Optimal ranking in networks with community structure

Huafeng Xie
New Media Lab,
The Graduate Center, CUNY
365 5th Ave, RM7388
New York, NY 10016
1.212.817.1964
hxie1@gc.cuny.edu

Koon-Kiu Yan
Department of Physics and Astronomy,
Stony Brook University,
Stony Brook, NY, 11794
1.631.344.3852
kyan@bnl.gov

Sergei Maslov
Department of Physics
Brookhaven National Laboratory
Upton, New York 11973, USA
1.631.344.3742
Maslov@bnl.gov

ABSTRACT

The World-Wide Web (WWW) with its enormous size (~ 10^{10} webpages) presents a challenge for efficient information retrieval and ranking. By effectively utilizing the topological information to rank the webpages, Google became the most popular tool on the web. One important feature of the WWW is that it exhibits a strong community structure in which groups of webpages (e.g. those devoted to a common topic or belonging to the same organization) are densely interconnected by hyperlinks. We study how such network architecture affects the average Google rank of individual communities. Using a mean-field approximation, we quantify how the average Google rank of community webpages depends on the degree to which it is isolated from the rest of the world in both incoming and outgoing directions, and α – the only intrinsic parameter of Google's PageRank algorithm[1]. Based on this expression we introduce a concept of a web-community being decoupled or conversely coupled to the rest of the network[2]. We proceed with empirical study of several internal web-communities within two US universities (UCLA and Long Island University). The predictions of our mean-field treatment were qualitatively verified in those real-life networks. Furthermore, the value α = 0.15 used by Google seems to be optimized for the degree of isolation of communities as they exist in the actual WWW.

Keywords

World-Wide Web, Community Structure, PageRank, Google, Optimization, Ranking, Mean-field

REFERENCES

[1] H. Xie, K. Yan, S. Maslov, arXiv: cond-mat/0409087.

[2] H. Xie, K. Yan, S. Maslov, arXiv: physics/0510107.