L17PageRank

April 5, 2015

1 PageRank

"Welcome, Harry, to Diagon Alley"
– Rubeus Hagrid
Source: http://img3.wikia.nocookie.net/__cb20140714004320/harrypotter/images/8/84/PottermoreDiagonAlley1.jpg

Today we'll study an algorithm that is probably important in your life: Google's PageRank.

By way of history: the World Wide Web starting becoming widely used in 1994.

By 1998 the Web had become an indispensable information resource. However the problem of effectively searching the Web for relevant information was not well addressed. A number of large search engines were available, with names that are now forgotten: Alta Vista, Lycos, Excite, and others. At present, most of them are no longer in existence, because Google emerged in 1998 and came to dominate Web search almost overnight.

The problem that Google solves better than the search engines of the mid 1990's concerns the ordering in which search results are presented. By displaying the most relevant pages for each query, Google makes its search results very useful. The algorithm that gave Google this advantage is called PageRank.

What was the insight that Brin and Page used?

From [BP98]: > PageRank can be thought of as a model of user behavior. We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page. The probability that the random surfer visits a page is its PageRank.

Intro and background material

1.1 Random Walks

1.2 PageRank

1.3 Computing PageRank: the Power Method