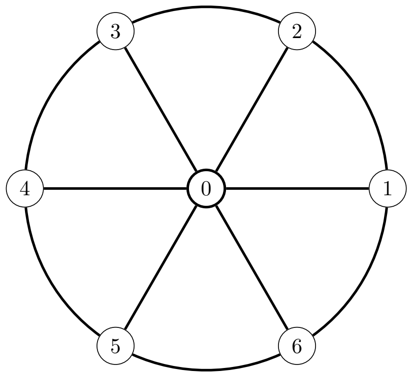
Internet Analytics (COM-308)
Problem Set 2

Problem 1

(a) Assume $G(V, E)$ is a connected, undirected, non-bipartite graph. We run a random walk $\{X_t\}$ on this graph, as seen in class. For some arbitrary edge $e \in E$, compute the probability that the random walk traverses edge e, asymptotically for $t \to \infty$.

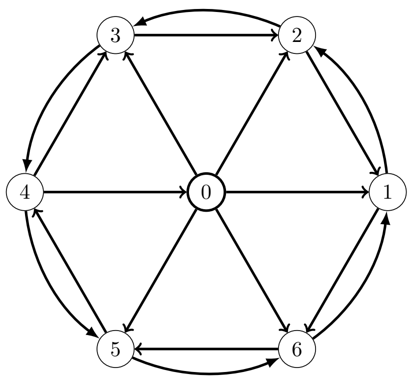
(b) To convince ourselves that the second formula in class for $\widehat{F}$ is correct, compute both the expectation under π of the numerator and of the denominator of the expression for the "self-normalized" estimator.

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



Undirected graph G_1.

(d) Consider the directed graph G_2 (Figure 2). This graph is a directed version of graph G_1. Does the random walk have a stationary distribution on this graph? If there is, compute it, and identify the node with the highest visiting probability; otherwise, explain why there is no stationary distribution.



Directed graph G_2.

(e) Check whether the stationary distribution π for the directed graph G_2 is proportional to either the node in–degrees and/or their out–degrees?

Problem 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.