

Assignment 7

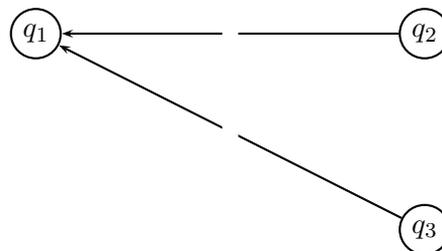
Remarks:

1. Please submit your work until **January, 24 23:59** on Ilias.
 2. If you need any **help**, send an email to `max.kisselew@ims.uni-stuttgart.de` or drop in at my office: 1.013 (Pfaffenwaldring 5b).
-

Exercise 1 (IIR 21) [1 P.]

What is ergodicity and why is it important for PageRank?

Exercise 2 (IIR 21) [3 P.]



For the web graph in the figure, compute PageRank, hub and authority scores for each of the three pages. Also give the relative ordering of the 3 nodes indicating any ties.

Assume that at each step of the PageRank random walk, we teleport to a random page with probability 0.1, with a uniform distribution over which particular page we teleport to. Normalize the hub and authority scores so that the maximum hub/authority score is 1.

Hint: Using symmetries to simplify and solving with linear equations might be easier than using iterative methods.

Exercise 3 (IIR 6) [3 P.]

One measure of the similarity of two vectors is the Euclidean distance between them: $|\vec{x} - \vec{y}| = \sqrt{\sum_{i=1}^M (x_i - y_i)^2}$. Given a query q and documents d_1, d_2, \dots , we may rank the documents d_i in order of increasing Euclidean distance from q . Show (by a mathematical proof) that if q and the d_i are all normalized to unit vectors, then the rank ordering produced by Euclidean distance is identical to that produced by cosine similarities.

Exercise 4 (IIR 8) [3 P.]

An unranked document retrieval approach is tested on a test set that consists of 300 documents. In response to a query 200 documents are retrieved of which 170 docs are relevant to the query and 30 not relevant. From the entire test corpus 190 documents are considered to be relevant for the mentioned query.

- (a) Calculate precision, recall, accuracy and (balanced) f-measure of the presented classifier.
- (b) Why do we usually have to face a tradeoff between precision and recall?

Exercise 5 (IIR 13-16) [5 P.]

As we have seen in chapter 14 there exist several types of classification algorithms.

- (a) List the classification algorithms we have seen in chapters 13, 14 and 15 and give their key properties.
- (b) Usually, we have dealt with only 2 classes in our examples. What changes with respect to the classification algorithms in (a) do we need to make if we want to classify more than 2 classes?
- (c) Explain the difference between classification and clustering.

Due date: Thursday, January 24, 2013, 23:59