fig. 7.20](#), which shows that for assortative and disassortative networks [Eq. 7.21](#) offers a poor fit to the data.
- As we have seen in [Fig. 7.10](#), the dependence of $k_{nn}(k)$ on k is rather complex, often changing trends for large k thanks to the structural cutoff. A linear fit ignores this inherent complexity. To illustrate the consequences of this phenomena, we calculated r and μ for the ten reference networks [TABLE 7.1](#). The results are plotted in [Fig. 7.19](#), indicating that while μ and r correlate for positive r , this correlation breaks down for negative r .
- As we discuss in **BOX 7.8**, the maximally correlated model has a vanishing r for large N , despite the fact that the network maintains its

NETWORK	N	r	μ
Internet	192,244	0.03	0.56
WWW	325,729	-0.05	-1.11
Power Grid	4,941	0.003	0.0
Mobile Phone Calls	36,595	0.21	0.33
Email	57,194	-0.08	-0.74
Science Collaboration	23,133	0.13	0.16
Actor Network	702,388	0.31	0.34
Citation Network	449,673	-0.02	-0.18
E Coli metabolism	1,039	-0.25	-0.76
Protein Interactions	2,018	-0.04	-0.1

Table 7.1
Degree correlations in reference networks

The table shows r and μ for the ten reference networks of [TABLE 4.1](#). Directed networks were made undirected to measure r and μ . Alternatively, we can use the directed correlation coefficient to characterize such directed networks **BOX 7.8**.

degree correlations. This suggests that the degree correlation coefficient r has difficulty detecting correlations characterizing large networks.

RELATIONSHIP BETWEEN μ AND r

If $k_{nn}(k)$ follows the scaling Eq. 7.10, then the sign of the degree coefficient r should agree with the sign of μ . This is supported by Fig. 7.20 as well. To show the origin of this behavior, next we derive a direct relationship between μ and r . To be specific we assume the validity of Eq. 7.10 and determine the value of r for a network with a given correlation exponent μ .

We start by determining a from Eq. 7.10. We can write the second moment of the degree distribution as

$$\langle k^2 \rangle = \langle k_{nn}(k)k \rangle = \sum_{k'} ak^{\mu+1}p_k = a\langle k^{\mu+1} \rangle,$$

which leads to

$$a = \frac{\langle k^2 \rangle}{\langle k^{\mu+1} \rangle}.$$

We now calculate r for a network with a given μ :

$$\begin{aligned} r &= \frac{\sum_{k'} kak^{\mu} q_k - \langle k^2 \rangle^2}{\sigma_r^2} = \frac{\sum_{k'} a k^{\mu+2} \frac{p_k}{\langle k \rangle} - \langle k^2 \rangle^2}{\sigma_r^2} = \frac{\frac{\langle k^2 \rangle \langle k^{\mu+2} \rangle}{\langle k^{\mu+1} \rangle} - \langle k^2 \rangle^2}{\sigma_r^2} = \\ &= \frac{1}{\sigma_r^2} \frac{\langle k^2 \rangle}{\langle k \rangle} \left(\frac{\langle k^{\mu+2} \rangle}{\langle k^{\mu+1} \rangle} - \frac{\langle k^2 \rangle}{\langle k \rangle} \right). \end{aligned} \quad (7.22)$$

For $\mu = 0$ the term in the last parenthesis vanishes, obtaining $r = 0$. Hence if $\mu = 0$ (neutral network), the network will be neutral based on r as well. For $k > 1$ Eq. 7.22 suggests that for $\mu > 0$ the parenthesis is positive, hence $r > 0$, and for $\mu < 0$ is negative, hence $r < 0$. Therefore, r and μ predict degree correlations of similar kind.

Therefore, if the degree correlation function follows Eq. 7.10, then the sign of the degree correlation exponent μ will determine the sign of the assortativity coefficient r :

$$\begin{aligned} \mu < 0 &\rightarrow r < 0 \\ \mu = 0 &\rightarrow r = 0 \\ \mu > 0 &\rightarrow r > 0. \end{aligned}$$

In summary, the degree correlation coefficient assumes that $k_{nn}(k)$ scales linearly with k , a hypothesis that lacks numerical and analytical support. Hence r forces a linear fit to $k_{nn}(k)$, giving occasionally rise to inconsistent results. While typically the sign of r and μ agree, overall r does not offer a natural characterization of the underlying degree correlations. An accurate characterization starts with e_{ij} , whose behavior is reasonably captured by $k_{nn}(k)$.

