Computation, Information, and Intelligence (COMS/ENGRI/INFO/COGST 172), Fall 2005
10/17/05: **Lecture 22 aid — PageRank and hubs 'n' authorities**

Agenda: Further discussion of PageRank and the hubs-and-authorities algorithm.
Announcements: An organizational suggestion: if you haven't already done so, it might be worthwhile to acquire a three-ring binder and a three-hole punch. Then, if one took notes for each lecture on loose-leaf paper, one could keep one's lecture notes appropriately interleaved with each lecture's handouts in the binder.

I. Reminder: the main PageRank equation Let ϵ be some number between 0 and 1.

$$\text{score}^{(t+1)}(d_j) = \frac{\epsilon}{n} + (1-\epsilon) \sum_{d \text{ pointing to } d_j} \frac{\text{score}^{(t)}(d)}{\text{outdegree}(d)}.$$

II. Some facts about probabilities

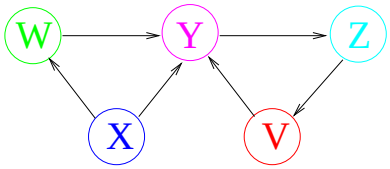
- The probability of a non-impossible event e_1 happening and then an event e_2 happening is the probability that e_1 happens *times* the probability that e_2 happens *given* that e_1 happened.
- The probability of either (but not both) of two *mutually exclusive alternative events* e_1 and e_2 happening is the probability of e_1 happening *plus* the probability of e_2 happening.
- The sum of the probabilities over all possible mutually exclusive alternatives for a given probabilistic choice must be 1.

III. The "random surfer" model Upon arriving at a document, the user either chooses to follow an existing hyperlink or to randomly jump to any document on the Web. The two cases have probability $(1-\epsilon)$ and ϵ, respectively (note that these sum to 1), and in either case, the choice among alternatives that then result is made uniformly at random.

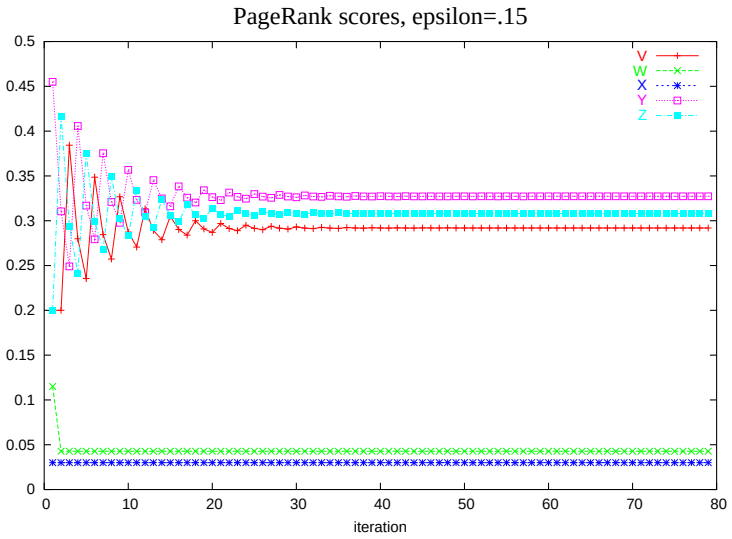
We then interpret $\text{score}^{(t)}(d_j)$ as the probability that the surfer is at document d_j at time t.

(OVER)

IV. Document set for comparison calculations We have to use a slightly different example than that from last time's handout because there's a slight but annoying technical problem in applying PageRank (even with "mini-links"/random jumps) when there are documents with zero out-degree.



V. PageRank results $\epsilon = 0.15$.



VI. Hubs-and-authorities results Since we are only interested in relative comparisons, we have not aligned the axes of the three plots.