ADJACENCY MATRIX

A complete description of a network requires us to keep track of its links. The simplest way to achieve this is to provide a complete list of the links. For example, the network of Figure 2.2 is uniquely described by listing its four links: {(1, 2), (1, 3), (2, 3), (2, 4)}. For mathematical purposes we often represent a network through its adjacency matrix. The *adjacency matrix* of a directed network of N nodes has N rows and N columns, its elements being:

$A_{ij} = 1$ if there is a link pointing from node j to node i
$A_{ij} = 0$ if nodes i and j are not connected to each other

The adjacency matrix of an undirected network has two entries for each link, e.g. link (1, 2) is represented as $A_{12} = 1$ and $A_{21} = 1$. Hence, the adjacency matrix of an undirected network is symmetric, $A_{ij} = A_{ji}$ (Figure 2.5b).

The degree k_i of node i can be directly obtained from the elements of the adjacency matrix. For undirected networks a node's degree is a sum over either the rows or the columns of the matrix, i.e.

$$k_i = \sum_{j=1}^{N} A_{ji} = \sum_{i=1}^{N} A_{ji} \ . \qquad (2.9)$$

For directed networks the sums over the adjacency matrix' rows and columns provide the incoming and outgoing degrees, respectively

$$k_i^{\text{in}} = \sum_{j=1}^{N} A_{ij} , \qquad k_i^{\text{out}} = \sum_{j=1}^{N} A_{ji} . \qquad (2.10)$$

Given that in an undirected network the number of outgoing links equals the number of incoming links, we have

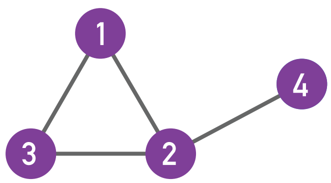
$$2L = \sum_{i=1}^{N} k_i^{\text{in}} = \sum_{i=1}^{N} k_i^{\text{out}} = \sum_{ij}^{N} A_{ij} . \qquad (2.11)$$

The number of nonzero elements of the adjacency matrix is $2L$, or twice the number of links. Indeed, an undirected link connecting nodes i and j appears in two entries: $A_{ij} = 1$, a link pointing from node j to node i, and $A_{ji} = 1$, a link pointing from i to j (Figure 2.5b).

(a) Adjacency matrix

$$A_{ij} = \begin{matrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{matrix}$$

(b) Undirected network



$$A_{ij} = \begin{matrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

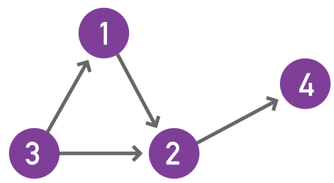
$$k_2 = \sum_{j=1}^{4} A_{2j} = \sum_{i=1}^{4} A_{i2} = 3$$

$$A_{ij} = A_{ji} \qquad A_{ii} = 0$$

$$L = \frac{1}{2} \sum_{i,j=1}^{N} A_{ij}$$

$$\langle k \rangle = \frac{2L}{N}$$

(c) Directed network



$$A_{ij} = \begin{matrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{matrix}$$

$$k_2^{\text{in}} = \sum_{j=1}^{4} A_{2j} = 2, \; k_2^{\text{out}} = \sum_{i=1}^{4} A_{i2} = 1$$

$$A_{ij} \neq A_{ji} \qquad A_{ii} = 0$$

$$L = \sum_{i,j=1}^{N} A_{ij}$$

$$\langle k^{\text{in}} \rangle = \langle k^{\text{out}} \rangle = \frac{L}{N}$$

Figure 2.5

The Adjacency Matrix

(a) The labeling of the elements of the adjacency matrix.

(b) The adjacency matrix of an *undirected network*. The figure shows that the degree of a node (in this case node 2) can be expressed as the sum over the appropriate column or the row of the adjacency matrix. It also shows a few basic network characteristics, like the total number of links, L, and average degree, $<k>$, expressed in terms of the elements of the adjacency matrix.

(c) The same as in **(b)** but for a *directed network*.

