DEGREE DISTRIBUTION

In a given realization of a random network some nodes gain numerous links, while others acquire only a few or no links (Figure 3.3). These differences are captured by the degree distribution, p_k, which is the probability that a randomly chosen node has degree k. In this section we derive p_k for a random network and discuss its properties.

BINOMIAL DISTRIBUTION

In a random network the probability that node i has exactly k links is the product of three terms [15]:

- The probability that k of its links are present, or p^k.
- The probability that the remaining (N-1-k) links are missing, or $(1-p)^{N-1-k}$.
- The number of ways we can select k links from N- 1 potential links a node can have, or

$$\binom{N-1}{k}.$$

Consequently the degree distribution of a random network follows the binomial distribution

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}. \tag{3.7}$$

The shape of this distribution depends on the system size N and the probability p (Figure 3.4). The binomial distribution (BOX 3.3) allows us to calculate the network's average degree $<k>$, recovering (3.3), as well as its second moment $<k^2>$ and variance σ_k (Figure 3.4).

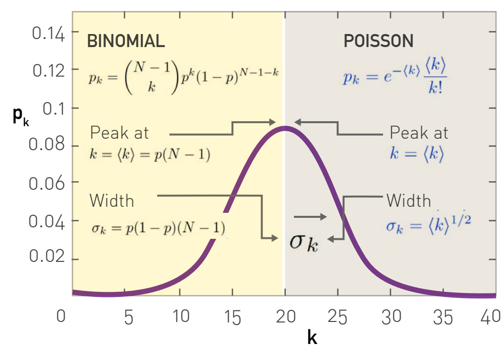

Figure 3.4
Binomial vs. Poisson Degree Distribution

The exact form of the degree distribution of a random network is the binomial distribution (left half). For $N \gg <k>$ the binomial is well approximated by a Poisson distribution (right half). As both formulas describe the same distribution, they have the identical properties, but they are expressed in terms of different parameters: The binomial distribution depends on p and N, while the Poisson distribution has only one parameter, $<k>$. It is this simplicity that makes the Poisson form preferred in calculations.

POISSON DISTRIBUTION

Most real networks are sparse, meaning that for them $<k> \ll N$ (Table 2.1). In this limit the degree distribution (3.7) is well approximated by the Poisson distribution (ADVANCED TOPICS 3.A)

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \tag{3.8}$$

which is often called, together with (3.7), the *degree distribution of a random network.*

The binomial and the Poisson distribution describe the same quantity, hence they have similar properties (Figure 3.4):

- Both distributions have a peak around $<k>$. If we increase p the network becomes denser, increasing $<k>$ and moving the peak to the right.

- The width of the distribution (dispersion) is also controlled by p or $<k>$. The denser the network, the wider is the distribution, hence the larger are the differences in the degrees.

When we use the Poisson form (3.8), we need to keep in mind that:

- The exact result for the degree distribution is the binomial form (3.7), thus (3.8) represents only an approximation to (3.7) valid in the $<k> \ll N$ limit. As most networks of practical importance are sparse, this condition is typically satisfied.

- The advantage of the Poisson form is that key network characteristics, like $<k>$, $<k^2>$ and σ_k , have a much simpler form (Figure 3.4), depending on a single parameter, $<k>$.

- The Poisson distribution in (3.8) does not explicitly depend on the number of nodes N. Therefore, (3.8) predicts that the degree distribution of networks of different sizes but the same average degree $<k>$ are indistinguishable from each other (Figure 3.5).

In summary, while the Poisson distribution is only an approximation to the degree distribution of a random network, thanks to its analytical simplicity, it is the preferred form for p_k. Hence throughout this book, unless noted otherwise, we will refer to the Poisson form (3.8) as the degree distribution of a random network. Its key feature is that its properties are independent of the network size and depend on a single parameter, the average degree $<k>$.

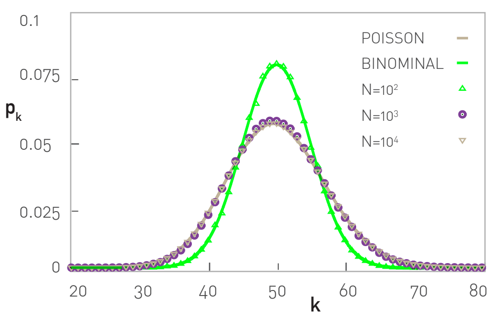

Figure 3.5

Degree Distribution is Independent of the Network Size

The degree distribution of a random network with $<k> = 50$ and $N = 10^2, 10^3, 10^4$.

