Introduction to Graph Theory and Random Walks on Graphs

1. Introduction

The intuitive notion of a **graph** is a figure consisting of points and lines adjoining these points. More precisely, we have the following definition: A graph is a set of objects called **vertices** along with a set of **unordered** pairs of vertices called **edges**. Note that each edge in a graph has no direction associated with it. If we wish to specify a direction, then we us the notion of a **directed graph** or **digraph**. The definition of a digraph is the same as that of a graph, except the edges are **ordered** pairs of edges. If (u, v) is an edge in a digraph, then we say that (u, v) is an edge from u to v. We also say that if (u, v) is an edge in a graph or digraph then u is **adjacent** to v (and v is adjacent from u in a digraph). Below are some examples of graphs and digraphs:

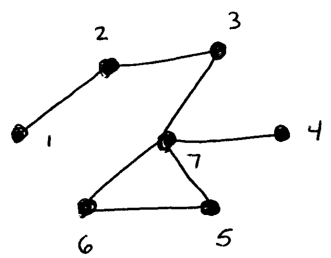

Fig. 1.

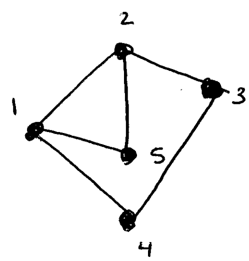

Fig 2.

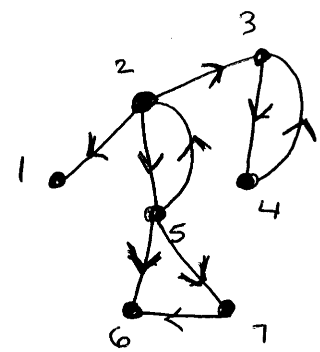

Fig. 3.

A **walk** in a graph or digraph is a sequence of vertices $v_1, v_2, \ldots, v_k$, not necessarily distinct, such that (v_i, v_{i+1}) is an edge in the graph or digraph. The **length** of a walk is number of edges in the path, equivalently it is equal to $k - 1$. A path is a walk with no repeated vertices except possibly the first and last vertex. A **cycle** is a path with $v_1 = v_k$. A graph is called **connected** if for each pair of vertices u and v, there is a path in G containing u and v. A digraph is called connected if the underlying graph is connected.

Example: In Fig. 1, v_1, v_2, v_3, v_7, v_5 is a path of length 4 from v_1 to v_5. In Fig. 2, v_1, v_2, v_3, v_4, v_1 is a cycle of length 4. In Fig. 3, v_2, v_5, v_7, v_6 is a path of length 3, but v_1, v_2, v_3 is not a path because (v_1, v_2) is not an edge.

2. Adjacency Matrices

Given a graph or digraph G with vertices $\{v_1, v_2, \ldots, v_n\}$, we define the **adjacency matrix** of G to be the matrix:

$$A = (a_{ij}) \text{ with } a_{ij} = \begin{cases} 1 \text{ if } (v_i, v_j) \text{ is an edge in } G \\ 0 \text{ otherwise} \end{cases}$$

Example: The adjacency matrices of the graphs and digraphs in the figures above are:

$$\text{Fig. 1: } A = \begin{pmatrix} 0&1&0&0&0&0&0\\ 1&0&1&0&0&0&0\\ 0&1&0&0&0&0&1\\ 0&0&0&0&0&0&1\\ 0&0&0&0&0&1&1\\ 0&0&0&0&1&0&1\\ 0&0&1&1&1&1&0 \end{pmatrix}, \text{ Fig. 2: } A = \begin{pmatrix} 0&1&0&1&1\\ 1&0&1&0&1\\ 0&1&0&1&0\\ 1&0&1&0&0\\ 1&1&0&0&0 \end{pmatrix} \text{ and Fig. 3: } A = \begin{pmatrix} 0&0&0&0&0&0&0\\ 1&0&1&0&1&0&0\\ 0&0&0&1&0&0&0\\ 0&0&1&0&0&0&0\\ 0&1&0&0&0&1&1\\ 0&0&0&0&0&0&0\\ 0&0&0&0&0&1&1 \end{pmatrix}$$

Note that if A is the adjacency matrix of a graph then $A^T = A$. This is not necessarily the case for digraphs. The main application of adjacency matrices is to determine the connectivity of a graph and the number of paths in a graph or digraph. In particular, we have the following results:

Theorem 1. *If A is the adjacency matrix of a graph or digraph G with vertices $\{v_1, \dots v_n\}$, then the i, j entry of A^k is the number of walks of length k from v_i to v_j.*

