How Google Works

1 Strategy Session

Today we are going to discuss how Google ranks websites. Although Google’s actual search algorithm is proprietary, we can develop a rudimentary version. Let’s pretend that you are Google. Here is what you do.

Step 1. Decide which websites are important.

Step 2. Read every website (so you know the content of every website).

Step 3. When someone does a search, you return the filter the websites according to the search and list the remaining in the order of their absolute ranking.

BIG PROBLEM! How do you decide which websites are important?

Idea 1. The websites that people click on are important.

Problem: You only knows websites that people click on when use your search engine. You do not have enough information!!

Idea 2. The websites that other websites link to are important.

Problem: This circumvents the problems from Idea 1 but is also problematic! If a nefarious individual sets up a million dummy sites that link to a particular site, the site will have links to it but may not be important.

Idea 3. Websites that important websites link to are important.

Problem: How can I figure out what websites are important? Isn’t this where we started?!?

2 A map of the web and its matrix representation

Look at the graph depicting the websites $A, B, C, D$ and the links between them. Google translates the “map” of the web into a matrix. At first, it simply counts how many arrow lead from one website to the others.

Please fill in the rest of the matrix:

\[
\begin{pmatrix}
0 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
\end{pmatrix}
\]

The matrix that you construct from a graph is called an adjacency matrix.
From the adjacency matrix, you normalize the matrix by averaging each column. Please fill in the rest of the matrix:

\[
\begin{bmatrix}
A & 0 & - & - & - \\
B & \frac{1}{2} & - & - & - \\
C & \frac{1}{3} & - & - & - \\
D & \frac{1}{3} & - & - & - \\
\end{bmatrix}
\]

A matrix whose columns sum to 1 is called a \textit{stochastic matrix}.

Using the stochastic matrix, Google performs a random walk on the graph.

3 \hspace{1em} \textbf{Experiment}

You are going to perform a random walk and record each occurrence of each website that you land on. You have a six-sided die that you are going to use for your random walk.

Step 1. Choose a website.

Step 2. Roll the die and use the graph to determine which website to move to. Record your new website.

Step 3. Repeat as many times as possible!

\textbf{Results.}
Which website is most important? Which least important?

If you searched for the phrase “Mathematician is the top-ranked job”, then list the websites in the order that they would be returned, according to your ranking.