BOX 7.7

AT A GLANCE: DEGREE CORRELATIONS

Degree Correlation Matrix e_{ij}
probability of finding a node with degrees i and j at the two ends of a link.

Neutral networks:

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

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

$\mu < 0$: Dissassortative

Degree Correlation Coefficient

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

$r > 0$: Assortative

$r = 0$: Neutral

$r < 0$: Dissassortative

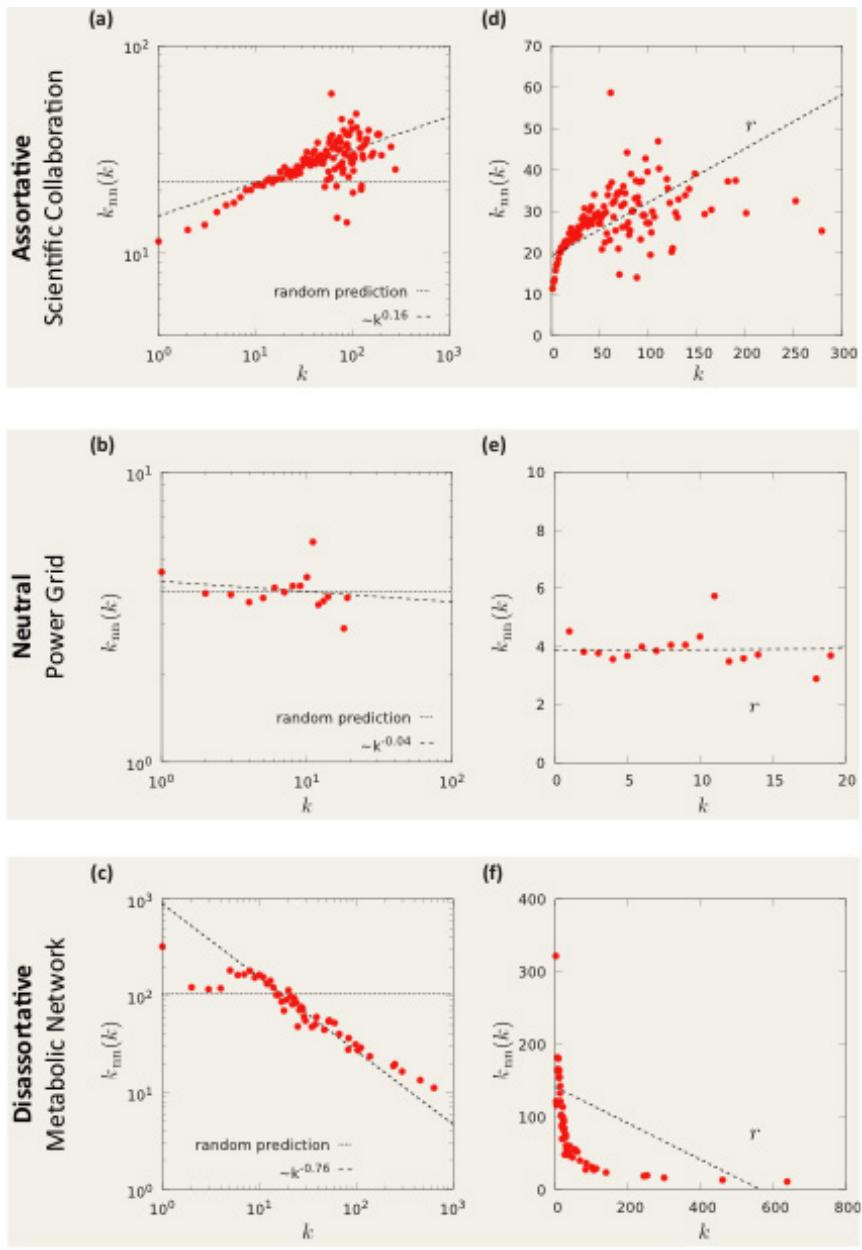


Figure 7.20
Degree correlation function

The degree correlation function k_{nn} for three real networks. The left panels show the cumulative function $k_{nn}(k)$ on a log-log plot to test the validity of Eq. 7.10. The right panels show $k_{nn}(k)$ on a lin-lin plot to test the validity of Eq. 7.21, i.e. the assumption that $k_{nn}(k)$ depends linearly on k , the hypothesis behind the correlation coefficient r . The slope of the dotted line corresponds to the correlation coefficient r . As the lin-lin plots illustrate, Eq. 7.21 offers a poor fit for assortative (d) and disassortative (f) networks.