Proof: The result proceeds by induction on k. Clearly, the case when $k = 1$ is true. Now suppose that the result is true for some $k > 1$, so that the entries of A^k are as claimed. Consider any walk of length $k+1$ from v_i to v_j. Then there must be a vertex v_l on this walk such that v_l is adjacent to v_j. If we delete v_j from this walk, then the remaining walk is a walk of length k from v_i to v_l. The number of such walks is given by i, l entry of A^k by induction. Now each such v_l corresponds to a 1 for the l, j entry of A. The result follows by considering the i, j entry of $A^{k+1} = A^k A$. $\square$

Theorem 2. *If A is the adjacency matrix of a graph G with vertices $\{v_1, \dots v_n\}$, then G is connected if and only if there is an integer k such that all the entries of $A + A^2 + \cdots + A^k$ are non-zero.*

Proof: Just note that the i, j entry of $A + A^2 + \cdots + A^k$ is the number of walks of length at most k from v_i to v_j. $\square$

Example: Using the adjacency matrix for figure 2, we have

$$A^5 = \begin{pmatrix} 18&34&10&26&22\\ 34&18&26&10&22\\ 10&26&4&20&14\\ 26&10&20&4&14\\ 22&22&14&14&16 \end{pmatrix}$$

The $(3, 2)$ entry of A^5 is 26. This means that there are 26 walks from v_3 to v_2 of length 5.

3. A Note on Probability

Suppose that $\{E_1, ..., E_n\}$ is a collection of **outcomes** or **events**. The probability that an event or number of events occurs is given by:

$$\frac{\#\text{ of favorable events}}{n}$$

where n is the number of total events.

Example: What is the probability of rolling a 1 or 2 on a six-sided die?

There are six possible outcomes $\{1, 2, 3, 4, 5, 6\}$. Of these, 1 and 2 are favorable (ie. meet our criteria). Thus the probability is $2/6$.

A **random walk** on the graph or digraph G is a random sequence of vertices $v_1, v_2, \ldots, v_k$ such that v_i, v_{i+1} is an edge in G.

Now suppose that A is the adjacency matrix of G and that v_i is a vertex. From the previous section, the number of walks of length k that start at v_i and end at v_j is given by the i, j entry in A^k. In particular the total number of walks of length k which start at v_i is the i-th row sum of A^k. ie. if $A^k = (b_{i,j})$ then this number is $b_{i,1} + b_{i,2} + \cdots + b_{i,m}$. Similarly, the total number of walks of length k which end at v_j is the j-th column sum of A^k.

Example: Given the digraph in figure 4, what is the probability that a random walk of length 6 ends at v_4?

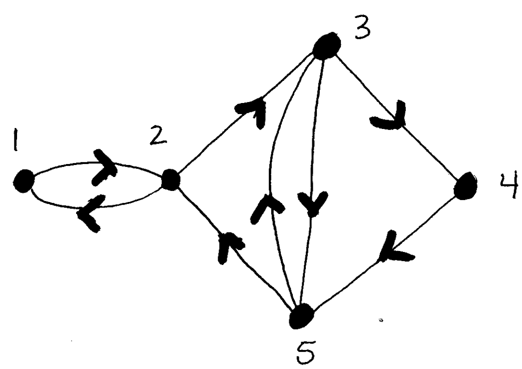

FIG. 4.

The adjacency matrix is:

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

and

$$A^6 = \begin{pmatrix} 2 & 3 & 4 & 2 & 4 \\ 3 & 6 & 7 & 4 & 6 \\ 4 & 6 & 8 & 5 & 7 \\ 2 & 4 & 4 & 4 & 5 \\ 4 & 7 & 9 & 4 & 8 \end{pmatrix}$$

The number of walks of length 6 is the sum of all entries in A^6, which is 122. Of these, the number that end at v_4 is the sum of column 4 of A^6, which is 19. Thus, the probability that a random walk of length 6 ends at v_4 is 19/122.

Example: Given the same digraph as the last example, what is the probability that a random walk of length 6 beginning at v_1, will end at v_5?

The number of walks of length 6 which begin at v_1 is the sum of row 1 of A^6, which is 15. Of the walks of length 6 which begin at v_1, there are 4 that end at v_5. Thus the probability that a random walk of length 6 beginning at v_1, will end at v_5 is 4/15.

Example: Given the same digraph as the first example, what is the probability that a random walk of length 6 starts at v_1 and ends at v_5?

Note that this is a different question than above because we are not assuming that our walk beings at v_1. The total number of walks of length 6 is 122 (from the first example), of these only 4 begin at v_1 and end at v_6 (from the second example). So this probability is 4/122. Another way to compute this is the probability that a random walk starts at v_1 is 15/122 and the probability that a random walk starts at v_1 and ends at v_5 is 4/15, so the probability that a random walk of length 6 starts at v_1 and ends at v_5 is $15/122 * 4/15 = 4/122$.