Nome e Cognome:

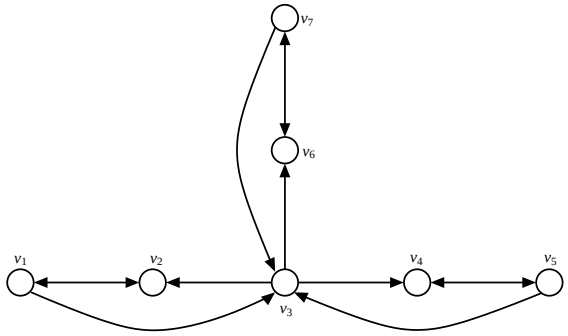
Matricola:

Ricerca dell'Informazione nel Web

Compito di esame dell'11 Gennaio 2011, *tempo a disposizione: 90 minuti*
5 punti/problema

Problema 1

1. What is the importance of the teleporting probability with respect to the convergence of pagerank?
2. We are given the following graph. Write down all the necessary equations needed to calculate the pagerank, for a general teleporting probability α.
3. Compute the pagerank of each node for teleporting probability $\alpha = 1/2$.



Problema 2

Show how we can compress the list `[2, 8, 17, 22, 30, 40, 52, 80]` using

1. Variable byte encoding.
2. γ encoding.

Problema 3

1. Give a linear-time algorithm for evaluating the query `Web AND Information AND NOT Retrieval` using standard inverted lists.
2. We have a 2-word query. For one term the postings list consists of the following 16 entries:
   `[5, 8, 10, 12, 15, 17, 19, 22, 28, 30, 32, 38, 40, 45, 60]`
   and for the other one by the two entries: `[11, 23]`

   Work out how many comparisons would be done to intersect the two postings lists with the following two strategies. Briefly justify your answers:

   (a) Using standard postings lists.
   (b) Using lists stored with skip pointers, with a skip length of $\sqrt{P}$, where P is the length of the postings list.

Problema 4

1. What are the roles of front queues and back queues in Mercator's crawler URL frontier scheme? Explain briefly how they work.
2. Usually when we start crawling we start with several seed pages. Why is this necessary?