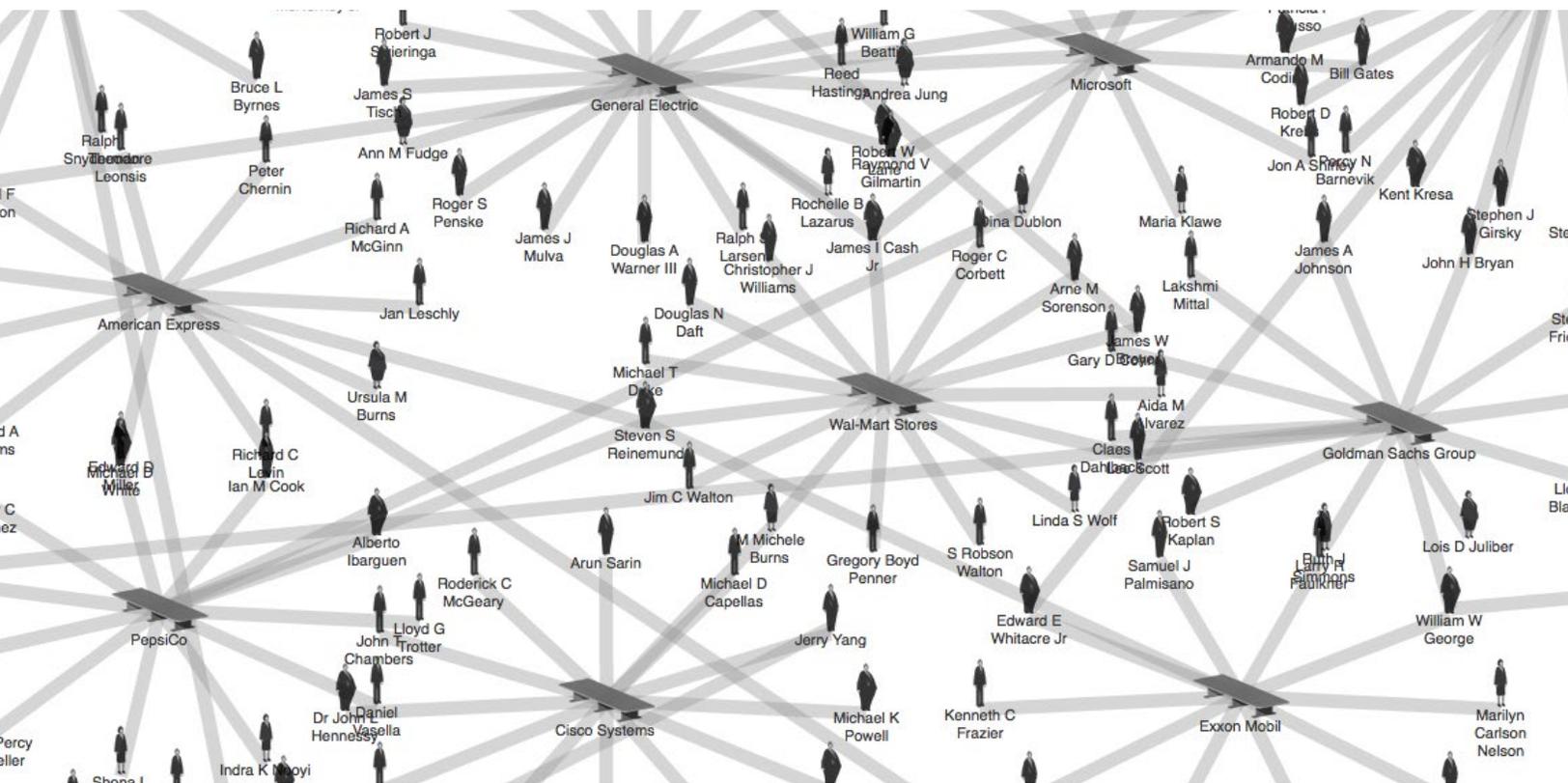


ALBERT-LÁSZLÓ BARABÁSI

NETWORK SCIENCE

DEGREE CORRELATION



ACKNOWLEDGEMENTS

MÁRTON PÓSFAI
GABRIELE MUSELLA
MAURO MARTINO
NICOLE SAMAY

ROBERTA SINATRA
SARAH MORRISON
AMAL HUSSEINI
PHILIPP HOEVEL

INDEX

Introduction	1
Assortativity and Disassortativity	2
Measuring Degree Correlations	3
Structural Cutoffs	4
Correlations in Real Networks	5
Generating Correlated Networks	6
The Impact of Degree Correlations	7
Summary	8
Homework	9
ADVANCED TOPICS 7.A	
Degree Correlation Coefficient	10
ADVANCED TOPICS 7.B	
Structural Cutoffs	11
Bibliography	12

Figure 7.0 (cover image)
TheyRule.net by Josh On

Created by Josh On, a San Francisco-based designer, the interactive website [TheyRule.net](#) uses a network representation to illustrate the interlocking relationship of the US economic class. By mapping out the shared board membership of the most powerful U.S. companies, it reveals the influential role of a small number of individuals who sit on multiple boards. Since its release in 2001, the project is interchangeably viewed as art or science.



This book is licensed under a
Creative Commons: CC BY-NC-SA 2.0.
PDF V24 18.09.2014

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 (Figure 7.1). 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. Their weddings (and breakups) has drawn countless hours of media coverage and sold millions of gossip magazines. Thanks to them we take for granted that celebrities marry each other. We rarely pause to ask: Is this normal? In other words, what is the true chance that a celebrity marries another celebrity?

Assuming that a celebrity could date anyone from a pool of about a hundred million (10^8) eligible individuals worldwide, the chances that their mate would be another celebrity from a generous list of 1,000 other celebrities is only 10^{-5} . Therefore, if dating were driven by random encounters, celebrities would never marry each other.

Even if we 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. 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 (Figure 7.1) and joint board memberships (Figure 7.0) are manifestations 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 Figure 7.2. 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 to 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 Figure 7.2: they almost exclusively interact with small-degree proteins.



Figure 7.1
Hubs Dating Hubs

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

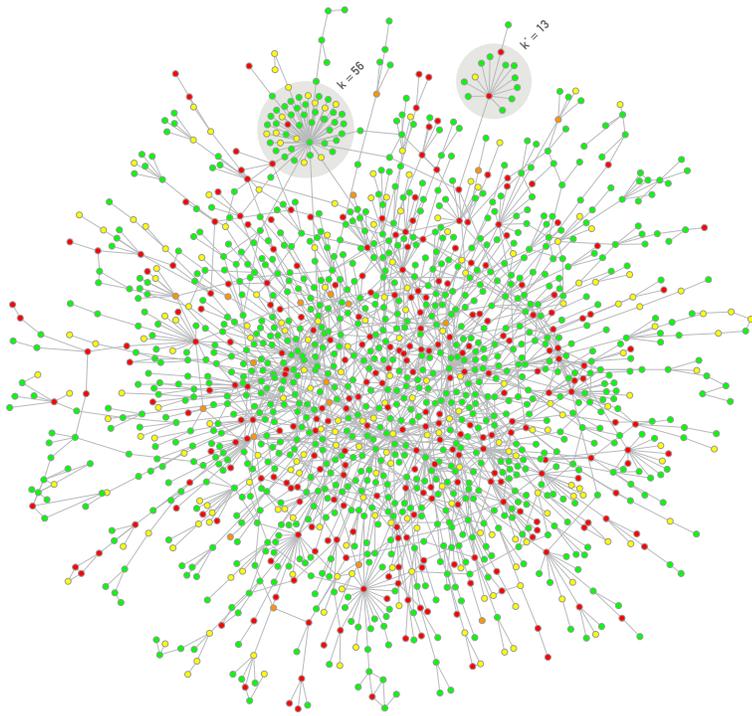


Figure 7.2

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. We highlighted the two largest hubs, with degrees $k = 56$ and $k' = 13$. They both connect to many small degree nodes and avoid linking to each other.

The network has $N = 1,870$ proteins and $L = 2,277$ 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 in TABLE 4.1 represents a later map, hence it contains more nodes and links than the network shown in this figure. Node color corresponds to the essentiality of each protein: the removal of the red nodes kills the organism, hence they are called lethal or essential proteins. In contrast the organism can survive without one of its green nodes. After [3].

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 nodes with degrees k and k' link to each other is

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

Equation (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. Indeed, if k and k' are large, so is $p_{k,k'}$. Consequently, the likelihood that hubs with degrees $k=56$ and $k' = 13$ have a direct link between them is $p_{k,k'} = 0.16$, 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 Figure 7.2, but we observe numerous direct links between small degree nodes.

Instead of linking to each other, the hubs highlighted in Figure 7.2 almost exclusively connect to degree one nodes. By itself this is not unexpected: We expect that a hub with degree $k = 56$ should link to $N_1 p_{1,56} \approx 12$ nodes with $k = 1$. The problem is that this hub connects to 46 degree one neighbors, i.e. four times the expected number.

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, connecting instead to many small degree nodes. While it is dangerous to derive 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 degree correlations and explore their impact on the network topology.

ASSORTATIVITY AND DISASSORTATIVITY

Just by the virtue of the many links they have, hubs are expected to link to each other. In some networks they do, in others they don't. This is illustrated in [Figure 7.3](#), that shows three networks with identical degree sequences but different topologies:

- **Neutral Network**

[Figure 7.3b](#) shows a network whose wiring is random. We call this network *neutral*, meaning that the number of links between the hubs coincides with what we expect by chance, as predicted by [\(7.1\)](#).

