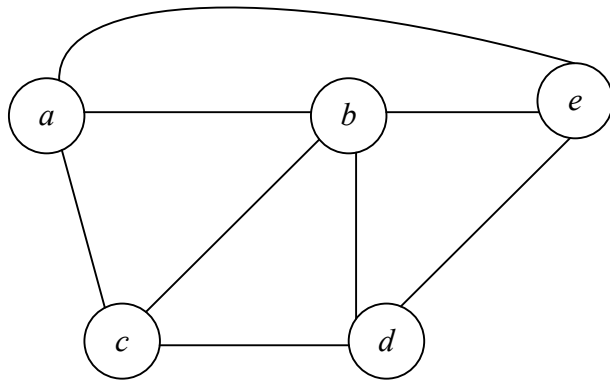


Graph Reachability and Connectedness

An **adjacency matrix** shows which vertices are *directly* connected: a 1 indicates that two vertices are connected, and a 0 indicates that they are not.

Example 1:



$$A = \begin{matrix} & \begin{matrix} a & b & c & d & e \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix} \end{matrix}$$

A **path** in a graph is a sequence of edges that connect one vertex to another. Since an adjacency matrix shows us the direct connections between vertices its entries show us the paths of length 1 between vertices in the graph.

The adjacency matrix A shows how many paths of length 1 there are from one vertex to another.

What if we wanted to know about paths of length 2? For instance in the graph above we have two paths of length 2 connecting a to b : the path $a-c-b$ and the path $a-e-b$. Could we build a matrix that contains a count of all the paths of length 2 between vertices, just like A shows us all the paths of length 1?

It turns out that this is exactly what A^2 shows us.

$$A^2 = \begin{matrix} & \begin{matrix} a & b & c & d & e \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix} & \begin{bmatrix} 3 & 2 & 1 & 3 & 1 \\ 2 & 4 & 2 & 2 & 2 \\ 1 & 2 & 3 & 1 & 3 \\ 3 & 2 & 1 & 3 & 1 \\ 1 & 2 & 3 & 1 & 3 \end{bmatrix} \end{matrix}$$

Notice that it correctly shows us the two paths of length 2 from a to b . It claims there are 3 paths of length 2 from a to d . Is it correct?

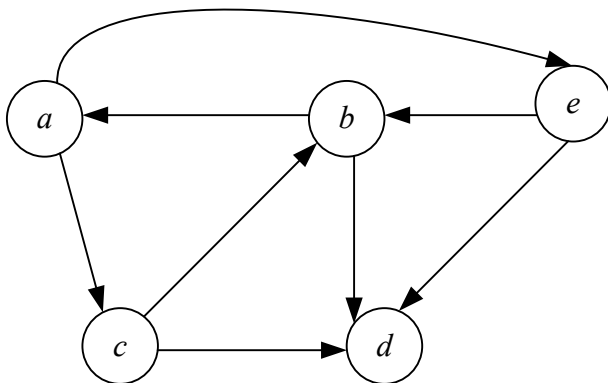
The adjacency matrix A^2 shows how many paths of length 2 there are from one vertex to another.

And the adjacency matrix A^n shows how many paths of length n there are from one vertex to another.

Notice that there are no 0s in A^2 . This means that there is at least one path of length 2 from each vertex to every other vertex in the graph so this graph is **connected**.

A connected graph is one in which every vertex can be reached from every other vertex.

Example 2: Here's a directed graph to consider.



$$A = \begin{matrix} & \begin{matrix} a & b & c & d & e \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix} & \begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix} \end{matrix}$$

Can every vertex be reached from every other vertex?

Since A contains 0s there isn't a direct connection from every vertex to every other vertex.

Let's check A^2 .

$$A^2 = \begin{matrix} & \begin{matrix} a & b & c & d & e \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix} & \begin{bmatrix} 0 & 2 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \end{matrix}$$

Since there are still 0s in it, A^2 shows us that unlike the graph above if all the vertices are connected it will take paths more than 2 edges long to do it. Let's consider some longer paths:

$$A^3 = \begin{matrix} & \begin{matrix} a & b & c & d & e \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix} & \begin{bmatrix} 2 & 0 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \end{matrix} \quad A^4 = \begin{matrix} & \begin{matrix} a & b & c & d & e \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix} & \begin{bmatrix} 0 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 2 & 0 \end{bmatrix} \end{matrix}$$

Each matrix still has 0s in it so in each matrix some vertices are not connected.

How long do we have to keep on multiplying the adjacency matrix?

Actually we can stop now. The reason is that the longest path we could possibly have in this matrix would be 4 edges long. Why? Because once we have traversed 4 edges we will have passed through every vertex it is possible to reach, and will either have reached our destination node or be unable to.

In general the longest path (without circuits) possible in a graph with n nodes is $n-1$ edges long. If you don't buy this result immediately draw graphs with 2, 3 and 4 vertices and observe what the longest path possible is in each.

The longest path (without circuits) possible in a graph with n vertices is $n-1$ edges long.

So we can stop multiplying adjacency matrices, but the answer to our question “Can every vertex be reached from every other vertex?” is not all in one place. Instead the information is sprinkled across the adjacency matrices. For example according to A^4 there are 0 paths from a to a , but according to A^3 there are 2 paths. Both are right of course: A^4 is right that there are 0 paths *of length 4* from a to a , and A^3 is right that there are 2 paths *of length 3* from a to a . So to answer the question if there are paths from all vertices to all other vertices we need to combine the information from all the adjacency matrices. We do this by adding them together.

Unlike multiplying them, adding matrices is very straightforward: we just add the elements at corresponding positions. The result is the **reachability matrix** R which shows how many ways there are to reach each vertex from every other vertex:

$$R = A + A^2 + A^3 + A^4 = \begin{matrix} & \begin{matrix} a & b & c & d & e \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix} & \begin{bmatrix} 0+0+2+0 & 0+2+0+0 & \dots & \dots & \dots \\ 1+0+2+2 & 0+0+2+0 & \dots & \dots & \dots \\ 0+1+0+0 & 1+0+0+2 & \dots & \dots & \dots \\ 0+0+0+0 & 0+0+0+0 & \dots & \dots & \dots \\ 0+1+0+0 & 1+0+0+2 & \dots & \dots & \dots \end{bmatrix} \end{matrix} = \begin{matrix} & \begin{matrix} a & b & c & d & e \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix} & \begin{bmatrix} 2 & 2 & 3 & 4 & 3 \\ 3 & 2 & 1 & 5 & 1 \\ 1 & 3 & 1 & 4 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 3 & 1 & 4 & 1 \end{bmatrix} \end{matrix}$$

For example R shows that there are 4 paths in all from a to d , and 5 paths of all lengths from b to d .

It also shows us there is no way to reach any other vertex from d so this graph is not connected. Vertex d is a Hotel California vertex there are many ways to reach it (4 from a , 5 from b , 4 from c , and 4 from e), but no way to leave once you arrive. More formally, vertices with no outbound edges are called *sinks*.

A reachability matrix shows the total number of paths of all lengths (up to $n-1$) between vertices.

The reachability matrix R is the sum of the adjacency matrices:

$$R = A + A^2 + A^3 + \dots + A^{n-1}$$