The University of Sydney
School of Computer Science

COMP5313/COMP4313—Large Scale Networks S1 2025
Week 5 - Hub and Authority

The goal of this tutorial is to understand what impact the ranking of Web pages based on hub and authority update rules.

Exercise 1: Computing Authority Scores

Show the values that you get if you run two rounds of computing hub and authority values on the network of Web pages in Figure 1. (That is, the values computed by the k-step hub-authority computation when we choose the number of steps k to be 2.)

Answer: Initially, we have for all $X \in \{C, D, E, F\}, hub(X) = 1$

- **Round 1:**
  - after the authority update rule: $auth(A) = 3$ and $auth(B) = 2$
  - after the hub update rule: $hub(C) = hub(D) = 3$, $hub(E) = 5$ and $hub(F) = 2$.
- **Round 2:**
  - after the authority update rule: $auth(A) = 11$ and $auth(B) = 7$
  - after the hub update rule: $hub(C) = hub(D) = 11$, $hub(E) = 18$ and $hub(F) = 7$.

Note that, in the above computation, we do not compute/initialize $hub(A), hub(B), auth(C), auth(D), auth(E), auth(F)$, as they will become zero after one round of hub-authority update.

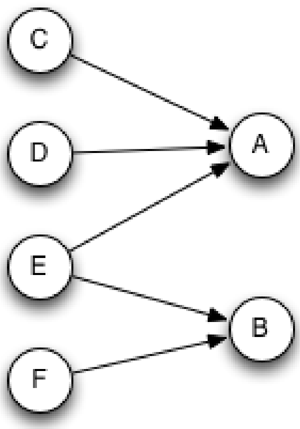

Figure 1: A simplified example of a Web graph

Now, show the values after the final normalization step, in which we divide each authority score by the sum of all authority scores, and divide each hub score by the sum of all hub scores. (We will call the scores obtained after this dividing-down step the normalized scores. It's fine to write the normalized scores as fractions rather than decimals.)

Answer:

- **Round 1:**
  - after the authority update rule: $auth(A) = 3$ and $auth(B) = 2$.
  - after the hub update rule: $hub(C) = hub(D) = 3$, $hub(E) = 5$ and $hub(F) = 2$.
- **Round 2:**
  - after the authority update rule: $auth(A) = 11$ and $auth(B) = 7$. After normalisation we have: $\widehat{auth}(A) = \frac{11}{18}$ and $\widehat{auth}(B) = \frac{7}{18}$.
  - after the hub update rule: $hub(C) = hub(D) = 11$, $hub(E) = 18$ and $hub(F) = 7$. After normalization we have $\widehat{hub}(C) = \widehat{hub}(D) = \frac{11}{47}$, $\widehat{hub}(E) = \frac{18}{47}$ and $\widehat{hub}(F) = \frac{7}{47}$.

(Duration: 15 min)

Exercise 2: Creating Web pages

Now we come to the issue of creating pages so as to achieve large authority scores, given an existing hyperlink structure.

In particular, suppose you wanted to create a new Web page X, and add it to the network in Figure 1, so that it could achieve a normalized authority score that is as large as possible. One thing you might try is to create a second page Y as well, so that Y links to X and thus confers authority on it. In doing this, it is natural to wonder whether it helps or hurts X's authority to have Y link to other nodes as well. Specifically, suppose you add X and Y to the network in Figure 1. In order to add X and Y to this network, one needs to specify what links they will have. Here are two options; in the first option, Y links only to X, while in the second option, Y links to other strong authorities in addition to X.

1. Add new nodes X and Y to Figure 1; create a single link from Y to X; create no links out of X.
2. Add new nodes X and Y to Figure 1; create links from Y to each of A, B, and X; create no links out of X.

For each of these two options, we would like to know how X fares in terms of its authority score. So, for each option, show the normalized authority values that each of A, B, and X get when you run the 2-step hub-authority computation on the resulting network (as in the former question). (That is, you should perform the normalization step where you divide each authority value down by the total.)

Answer:
In the first case, we have for both rounds: $hub(Y) = auth(X) = 1$. Thus, after normalization, we have $\widehat{hub}(Y) = \frac{1}{48}$ and $\widehat{auth}(X) = \frac{1}{19}$.

The second case, leads to the following scores, whose non-normalized values are depicted in Figure 2.