- **Assortative Network**

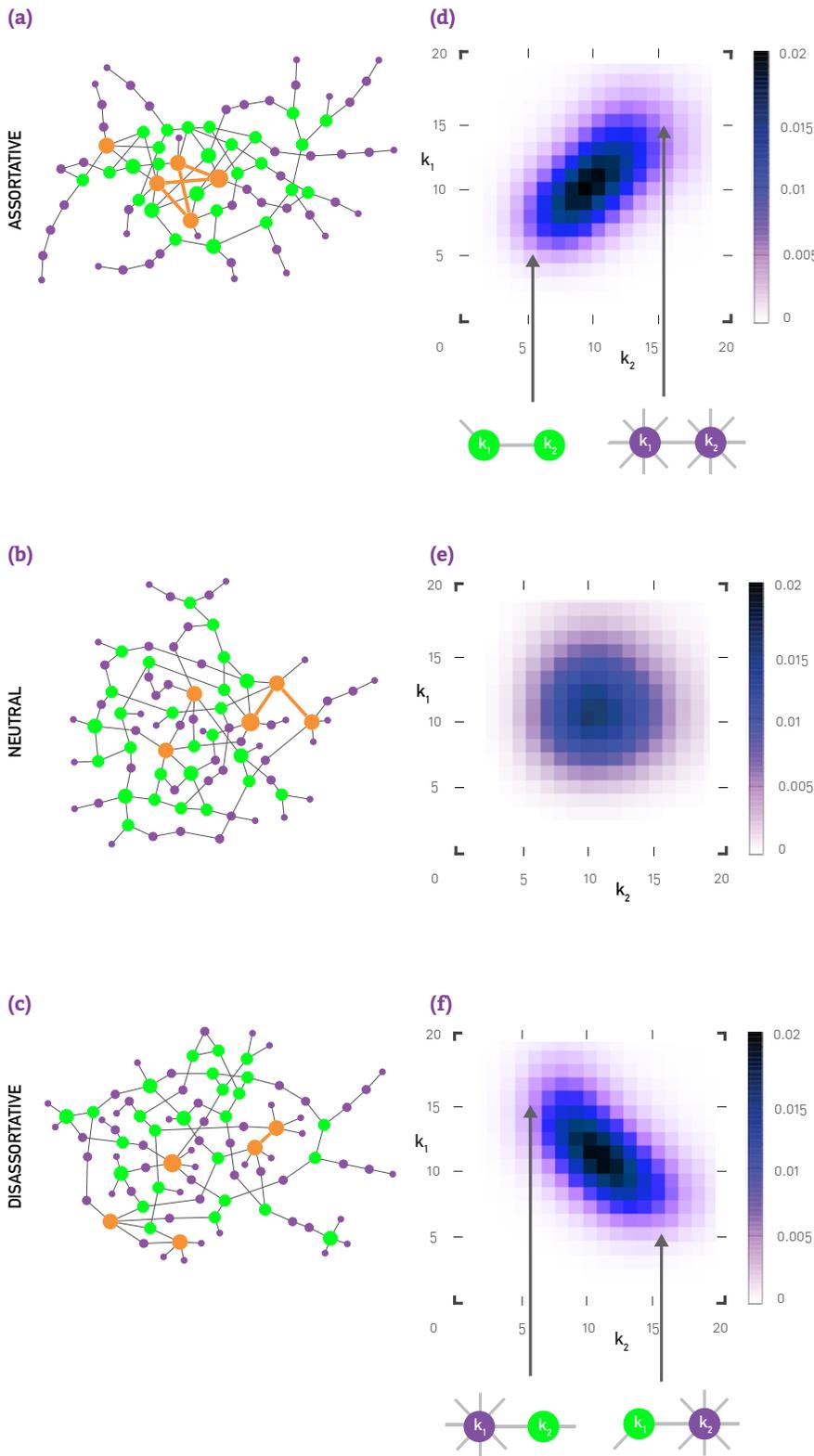
The network of [Figure 7.3a](#) has precisely the same degree sequence as the one in [Figure 7.3b](#). Yet, the hubs in [Figure 7.3a](#) tend to link to each other and avoid linking to small-degree nodes. At the same time the small-degree nodes tend to connect to other small-degree nodes. Networks displaying such trends are *assortative*. An extreme manifestation of this pattern is a perfectly assortative network, in which each degree- k node connects only to other degree- k nodes ([Figure 7.4](#)).

- **Disassortative Network**

In [Figure 7.3c](#) the hubs avoid each other, linking instead 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, the number of links between nodes of degrees k and k' deviates from [\(7.1\)](#).

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 in orange the five highest degree nodes and the direct links between them.

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

(a,d) Assortative Networks

For assortative networks e_{ij} is high along the main diagonal. This indicates that nodes of comparable degree tend to link to each other: small-degree nodes to small-degree nodes and hubs to hubs. Indeed, the network in (a) has numerous links between its hubs as well as between its small degree nodes.

(b,e) Neutral Networks

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

(c,f) Disassortative Networks

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. Consequently these networks have a hub and spoke character, as seen in (c).

The information about potential degree correlations is captured by 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 is normalized, i.e.

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

In (5.27) we derived the probability q_k that there is a degree- k node at the end of the randomly selected link, obtaining

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

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 (7.5). Note that (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} encodes all information about potential degree correlations, we start with its visual inspection. Figures 7.3d,e,f show 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 (Figure 7.3e). In contrast, assortative networks show high correlations along the main diagonal, indicating that nodes predominantly connect to nodes with comparable degree. Therefore low-degree nodes tend to link to other low-degree nodes and hubs to hubs (Figure 7.3d). In disassortative networks e_{ij} displays the opposite trend: it has high correlations along the secondary diagonal. Therefore high-degree nodes tend to connect to low-degree nodes (Figure 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 $k_{\max}^2/2$ 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 degree correlations. This is the goal of the subsequent sections.

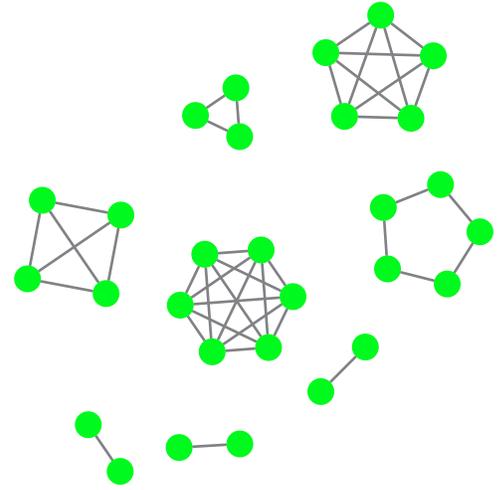


Figure 7.4
Perfect Assortativity

In a perfectly assortative network each node links only to nodes with the same degree. Hence $e_{jk} = \delta_{jk} q_k$, where δ_{jk} is the Kronecker delta. In this case all non-diagonal elements of the e_{jk} matrix are zero. The figure shows such a perfectly assortative network, consisting of complete k -cliques.

MEASURING DEGREE CORRELATIONS

While e_{ij} contains the complete information about the degree correlations characterizing a particular network, it is difficult to interpret its content. In this section we introduce the degree correlation function that offers a simpler way to quantify 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 (Figure 7.5)

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

The *degree correlation function* calculates (7.6) for all nodes with degree k [4, 5]

$$k_{\text{nn}}(k) = \sum_{k'} k' P(k'|k), \quad (7.7)$$

where $P(k'|k)$ is the conditional probability that following a link of a k -degree node we reach a degree- k' node. Therefore $k_{\text{nn}}(k)$ is the average degree of the neighbors of all degree- k nodes. To quantify degree correlations we inspect the dependence of $k_{\text{nn}}(k)$ on k .

- **Neutral Network**

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

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

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

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

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 the global network characteristics $\langle k \rangle$ and $\langle k^2 \rangle$. So plotting $k_{\text{nn}}(k)$ in function of k should result in a horizontal line at $\langle k^2 \rangle / \langle k \rangle$, as observed for

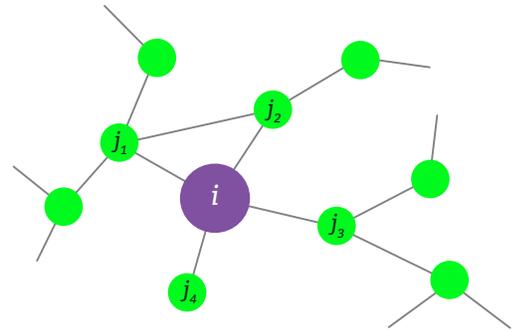


Figure 7.5

Nearest Neighbor Degree: k_{nn}

To determine the degree correlation function $k_{\text{nn}}(k_i)$ we calculate the average degree of a node's neighbors. The figure illustrates the calculation of $k_{\text{nn}}(k_i)$ for node i . 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 $k_{\text{nn}}(4) = (4 + 3 + 3 + 1)/4 = 2.75$.

the power grid (Figure 7.6b). Equation (7.9) also captures an intriguing property of real networks: our friends are more popular than we are, a phenomenon called the *friendship paradox* (BOX 7.1).

- **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).

The behavior observed in Figure 7.6 prompts us to approximate the degree correlation function with [4]

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

If the scaling (7.10) holds, then the nature of degree correlations is determined by the sign of the correlation exponent μ :

- **Assortative Networks: $\mu > 0$**

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

- **Neutral Networks: $\mu = 0$**

According to (7.9) $k_{nn}(k)$ is independent of k . Indeed, for the power grid we obtain $\mu = 0.04 \pm 0.05$, which is indistinguishable from zero (Figure 7.6b).

- **Disassortative Networks: $\mu < 0$**

For the metabolic network we obtain $\mu = -0.76 \pm 0.04$ (Figure 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 predict the impact of degree correlations on various network characteristics (SECTION 7.6). Yet, it is often convenient to use 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 (7.10), or using the degree correlation coefficient introduced in BOX 7.2.

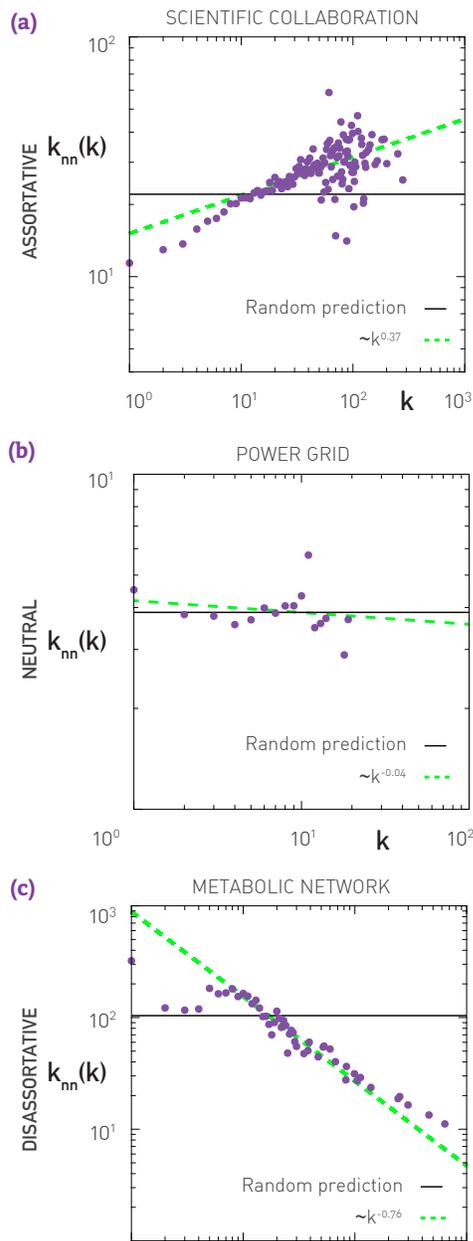


Figure 7.6
Degree Correlation Function

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 the scaling law (7.10).

(a) **Collaboration Network**

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, in line with (7.9) for neutral networks.

(c) **Metabolic Network**

The decreasing $k_{nn}(k)$ documents the network’s disassortative nature.

On each panel the horizontal line corresponds to the prediction (7.9) and the green dashed line is a fit to (7.10).

BOX 7.2

DEGREE CORRELATION COEFFICIENT

If we wish to characterize degree correlations using a single number, we can use either μ or the *degree correlation coefficient*. Proposed by Mark Newman [8,9], the degree correlation coefficient is defined as

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma^2} \quad (7.11)$$

with

$$\sigma^2 = \sum_k k^2 q_k - \left[\sum_k k q_k \right]^2. \quad (7.12)$$

Hence r is the Pearson correlation coefficient between the degrees found 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 scientific 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 we have $r = 0$.

The assumption behind the degree correlation coefficient is that $k_{nn}(k)$ depends linearly on k with slope r . In contrast the correlation exponent μ assumes that $k_{nn}(k)$ follows the power law (7.10). Naturally, both cannot be valid simultaneously. The analytical models of SECTION 7.7 offer some guidance, supporting the validity of (7.10). As we show in ADVANCED TOPICS 7.A, in general r correlates with μ .

BOX 7.1

FRIENDSHIP PARADOX

The friendship paradox makes a suprising statement: *On average my friends are more popular than I am* [6,7]. This claim is rooted in (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 a random network, for which $\langle k^2 \rangle = \langle k \rangle (1 + \langle k \rangle)$. According to (7.9) $k_{nn}(k) = 1 + \langle k \rangle$. Therefore the average degree of a node's neighbors is always higher than the average degree of a randomly chosen node, which is $\langle k \rangle$.

