CONNECTEDNESS

A phone would be of limited use as a communication device if we could not call any valid phone number; email would be rather useless if we could send emails to only certain email addresses, and not to others. From a network perspective this means that the network behind the phone or the Internet must be capable of establishing a path between *any* two nodes. This is in fact the key utility of most networks: they ensure *connectedness*. In this section we discuss the graph-theoretic formulation of connectedness.

In an undirected network nodes i and j are *connected* if there is a path between them. They are *disconnected* if such a path does not exist, in which case we have $d_{ij} = \infty$. This is illustrated in Figure 2.15a, which shows a network consisting of two disconnected clusters. While there are paths between any two nodes on the same cluster (for example nodes 4 and 6), there are no paths between nodes that belong to different clusters (nodes 1 and 6).

A *network is connected* if all pairs of nodes in the network are connected. A *network is disconnected* if there is at least one pair with $d_{ij} = \infty$. Clearly the network shown in Figure 2.15a is disconnected, and we call its two subnetworks *components* or *clusters*. A *component* is a subset of nodes in a network, so that there is a path between any two nodes that belong to the component, but one cannot add any more nodes to it that would have the same property.

If a network consists of two components, a properly placed single link can connect them, making the network connected (Figure 2.15b). Such a link is called a *bridge*. In general a bridge is any link that, if cut, disconnects the network.

While for a small network visual inspection can help us decide if it is connected or disconnected, for a network consisting of millions of nodes connectedness is a challenging question. Mathematical and algorithmic tools can help us identify the connected components of a graph. For example, for a disconnected network the adjacency matrix can be rearranged into a block diagonal form, such that all nonzero elements in the matrix are contained in square blocks along the matrix' diagonal and all other elements are zero (Figure 2.15a). Each square block corresponds to a component. We can use the tools of linear algebra to decide if the adjacency matrix is block diagonal, helping us to identify the connected components.

In practice, for large networks the components are more efficiently identified using the BFS algorithm (BOX 2.6).

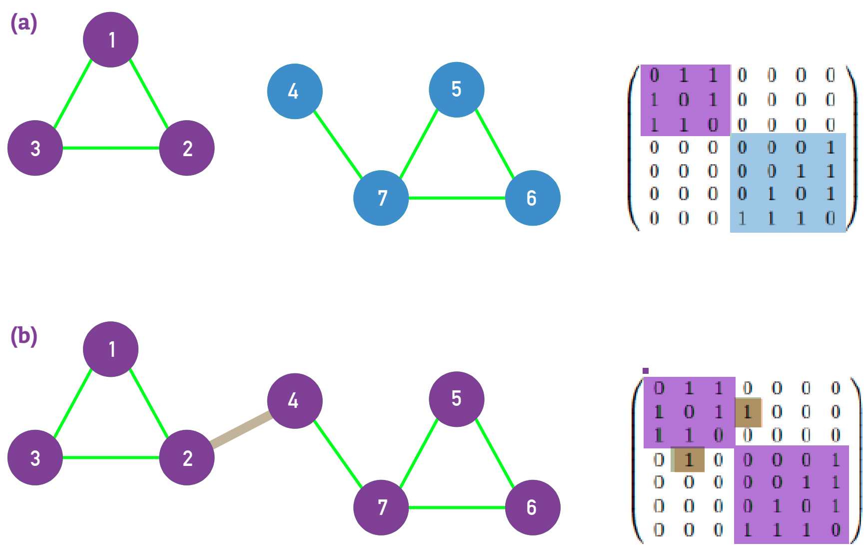

Figure 2.15
Connected and Disconnected Networks

(a) A small network consisting of two disconnected components. Indeed, there is a path between any pair of nodes in the (1,2,3) component, as well in the (4,5,6,7) component. However, there are no paths between nodes that belong to the different components.

The right panel shows the adjacently matrix of the network. If the network has disconnected components, the adjacency matrix can be rearranged into a block diagonal form, such that all nonzero elements of the matrix are contained in square blocks along the diagonal of the matrix and all other elements are zero.

(b) The addition of a single link, called a *bridge*, shown in grey, turns a disconnected network into a single connected component. Now there is a path between every pair of nodes in the network. Consequently the adjacency matrix cannot be written in a block diagonal form.

BOX 2.6

FINDING THE CONNECTED COMPONENTS OF A NETWORK

1. Start from a randomly chosen node i and perform a BFS (BOX 2.5). Label all nodes reached this way with $n = 1$.

2. If the total number of labeled nodes equals N, then the network is connected. If the number of labeled nodes is smaller than N, the network consists of several components. To identify them, proceed to step 3.

3. Increase the label $n \rightarrow n + 1$. Choose an unmarked node j, label it with n. Use BFS to find all nodes reachable from j, label them all with n. Return to step 2.

CLUSTERING COEFFICIENT

The clustering coefficient captures the degree to which the neighbors of a given node link to each other. For a node i with degree k_i the *local clustering coefficient* is defined as [12]

$$C_i = \frac{2L_i}{k_i(k_i-1)} \tag{2.15}$$

where L_i represents the number of links between the k_i neighbors of node i. Note that C_i is between 0 and 1 (Figure 2.16a):

- $C_i = 0$ if none of the neighbors of node i link to each other.
- $C_i = 1$ if the neighbors of node i form a complete graph, i.e. they all link to each other.
- C_i is the probability that two neighbors of a node link to each other. Consequently $C = 0.5$ implies that there is a 50% chance that two neighbors of a node are linked.

In summary C_i measures the network's local link density: The more densely interconnected the neighborhood of node i, the higher is its local clustering coefficient.

The degree of clustering of a whole network is captured by the *average clustering coefficient*, $\langle C \rangle$, representing the average of C_i over all nodes i = 1, ..., N [12],

$$\langle C \rangle = \frac{1}{N}\sum_{i=1}^{N} C_i . \tag{2.16}$$

In line with the probabilistic interpretation $\langle C \rangle$ is the probability that two neighbors of a randomly selected node link to each other.

While (2.16) is defined for undirected networks, the clustering coefficient can be generalized to directed and weighted [13, 14, 15, 16] networks as well. In the network literature we may encounter the *global clustering coefficient* as well, discussed in ADVANCED TOPICS 2.A.

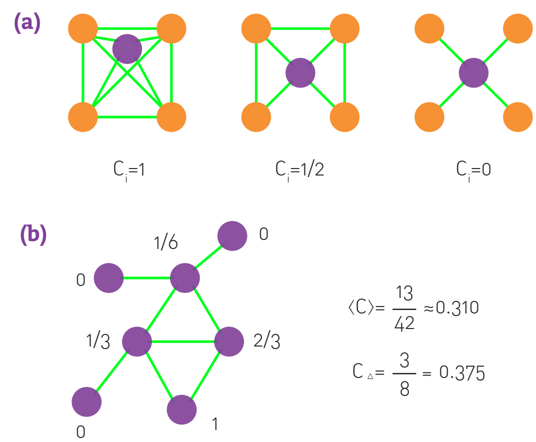

Figure 2.16
Clustering Coefficient

(a) The local clustering coefficient, C_i, of the central node with degree $k_i = 4$ for three different configurations of its neighborhood. The local clustering coefficient measures the local density of links in a node's vicinity.

(b) A small network, with the local clustering coefficient of each nodes shown next to it. We also list the network's average clustering coefficient $\langle C \rangle$, according to (2.16), and its global clustering coefficient C_Δ, defined in SECTION 2.12, Eq. (2.17). Note that for nodes with degrees k_i = 0,1, the clustering coefficient is zero.