SECTION 2.5

REAL NETWORKS ARE SPARSE

In real networks the number of nodes (N) and links (L) can vary widely. For example, the neural network of the worm C. elegans, the only fully mapped nervous system of a living organism, has $N = 302$ neurons (nodes). In contrast the human brain is estimated to have about a hundred billion ($N \approx 10^{11}$) neurons. The genetic network of a human cell has about 20,000 genes as nodes; the social network consists of seven billion individuals ($N \approx 7\times10^9$) and the WWW is estimated to have over a trillion web documents ($N > 10^{12}$).

These wide differences in size are noticeable in Table 2.1, which lists N and L for several network maps. Some of these maps offer a complete wiring diagram of the system they describe (like the actor network or the E. coli metabolism), while others are only samples, representing a subset of the full network (like the WWW or the mobile call graph).

Table 2.1 indicates that the number of links also varies widely. In a network of N nodes the number of links can change between $L = 0$ and L_{max}, where

$$L_{max} = \binom{N}{2} = \frac{N(N-1)}{2} \tag{2.12}$$

is the total number of links present in a *complete graph* of size N (Figure 2.6). In a complete graph each node is connected to every other node.

In real networks L is much smaller than L_{max}, reflecting the fact that most real networks are sparse. We call a network *sparse* if $L \ll L_{max}$. For example, the WWW graph in Table 2.1 has about 1.5 million links. Yet, if the WWW were to be a complete graph, it should have $L_{max} \approx 5\times10^{10}$ links according to (2.12). Consequently the web graph has only a 3×10^{-5} fraction of the links it could have. This is true for all of the networks in Table 2.1: One can check that their number of links is only a tiny fraction of the expected number of links for a complete graph of the same number of nodes.

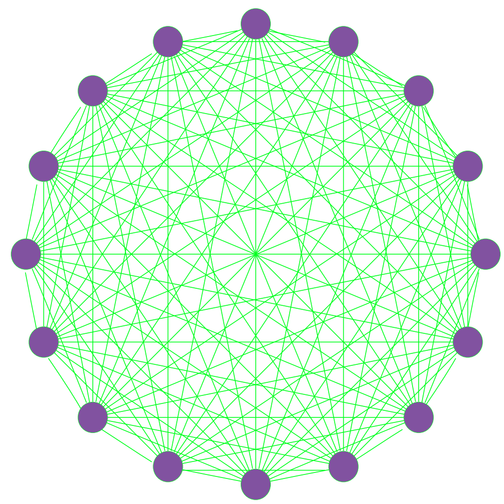


Figure 2.6
Complete Graph

A complete graph with $N = 16$ nodes and $L_{max} = 120$ links, as predicted by (2.12). The adjacency matrix of a complete graph is $A_{ij} = 1$ for all $i, j = 1, \ldots, N$ and $A_{ii} = 0$. The average degree of a complete graph is $\langle k \rangle = N - 1$. A complete graph is often called a *clique*, a term frequently used in community identification, a problem discussed in CHAPTER 9.

The sparsity of real networks implies that the adjacency matrices are also sparse. Indeed, a complete network has $A_{ij} = 1$, for all (i, j), i.e. each of its matrix elements are equal to one. In contrast in real networks only a tiny fraction of the matrix elements are nonzero. This is illustrated in Figure 2.7, which shows the adjacency matrix of the protein-protein interaction network listed in Table 2.1 and shown in Figure 2.4a. One can see that the matrix is nearly empty.

Sparseness has important consequences on the way we explore and store real networks. For example, when we store a large network in our computer, it is better to store only the list of links (i.e. elements for which $A_{ij} \neq 0$), rather than the full adjacency matrix, as an overwhelming fraction of the A_{ij} elements are zero. Hence the matrix representation will block a huge chunk of memory, filled mainly with zeros (Figure 2.7).

