COMP38120 Workshop 5: PageRank (all group work)

1. What's the PageRank for site X?

Connect to http://pr.eyedomain.com/

This checks the last publicly-discoverable PageRank for a large number of sites. Google has now removed PageRank scores from public access, although they still use the scores for their own purposes.

Check the PageRank for a few well-known sites (the results here are given in whole numbers), see how they compare. Check also for some ostensibly lesser-known sites. There's a limit on the number of queries you can make to this service (between 40 and 200).

Next, gather a few queries from the group to put to Google, some general, some more specific, some informational, some transactional, etc. Avoid giving "pure" navigational queries with elements of URLs in them.

For each query, take the top 2 or 3 results and check the PageRank of each using the above tool (you will have to check the top-level site rather than an individual page, but for general queries this may be OK, let's see). Note to what extent the PageRank is apparently being used to rank the results (we cannot tell exactly, given that the set of features Google uses and how they are weighted are not revealed). Gather some observations to report back to the class.

For a small sample of sites (from any you have used/found above), check how many outgoing links they have, using: http://www.getrank.org/tools/link-extractor/

Record any interesting observations.

2. Modifying PageRank and Web graph properties

Connect to: https://learnforeverlearn.com/pagerank/

This demo allows you to set up and manipulate a graph, and to explore the effects of changing PageRank values as progress is made towards convergence. It also allows you to change the value of the damping factor and observe the effect. We haven't explored the underlying matrix manipulations as this would demand more time than available. However, you should be aware of how a graph can be represented as a matrix and that it can therefore be subject to matrix operations. You can hide the matrix display, or keep it if you wish to observe in detail what is going on as nodes are inserted/deleted and links established.

Task 1: Collaborate to modify the existing graph by introducing other links and/or nodes, observe changes. On the right, you can click on "Explore convergence" and then use left-right arrows to step through how values for each node change at each iteration. Note that as you introduce changes, the matrix displays will change, if you have them visible. Mousing over blue text or over matrix entries will bring-up a mini-explanation.

Task 2: Ensure the matrix displays are visible. Start with an empty graph and add 4 nodes, no links. Observe how the Dangling Nodes Matrix is filled, mouse over one of its entries to get the explanation. Now from node A, draw a link to node B. Observe the changes in the matrices. You will see a "1" appear in the hyperlink matrix and node A should no longer have a dangling node value: it has given all its PR mass to node B. Now add another link from node A to node C. You should see the hyperlink matrix update to reflect the division of the PR mass into two (so, 1/2 for each node linked to). Continue to add more nodes and links (max 14 nodes allowed), observing at each stage what is happening in convergence.

Connect to: netlogoweb.org

Select NetLogo Web (if you wish, at a later stage, you can download the desktop version for your own use, but do not do this at present in this workshop).

This demo also allows you to set up and manipulate a graph, and to explore the effects of changing PageRank values as progress is made towards convergence. It also allows you to change the value of the damping factor and observe the effect.

Search the models library for 'page' and select the PageRank model. This offers 3 examples. Select **example 1** or **example 2**, then click on 'setup'. You can then either inspect a step at a time (click 'step') or click 'go' to run (click 'go' again to stop). The speed can be adjusted by the slider at the top. You can alter the number of random surfers, and you can choose between a random surfer method or a diffusion method.

The programming language in use here is Logo, which some of you may be familiar with. It is rather straightforward to program in. So, you can have a go in your group (or individually, if you have your own device: this application is claimed to run OK on tablets and mobiles). You may have enough knowledge of Logo to be able to add one or more other examples. If not, just edit an existing example, and it would be best to edit **example 2**.

Task: Edit the NetLogo code to create different graphs. Edit the code for **example 2**, which you will find under the code line:
to create-network-example-2

Keep your graph sizes small, e.g., 5-6 nodes, as in the Web version the display cannot be manipulated and so larger graphs can be hard to follow. Here is one to try (note different command for creating links from a node to several others):

```
create-pages 5
  ask page 0 [ create-links-to (turtle-set page 1 page 2) ]
  ask page 1 [ create-link-to page 4 ]
  ask page 2 [ create-link-to page 4 ]
  ask page 3 [ create-links-to (turtle-set page 0 page 2 page 4) ]
  ask page 4 [ create-link-to page 1 ]
```

Recompile the code. For this graph, use:
damping factor=1
calculation method = diffusion
number of surfers = 1

Click 'setup' to load the newly compiled code. Observe that the probability mass is evenly distributed across all nodes (each has 0.2). (You can keep pressing setup to generate a different display if the one you get is not so clear.)
Click 'step' once, observe redistribution of mass. One page has gone to zero already.
Click 'step' again, observe. Another page has gone to zero.
Click 'step' again, observe. A third page has gone to zero. From now on, probability mass will be swapped between the two remaining nodes.

Try out some other graphs, but I advise you to **design** these first rather than attempt to code them directly on the fly. By varying a page's number of in/out links you will be able to see how PageRank is amassed by some pages and lost by others. See what happens if you have a dangling page (no out-links).