3 Mathematics of Networks

Networks: An Introduction,
M. E. J. Newman, Oxford University Press (2010), Chapter 6.

The structure and function of complex networks,
M. E. J. Newman, SIAM review (2003), Sec. I.

3.1 Representation of Networks

In its most general form a network is a graph, i.e. an ordered set of vertices and edges:

$$\mathcal{N} = \{\mathcal{V}, \mathcal{E}\} \tag{1}$$

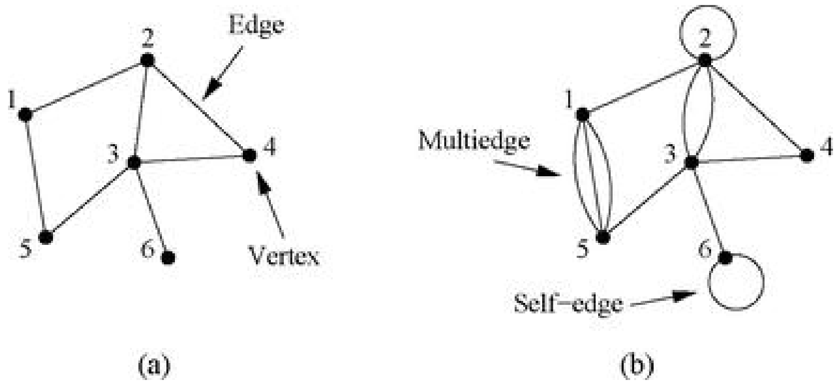


Figure 1: (a) A simple graph, i.e., one having no multiedges or self-edges. (b) A network with both multiedges and self-edges.

The set of vertices is simply a list of the nodes present in the network. For the example graph in Fig. 1 (a) this is:

$$\mathcal{V} = \{1, 2, 3, 4, 5, 6\}.$$

A compact way to write and store the set of edges is in the form of the edge-list. Figure 1 (a) is a *simple graph*, meaning there aren't any pairs of vertices connected by multiple edges, the edges are all *un-weighted* (having the same strength as each other) and not directed: i.e. all edges are two-way, e.g.:

$$\mathcal{E} = \{\{1,2\},\ \{1,5\},\ \{2,3\},\ \{2,4\},\ \{3,4\},\ \{3,5\},\ \{3,6\}\};$$

where we only need to write each edge once. For directed graphs we have to specify the connected vertices and also the *order*, and hence must write both (i, j) and (j, i) wherever a two-way connection exists.

Computationally it is also sometimes convenient to write this as a list of edges for each node:

$$\mathcal{E} = \{1 : \{2,5\},\ 2 : \{1,3,4\},\ 3 : \{2,4,5,6\},\ 4 : \{2,3\},\ 5 : \{1,3\},\ 6 : \{3\}\}.$$

This can be more compact for directed graphs and allows us to quickly access information about a particular node, including its *degree* and *network neighbours*.

3.2 The Adjacency Matrix

One of the most useful representations of the connections in a network is the *adjacency matrix*. For a network with N nodes this is an $N \times N$ matrix, having a row and a column for each node, where the element in the ith row and jth column is zero where no connection exists between node i and j, but takes the value 1 where such a connection exists, i.e.:

$$A_{ij} = \begin{cases} 1 & \text{if an edge } \{i, j\} \text{ exists} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

It is also possible to represent multiedges and self-edges using an adjacency matrix. Self-edges count as two links in this case, one from each vertex to the other. Numbers other than one can also be used to indicate edge weights in weighted graphs. For the networks in Figures 1 (a) and (b) the adjacency matrices are written:

$$A_{(a)} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad A_{(b)} = \begin{pmatrix} 0 & 1 & 0 & 0 & 3 & 0 \\ 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 2 \end{pmatrix}$$

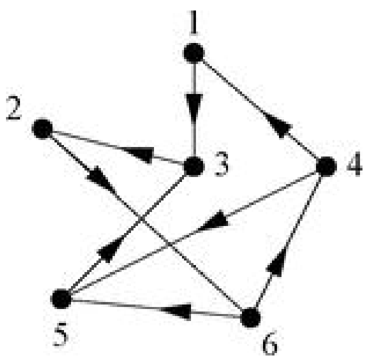
Note that for the case of a simple *un-directed* graph A is *symmetric* about the diagonal, having the property:

$$A_{ij} = a_{ji}. \tag{3}$$

For *directed networks* the adjacency matrix is defined as:

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge } \textit{from } j \textit{ to } i \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Directed networks do not generally have symmetric adjacency matrices, as shown below for the graph in Figure 2.



$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Figure 2: A directed network.

