Computation, Information, and Intelligence (ENGRI/CS/INFO/COGST 172), Spring 2007

3/7/07: **Lecture 20 aid — PageRank**

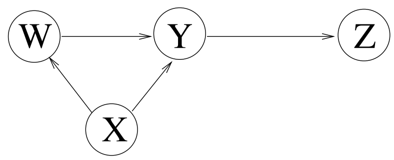
Topics: PageRank (Brin and Page, 1998[1]), "the" Google link-based ranking algorithm.

I. Definitions and conventions For a document d, we define:

$\mathrm{To}(d)$:	the set of documents that link to d.
$\mathrm{From}(d)$:	the set of documents that are linked to by d.

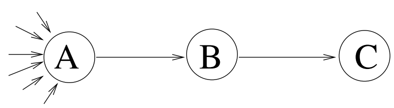
The size of these two quantities correspond exactly to the *in-degree* and *out-degree* of d, respectively.

II. Example: in-degree vs. "prestige"



Note that the in-degree of W is the same as the in-degree of Z.

III. Example: propagation of "prestige"



IV. PageRank We give an explicitly iterated version here. Let ϵ be some constant between 0 and 1, exclusive.

- For every d_j in the n-document corpus, set $\mathrm{PR}^{(0)}(d_j)$ to $1/n$.
- Let i be increasing from 1 on, until it's the case that the set of PageRank scores "converges" (the change in the set of scores between one value of i and and the next is sufficiently small): set

$$\mathrm{PR}^{(i)}(d_j) = (1-\epsilon)\left[\sum_{d\in \mathrm{To}(d_j)} \boxed{\mathrm{PR}^{(i-1)}(d)} \times \frac{1}{\mathrm{outdegree}(d)}\right] + \epsilon \times \frac{1}{n}.$$

Handy fact: assuming the link structure of the network under consideration has certain nice properties, the sum of the PageRank scores over all documents d_j in the corpus at any given iteration i is always 1.

---

[1]Beware of the typo in the PageRank equation given in the original paper.