2009 Paper 8 Question 9

Information Retrieval

The PageRank R of a webpage u is defined as:

$$R(u) = (1 - q) + q \sum_{v \in B_u} \frac{R(v)}{N_v}$$

Here, B_u is the set of pages that points to u, N_u is the number of pages that u points to, and q is the probability of staying locally on the webpage.

(a) Explain the concept of PageRank, and how it is calculated.   [4 marks]

(b) Why is it relevant for web search?   [3 marks]

(c) Give, and briefly explain, the corresponding matrix notation of the PageRank computation.   [3 marks]

(d) Give the linkage matrix A of the network given in the diagram below.   [5 marks]
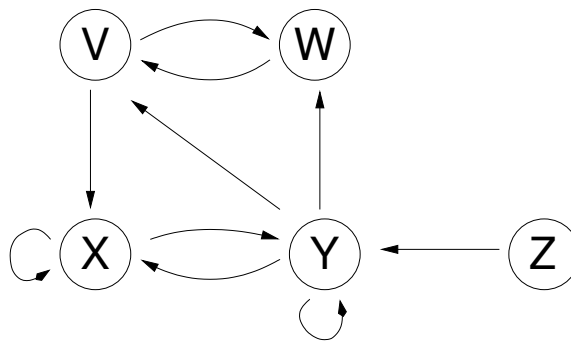
(e) Show the final matrix that will be subjected to the PageRank calculation, if $q = 0.8$ is used.   [5 marks]