

1. Summary

This paper focuses on the impact of search engines on page popularity. Search engines usually treat “relevance” as “popularity” and contribute to more severe polarization on “popularity” (“rich-get-richer”). Experiments and theories observe this trend and estimate the impact on new pages.

Following the background introduction including PageRank and popularity in Session 1 and 2, Session 3 discusses experiments about popularity evolution. In setup, 154 websites were downloaded (about 7.8 million pages in common) twice over a period of 7 months. PageRank and total number of incoming links are calculated for each page and used as measure of popularity. Similar phenomenon that *only* popular pages become more popular could be seen in the viewpoint of absolute or relative popularity increases.

In Section 4 and 5, how much bias introduced by search engines to the popularity of Web pages are studied by comparing two theoretical models: the random-surfer model and the search-dominant model. Two important concepts “popularity” and “visit popularity” are defined for derivation of popularity evolution over time. Under the proposition of visit popularity is proportional to popularity, the popularity evolution depicted by random-surfer model (Session 4) undergoes infant, expansion and maturity stages. Using Google whose history was not affected too much by popularity-based rankings and also available from Nielsen-NetRatings, the fitted curve generally follows the S-curve predicted by the random-surfer model.

On the other hand, the search-dominant model (Session 5) assumes that the search-dominant model takes the proposition that visit popularity is proportional to $(\text{popularity})^{9/4}$ with the introduction of rank (PageRank?) for formula derivation, and the derived formula shows that it takes several orders of magnitude more time for a *new* (not very popular at the beginning) page to obtain popularity one accompanied with a more sudden change. And another important reason is that the user traffic is *directed* to popular pages while pages of high quality but low initial popularity has very few chances to get user’s attention.

2. Comments

The paper is clear and reader-friendly. Experiments and theories are complementary to supports the authors’ claim. And they are designed/derived carefully with detailed. Moreover, it is really smart to come out with the idea to use Google as an approximate case of the random-surfer model since it is the first search engine using popularity-based rankings.

3. Questions

(1) How will the popularity growth be delayed if damping factor of PageRank varies (in order to fit real lives)? Is it still a change in orders of magnitude?

(2) The ranking algorithms has been changed a lot for many search engines and not all of them are pure popularity-based (e.g., Google does not use PageRank as the only ranking algorithm). What can we expect of their influence on the growth of webpage popularity?