3.3 Using the Adjacency Matrix

3.3.1 The Degree Sequence

The degree k_i of a node i is the sum over all its adjacent *edges*, which can be obtained by summing over the ith row of the adjacency matrix:

$$k_i = \sum_{j=1}^{N} A_{ij} \tag{5}$$

Therefore multiplying A by the length-N unit vector $\mathbf{1}$ gives a vector $\mathbf{k}$ corresponding to node degrees. For example, for the graph in Figure 1 (a):

$$A\mathbf{1} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 2 \\ 2 \\ 1 \end{pmatrix} = \mathbf{k}.$$

This vector is a list of node degrees known as the *degree sequence*, which is often ordered by degree: $\mathcal{K} = \{4, 3, 2, 2, 2, 1\}$. For *directed* graphs this gives the *in-degree*, due to the definition of A given in (4) that there are non-zero entries on row i and column j when there is an *incoming* link from j to i. To obtain the *out-degree* we must therefore use the transpose A^T instead of A.

3.3.2 Length-n Paths

Common non-zero entries between rows i and j in the adjacency matrix occur wherever nodes i and j share a common neighbour k (such that $A_{ik} = A_{jk} = 1$ for simple graphs). Therefore the number of paths of length two between i and j can be found by summing over all such neighbours of i and j. For *undirected* graphs amounts to multiplying row i by column j, since rows and columns of A are symmetric in this case. Doing this for all nodes and multiplying the adjacency matrix of a graph by itself gives all paths of length two between any pair of vertices:

$$AA = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 2 & 1 & 0 & 0 \\ 0 & 3 & 1 & 1 & 2 & 1 \\ 2 & 1 & 4 & 1 & 0 & 0 \\ 1 & 1 & 1 & 2 & 1 & 1 \\ 0 & 2 & 0 & 1 & 2 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

Note that, for simple graphs, the diagonal of A^2 contains the degree vector $\mathbf{k}$, so that $A^2_{ii} = [A\mathbf{1}]_i$. In general the diagonal represents paths of length two which go out from node i to nearest neighbours and back to node i. The above results can be generalised for A^n, representing all paths of length n between all pairs of nodes. For *directed* networks A^n gives the number of *inward paths* of length-n, and $(A^T)^n$ the number of *outgoing paths*.

For the example for the network in Figure 2:

$$A^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad (A^T)^2 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

3.3.3 Clustering (*Transitivity*)

The above property can be used to help calculate the degree of *clustering* on graphs (also known as *"transitivity"*). This is the *"friend of a friend is also a friend"* effect, whereby in non-random graphs there is often a higher than random probability of an edge existing between i and j if both share a common network neighbour k. Transitivity is defined as the fraction of closed length-2 paths in the graph. Since each triangle can be made by adding an extra edge to three possible length-2 paths (*triads*):

$$c = \frac{3\#(\mathit{triangles})}{\#(\mathit{triads})} \tag{6}$$

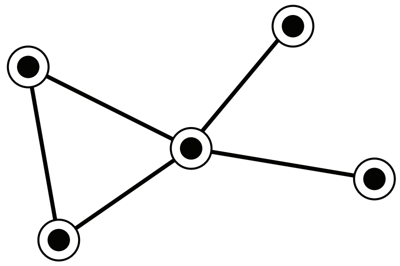


Figure 3: A network with one triangle and eight triads, giving a clustering coefficient $c = 3 \times \frac{1}{8} = \frac{3}{8}$.

The number of triangles associated with a node i is half the number of length-3 paths from node i back to itself (as each path can be taken in either direction). Since there are three nodes per triangle the sum of diagonals of A^3 (i.e. the *trace* of the matrix) is six times the total number of triangles in the network. Similarly the sum of off-diagonal elements in the A^2 matrix is twice the number of triads, hence:

$$c = 3\frac{\mathit{trace}(A^3)/6}{\sum\limits_{i \neq j} A^2_{ij}/2} = \frac{\mathit{trace}(A^3)}{\sum\limits_{i \neq j} A^2_{ij}} \tag{7}$$

For the example in Figure 3: $A = \begin{pmatrix} 0&1&1&0&0 \\ 1&0&1&0&0 \\ 1&1&0&1&1 \\ 0&0&1&0&0 \\ 0&0&1&0&0 \end{pmatrix}$, $A^2 = \begin{pmatrix} 2&1&1&1&1 \\ 1&2&1&1&1 \\ 1&1&4&0&0 \\ 1&1&0&1&1 \\ 1&1&0&1&1 \end{pmatrix}$,

$$A^3 = \begin{pmatrix} 2&3&5&1&1 \\ 3&2&5&1&1 \\ 5&5&2&4&4 \\ 1&1&4&0&0 \\ 1&1&4&0&0 \end{pmatrix}, \quad c = \frac{\mathit{trace}(A^3)}{\sum\limits_{i \neq j} A^2_{ij}} = \frac{6}{16} = \frac{3}{8}.$$