The gap between $\langle k \rangle$ and our friends' degree can be particularly large in scale-free networks, for which $\langle k^2 \rangle / \langle k \rangle$ significantly exceeds $\langle k \rangle$ (Figure 4.8). Consider for example the actor network, for which $\langle k^2 \rangle / \langle k \rangle = 565$ (Table 4.1). In this network the average degree of a node's friends is hundreds of times the degree of the node itself.

The friendship paradox has a simple origin: We are more likely to be friends with hubs than with small-degree nodes, simply because hubs have more friends than the small nodes.

STRUCTURAL CUTOFFS

Throughout this book we assumed that networks are *simple*, meaning that there is at most one link between two nodes (Figure 2.17). 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. Similarly, in the actor network we connect two actors with a single link if they acted in the same movie, independent of the number of joint movies. 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 Figure 7.7a, whose two largest hubs have degrees $k = 55$ and $k' = 46$. In a network with degree correlations $e_{kk'}$ the expected number of links between k and k' is

$$E_{kk'} = e_{kk'} \langle k \rangle N. \quad (7.13)$$

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

$$E_{kk'} = \frac{k p_k k' p_{k'}}{\langle k \rangle} N = \frac{55 \cdot 46}{\frac{300 \cdot 300}{3}} 300 = 2.8. \quad (7.14)$$

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 can have only one link between them, causing a conflict between degree correlations and the scale-free property. The goal of this section is to understand the origin and the consequences of this conflict.

For small k and k' (7.14) 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 does (7.14) predict multiple links. As we show in ADVANCED TOPICS 7.B, k_s , called *structural cutoff*, scales as

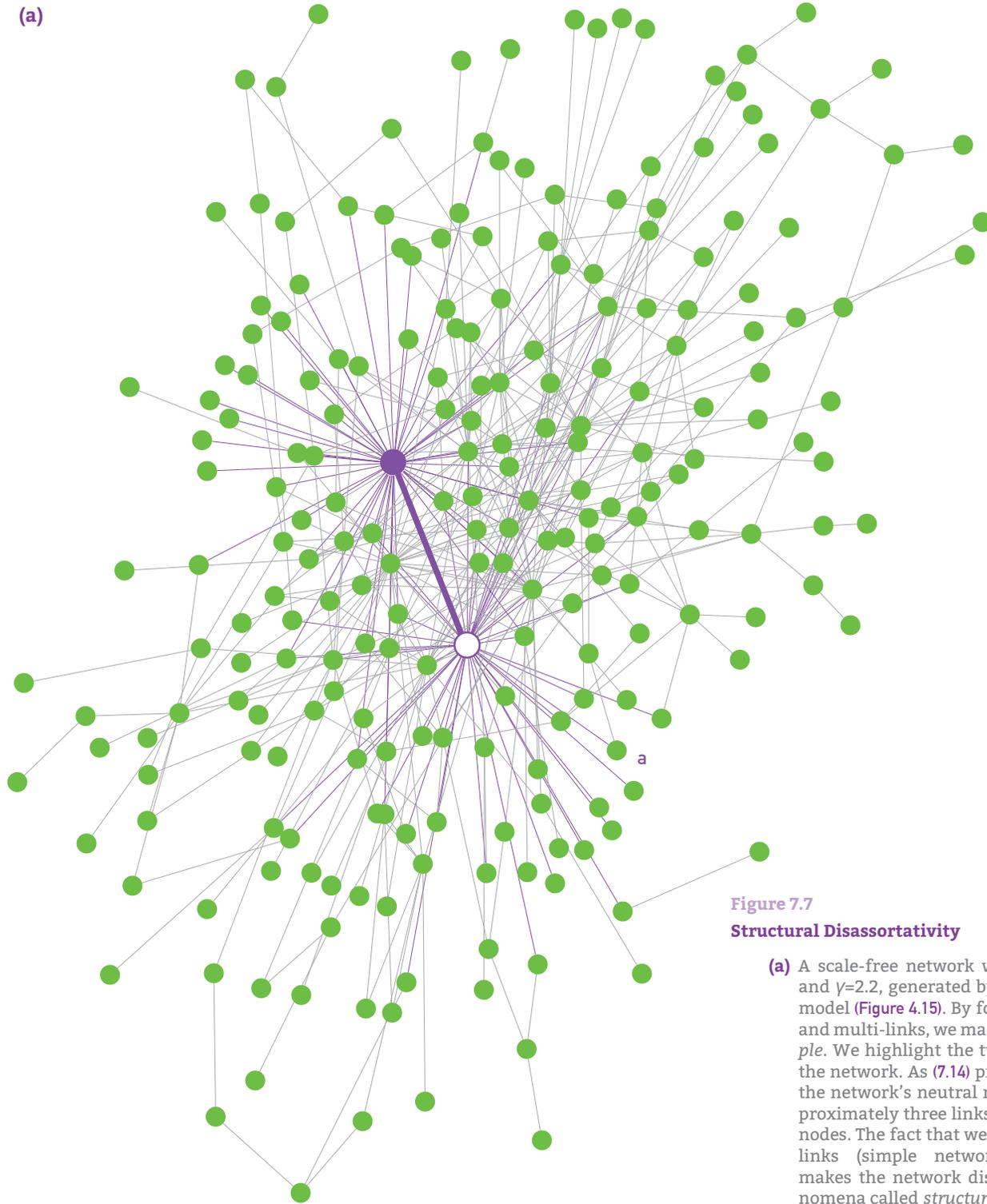
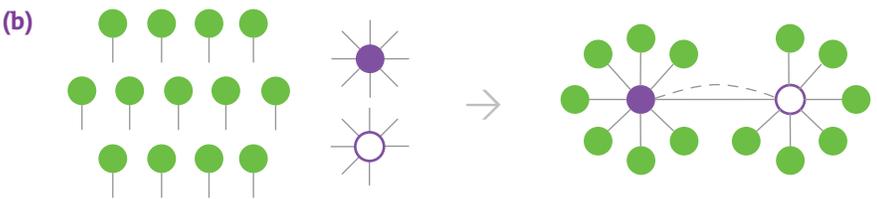


Figure 7.7
Structural Disassortativity

(a) A scale-free network with $N=300$, $L=450$, and $\gamma=2.2$, generated by the configuration model (Figure 4.15). By forbidding self-loops and multi-links, we made the network *simple*. We highlight the two largest nodes in the network. As (7.14) predicts, to maintain the network's neutral nature, we need approximately three links between these two nodes. The fact that we do not allow multi-links (simple network representation) makes the network disassortative, a phenomena called *structural disassortativity*.



(b) To illustrate the origins of structural correlations we start from a fixed degree sequence, shown as individual stubs on the left. Next 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 multi-links, there can only be one link between these two nodes, making the network structurally disassortative.

$$k_s(N) \sim (\langle k \rangle N)^{1/2}. \quad (7.15)$$

In other words, nodes whose degree exceeds (7.15) have $E_{kk'} > 1$, a conflict that as we show below gives rise to degree correlations.

To understand the consequences of the structural cutoff we must first ask if a network has nodes whose degrees exceeds (7.15). For this we compare the structural cutoff, k_s , with the natural cutoff, k_{\max} , which is the expected largest degree in a network. According to (4.18), for a scale-free network $k_{\max} \sim N^{1/\gamma}$. Comparing k_{\max} to k_s allows us to distinguish two regimes:

- **No Structural 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.

- **Structural Disassortativity**

For scale-free 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.

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

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

Finally, how can we decide whether the correlations observed in a particular network are a consequence of structural disassortativity, or are generated by some unknown process that leads to degree correlations? Degree-preserving randomization (Figure 4.17) helps us distinguish these two possibilities:

- (i) **Degree Preserving Randomization with Simple Links (R-S)**

We apply degree-preserving randomization to the original network

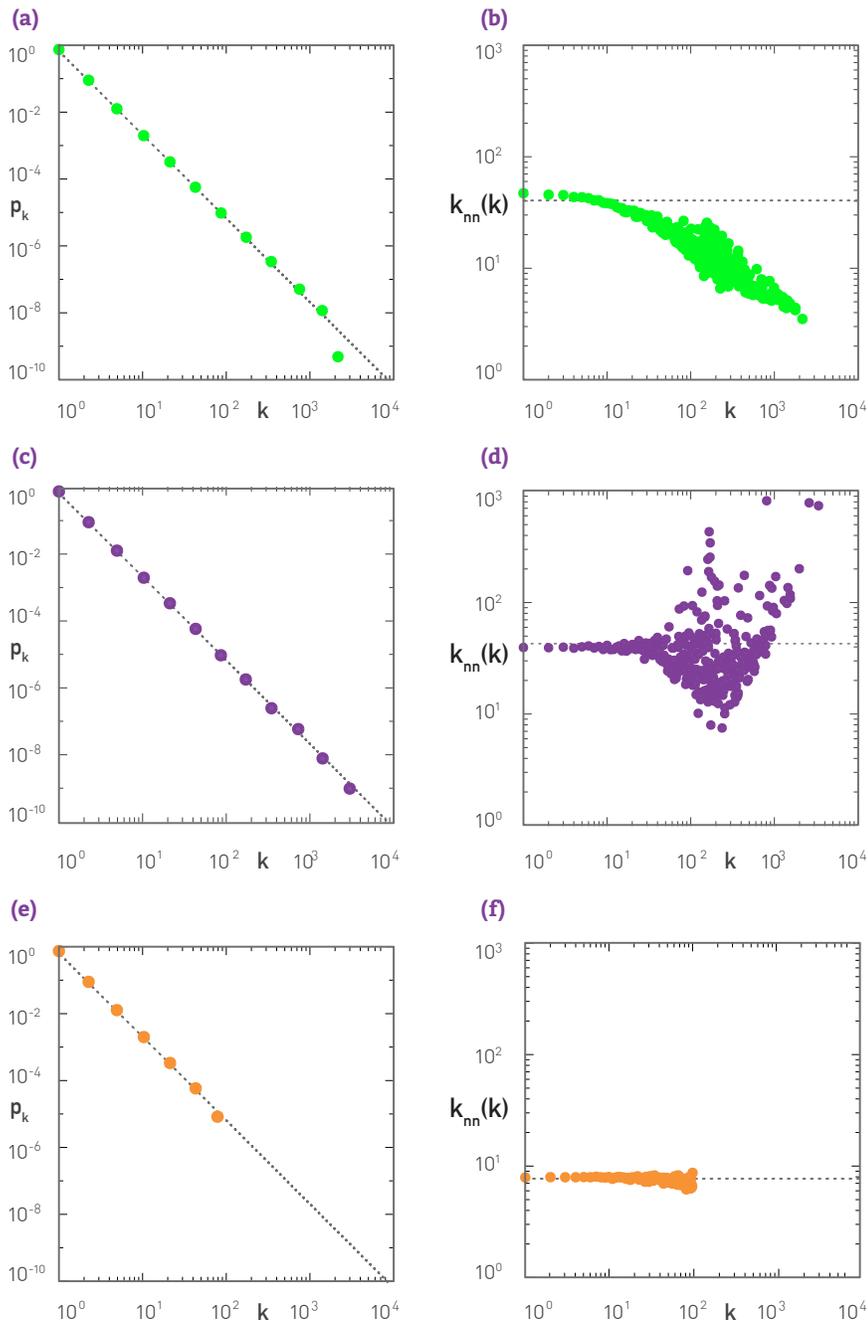


Figure 7.8
Natural and Structural Cutoffs

The figure illustrates the tension between the scale-free property and degree correlations. We show 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 (Figure 4.15).

(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.

