# Computational Methods in Internet Economy
## TUTORIAL 2

Announcement date:     19/12/2011
Submission date:         **13/01/2012**
Contact person:          Josif Grabocka, room C142 Spl
Submission:              cmie@ismll.de

## 1. Graph Data

Download from http://snap.stanford.edu/data/web-Stanford.html the Stanford web graph, in which nodes represent pages from Stanford University (stanford.edu) and directed edges represent hyperlinks between them.

## 2. Strongly Connected Components

Implement and run the algorithm for finding the strongly connected components. Report the size of the:
- Giant strongly connected component (SCC with maximum size)
- IN and OUT
- Tendrils and Tubes (give a single size for both)
- Disconnected components (give a single size for all of them collectively)

If the size of the directed graph prohibits the execution of the algorithm in your computer, you can apply sampling to both nodes/edges. Please report whether you used sampling and the size of the sample. (Hints: use the adjacency list representation in order to save space and reduce the execution time of the DFS procedure.)

## 3. PageRank

Implement the PageRank algorithm. Perform 30 times a sample of the nodes of the Stanford web graph having in each sample 10 000 nodes. Pick the nodes in each sample at random. In each sample, keep

only the edges which connect nodes that have been included in the sample (i.e., do not keep an edge (u,v), if either u or v is not in the sample).

Apply PageRank in each sample and rank all nodes in it. Apply a second, separate ranking of the nodes by sorting nodes by decreasing in-degree (the number of incoming edges). For each sample, measure the Kendall tau rank correlation coefficient (http://en.wikipedia.org/wiki/ Kendall_tau_rank_correlation_coefficient) between the two rankings (one with PageRank and one with decreasing in-degree). Report the average Kendall tau rank correlation coefficient and the standard deviation, using all 30 samples. Explain the meaning of your result.

If the size of the directed graph prohibits the execution of the algorithm in your computer, you can use samples of smaller sizes (such as 5 000 or 1 000).

## 4. Submit

Please submit the report and the source code to cmie@ismll.de .