BOX 7.9

CORRELATION COEFFICIENT FOR DIRECTED NETWORKS

To measure correlations in directed networks we must take into account that each node i is characterized by an incoming k_i^{in} and an outgoing k_i^{out} degree. Hence, we can define four degree correlation coefficients, $r_{in, in}$, $r_{in, out}$, $r_{out, in}$, $r_{out, out}$ capturing all possible combinations between the incoming and outgoing degrees of two nodes linked to each other Figs. 7.12 a-d. Formally we have [14].

$$r_{\alpha, \beta} = \frac{\sum_{jk} jk(e_{jk}^{\alpha, \beta} - q_j^\alpha q_k^\beta)}{\sigma^\alpha \sigma^\beta}, \quad (7.23)$$

where α and β refer to the in and out indices. To illustrate the use of Eq. 7.23, we show in Fig. 7.21e the four correlation coefficients for the five directed reference networks TABLE 7.1. For a complete characterization of degree correlations, it is desirable to measure the four $k_{nn}(k)$ functions as well BOX 7.2.

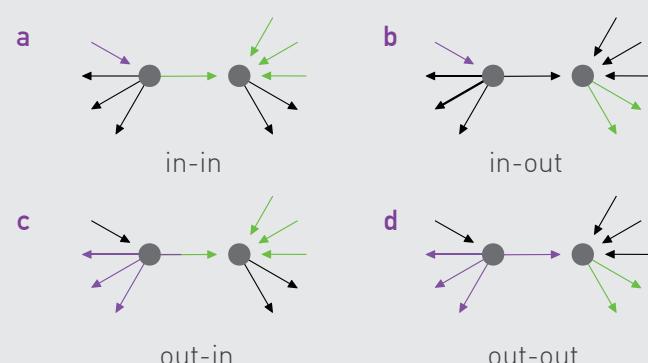
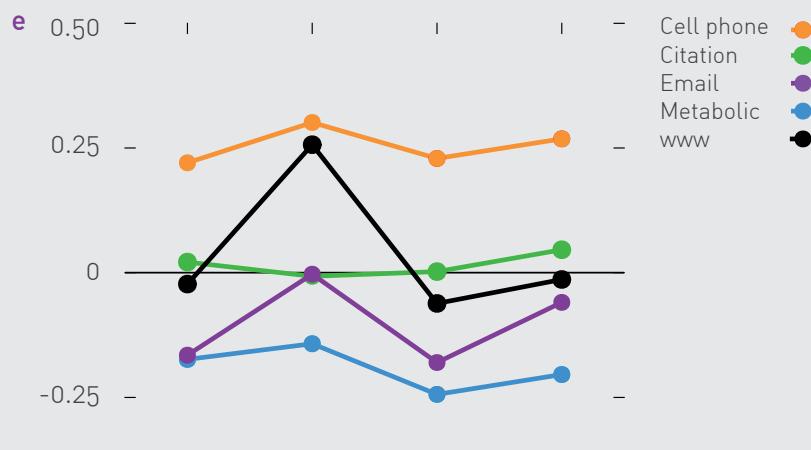


Figure 7.21

Panels (a)-(d) illustrate in red and green the (α, β) indices that define the appropriate correlation coefficient for directed networks. (e) The correlation profile of the five directed reference networks, indicating, for example, that while citation networks have negligible correlations, all four correlation coefficients document strong assortative behavior for cell phone calls and strong disassortative behavior for metabolic networks. The case of the WWW is particularly interesting: while three of its correlation coefficients are close to zero, there is a strong assortative tendency for the (in, out) combinations.

ADVANCED TOPICS 7.B

STRUCTURAL CUTOFFS

As discussed in **SECT. 7.3**, there is a fundamental conflict between the scale-free property and degree correlations, which leads to a structural cutoff in simple networks. In this section we derive [Eq. 7.16](#), providing the system size dependence of the structural cutoff [11]. We start by defining

$$r_{kk'} = \frac{E_{kk'}}{m_{kk'}}, \quad (7.24)$$

where $E_{kk'}$ is the number of links between nodes of degrees k and k' , and

$$m_{kk'} = \min\{kN_k, k'N_{k'}, N_kN_{k'}\} \quad (7.25)$$

is the largest possible value of $E_{kk'}$. If multiple links are allowed, $m_{kk'}$ is simply $m_{kk'} = \min\{kN_k, k'N_{k'}, N_kN_{k'}\}$. The origin of [Eq. 7.25](#) is explained in [Fig. 7.22](#). Consequently, we can write the $r_{kk'}$ ratio as

$$r_{kk'} = \frac{E_{kk'}}{m_{kk'}} = \frac{\langle k \rangle P(k, k')}{\min\{kP(k), k'P(k'), NP(k)P(k')\}}. \quad (7.26)$$

