Computation, Information, and Intelligence (ENGRI/CS/INFO/COGST 172), Spring 2007
3/12/07: **Lecture 22 aid — More on PageRank; hubs and authorities**

Topics: finish the random-surfer model; hubs and authorities (Kleinberg, 1998).

I. Reminder: The “random surfer” model At the very beginning ($i = 0$), the user picks uniformly at random some document to start looking at. Upon arriving at a document, the user either chooses to follow an existing hyperlink from it, or to randomly jump to any document on the Web. The two cases have probability $(1 - \epsilon)$ and ϵ, respectively, and in either case, the choice among alternatives that then result is made uniformly at random.

II. The hubs and authorities algorithm We consider a simplified[1] and explicitly iterated version here. For a given query, we consider *only* those documents in a *root set*: those documents retrieved by a content-based IR system in response to the query, *plus* those that link to or are linked from the documents that the content-based system produced. We write $d_1, d_2, \ldots, d_r$ for the documents in the rootset, so r is the size of that set. For each d_j (in the root set), we wish to compute its *authority score* $\mathrm{A}^{(i)}(d_j)$ and its *hub score* $\mathrm{H}^{(i)}(d_j)$.

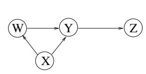
1. *Initialization:* For every d_j in the root set, set $\mathrm{H}^{(0)}(d_j)$ to $1/r$.

Let i be increasing from 1 on, until it’s the case that the set of authority and hub scores “converge”:

2. *Compute temporary authority scores*:
   For every d_j in the root set, set $\mathrm{TempA}(d_j)$ to $\sum_{d_k \text{ in } \mathrm{To}(d_j)} \mathrm{H}^{(i-1)}(d_k)$.
3. *Get the authority scores by sum-normalizing the temporary ones:*
   Calculate $\mathrm{authnorm} = \sum_{k=1}^{r} \mathrm{TempA}(d_k)$;
   then, for every root-set d_j, set $\mathrm{A}^{(i)}(d_j)$ to $\mathrm{TempA}(d_j)/\mathrm{authnorm}$.

4. *Compute temporary hub scores:*
   For every document d_j in the root set, set $\mathrm{TempH}(d_j)$ to $\sum_{d_k \text{ in } \mathrm{From}(d_j)} \mathrm{A}^{(i)}(d_k)$.
5. *Get the hub scores by sum-normalizing the temporary ones:*
   Calculate $\mathrm{hubnorm} = \sum_{k=1}^{r} \mathrm{TempH}(d_k)$;
   then, for every root-set d_j, set $\mathrm{H}^{(i)}(d_j)$ to $\mathrm{TempH}(d_j)/\mathrm{hubnorm}$.

III. Example calculations Convergence has not been reached at the end of the table (you can check).



		W		X		Y		Z	
		auth	(hub)	auth	(hub)	auth	(hub)	auth	(hub)
a.	Init	—	(1/4)	—	(1/4)	—	(1/4)	—	(1/4)
b.	TempAs	1/4	”	0	”	2/4	”	1/4	”
c.	TrueAs	$1/4$	”	0	”	$1/2$	”	$1/4$	”
d.	TempHs	”	$(1/2)$	”	$(3/4)$	”	$(1/4)$	”	(0)
e.	TrueHs	”	$(1/3)$	”	$(1/2)$	”	$(1/6)$	”	(0)
f.	TempAs	$1/2$	”	0	”	$5/6$	”	$1/6$	”
g.	TrueAs	$1/3$	”	0	”	$5/9$	”	$1/9$	”

[1]We’re using sum-normalization rather than vector-length-normalization (considering the scores to form the components of a vector) to make (your) calculations a little easier.