Small Networks: Binomial
For a small network ($N = 10^2$) the degree distribution deviates significantly from the Poisson form (3.8), as the condition for the Poisson approximation, $N \gg <k>$, is not satisfied. Hence for small networks one needs to use the exact binomial form (3.7) (green line).

Large Networks: Poisson
For larger networks ($N = 10^3, 10^4$) the degree distribution becomes indistinguishable from the Poisson prediction (3.8), shown as a continuous grey line. Therefore for large N the degree distribution is independent of the network size. In the figure we averaged over 1,000 independently generated random networks to decrease the noise.

REAL NETWORKS ARE NOT POISSON

As the degree of a node in a random network can vary between 0 and N-1, we must ask, how big are the differences between the node degrees in a particular realization of a random network? That is, can high degree nodes coexist with small degree nodes? We address these questions by estimating the size of the largest and the smallest node in a random network.

Let us assume that the world's social network is described by the random network model. This random society may not be as far fetched as it first sounds: There is significant randomness in whom we meet and whom we choose to become acquainted with.

Sociologists estimate that a typical person knows about 1,000 individuals on a first name basis, prompting us to assume that $\langle k \rangle \approx 1{,}000$. Using the results obtained so far about random networks, we arrive to a number of intriguing conclusions about a random society of $N \simeq 7 \times 10^9$ of individuals (ADVANCED TOPICS 3.B):

- The most connected individual (the largest degree node) in a random society is expected to have $k_{max} = 1{,}185$ acquaintances.

- The degree of the least connected individual is $k_{min} = 816$, not that different from k_{max} or $\langle k \rangle$.

- The dispersion of a random network is $\sigma_k = \langle k \rangle^{1/2}$, which for $\langle k \rangle = 1{,}000$ is $\sigma_k = 31.62$. This means that the number of friends a typical individual has is in the $\langle k \rangle \pm \sigma_k$ range, or between 968 and 1,032, a rather narrow window.

Taken together, in a random society all individuals are expected to have a comparable number of friends. Hence if people are randomly connected to each other, we lack outliers: There are no highly popular individuals, and no one is left behind, having only a few friends. This suprising conclusion is a consequence of an important property of random networks: *in a large random network the degree of most nodes is in the narrow vicinity of* $\langle k \rangle$

(BOX 3.4).

This prediction blatantly conflicts with reality. Indeed, there is extensive evidence of individuals who have considerably more than 1,185 acquaintances. For example, US president Franklin Delano Roosevelt's appointment book has about 22,000 names, individuals he met personally [16, 17]. Similarly, a study of the social network behind Facebook has documented numerous individuals with 5,000 Facebook friends, the maximum allowed by the social networking platform [18]. To understand the origin of these discrepancies we must compare the degree distribution of real and random networks.

In Figure 3.6 we show the degree distribution of three real networks, together with the corresponding Poisson fit. The figure documents systematic differences between the random network predictions and the real data:

- The Poisson form significantly underestimates the number of high degree nodes. For example, according to the random network model the maximum degree of the Internet is expected to be around 20. In contrast the data indicates the existence of routers with degrees close to 10^3.

- The spread in the degrees of real networks is much wider than expected in a random network. This difference is captured by the dispersion σ_k (Figure 3.4). If the Internet were to be random, we would expect σ_k = 2.52. The measurements indicate $\sigma_{internet}$ = 14.14, significantly higher than the random prediction. These differences are not limited to the networks shown in Figure 3.6, but all networks listed in Table 2.1 share this property.

In summary, the comparison with the real data indicates that the random network model does not capture the degree distribution of real networks. In a random network most nodes have comparable degrees, forbidding hubs. In contrast, in real networks we observe a significant number of highly connected nodes and there are large differences in node degrees. We will resolve these differences in CHAPTER 4.

BOX 3.4

WHY ARE HUBS MISSING?

To understand why hubs, nodes with a very large degree, are absent in random networks, we turn to the degree distribution (3.8).

