Informatics researchers throttle notion of search engine dominance

August 7 2006

Search engines are not biased toward popular Web sites, and may even be egalitarian in the way they direct traffic, say Indiana University School of Informatics researchers. Their study, "Topical interests and the mitigation of search engine bias," in the Aug. 7-11 issue of the *Proceedings of the National Academy of Sciences*, challenges the view of a Web-dominating "Googlearchy" in which search engines like Google push all Web traffic to established, mainstream Web sites.

"Empirical data do not support the idea of a vicious cycle amplifying the rich-get-richer dynamic of the Web," said Filippo Menczer, associate professor of informatics and computer science. "Our study demonstrates that popular sites receive on average far less traffic than predicted by the Googlearchy theory and that the playing field is more even."

Menczer was joined in the study by IU post-doctoral fellow Santo Fortunato; Alessandro Flammini, assistant professor of informatics; and Alessandro Vespignani, professor of informatics.

The IU team pooled their expertise in Web mining, networks and complex systems to collect empirical data from various search engines. In one scenario, users browsed the Web using only random links. In another, users visited only pages returned by the search engines. The researchers also studied the way in which search engines have influenced the Web's evolution.

"A simple ranking mechanism provides an elegant model to understand

the genesis of a broad class of complex systems, including social and technological networks such as the Internet and the World Wide Web," Fortunato said. "These networks possess a peculiar 'long-tail'TM structure in which a few nodes attract a great majority of connections."

The long tail structure of the Web is commonly explained through rich-get-richer models that require knowledge of the prestige of each node in the network. However, those who create and link Web pages may not know the prestige values of target pages.

In another study, "Scale-Free Network Growth by Ranking," (May 27 Physical Review Letters), the Menczer, Fortunato, and Flammini showed that for a search engine to give rise to a long tail network, it must simply sort nodes according to any prestige measure, even if the exact values are unknown. If new nodes are linked to old ones according to their ranking order, a long tail emerges.

"By sorting results, search engines give us a simple mechanism to interpret how the Web grows and how traffic is distributed among Web sites," said Menczer.

The ranking model can help understand the dynamics of other complex networks besides the Web. For example, in a social system, one may be able to tell which of two people is richer without knowing their bank account balance. Such a criterion might explain the frequency and robustness of the complex structure observed in many real networks.

Source: Indiana University

https://phys.org/news/2006-08-informatics-throttle-notion-dominance.html