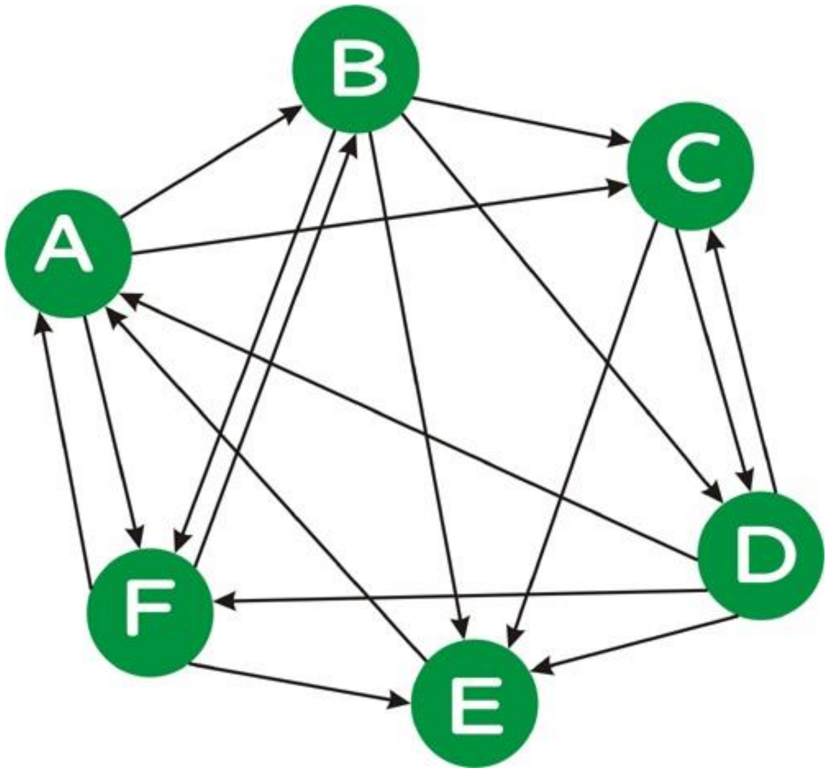
Q1) PageRank and Markov Chains

Consider the following directed graph:



a) Treat the above graph as a Markov chain, assuming a uniform distribution on the edges outgoing from each vertex. (In this problem part, you should *not* use any "teleportation.") Give the state transition matrix P of this Markov chain.

b) Compute the stationary distribution of this Markov Chain. This is a distribution π over the vertices such that

$$\pi = \pi P.$$

Note: In order to solve for π, you will need to solve six equations in six unknowns. Feel free to use a tool such as MatLab, if you like; otherwise, solve the equations by hand, eliminating one variable at a time. Also recall from class that the six equations given from $\pi = \pi P$ are not linearly independent; you will need to use five of these equations, together with the equation which specifies that the sum of the π probabilities must be 1.

c) Starting with the uniform distribution

$$\pi^{(0)} = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$$

as an initial "guess", multiply $\pi^{(0)}$ by P to obtain a new "guess" $\pi^{(1)}$. Repeat this process, obtaining $\pi^{(n)}$ from $\pi^{(n\text{-}1)}$ via

$$\pi^{(n)} = \pi^{(n\text{-}1)} \, \mathrm{P}$$

until each of the π values are accurate within two decimal places (i.e, $\pm$ 0.01) of the values you solved for above. How many iterations are required?

d) Consider the PageRank formula as described in class and at the Wikipedia PageRank page. In particular, consider the following PageRank formula described on that page

$$PR(p_i) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

(formula image courtesy Wikipedia). Let d = 0.85 be the damping factor.

Demonstrate that for the graph above, this formula is equivalent to computing the stationary distribution of a Markov chain described by transition matrix P', where each entry p'_{ij} in P' is obtained from the corresponding entry p_{ij} in P as follows:

$$p'_{ij} = (1\text{-}d)/N + d\, p_{ij}.$$

Using the matrix P' and your code from the problem part above, solve for the PageRank values of each vertex. (Start with a uniform distribution for $\pi^{(0)}$ and repeatedly multiply by P' until the π values "converge", e.g., they no longer change in the second decimal place). How do the PageRank values compare to the original stationary distribution values you computed above (and why)?

Note: There is no sink node in this graph so the third part of formula discussed in class is not written here. (Page rank of sink nodes will be 0 since there is no sink node so third part of formula will be zero)