We first note that the $1/k!$ term in (3.8) significantly decreases the chances of observing large degree nodes. Indeed, the Stirling approximation

$$k! \sim \left[\sqrt{2\pi k}\right]\left(\frac{k}{e}\right)^k$$

allows us rewrite (3.8) as

$$p_k = \frac{e^{-\langle k \rangle}}{\sqrt{2\pi k}}\left(\frac{e\langle k \rangle}{k}\right)^k. \quad (3.9)$$

For degrees $k > e\langle k \rangle$ the term in the parenthesis is smaller than one, hence for large k both k-dependent terms in (3.9), i.e. $1/\sqrt{k}$ and $(e\langle k \rangle/k)^k$ decrease rapidly with increasing k. Overall (3.9) predicts that in a random network the chance of observing a hub decreases faster than exponentially.

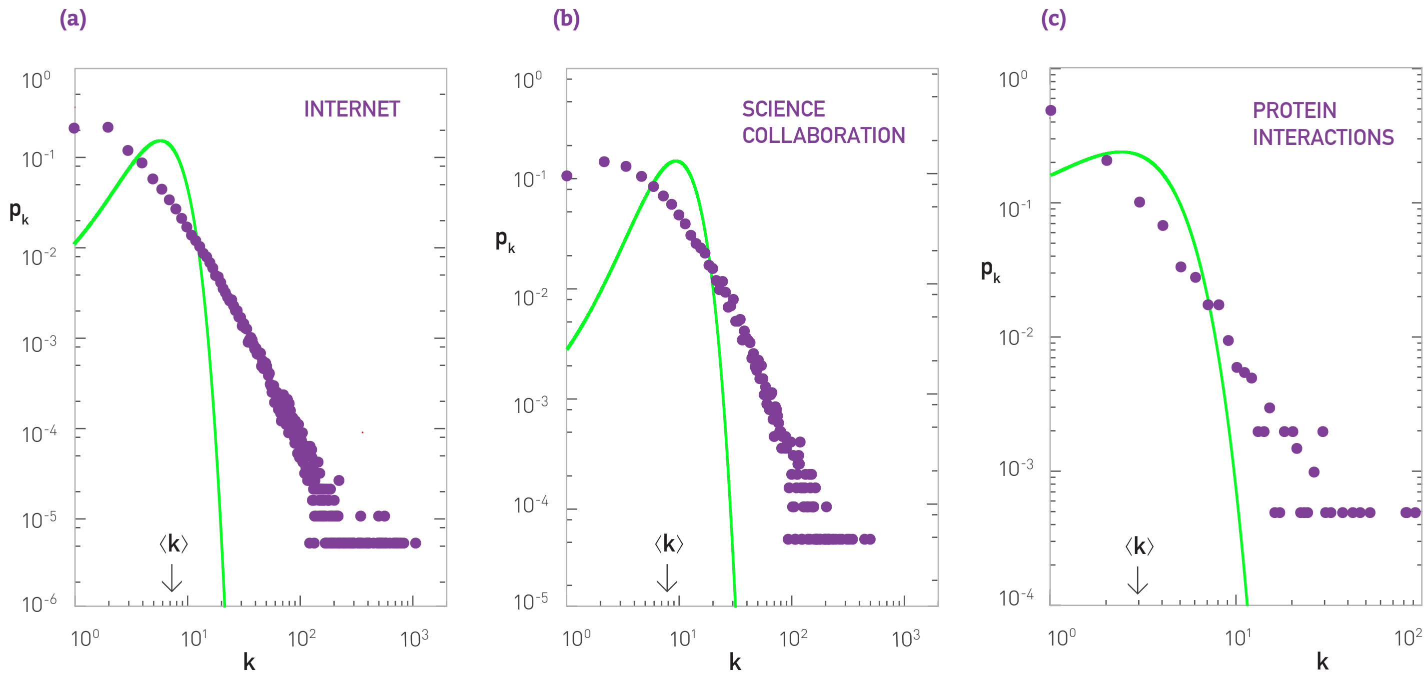

Figure 3.6

Degree Distribution of Real Networks

The degree distribution of the (a) Internet, (b) science collaboration network, and (c) protein interaction network (Table 2.1). The green line corresponds to the Poisson prediction, obtained by measuring $<k>$ for the real network and then plotting (3.8). The significant deviation between the data and the Poisson fit indicates that the random network model underestimates the size and the frequency of the high degree nodes, as well as the number of low degree nodes. Instead the random network model predicts a larger number of nodes in the vicinity of $<k>$ than seen in real networks.