APPLICATION OF THE H-INDEX FOR RANKING SEARCH RESULTS

JUDIT BAR-ILAN, BAR-ILAN UNIVERSITY, ISRAEL

MARK LEVENE, BIRKBECK UNIVERSITY OF LONDON
CITATION NETWORKS

• Nodes – publications/authors/journals
• Directed links – cited by
WEB GRAPH

- Nodes – web pages
- Directed links – hypertext links
RANKING OF NODES

• Count number of incoming links
  • Works OK for citations
  • Problematic for the Web
    • Solution – assign “importance” to nodes
    • Links from more “important” nodes count more
  • PageRank (Page & Brin, 1998)
    • Also explored for citation based ranking (Pinski & Narin, 1976)
    • Complex computation involving the whole web graph
BORROWING FROM BIBLIOMETRICS

• h-index for individuals (Hirsch, 2005)

• h-index for single publications (Schubert, 2009)
  • A publication has \( h_p \) index \( h_p \), if \( h_p \) of the publication that cite it received at least \( h_p \) citations, and \( h_p \) is maximal

• Can be applied to the Web
  • \( h_w \) index
$\text{hw index} = 3$
HOW TO RANK WITH THE \( hw \) INDEX?

Basic idea

- Have a list of search results relevant to the query
  - Like in the HITS algorithm (Kleinberg, 1998)
- Compute the \( hw \) index of each search result
- Rank from highest to lowest
ADVANTAGES & LIMITATIONS

Plusses

• Simple, involves local computation

Minuses

• Multiple web pages with the same $hw$ index
  • Can be partially solved using variants similar to the rational $h$-index

• More prone to manipulation than PageRank
SUMMARY

• Demonstrates how IR can benefit from bibliometrics
• Also in the other direction
  • Eigenfactor, SJR