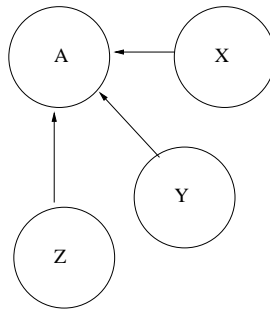


Determining Relevance: Weblinking

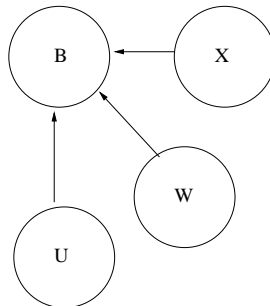
Linguistics 445/515

Autumn 2008

Example of how pages link In this example, pages X, Y, and Z all link to page A.



Now, are these links any better or worse than the links to page B?

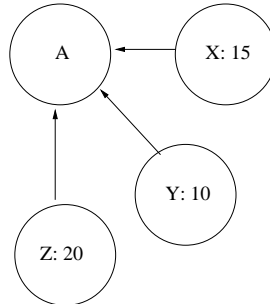


That depends on how popular, or authoritative, the links are.

Figuring popularity

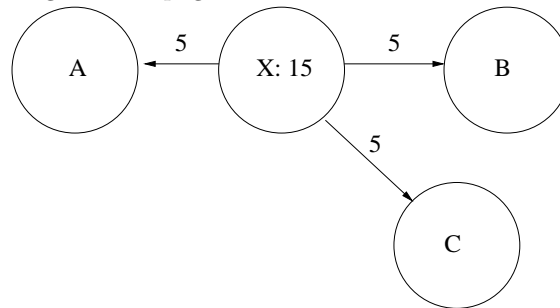
- So, in order to compare how popular website A is as compared to how popular website B is, we can add up how popular each incoming site is.
- It's like each site that links to A gets to vote for A, but they get so many votes based on how popular each one of them is.

In this case, X casts 15 votes for A, Y casts 10, and Z casts 20:



But wait, now A has 45 votes. That's too many!

Factoring out outgoing links The solution google uses is to spread out each page's votes through all the pages it links to.



So, after spreading votes out among their different webpages, let's say A's final score is: 12. What does that mean?

- On its own, the number means nothing.
- But if we compare the number with website B, which we'll say has a score of 10, we can see that A is more "authoritative".

Ranking with weblinks To sum, there are two main things to consider when calculating a ranking for a website based on its weblinks:

- Links coming in
- Links going out

The formula (for google) is as follows, where $R(A)$ means *rank of page A*; $C(X)$ means *number of pages going out of X*

$$R(A) = \frac{R(X)}{C(X)} + \frac{R(Y)}{C(Y)} + \frac{R(Z)}{C(Z)}$$

Explanation:

1. We add up all the pages *coming into* page A because to know how popular A is, we need to know how popular everyone else thinks it is.
2. We divide by the pages *going out* of X, Y, and Z because we're spreading out its weight among all the pages they link to. i.e. If we didn't divide, page A would have a huge ranking.

This tells us how "popular" a site is, which is one factor used in ranking results