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

(e,f) If we impose an upper cutoff by removing all nodes with $k \geq k_s \approx 100$, as predicted by (7.15), the network becomes neutral, as seen in (f).

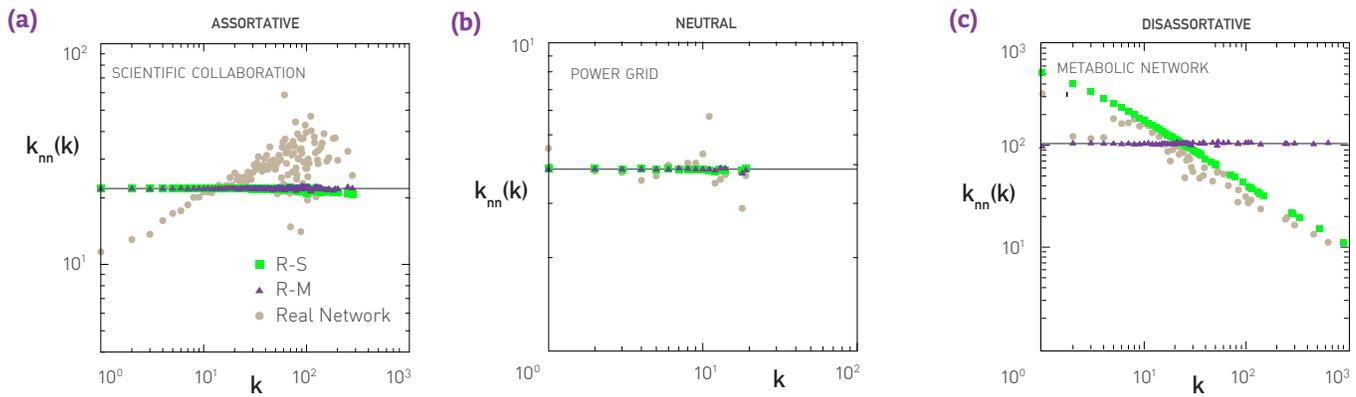


Figure 7.9
Randomization and Degree Correlations

and at each step we make sure that we do *not* permit more than one link between a pair of nodes. On the algorithmic side this means that each rewiring that generates multi-links 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, 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 sometimes useful to 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 it leads to multi-links. This process eliminates all degree correlations.

We performed the randomizations discussed above for three real networks. As Figure 7.9a shows, the assortative nature of the scientific collaboration network disappears under both randomizations. This indicates that the assortative correlations of the collaboration network is not linked to its scale-free nature. In contrast, for the metabolic network the observed disassortativity remains unchanged under *R-S* (Figure 7.9c). Consequently the disassortativity of the metabolic network is structural, being induced by its degree distribution.

In summary, the scale-free property can induce disassortativity in simple networks. Indeed, in neutral or assortative networks we expect multiple links between the hubs. If 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 multiple links between the nodes.

To uncover the origin of the observed degree correlations, we must compare $k_{nn}(k)$ (grey symbols), with $k_{nn}^{R-S}(k)$ and $k_{nn}^{R-M}(k)$ obtained after degree-preserving randomization. Two degree-preserving randomizations are informative in this context:

Randomization with Simple Links (R-S):

At each step of the randomization process we check that we do not have more than one link between any node pairs.

Randomization with Multiple Links (R-M):

We allow multi-links during the randomization processes.

We performed these two randomizations for the networks of Figure 7.6. The *R-M* procedure always generates a neutral network, consequently $k_{nn}^{R-M}(k)$ is always horizontal. The true insight is obtained when we compare $k_{nn}(k)$ with $k_{nn}^{R-S}(k)$, helping 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. Consequently the assortativity 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 disassortativity is induced by its scale-free property. Hence the observed degree correlations are structural.

CORRELATIONS IN REAL NETWORKS

To understand the prevalence of degree correlations we need to inspect the correlations characterizing real networks. In [Figure 7.10](#) we show the $k_{nn}(k)$ function for the ten reference networks, observing several patterns:

- **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 $k_{nn}^{R-S}(k)$, 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 $k_{nn}^{R-S}(k)$ 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 $k_{nn}^{R-S}(k)$ is indistinguishable from $k_{nn}(k)$, indicating that we are observing structural disassortativity, rooted in the scale-free nature of these networks (Figure 7.10 g,h).

- **WWW**

The decaying $k_{nn}(k)$ implies disassortative correlations (Figure 7.10i). The randomized $k_{nn}^{R-S}(k)$ 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 $k \leq 20$ 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 $k \gg k_s$.

In summary, Figure 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 is the only truly neutral network. 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 assortative 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 mechanisms have been proposed to explain the origin of the observed assortativity. For example, the tendency of individuals to form communities, the topic of CHAPTER 9, can induce assortative correlations [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], indicates that individuals tend to associate with other individuals of similar background and characteristics, hence individuals with comparable degree tend to know each other. This degree-homophily may be responsible for the celebrity marriages as well (Figure 7.1).

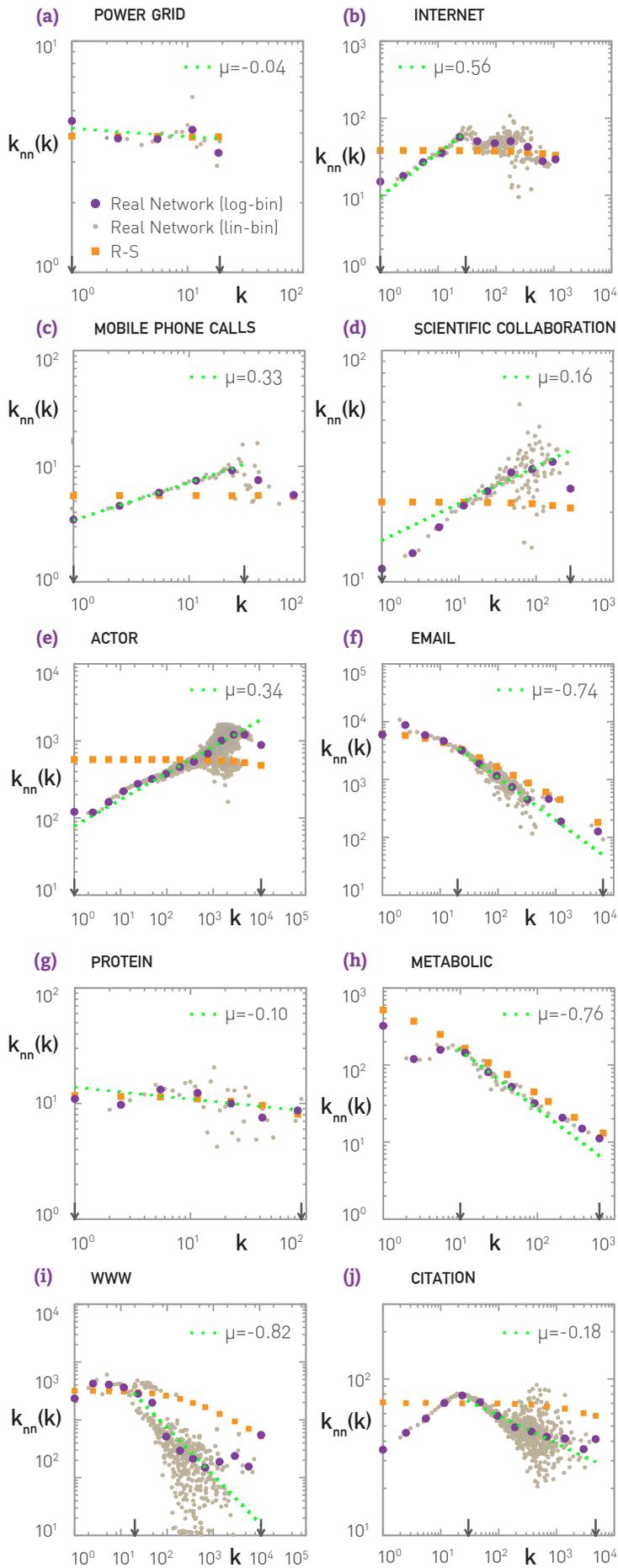


Figure 7.10
Randomization and Degree Correlations

The degree correlation function $k_{nn}(k)$ for the ten reference networks (Table 4.1). The grey symbols show the $k_{nn}(k)$ function using linear binning; purple circles represent the same data using log-binning (SECTION 4.11). The green dotted line corresponds to the best fit to (7.10) within the fitting interval marked by the arrows at the bottom. Orange squares represent $k_{nn}^{R-S}(k)$ obtained for 100 independent degree-preserving randomizations, while maintaining the *simple* character of these networks. 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 (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 [14]. We therefore define four degree correlation functions, $k_{\text{nn}}^{\alpha,\beta}(k)$, where α and β refer to the *in* and *out* indices (Figures 7.11 a-d). In Figure 7.11e we show $k_{\text{nn}}^{\alpha,\beta}(k)$ for citation networks, indicating a lack of *in-out* correlations and the presence of assortativity for small k for the other three correlations (*in-in*, *out-in*, *out-out*).

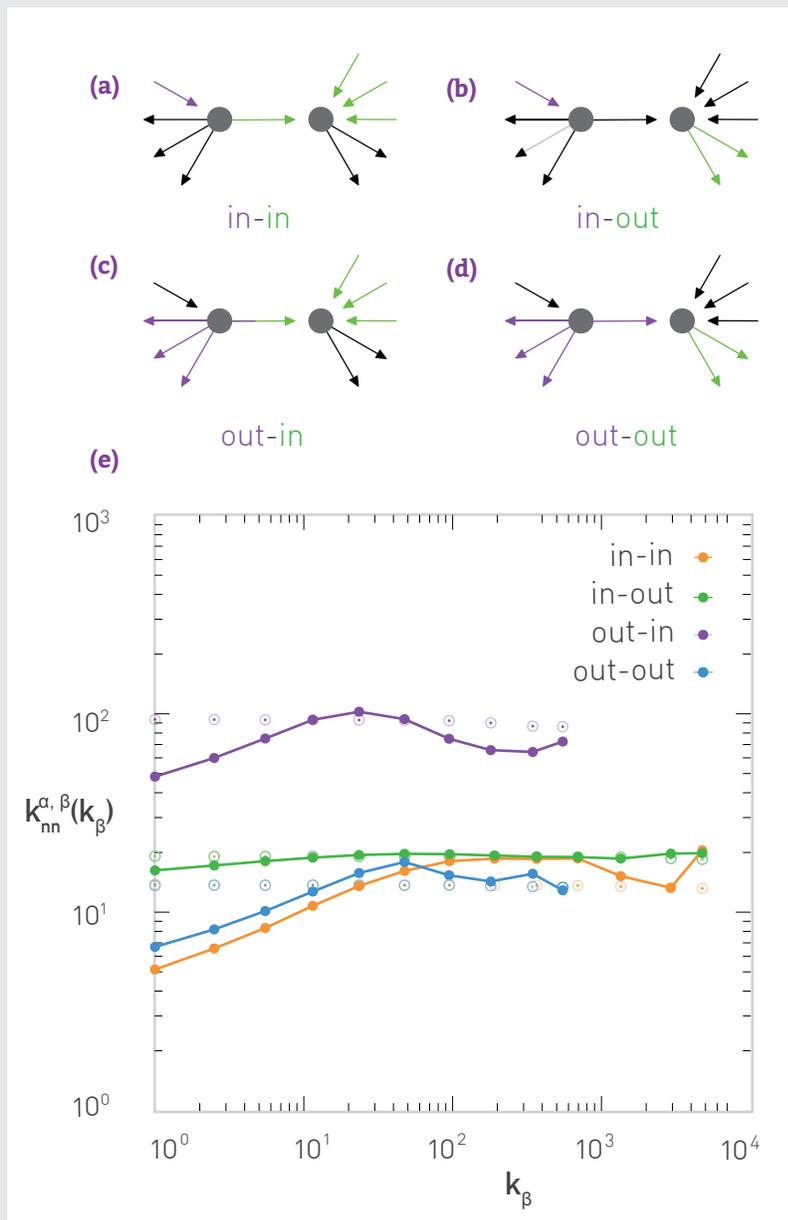


Figure 7.11
Correlations in Directed Network

(a)-(d) The four possible correlations characterizing a directed network. We show in purple and green the (α, β) indices that define the appropriate correlation function [14]. For example, (a) describes the $k_{\text{nn}}^{\text{in,in}}(k)$ correlations between the in-degrees of two nodes connected by a link.

