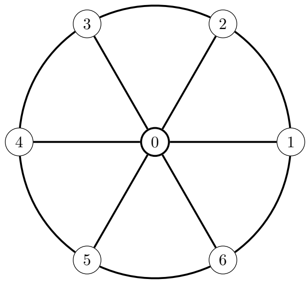
Internet Analytics (COM-308)

Homework Set 2

Exercise 1

(a) Consider the undirected graph G_1 (Figure 1). Find the stationary distribution of a random walk on this graph. Which node has the highest visiting probability?



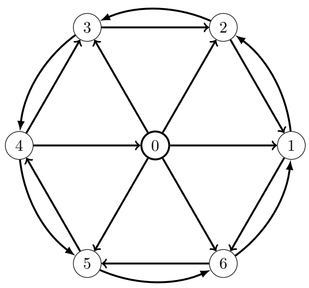
Undirected graph G_1

We have seen that a random walk on an undirected graph G has a stationary distribution $\pi(.) \propto d(.)$. Assume π_i is the visiting probability of node i in the stationary regime. Therefore, we have

$$\pi = (\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6) = (\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}),$$

where the node 0 has the highest visiting probability.

(b) Consider the directed graph G_2 (Figure 2). This graph is a directed version of graph G_1. Does the random walk also possess a stationary distribution on G_2? If yes, compute it. Which node has the highest visiting probability? If no, justify why.



Directed graph G_2

The random walk on this graph is both irreducible and aperiodic. Therefore, there is a stationary probability distribution. To find the stationary distribution $\pi = (\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6)$ we need to solve $\pi = \pi P$, expressed by the following set of equations:

$$\begin{aligned}
&\pi_0 = \pi_4/3 \\
&\pi_1 = \pi_0/5 + \pi_2/2 + \pi_6/2 \\
&\pi_2 = \pi_6 = \pi_0/5 + \pi_1/2 + \pi_3/2 \\
&\pi_3 = \pi_5 = \pi_0/5 + \pi_2/2 + \pi_4/3 \\
&\pi_4 = \pi_3/2 + \pi_5/2 = \pi_3 = \pi_5 \\
&\pi_0 + \pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 + \pi_6 = 1.
\end{aligned}$$

By solving this system of equations, we obtain the stationary distribution

$$\pi = (\pi_0, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6) = (\frac{1}{21}, \frac{19}{105}, \frac{6}{35}, \frac{3}{21}, \frac{3}{21}, \frac{3}{21}, \frac{6}{35}),$$

where the node 1 has the highest visiting probability.

(c) Verify whether the stationary distribution π for the directed graph G_2 is proportional to the node in–degrees or/and out–degrees?

No, it is not proportional. The nodes $1, 2, 3, 5$ and 6 have the same in–degrees and out–degrees, but they do not have the same visiting probabilities.

Note that in the graph G_2 the high degree node 0 has many outgoing edges, but it has just one incoming edge. In this case, once the RW leaves node 0, it takes a lot of time to visit it again.

Exercise 2

The conductance Φ of a graph measures how well different node subsets are connected to their complements. We saw in class that this has connections to the mixing time of a random walk on the graph.

We want to gain some intuition about this measure through examples. For this, assume n is even and compute the conductance of the following three graphs:

(a) the cycle C_n.

(b) the complete graph K_n.

(c) two copies of $K_{n/2}$ connected by a single edge.

As seen in class, the conductance of a graph $G(V, E)$, with n nodes and m edges, is defined as

$$\Phi = \min_{S \subset V} \frac{|\delta S|}{2m\pi(S)\pi(S')}, \tag{1}$$

where S is a non empty set of nodes, $S' = V \setminus S$ and δS the set of edges between S and S'.

(a) Note that the number of nodes n needs to be odd in order to have a stationary distribution π.
$\Phi = \dfrac{2}{2m\frac{1}{2}\frac{1}{2}} = \dfrac{4}{m} = \dfrac{4}{n}$

(b) $\Phi = \min_{S \subset V} \dfrac{|S||n-S|}{2m\frac{|S|}{n}\frac{n-|S|}{n}} = \dfrac{n}{n-1}$

(c) $\Phi = \dfrac{2}{m} = \dfrac{2}{\frac{n}{2}(\frac{n}{2}-1)+1} = \dfrac{8}{n^2-2n+4}$

As expected, the graph with the lowest conductance is the graph with two copies of $K_{n/2}$ connected by a single edge. A random walk on this graph converges slowly because it is very hard to get from one $K_{n/2}$ to the other $K_{n/2}$.