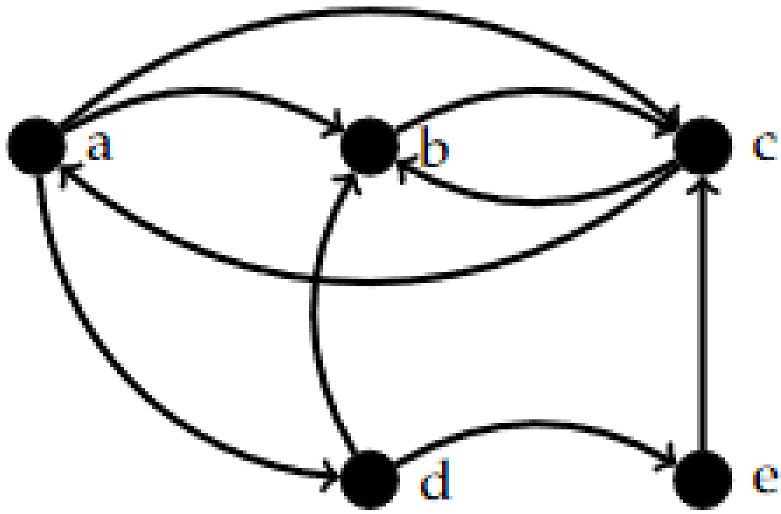
Exercise P22: Consider the graph of five web pages a, b, c, d, e, shown below, where each page links to some of the other pages as show. The goal of this problem, is to comput the Google PageRank of the pages., In factm, this proble, ios about the essential idea which created Google, only about 25 years ago.

Figure 1: P22 graph



a. Start by asking Google to document itself. Go to Google Scholar and search for 'pagerank citation ranking'. The firt lin, the one with highest PageRank, will be a technical prepring by Page, Brin Motwani, an Winograd; it has more tha 150000 citations., Download the 17 page PDF and read it. It decrinbes a simpler time, The main ideas are about a web as a directed graph, as in the example shoen above, that the limmiting probability that a ceraint random web surfer visits any partivular page is its PageRank, an finally that a asearch engine should report results in the Page Rank order. We can regard PageRank as simply an eigenvector of a certina matrix A, a transition probability matrix for a Markov chain. To build A we will start with the adjacency matrix G of the web. Were is how to dtart with a web as directred graph and construct G,A and the eigen vector of PageRank values:
Let W be a connecte, direcred grpha of n webpages, Index the pages 1 through n. Let G, be the nxn connectivity matrix W that is,

Th=e number of nonzeroes in G is the total number of hyperlinks in W. Let c_j be the columns sums of G;

$$c_j = \sum_{i=1}^{n} g_{ij}$$

Tghe value of c_j gives the outdegreee of the jth page. Let p be the fraction of the time that the random walk follows some link, so $1 - p$j is the gracion of time that an arbitratry page is chose. Define $\sigma = (1 - p)/n$.Now let A be the nxn matric whose entris are,

$$a_{ij} = \frac{pg_{ij}}{c \quad j} + \sigma$$

The matrix A is the transition probability matrix of the random-surfer Markov chain. An old theorem says the larges eigenvalus of A is wqual to on and the corresponding eigen vector, which satisfies

$$Ax = x$$

is uniq1ue up to a scaling factor, and has positive entris, Choose the scaling so that,

$$1 = \sum_{i=1}^{n} x_i$$

the entris of x are the Page Ranks of webpages in W.
For a realsitive web, G is huge and sparse because mose pairs of webpages are not connecred by a single link. Google suggests $p = .85$. Matrix A is not spare because mose entris are equal to the small constant $\sigma > 0$. The 'old' Perron=Frobenius Theorem says that if all entries ina matrix are positie then all the entries of the eigenvectro associated to the larges eigenvalue can ve chose to be posoitb.

1. b. Folling my descriptino of the process, compute G for the web shown above.

Solution:
From the passage above we know that G is the adjacency matrix for the figure 1. Building the adjacency matrix we get,

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

c. Continueing from my description, build A from G using $p = .85$ as sugestged. Finnaly, use Matlabs eig to compute the PageRanks. Which page gets the highest PageRank? explain in intuative terms.

Solution:
The following Matlab script computes A from G, and uses the eig() function to compute the PageRanks