(e) The $k_{\text{nn}}^{\alpha,\beta}(k)$ correlation function for citation networks, a directed network. For example $k_{\text{nn}}^{\text{in,in}}(k)$ is the average in-degree of the in-neighbors of nodes with in-degree k_{in} . These functions show a clear assortative tendency for three of the four functions up to degree $k = 100$. The empty symbols capture the degree randomized $k_{\text{nn}}^{\alpha,\beta}(k)$ for each degree correlation function (R-S randomization).

GENERATING CORRELATED NETWORKS

To explore the impact of degree correlations on various network characteristics we must first understand the correlations characterizing the network models discussed thus far. It is equally important to develop algorithm that can generate networks with tunable correlations. As we show in this section, given the conflict between the scale-free property and degree correlations, this is not a trivial task.

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 (7.9), predicting $\mu = 0$ for any $\langle k \rangle$ and N .

Configuration Model

The configuration model (Figure 4.15) is also neutral, independent of the 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 resolved by the multiple links between them. If, however, we force the network to be simple, then the generated network will develop structural disassortativity (Figure 7.8).

Hidden Parameter Model

In the model e_{jk} is proportional to the product of the randomly chosen hidden variables η_j and η_k (Figure 4.18). Consequently the network is technically uncorrelated. However, if we do not allow multi-links, for scale-free networks we again observe structural disassortativity. Analytical calculations indicate that in this case [18]

$$k_{nn}(k) \sim k^{-1}, \quad (7.16)$$

i.e. the degree correlation function follows (7.10) with $\mu = -1$.

Taken together, the static models explored so far generate either neutral networks, or networks characterized by structural disassortativity following (7.16).

DEGREE CORRELATIONS IN EVOLVING NETWORKS

To understand the emergence (or the absence) of degree correlations in growing networks, we start with the initial attractiveness model (SECTION 6.5), which includes as a special case the Barabási-Albert model.

Initial Attractiveness Model

Consider a growing network in which preferential attachment follows (6.23), i.e. $\Pi(k) \sim A + k$, where A is the initial attractiveness. Depending on the value of A , we observe three distinct scaling regimes [15]:

(i) Disassortative Regime: $\gamma < 3$

If $-m < A < 0$ we have

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

Hence the resulting network is disassortative, $k_{nn}(k)$ decaying following the power-law [15, 16]

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

(ii) Neutral Regime: $\gamma = 3$

If $A = 0$ the initial attractiveness model reduces to the Barabási-Albert model. In this case

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

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

(iii) Weak Assortativity: $\gamma > 3$

If $A > 0$ 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 (7.10).

In summary, (7.17) - (7.20) indicate that the initial attractiveness model generates rather complex degree correlations, from disassortativity to weak assortativity. Equation (7.19) also shows that the network generated by the Barabási-Albert model is neutral. Finally, (7.17) predicts a power law k -dependence for $k_{nn}(k)$, offering analytical support for the empirical scaling (7.10).

Bianconi-Barabási Model

With a uniform fitness distribution the Bianconi-Barabási model generates a disassortative network [5] (Figure 7.12). The fact that the randomized version of the network is also disassortative indicates that the model's disassortativity is structural. Note, however, that the real $k_{nn}(k)$

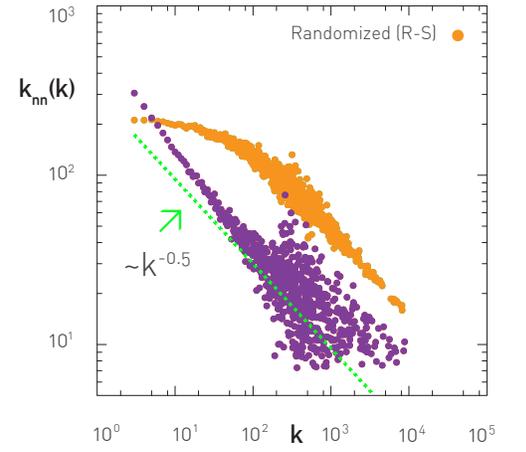
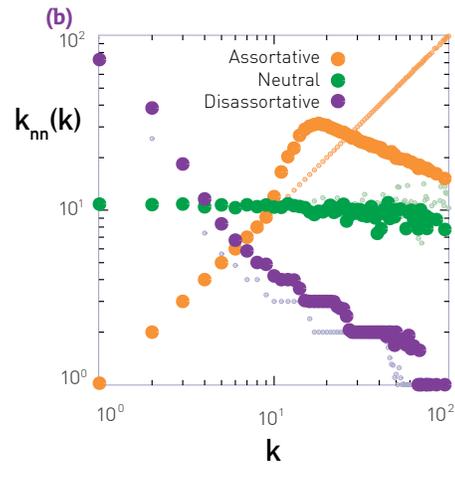
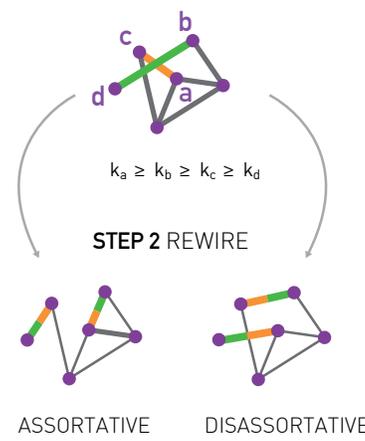


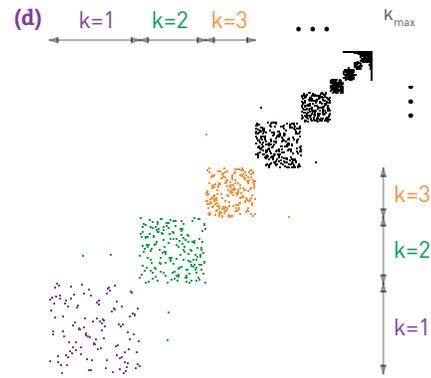
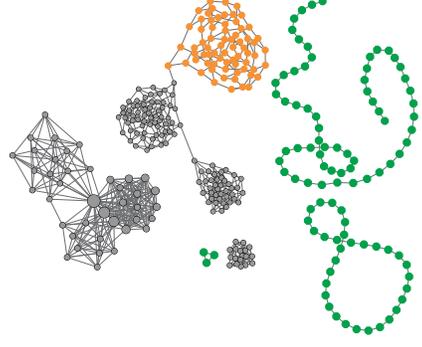
Figure 7.12
Correlations in the Bianconi-Barabási Model

The degree correlation function of the Bianconi-Barabási model for $N = 10,000$, $m = 3$ and uniform fitness distribution (SECTION 6.2). As the green dotted line indicates, following (7.10) indicates, the network is disassortative, consistent with $\mu = -0.5$. The orange symbols correspond to $k_{nn}^{R-S}(k)$. 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 $k_{nn}(k)$ suggests that structural effects cannot fully account for the observed degree correlation.

(a) STEP 1 LINK SELECTION



(c) ASSORTATIVE



(e) DISASSORTATIVE

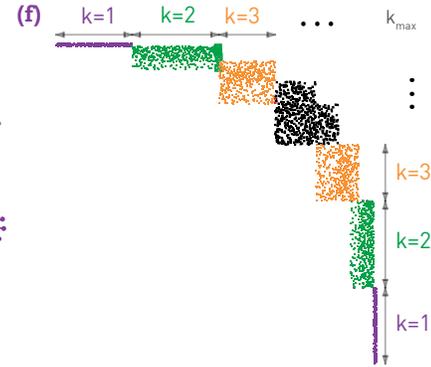
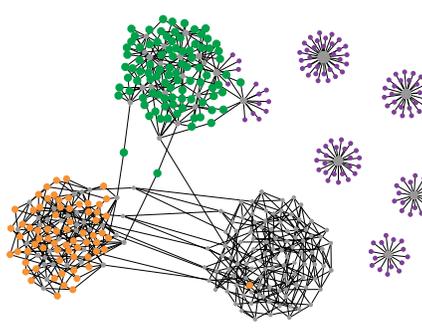


Figure 7.13
Xulvi-Brunet & Sokolov Algorithm

The algorithm generates networks with *maximal degree correlations*.

- (a)** The basic steps of the algorithm.
- (b)** $k_{nn}(k)$ for networks generated by the algorithm for a scale-free network with $N = 1,000$, $L = 2,500$, $\gamma = 3.0$.
- (c, d)** A typical network configuration and the corresponding A_{ij} matrix for the maximally assortative network generated by the algorithm, where the rows and columns of A_{ij} were ordered according to increasing node degrees k .
- (e, f)** Same as in (c, d) for a maximally disassortative network.

The A_{ij} matrices (d) and (f) capture the inner regularity of networks with maximal correlations, consisting of blocks of nodes that connect to nodes with similar degree in (d) and of blocks of nodes that connect to nodes with rather different degrees in (f).

and the randomized $k_{nn}^{R-S}(k)$ do not overlap, indicating that the disassortativity of the model is not fully explained by its scale-free nature.

TUNING DEGREE CORRELATIONS

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

- **Step 2A: Assortative**

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

By iterating these steps we gradually enhance the network's assortative (Step 2A) or disassortative (Step 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.

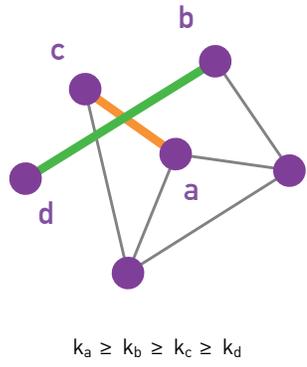
The correlations characterizing the networks generated by this algorithm converge to the maximal (assortative) or minimal (disassortative) value that we can reach for the given degree sequence (Figure 7.13b). The model has no difficulty creating disassortative correlations (Figures 7.13e,f). In the assortative limit simple networks display a mixed $k_{nn}(k)$: assortative for small k and disassortative for high k (Figures 7.13b). This is a consequence of structural cutoffs: For scale-free networks the system is unable to sustain assortativity for high k . The observed behavior is reminiscent of the $k_{nn}(k)$ function of citation networks (Figure 7.10j).

The version of the Xalvi-Brunet & Sokolov algorithm introduced in Figure 7.13 generates maximally assortative or disassortative networks. We can tune the magnitude of the generated degree correlations if we use the algorithm discussed in Figure 7.14.

In summary, static models, like the configuration or hidden parameter model, 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 (7.16) and (7.18), offering the analytical form of the degree correlation function for the hidden parameter model and for a growing network, in both case predicting a power-law k -dependence. These results offer analytical backing for the scaling hypothesis (7.10), indicating that both structural and dynamical effects can result in a degree correlation function that follows a power law.