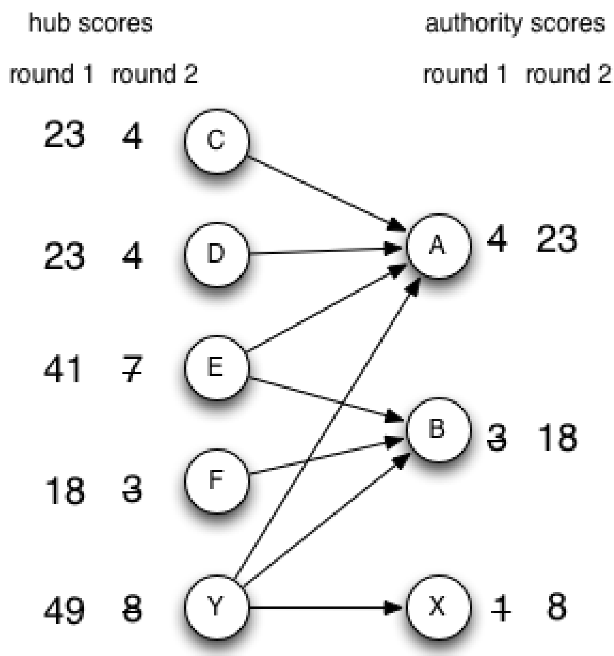

Figure 2: Addition of new pages (case 2)

- **Round 1:**
  - after the authority update rule: $auth(A) = 4$, $auth(B) = 3$ and $auth(X) = 1$.
  - after the hub update rule: $hub(C) = hub(D) = 4$, $hub(E) = 7$, $hub(F) = 3$ and $hub(Y) = 8$.
- **Round 2:**
  - after the authority update rule: $auth(A) = 23$ and $auth(B) = 18$ and $auth(X) = 8$. After normalisation we have: $\widehat{auth}(A) = \frac{23}{49}$ and $\widehat{auth}(B) = \frac{18}{49}$ and $\widehat{auth}(X) = \frac{8}{49}$.
  - after the hub update rule: $hub(C) = hub(D) = 23$, $hub(E) = 41$, $hub(F) = 18$ and $hub(Y) = 49$. After normalisation we have: $\widehat{hub}(C) = \widehat{hub}(D) = \frac{23}{154}$, $\widehat{hub}(E) = \frac{41}{154}$, $\widehat{hub}(F) = \frac{18}{154}$ and $\widehat{hub}(Y) = \frac{49}{154}$.

(Duration: 20 min)

Exercise 3: Creating Hyperlinks

Suppose instead of creating two pages, you create three pages X, Y, and Z. Describe a strategy for adding three nodes X, Y, and Z to the network in Figure 1, with choices of links out of Y and Z while X has no out links, so that when you run the 2-step hub-authority computation (as in former questions), and then rank all pages by their authority score, node X shows up in second place.

Answer: The solution is to link both Y and Z to A to pull a maximum hub scores and to X to convert this hub score into X's authority score as depicted in Figure 3.

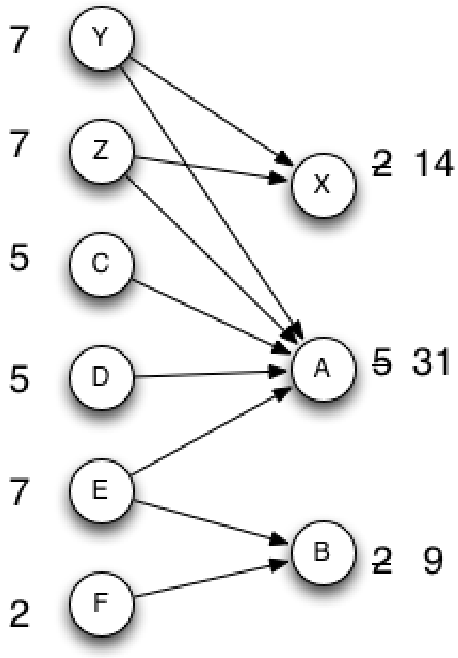

Figure 3: A strategy of adding outgoing edges to X, Y , Z so that X authority gets ranked as the second highest.

Is there another strategy of choosing X, Y and Z outgoing edges so that X shows up in first place?

Answer: No. As a result, the 2nd highest score obtained by the former strategy is the best X can achieve.

Hint: at the end of 2-step hub-authority computation, $auth(X)$ is at most $hub(Y) + hub(Z)$. If Y and Z link to neither A nor B, then $hub(Y)$ and $hub(Z)$ will be very small (see case 1 of Exercise 2). Thus, to get a large value of $hub(Y) + hub(Z)$, Y and Z must link to A and/or B. If Y and Z are linked to A (as shown in Figure 1), then $hub(Y) + hub(Z)$ contributes to $auth(A)$, moreover $hub(C)+hub(D)+hub(E)$ also contributes to $auth(A)$; thus, X will not rank higher than A. Similarly, if Y and Z are linked to B, then X will not rank higher than B.
(Duration: 15 min)