Week 9: Link Analysis

# Assignment 1: PageRank

a) Compute the pageRank score for the following network. Start by writing down the adja- cency matrix. Assume a random jump probability of 0.15.

|  |  |  |  |
| --- | --- | --- | --- |
| adjacency matrix | A | B | C |
| A | 0 | 1 | 0 |
| B | 1 | 0 | 0 |
| C | 0 | 1 | 0 |

|  |  |  |  |
| --- | --- | --- | --- |
| transition matrix | A | B | C |
| A | 0.05 | 0.9 | 0.05 |
| B | 0.9 | 0.05 | 0.05 |
| C | 0.05 | 0.9 | 0.05 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **A** | **B** | **C** |
| **t = 0** | 0.33 | 0.33 | 0.33 |
| **t = 1** | 0.33 | 0.6105 | 0.0495 |
| **t = 2** | 0.5684 | 0.3721 | 0.0495 |
| **t = 3** | 0.3658 | 0.5747 | 0.0495 |
| **t = 4** | 0.538 | 0.4025 | 0.0495 |
| **t = 5** | 0.3916 | 0.5489 | 0.0495 |
| **t = 6** | 0.5161 | 0.4244 | 0.0495 |
| **t = 7** | 0.4103 | 0.5302 | 0.0495 |
| **t = 8** | 0.5002 | 0.4403 | 0.0495 |
| **t = 9** | 0.4238 | 0.5167 | 0.0495 |
| **t = 10** | 0.4887 | 0.4518 | 0.0495 |
| **t = 11** | 0.4335 | 0.507 | 0.0495 |
| **t = 12** | 0.4804 | 0.4601 | 0.0495 |
| **t = 13** | 0.4405 | 0.5 | 0.0495 |
| **t = 14** | 0.4745 | 0.466 | 0.0495 |
| **t = 15** | 0.4456 | 0.4949 | 0.0495 |
| **t = 16** | 0.4701 | 0.4704 | 0.0495 |
| **t = 17** | 0.4493 | 0.4912 | 0.0495 |
| **t = 18** | 0.467 | 0.4735 | 0.0495 |
| **t = 19** | 0.452 | 0.4885 | 0.0495 |

b) PageRank is initialized with identical scores for each web page. What would be a better method to initialize pageRank so that less iterations are necessary?

Consider amount of links to a node. One possibility would be to increase the proportion with each incoming link, e.g. 0.4, 0.4, 0.2 for A (2 shares), B (2 shares), and C (1 share) from task a).

# Assignment 2: (Programming) Link Analysis

Print the ranking of Wikipedia titles and snippets in Wikipedia together with its NDCG value for the queries:

a) “ LINKTO schnitzelmitkartoffelsalat ”

query: LINKTO schnitzelmitkartoffelsalat

TODO