3.3.4 Cocitation and Bibliographic Matrices

Consider the two networks in Figure 4 below (where (a) is (b) with the directions reversed). In both cases $A^2 = [0]$, since there are no paths of length-2. Elements

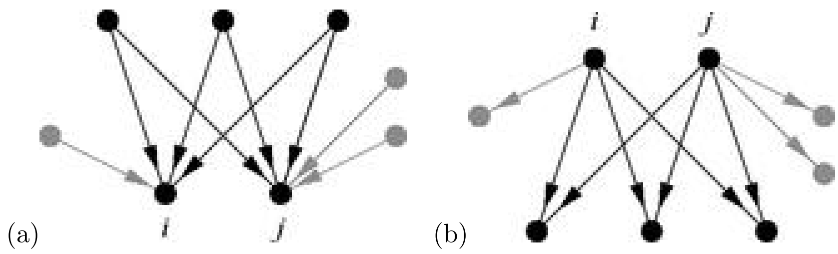


Figure 4: Directed graphs representing citation.

of row i represent incoming links to node i, so in order to obtain the total number of common *incoming* links to both i and j we must multiply the ith row of A by the transpose of the jth row. That is multiplying A by its transpose A^T results in a matrix whose off-diagonal A_{ij}th element shows the number of common links to both i and j. This is called the *cocitation* matrix $C = AA^T$, and in the case of (a):

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 4 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

In networks where vertices are papers and edges are citations from another, the value of off-diagonal elements of the *cocitation* matrix is often a good indicator that the works deal with related topics. Note that the diagonal elements are the *in-degree* of incoming edges.

A similar matrix can be defined for outgoing citations in the bibliography, known as the *bibliographic coupling matrix* $B = A^TA$, which shows the similarity of papers based on them citing many of the same sources.

In the example above case the cocitation matrix of (a) is the same as the bibliographic matrix of (b), due to the symmetry of the two networks, i.e. $B_b = A_b^TA_b = (A_a^T)^T(A_a^T) = A_aA_a^T = C_a$.

3.4 Representation of Bipartite Graphs

The adjacency matrix for a bipartite graphs with N nodes and G groups has the following form:

$$A = \begin{pmatrix} O & E \\ E^T & O \end{pmatrix}, \tag{8}$$

where E is a $G \times N$ matrix called the *incidence matrix*, defined by:

$$E_{ij} = \begin{cases} 1 & \text{if node } j \text{ is in group } i \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

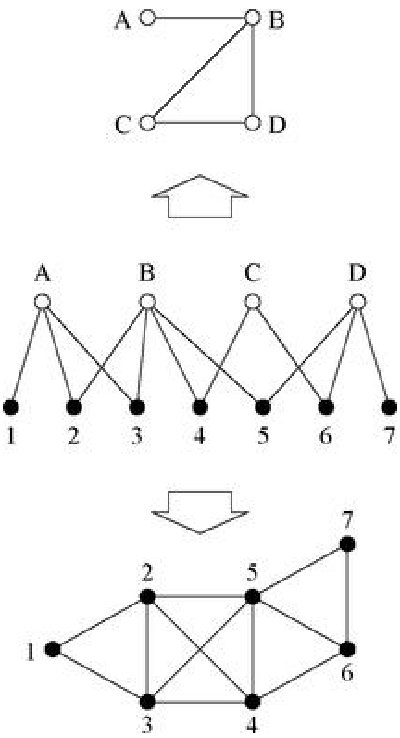


Figure 5: A bipartite graph and projections on to the two sets of vertices.

For example, for the central graph in Figure 5:

$$E = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

To obtain a one-mode projection for the links between only nodes or groups we can multiply E by its transpose E^T to obtain either $P = E^T E$ for the projection onto nodes (columns in E), or $P' = EE^T$ for a group-node projection (or *hypergraph*), where links are *overlaps* between groups.

For the example above:

$$P = E^T E = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 2 & 1 & 1 & 0 & 0 \\ 1 & 2 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 3 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 & 2 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}, \quad P' = EE^T = \begin{pmatrix} 3 & 2 & 0 & 0 \\ 2 & 4 & 1 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 1 & 2 & 3 \end{pmatrix}$$

These is similar to adjacency matrices except that the off-diagonal elements are *number* of common connections. Diagonal elements in P give the number of groups of which each vertex is a member, and in P' gives the number of members of each group.