(a)

STEP 1 LINK SELECTION



STEP 2 REWIRE

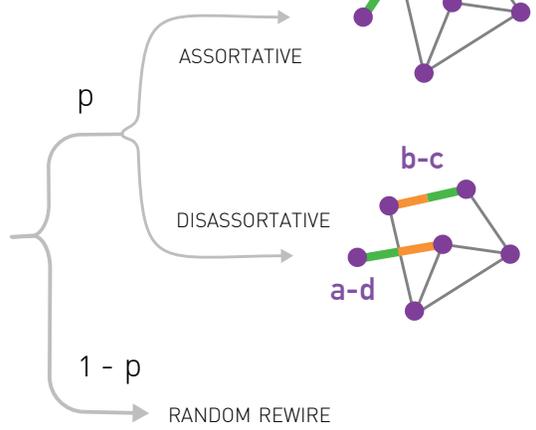


Figure 7.14

Tuning Degree Correlations

We can use the Xalvi-Brunet & Sokolov algorithm to tune the magnitude of degree correlations.

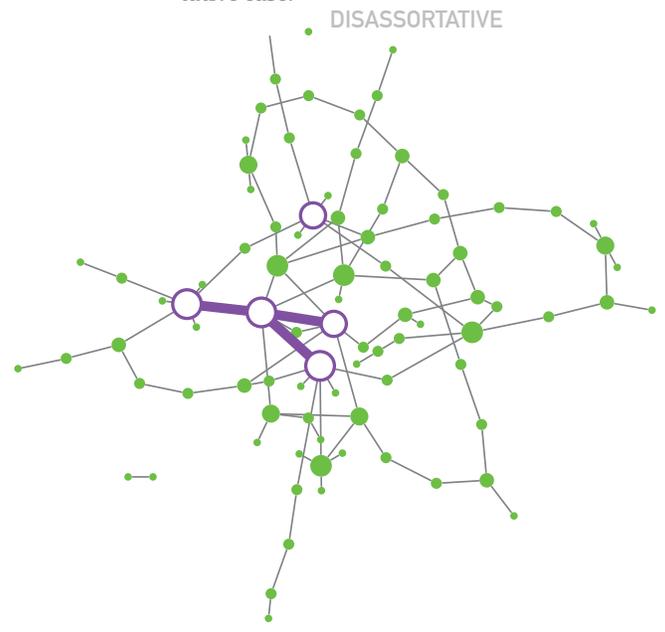
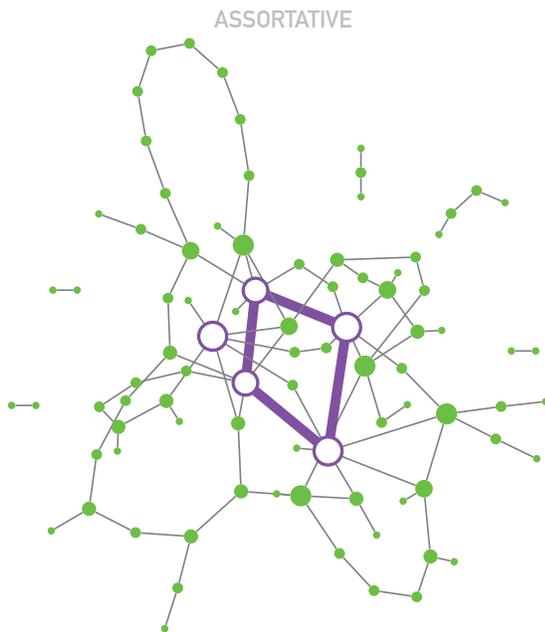
(a) We execute the deterministic rewiring step with probability p , and with probability $1 - p$ we randomly pair the a, b, c, d nodes with each other. For $p = 1$ we are back to the algorithm of Figure 7.13, generating maximal degree correlations; for $p < 1$ the induced noise tunes the magnitude of the effect.

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

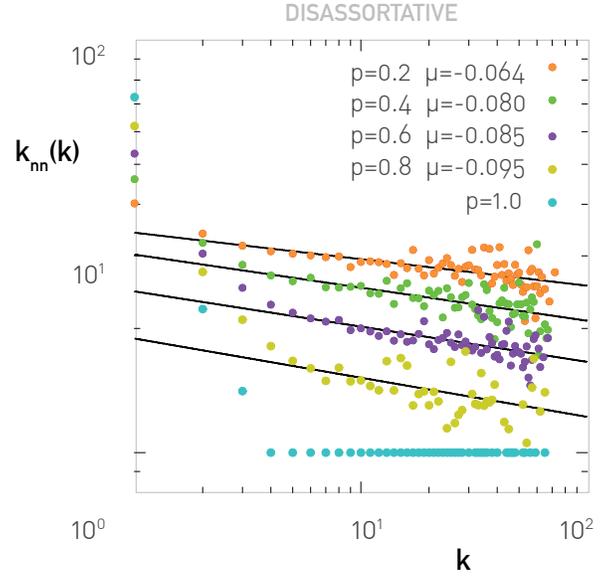
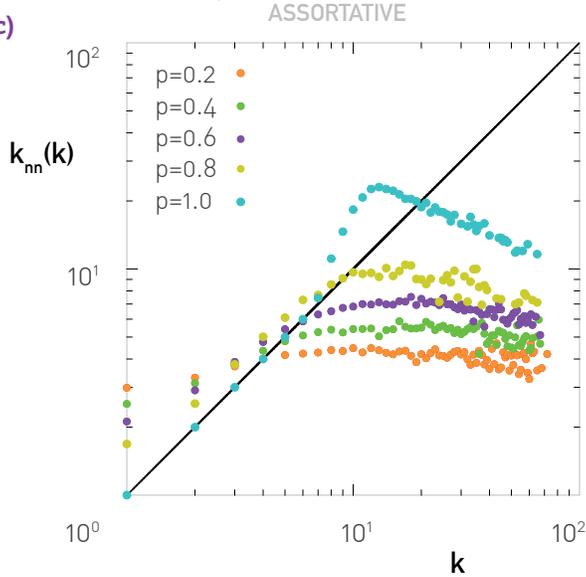
(c) The $k_{nn}(k)$ functions for various p values for a network with $N = 10,000$, $\langle k \rangle = 1$, and $\gamma = 3.0$.

Note that the correlation exponent μ depends on the fitting region, especially in the assortative case.

(b)



(c)



THE IMPACT OF DEGREE CORRELATIONS

As we have seen in Figure 7.10, most real networks are characterized by some degree correlations. Social networks are assortative; biological networks display structural disassortativity. 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? This section addresses these important questions.

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 (SECTION 3.6). Figure 7.15 shows the relative size of the giant component for networks with different degree correlations, documenting several patterns [8, 19, 20]:

- **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.

These changes in the size and the structure of the giant component have implications to the spread of diseases [22, 23, 24], the topic of CHAPTER 10. Indeed, as we have seen in Figure 7.10, social networks tend to be assortative. The high degree nodes therefore form a giant component that acts as

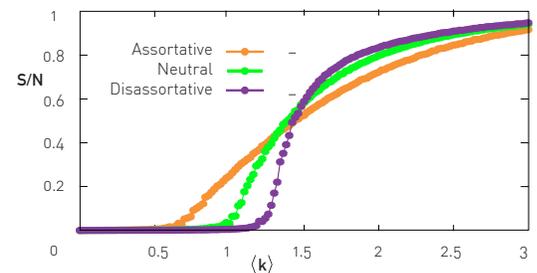


Figure 7.15
Degree Correlations and the Phase Transition Point

Relative size of the giant component for an Erdős-Rényi network of size $N=10,000$ (green curve), which is then rewired using the Xalvi-Brunet & Sokolov algorithm with $p = 0.5$ to induce degree correlations (Figure 7.14). 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 $\langle k \rangle$. Each point represents an average over 10 independent runs.

a “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, the removal of a network’s hubs fragments a network. In assortative networks hub removal makes less damage because the hubs form a core group, hence many of them are redundant. Hub removal 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:

- 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].

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

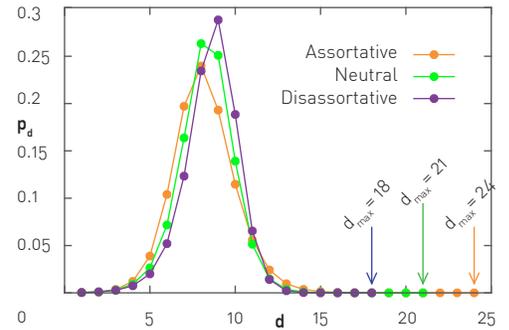


Figure 7.16
Degree Correlations and Path Lengths

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.

BOX 7.4

VERTEX COVER AND MUSEUM GUARDS

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 an 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].

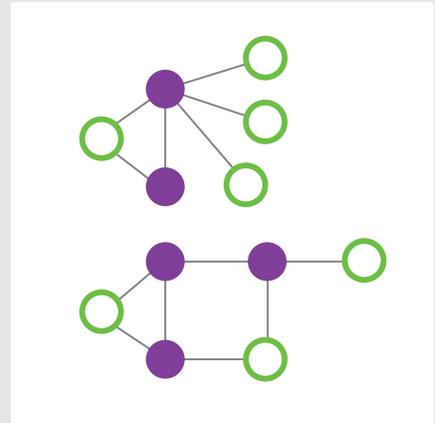


Figure 7.17
The Minimum Cover

Formally, a *vertex cover* of a network is a set C of nodes such that each link of the network 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 minimum vertex covers in two small networks, where the set C is shown in purple. We can check that if we turn any of the purple nodes into green nodes, at least one link will not connect to a purple node.

SUMMARY

Degree correlations were first discovered in 2001 in the context of the Internet by Romualdo Pastor-Satorras, Alexei Vazquez, and Alessandro Vespignani [4, 5], who also introduced the degree correlation function $k_{nn}(k)$ and the scaling (7.10). A year later Kim Sneppen and Sergey Maslov used the full $p(k_i, k_j)$, related to the e_{ij} matrix, to characterize the degree correlations of protein interaction networks [32]. In 2003 Mark Newman introduced the degree correlation coefficient [8, 9] together with the assortative, neutral, and disassortative distinction. These terms have their roots in social sciences [13]:

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 who 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.

Taken together, there are several reasons why we care about degree correlations in networks (BOX 7.5):

- Degree correlations are present in most real networks (SECTION 7.5).