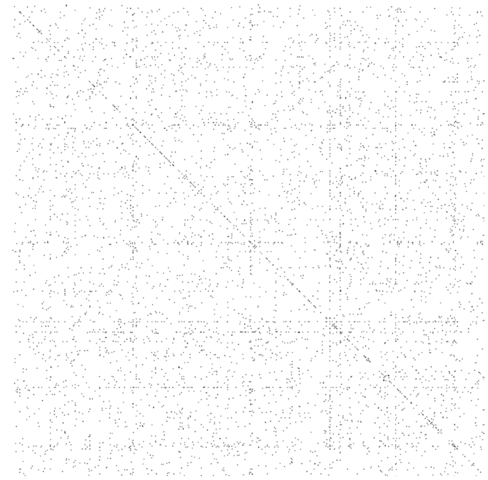


Figure 2.7

The Adjacency Matrix is Sparse

The adjacency matrix of the yeast protein-protein interaction network, consisting of 2,018 nodes, each representing a yeast protein (Table 2.1). A dot is placed on each position of the adjacent matrix for which $A_{ij} = 1$, indicating the presence of an interaction. There are no dots for $A_{ij} = 0$. The small fraction of dots illustrates the sparse nature of the protein-protein interaction network.

WEIGHTED NETWORKS

So far we discussed only networks for which all links have the same weight, i.e. $A_{ij} = 1$. In many applications we need to study *weighted networks*, where each link (i, j) has a unique weight w_{ij}. In mobile call networks the weight can represent the total number of minutes two individuals talk with each other on the phone; on the power grid the weight is the amount of current flowing through a transmission line.

For *weighted networks* the elements of the adjacency matrix carry the weight of the link as

$$A_{ij} = w_{ij} \; . \tag{2.13}$$

Most networks of scientific interest are weighted, but we can not always measure the appropriate weights. Consequently we often approximate these networks with an unweighted graph. In this book we predominantly focus on unweighted networks, but whenever appropriate, we discuss how the weights alter the corresponding network property (BOX 2.3).

BOX 2.3

METCALFE'S LAW: THE VALUE OF A NETWORK

Metcalfe's law states that the *value of a network* is proportional to the square of the number of its nodes, i.e. N^2. Formulated around 1980 in the context of communication devices by Robert M. Metcalfe [9], the idea behind Metcalfe's law is that the more individuals use a network, the more valuable it becomes. Indeed, the more of your friends use email, the more valuable the service is to you.

During the Internet boom of the late 1990s Metcalfe's law was frequently used to offer a quantitative valuation for Internet companies. It suggested that the value of a service is proportional to the number of connections it can create, which is the square of the number of its users. In contrast the cost grows only linearly with N. Hence if the service attracts sufficient number of users, it will inevitably become profitable, as N^2 will surpass N at some large N (Figure 2.8). Metcalfe's Law therefore supported a "build it and they will come" mentality [10], offering credibility to growth over profits.

Metcalfe's law is based on (2.12), telling us that if *all links* of a communication network with N users are equally valuable, the total value of the network is proportional to $N(N - 1)/2$, that is, roughly, N^2. If a network has $N = 10$ consumers, there are $L_{max} = 45$ different possible connections between them. If the network doubles in size to $N = 20$, the number of connections doesn't merely double but roughly quadruples to 190, a phenomenon called *network externality* in economics.

Two issues limit the validity of Metcalfe's law:

(a) Most real networks are sparse, which means that only a very small fraction of the links are present. Hence the value of the network does not grow like N^2, but increases only linearly with N.

(b) As the links have weights, not all links are of equal value. Some links are used heavily while the vast majority of links are rarely utilized.

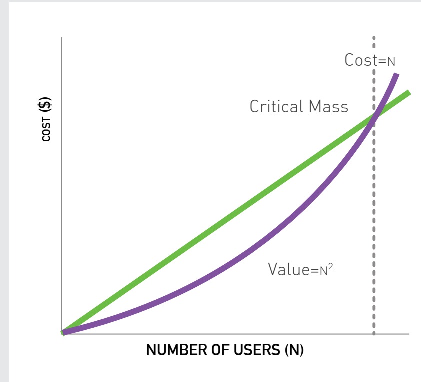


Figure 2.8
Metcalfe's Law

According to Metcalfe's law the *cost* of network based services increases linearly with the number of nodes (users or devices). In contrast the *benefits* or *income* are driven by the number of links L_{max} the technology makes possible, which grows like N^2 according to (2.12). Hence once the number of users or devices exceeds some *critical mass*, the technology becomes profitable.