As $m_{kk'}$ is the maximum of $E_{kk'}$, $r_{kk'}$ must be smaller than or equal to one for any k and k' . Yet, for some networks and for some k, k' pairs $r_{kk'}$ becomes larger than one. This is clearly non-physical and signals some conflict in the network configuration. Strictly speaking, in simple networks degree pairs for which $r_{kk'} > 1$ cannot exist. Hence, we define the structural cut off k_s as the solution of the equation

$$r_{k_s k_s} = 1. \quad (7.27)$$

Note that as soon as $k > NP(k')$ and $k' > NP(k)$, the effects of the restriction on the multiple links are already felt, turning the expression for $r_{kk'}$ into

$$r_{kk'} = \frac{\langle k \rangle P(k, k')}{N p_k p_{k'}}. \quad (7.28)$$

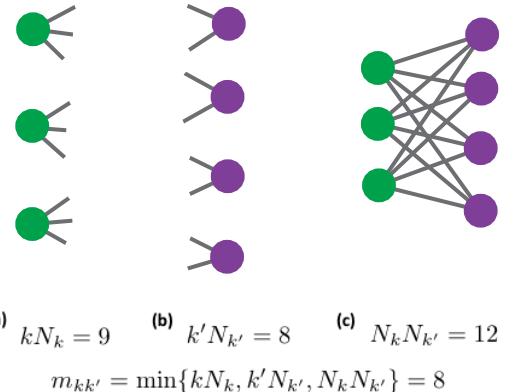


Figure 7.22
Correlation between r and N

Illustrating the maximum number of links one can have between two groups of nodes. The figure shows two groups of nodes, with degree $k=3$ and $k'=2$. The total number of links between these two groups must not exceed

- (a) The total number of links available in $k=3$ group, which is $kN_k=9$;
- (b) The total number of links available in $k'=2$ group, which is $k'N_{k'}=8$;
- (c) The total number of links one can potentially have between the two groups, which is $N_kN_{k'}=12$.

In the example shown above the smallest of the three is $k'N_{k'}=8$ of (b). The resulting configuration is shown on the top right. One can see that in this configuration, one link in the $k=3$ class remains unpaired.

For scale-free networks these conditions are fulfilled in the region $k, k' > (aN)^{1/(\gamma+1)}$, where a is a constant that depends on the function p_k . Note that this value is below the natural cut off. As a consequence, this scaling provides a lower bound for the structural cut off, in the sense that whenever the cut off of the degree distribution falls below this limit, the condition $r_{kk'} < 1$ is always satisfied.

For neutral networks the joint distribution factorizes as

$$P(k, k') = \frac{kk' p_k p_{k'}}{\langle k \rangle^2}. \quad (7.29)$$

Hence, the ratio $r_{kk'}$ of Eq. 7.28 takes the form

$$r_{kk'} = \frac{kk'}{\langle k \rangle N}. \quad (7.30)$$

Therefore, the structural cutoff needed to preserve the condition $r_{kk'} \leq 1$ has the form [35, 36, 37]

$$k_s(N) \sim (\langle k \rangle N)^{1/2}, \quad (7.31)$$

which is Eq. 7.16. Note that Eq. 7.31 is independent of the degree distribution of the underlying network. Consequently, for a scale-free network $k_s(N)$ is independent of the degree exponent γ .

BIBLIOGRAPHY

[1] P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, RS Judson, JR Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, J. M. Rothberg. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403: 623–627, 2000.

[2] I. Xenarios, D. W. Rice, L. Salwinski, M. K. Baron, E. M. Marcotte, D. Eisenberg DIP: the database of interacting proteins. *Nucleic Acids Res.* 28: 289–29, 2000.

[3] H. Jeong, S.P. Mason, A.-L. Barabási, Z.N. Oltvai, *Nature* 411: 41-42, 2001.

[4] R. Pastor-Satorras, A. Vázquez, and A. Vespignani. Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.* 87: 258701, 2001.

[5] A. Vazquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of Internet. *Phys. Rev. E* 65: 066130, 2002.

[6] S.L. Feld. Why your friends have more friends than you do. *American Journal of Sociology* 96: 1464–1477, 1991.

[7] E.W. Zuckerman and J.T. Jost. What makes you think you're so popular? Self evaluation maintenance and the subjective side of the “friendship paradox”. *Social Psychology Quarterly* 64: 207–223, 2001.