BOX 7.5

AT A GLANCE: DEGREE CORRELATIONS

Degree Correlation Matrix e_{ij}

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

$\mu < 0$: Disassortative

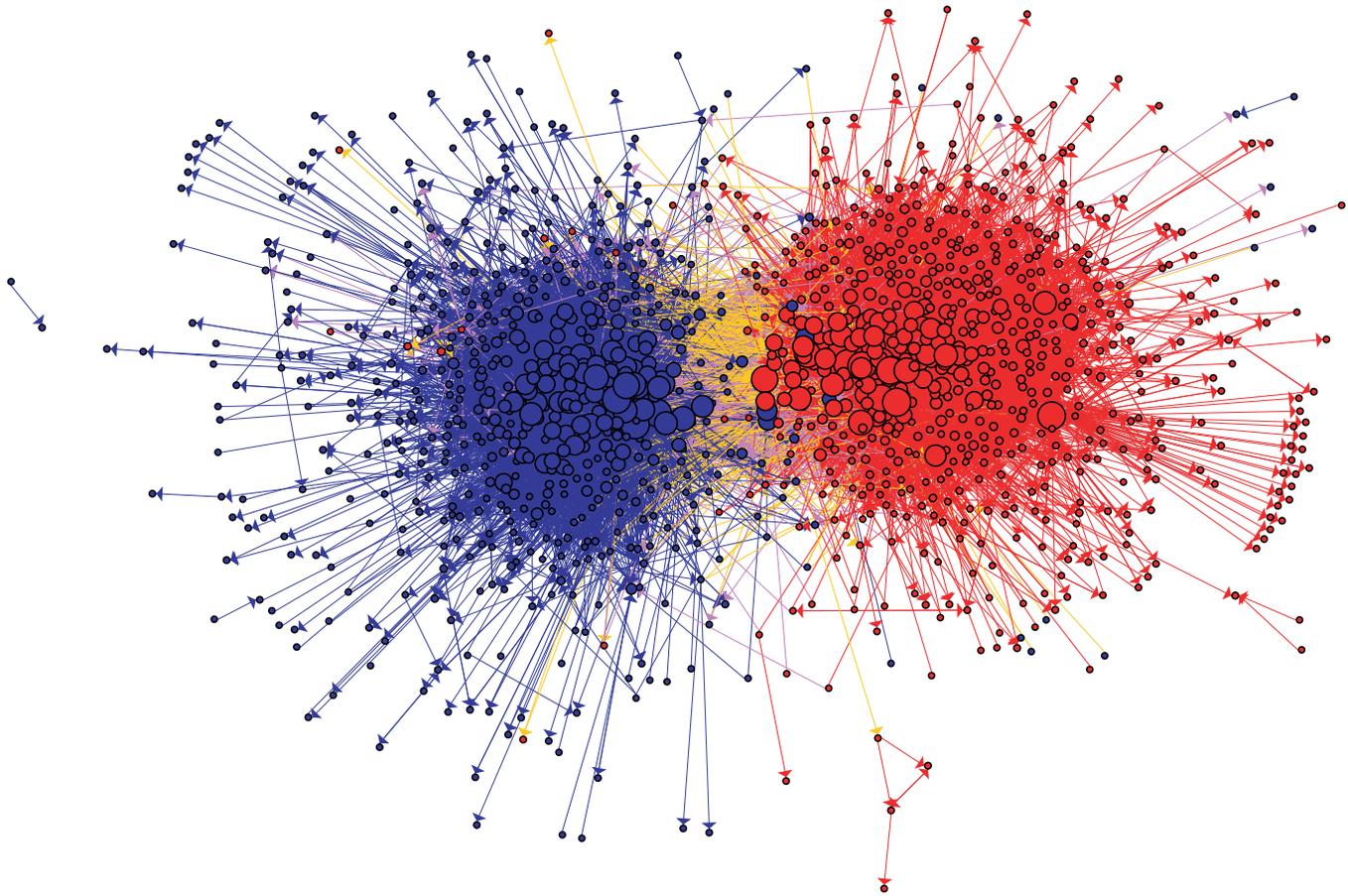
Degree Correlation Coefficient

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

$r > 0$: Assortative

$r = 0$: Neutral

$r < 0$: Disassortative



- Once present, degree correlations change a network's behavior (SECTION 7.7).
- Degree correlations force us to move beyond the degree distribution, representing quantifiable patterns that govern the way nodes link to each other that are not captured by p_k alone.

Figure 7.18
Politics is Never Neutral

Despite the considerable effort devoted to characterizing degree correlations, our understanding of the phenomena remains incomplete. For example, while in SECTION 7.6 we offered an algorithm to tune degree correlations, the problem is far from being fully resolved. Indeed, the most accurate description of a network's degree correlations is contained in the e_{ij} matrix. Generating networks with an arbitrary e_{ij} remains a difficult task.

Finally, in this chapter we focused on the $k_{nn}(k)$ function, which captures two-point correlations. In principle higher order correlations are also present in some networks (BOX 7.6). The impact of such three or four point correlations remains to be understood.

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.

After [33].

BOX 7.6

TWO-POINT, THREE-POINT CORRELATIONS

The complete degree correlations characterizing a network are determined by the conditional probability $P(k^{(1)}, k^{(2)}, \dots, k^{(k)}|k)$ that a node with 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 correlation discussed in this chapter, being the conditional probability $P(k'|k)$ that a node with degree k is connected to a node with degree k' . For uncorrelated networks this conditional probability is independent of k , i.e. $P(k'|k) = k' p_{k'}/\langle k \rangle$ [18]. As the empirical evaluation of $P(k'|k)$ in real networks is cumbersome, it is more practical to analyze the degree correlation function $k_{nn}(k)$ defined in (7.7).

Three-point Correlations

Correlations involving three nodes are determined by $P(k^{(1)}, k^{(2)}|k)$. This conditional probability is connected to the clustering coefficient. Indeed, the average clustering coefficient $C(k)$ [22, 23] can be formally written as the probability that a degree- k node 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 $C(k)$ is independent of k , following

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

HOMEWORK

7.1. Detailed Balance for Degree Correlations

Express the joint probability $e_{kk'}$, the conditional probability $P(k'|k)$ and the probability $q_{k'}$ discussed in this chapter, in terms of number of nodes N , average degree $\langle k \rangle$, number of nodes with degree k , N_k , and the number of links connecting nodes of degree k and k' , $E_{kk'}$ (note that $E_{kk'}$ is twice the number of links when $k = k'$). Based on these expressions, show that for any network we have

$$e_{kk'} = q_{k'} P(k'|k).$$

7.2. Star Network

Consider a star network, where a single node is connected to $N - 1$ degree one nodes. Assume that $N \gg 1$.

- What is the degree distribution p_k of this network?
- What is the probability q_k that moving along a randomly chosen link we find at its end a node with degree k ?
- Calculate the degree correlation coefficient r for this network. Use the expressions of $e_{kk'}$ and $P(k'|k)$ calculated in HOMEWORK 7.1.
- Is this network assortative or disassortative? Explain why.

7.3. Structural Cutoffs

Calculate the structural cutoff k_s for the undirected networks listed in Table 4.1. Based on the plots in Figure 7.10, predict for each network whether k_s is larger or smaller than the maximum expected degree k_{\max} . Confirm your prediction by calculating k_{\max} .

7.4. Degree Correlations in Erdős-Rényi Networks

Consider the Erdős-Rényi $G(N,L)$ model of random networks, introduced in CHAPTER 2 (BOX 3.1 and SECTION 3.2), where N labeled nodes are connected with L randomly placed links. In this model, the probability that there is a link connecting nodes i and j depends on the existence of a link between nodes l and s .

- (a) Write the probability that there is a link between i and j , e_{ij} and the probability that there is a link between i and j conditional on the existence of a link between l and s .
- (b) What is the ratio of such two probabilities for small networks? And for large networks?
- (c) What do you obtain for the quantities discussed in (a) and (b) if you use the Erdős-Rényi $G(N,p)$ model?

Based on the results found for (a)-(c) discuss the implications of using the $G(N,L)$ model instead of the $G(N,p)$ model for generating random networks with small number of nodes.

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 [8, 9]. The use of a single number to characterize degree correlations is attractive, as it offers a way to compare the correlations observed in networks of different nature and size. Yet, to effectively use r we must be aware of its origin.

The hypothesis behind the correlation coefficient r implies 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 (7.10), which assumes a power law dependence on k . Equation (7.21) raises several issues:

- The initial attractiveness model predicts a power law (7.18) or a logarithmic k -dependence (7.20) for the degree correlation function. A similar power law is derived in (7.16) for the hidden parameter model. Consequently, r forces a linear fit to an inherently nonlinear function. This linear dependence is not supported by numerical simulations or analytical calculations. Indeed, as we show in **Figure 7.19**, (7.21) offers a poor fit to the data for both assortative and disassortative networks.
- As we have seen in **Figure 7.10**, the dependence of $k_{nn}(k)$ on k is complex, often changing trends for large k thanks to the structural cutoff. A linear fit ignores this inherent complexity.
- The maximally correlated model has a vanishing r for large N , despite the fact that the network maintains its degree correlations (**BOX 7.7**). This suggests that the degree correlation coefficient has difficulty detecting correlations characterizing large networks.

NETWORK	N	r	μ
Internet	192,244	0.02	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 the estimated r and μ for the ten reference networks. 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**).

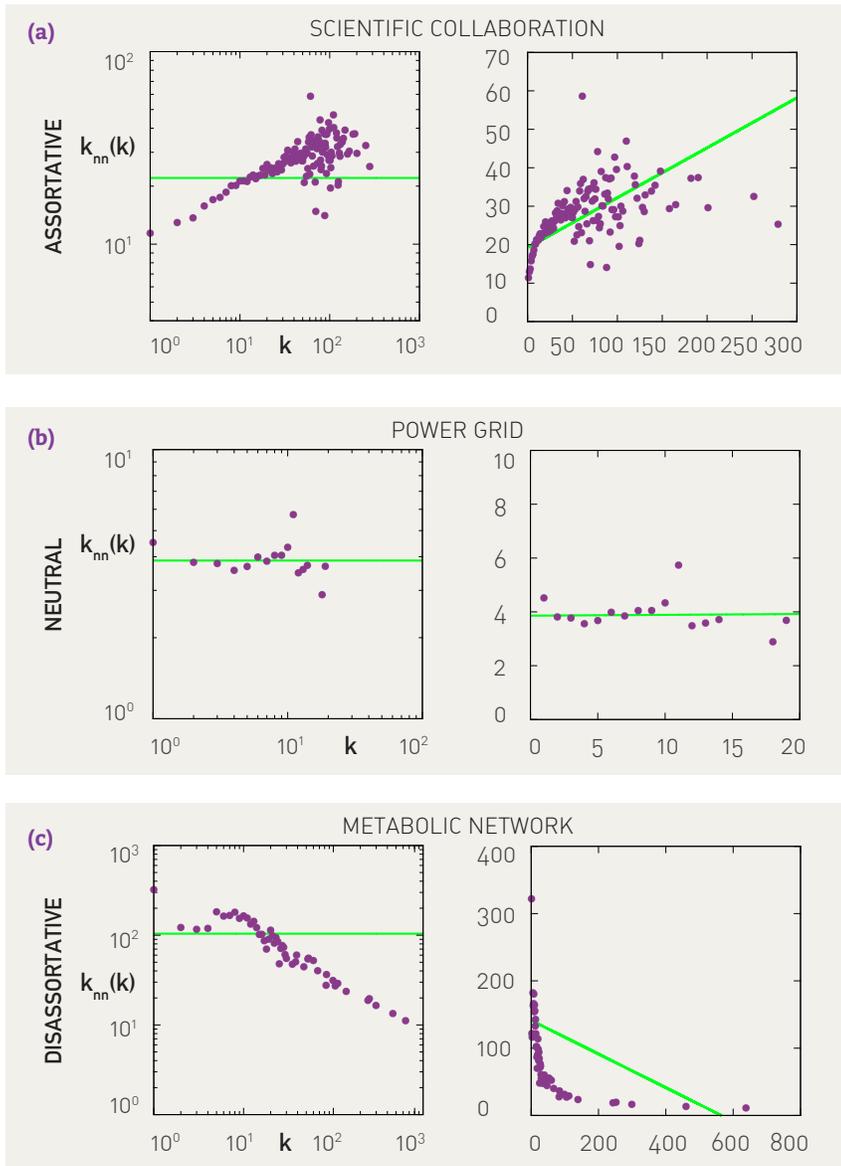


Figure 7.19
Degree Correlation Function

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

Relationship Between μ and r

On the positive side, r and μ are not independent of each other. To show this we calculated r and μ for the ten reference networks (TABLE 7.1). The results are plotted in Figure 7.20, indicating that μ and r correlate for positive r . Note, however, that this correlation breaks down for negative r . To understand the origin of this behavior, next we derive a direct relationship between μ and r . To be specific we assume the validity of (7.10) and determine the value of r for a network with correlation exponent μ .

