Problem 3 - Random Walks

3^{rd} December 2018 - [redacted]

Problem: Compute the access time for the complete graph K_n and for the path P_n.

First of all, let's define the access time. In a graph G with $i, j \in V(G)$, the access time $H(i, j)$ is defined as the expected number of steps to reach j from i in a random walk.

Access time for K_n

Consider the complete graph $G = K_n$ with n vertices, named $V(G) = \{0, 1, ..., n-1\}$. Note that, as every vertex is connected to all others, it does not matter from where we begin the random walk. Using that, we can easily see that the access time neither depend on the destination. Therefore we can say that

$$H(i, j) = H(0, 1)$$

so we chose arbitrarily our random walk to be from node 0 to 1.

Let's call P_t the probability to first reach node 1 in the exactly t-th step, that is, the random walk starts at node 0 and after t jumps, node 1 is reached for the first time.

$$P_1 = \frac{1}{n-1}$$

$$P_2 = \frac{n-2}{n-1} \cdot \frac{1}{n-1}$$

$$P_3 = \left(\frac{n-2}{n-1}\right)^2 \cdot \frac{1}{n-1}$$

$$\cdots$$

$$P_t = \left(\frac{n-2}{n-1}\right)^{t-1} \cdot \frac{1}{n-1}$$

Now, to compute the expected access time, we need to find the expectation. First remember the geometric series

$$\sum_{n=1}^{\infty} n \cdot x^{n-1} = \frac{1}{(1-x)^2} \tag{1}$$

$$H(i,j) = H(0,1) = \sum_{t=1}^{\infty} t \cdot P_t = \sum_{t=1}^{\infty} t \cdot \left(\frac{n-2}{n-1}\right)^{t-1} \cdot \frac{1}{n-1} = \frac{1}{n-1} \sum_{t=1}^{\infty} t \cdot \left(\frac{n-2}{n-1}\right)^{t-1} \overset{(1)}{=} \frac{1}{n-1} \cdot \frac{1}{(1 - \frac{n-2}{n-1})^2} = n-1$$

Therefore, the access time for the complete graph K_n is precisely $n-1$ and does not depend on i or j.

Access time for P_n

This graph is called the *path graph* and it consists of one single path of n vertices, named $V(G) = \{0, 1, ..., n-1\}$ as before. It is connected and 2 vertices are pendant (they have degree 1) while the other ones have degree 2. The order of the vertices in $V(G)$ are such that to construct the graph, we walk through the path $P = \{0, 1, ..., n-1\}$, in this order.

As one theorem says, in any connected graph G, if the random walk just visited vertex $i \in V(G)$, then the expected number of steps before it returns to i is

$$H(i, i) = \frac{2m}{d(i)}$$

This is usually called the *return time*. Let us focus on one of the pendant vertices of P_n, say the node $n - 1 = k$. Its degree is 1, so its return time is $2m = 2k$. This is the expected number of steps of a random walk if we start in the last vertex and want to return to it. So

$$H(k, k) = 2k$$

We can look this situation in another way to be able to get to the same result without proving the previous theorem if $G = P_n$. Let X be a random variable which indicates the number of edges used (with repetition). Then, we can write

$$X = \sum_{e \in E(G)} number\ of\ times\ e\ is\ crossed$$

$$E(X) = expected\ number\ of\ times\ \{k-1, k\}\ is\ crossed + \sum_{e \neq \{k-1,k\}} expected\ number\ of\ times\ e\ is\ crossed$$

Note that the first term is constant and its value is 2. This leads us to a recursive formula such that

$$H(k, k) = E(X) = 2 + H(k - 1, k - 1)$$

This is done until we reach the edge $\{k - k, k - k + 1\} = \{0, 1\}$. Thus, the final result is $H(k, k) = 2k$.

Now, as there is only one possibility for the first step, we can say that the access time from $k - 1$ to k is the return time for k subtracting the last step (remember that the random walk is memoryless):

$$H(k - 1, k) = H(k - 1, k - 1) + 1 = 2(k - 1) + 1 = 2k - 1$$

Now we want to find the access time from two arbitrarily vertices of P_n.

Consider $H(i, j)$ where $0 \leq i < j \leq k = n - 1$. Note that if we want to find $H(j, i)$ we can renumber the vertices the other way around and then compute $H(i, j)$ keeping the condition of $i < j$.

At this point we can omit the part of the graph **after** j, i.e. the vertices $V' = \{j+1, j+2, ..., k\}$, because starting the random walk in $i < j$ is impossible to get to a vertex $v \in V'$ without getting previously in j (which would finish our random walk). In order to reach j we first need to reach node $j - 1$. Therefore we can write the expression in the following way:

$$H(i, j) = H(i, j - 1) + H(j - 1, j)$$

As we know that $H(k - 1, k) = 2k - 1$ when k is a pendant vertex,

$$H(i, j) = H(i, j - 1) + 2j - 1$$

We can apply the same reasoning for the node $j - 1$,

$$H(i, j) = H(i, j - 2) + H(j - 2, j - 1) + H(j - 1, j) = H(i, j - 2) + 2(j - 1) - 1 + 2j - 1$$

and we can iterate like this until we reach node i. We can write this iteration in the following way:

$$H(i, j) = \sum_{l=0}^{j-i-1} 2(j - l) - 1 = (j - i) \cdot \frac{2j - 1 + 2i + 2 - 1}{2} = (j - i)(j + i) = j^2 - i^2$$

Finally, we can conclude that the access time for P_n is $j^2 - i^2$.