[8] M. E. J. Newman. Assortative mixing in networks, *Phys. Rev. Lett.* 89: 208701, 2002.

[9] M. E. J. Newman. Mixing patterns in networks. *Phys. Rev. E* 67: 026126, 2003.

[10] S. Maslov, K. Sneppen, and A. Zaliznyak, Pattern detection in complex networks: Correlation profile of the Internet, e-print cond-mat/0205379, 2002.

[11] M. Boguna, R. Pastor-Satorras, A. Vespignani. Cut-offs and finite size effects in scale-free networks. *Eur. Phys. J. B* 38: 205, 2004.

[12] M. E. J. Newman and Juyong Park. Why social networks are different from other types of networks, arXiv:cond-mat/0305612v1.

[13] M. McPherson, L. Smith-Lovin, J. M. Cook. Birds of a feather: homophily in social networks. *Annual Review of Sociology* 27:415-444, 2001.

[14] J. G. Foster, D. V. Foster, P. Grassberger, M. Paczuski. Edge direction and the structure of networks. *PNAS* 107: 10815, 2010.

[15] A. Barrat and R. Pastor-Satorras. Rate equation approach for correlations in growing network models. *Phys. Rev. E* 71, 036127, 2005.

[16] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Adv. Phys.* 51: 1079, 2002.

[17] J. Berg and M. Lässig. Correlated random networks. *Phys. Rev. Lett.* 89: 228701, 2002

[18] M. Boguñá and R. Pastor-Satorras. Class of correlated random networks with hidden variables. *Phys. Rev. E* 68: 036112, 2003.

[19] R. Xulvi-Brunet and I. M. Sokolov. Reshuffling scale-free networks: From random to assortative. *Phys. Rev. E* 70: 066102, 2004.

[20] R. Xulvi-Brunet and I. M. Sokolov. Changing correlations in networks: assortativity and dissimilarity. *Acta Phys. Pol. B* 36: 1431, 2005.

[21] J. Menche, A. Valleriani, and R. Lipowsky. Asymptotic properties of degree-correlated scale-free networks. *Phys. Rev. E* 81: 046103, 2010.

[22] V. M. Eguíluz and K. Klemm. Epidemic threshold in structured scale-free networks. *Phys. Rev. Lett.* 89:108701, 2002.

[23] M. Boguñá and R. Pastor-Satorras. Phys. Epidemic spreading in correlated complex networks. *Rev. E* 66: 047104, 2002.

[24] M. Boguñá, R. Pastor-Satorras, and A. Vespignani. Absence of epidemic threshold in scale-free networks with degree correlations. *Phys. Rev. Lett.* 90: 028701, 2003.

[25] A. Vázquez and Y. Moreno. Resilience to damage of graphs with degree correlations. *Phys. Rev. E* 67: 015101R, 2003.

[26] S.J.Wang,A.C.Wu,Z.X.Wu,X.J.Xu, and Y.H.Wang. Response of degree-correlated scale-free networks to stimuli. *Phys. Rev. E* 75: 046113, 2007.

[27] F. Sorrentino, M. Di Bernardo, G. Cuellar, and S. Boccaletti. Synchronization in weighted scale-free networks with degree-degree correlation. *Physica D* 224: 123, 2006.

[28] M. Di Bernardo, F. Garofalo, and F. Sorrentino. Effects of degree correlation on the synchronization of networks of oscillators. *Int. J. Bifurcation Chaos Appl. Sci. Eng.* 17: 3499, 2007.

[29] A. Vazquez and M. Weigt. Computational complexity arising from degree correlations in networks, arXiv:cond-mat/0207035, 2002.

[30] M. Posfai, Y Y. Liu, J-J Slotine, A.-L. Barabási. Effect of correlations on network controllability, *Scientific Reports* 3: 1067, 2013.

[31] M. Weigt and A. K. Hartmann. The number of guards needed by a museum: A phase transition in vertex covering of random graphs. *Phys. Rev. Lett.* 84: 6118, 2000.

[32] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Science* 296: 910–913, 2002.

[33] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: homophily in social networks. *Annual Review of Sociology* 27: 415–444, 2001.

[34] L. Adamic and N. Glance, The political blogosphere and the 2004 U.S. election: Divided they blog (2005).

[35] J. Park and M. E. J. Newman. The origin of degree correlations in the Internet and other networks. *Phys. Rev. E* 66: 026112, 2003.

[36] F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6: 125, 2002.

[37] Z. Burda and Z. Krzywicki. Uncorrelated random networks. *Phys. Rev. E* 67: 046118, 2003.