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

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

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 :

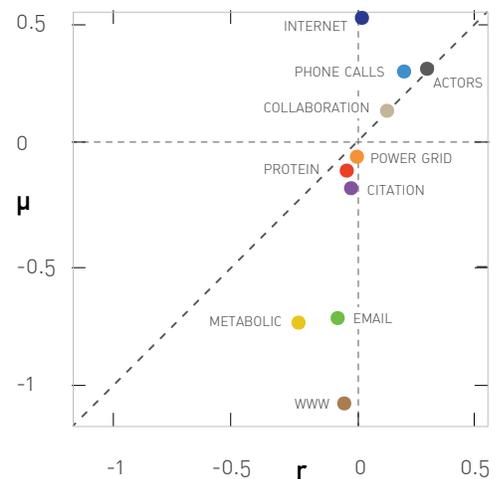


Figure 7.20
Correlation Between r and μ

To illustrate the relationship between r and μ , we estimated μ by fitting the $k_{nn}(k)$ function to (7.10), whether or not the power law scaling was statistically significant.

$$r = \frac{\sum_k k a k^\mu q_k - \frac{\langle k^2 \rangle^2}{\langle k \rangle^2}}{\sigma_r^2} = \frac{\sum_k a k^{\mu+2} \frac{P_k}{\langle k \rangle} - \frac{\langle k^2 \rangle^2}{\langle k \rangle^2}}{\sigma_r^2} = \frac{\langle k^2 \rangle \langle k^{\mu+2} \rangle - \langle k \rangle^2 \langle k^2 \rangle^2}{\sigma_r^2} = \frac{1}{\sigma_r^2} \frac{\langle k^2 \rangle}{\langle k \rangle} \left(\frac{\langle k^{\mu+2} \rangle}{\langle k^{\mu+1} \rangle} - \frac{\langle k^2 \rangle}{\langle k \rangle} \right). \quad (7.22)$$

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 $k > 1$ (7.22) suggests that for $\mu > 0$ the parenthesis is positive, hence $r > 0$, and for $\mu < 0$ the parenthesis is negative, hence $r < 0$. Therefore r and μ predict degree correlations of similar kind.

In summary, if the degree correlation function follows (7.10), then the sign of the degree correlation exponent μ will determine the sign of the coefficient r :

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

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. We therefore define four degree correlation coefficients, $r_{\text{in,in}}$, $r_{\text{in,out}}$, $r_{\text{out,in}}$, $r_{\text{out,out}}$, capturing all possible combinations between the incoming and outgoing degrees of two connected nodes (Figures 7.21a-d). Formally we have [14]

$$r_{\alpha,\beta} = \frac{\sum_{jk} jk (e_{jk}^{\alpha,\beta} - q_{\leftarrow j}^\alpha q_{\rightarrow k}^\beta)}{\sigma_\leftarrow^\alpha \sigma_\rightarrow^\beta}, \quad (7.23)$$

where α and β refer to the *in* and *out* indices and $q_{\leftarrow j}^\alpha$ is the probability of finding a node with α -degree j by following a random link backward and $q_{\rightarrow k}^\beta$ is the probability of finding a β -link with degree k by following a random link forward. σ_\leftarrow^α and σ_\rightarrow^β are the corresponding standard deviations. To illustrate the use of (7.23), in Figure 7.21e we show the four correlation coefficients for the five directed reference networks (TABLE 7.1). Note, however, that for a complete characterization of degree correlations it is desirable to measure the four $k_{\text{nn}}(k)$ functions as well (BOX 7.3).

In summary, the degree correlation coefficient assumes that $k_{\text{nn}}(k)$ scales linearly with k , a hypothesis that lacks numerical and analytical support. Analytical calculations predict the power-law form (7.10) or the weaker logarithmic dependence (7.20). Yet, in general the sign of r and μ do agree. Consequently, we can use r to get a quick sense of the nature of the potential correlations present in a network. Yet, the accurate characterization of the underlying degree correlations requires us to measure $k_{\text{nn}}(k)$.

BOX 7.7

THE PROBLEM WITH LARGE NETWORKS

The Xalvi-Brunet & Sokolov algorithm helps us calculate the maximal (r_{min}) and the minimal (r_{max}) correlation coefficient for a scale-free network, obtaining [21]

$$r_{\text{min}} \sim \begin{cases} -c_1(\gamma, k_0) & \text{for } \gamma < 2 \\ -N^{(2-\gamma)/(\gamma-1)} & \text{for } 2 < \gamma < 3 \\ -N^{(\gamma-4)/(\gamma-1)} & \text{for } 3 < \gamma < 4 \\ -c_2(\gamma, k_0) & \text{for } 4 < \gamma, \end{cases}$$

$$r_{\text{max}} \sim \begin{cases} -N^{(-\gamma-2)/(\gamma-1)} & \text{for } 2 < \gamma < \gamma_r \\ N^{-1/(\gamma^2-1)} & \text{for } \gamma_r < \gamma < 3, \end{cases}$$

where $\gamma_r \approx \frac{1}{2} + \sqrt{17/4} \approx 2.56$.

These expressions indicate that:

- (i) For large N both r_{min} and r_{max} vanish, even though the corresponding networks were rewired to have maximal correlations. Consequently the correlation coefficient r is unable to capture the correlations present in large networks.
- (ii) Scale-free networks with $\gamma < 2.6$ always have negative r . This is a consequence of structural correlations (SECTION 7.4).

Given r 's limitations, we must inspect $k_{\text{nn}}(k)$ to best characterize a large network's degree correlations.

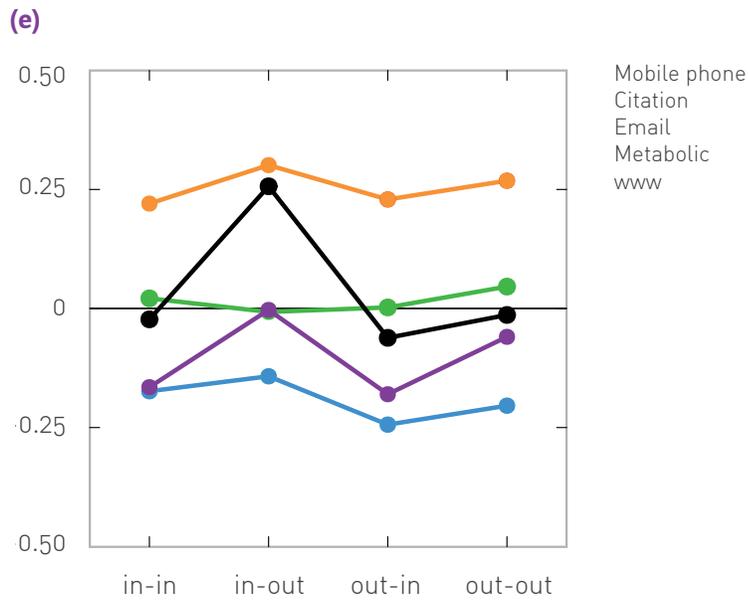
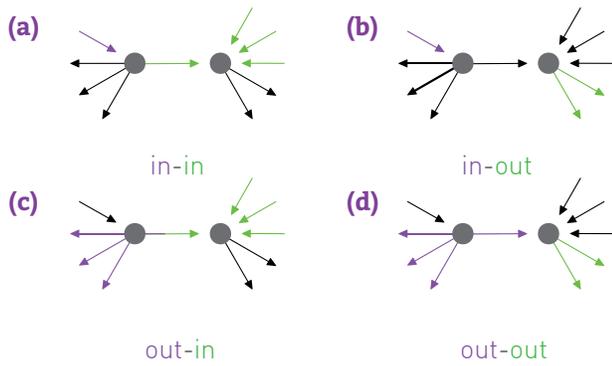


Figure 7.21
Directed Correlation

(a)-(d) The purple and green links indicate the α , β indices that define the appropriate correlation coefficient for a directed network.

(e) The correlation profile of the five directed networks. While citation networks have negligible correlations, all four correlation coefficients document strong assortative behavior for mobile phone calls and strong disassortative behavior for metabolic networks. The case of the WWW is interesting: while three of its correlation coefficients are close to zero, there is a strong assortative tendency for the *in-out* degree combination.

ADVANCED TOPICS 7.B

STRUCTURAL CUTOFFS

As discussed in SECTION 7.4, the fundamental conflict between the scale-free property and degree correlations leads to a structural cutoff in simple networks. In this section we derive (7.15), calculating how the structural cutoff depends on the system size N [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' for $k \neq k'$ and twice the number of connecting links for $k = k'$, and

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

is the largest possible value of $E_{kk'}$. The origin of (7.25) is explained in Figure 7.22. Consequently, we can write $r_{kk'}$ as

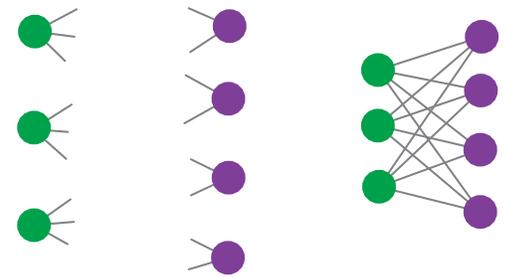
$$r_{kk'} = \frac{E_{kk'}}{m_{kk'}} = \frac{\langle k \rangle e_{kk'}}{\min\{kP(k), k'P(k'), NP(k)P(k')\}}. \quad (7.26)$$

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

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

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

$$r_{kk'} = \frac{\langle k \rangle e_{kk'}}{Np_k p_{k'}}. \quad (7.28)$$



(a) $kN_k = 9$ (b) $k'N_{k'} = 8$ (c) $N_k N_{k'} = 12$
 $mN_{kk'} = \min\{kN_k, k'N_{k'}, N_k N_{k'}\} = 8$

Figure 7.22
Calculating $m_{kk'}$

The maximum number of links one can have between two groups. 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 place between the two groups, which is $N_k N_{k'}$.

In the example shown above the smallest of the three is $k'N_{k'} = 8$ of (b).

For scale-free networks these conditions are fulfilled in the region $k, k' > (aN)^{1/(\gamma+2)}$, where a is a constant that depends on p_k . Note that this value is below the natural cutoff. Consequently this scaling provides a lower bound for the structural cutoff, in the sense that whenever the cutoff of the degree distribution falls below this limit, the condition $r_{kk'} < 1$ is always satisfied.

For neutral networks the joint distribution factorizes as

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

Hence, the ratio (7.28) becomes

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

Therefore, the structural cutoff needed to preserve the condition $r_{kk'} \leq 1$ has the form [11, 34, 35, 36]

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

which is (7.15). Note that (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, and Z.N. Oltvai. Lethality and centrality in protein networks. *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. Detection of topological pattern in complex networks: Correlation profile of the Internet. *Physica*

A, 333: 529-540, 2004.

[11] M. Boguna, R. Pastor-Satorras, and 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. *Phys. Rev. E*, 68: 036122, 2003.

[13] 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.

[14] J. G. Foster, D. V. Foster, P. Grassberger, and 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 dissortativity. *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. Epidemic spreading in correlated complex networks. *Phys. 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. *Phys. Rev. E*, 67: 027101, 2003.

[30] M. Posfai, Y Y. Liu, J-J Slotine, and 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] L. Adamic and N. Glance. The political blogosphere and the 2004 U.S. election: Divided they blog (2005).

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

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

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