

Ph.D. title: Quantitative analysis of networked environments to improve performance of information systems

Author: Vaclav Petricek

Review by: Ingemar J. Cox

Affiliation: University College London

Thesis Summary

This thesis is concerned with the quantitative analysis of networks. In particular, the thesis examines three different networks – a citation network, the network structure of various e-government websites, and social networks such as Flickr and Yahoo!360. In the case of citation networks, a comparative analysis is performed between a manually constructed citation network, DBLP, and an automatically constructed network, CiteSeer, specifically to determine whether there are any systematic biases in the creation of CiteSeer.

In the case of e-government websites, a comparative analysis is performed in an attempt to quantify the navigability of these sites. The analysis is supported by a user study.

Finally, in the case of social networks, information from the social networks is used to bootstrap recommendation systems during the period when the recommender system has little or no knowledge of a user.

Main contributions

The analysis on information systems based on network properties is very topical and interesting. Vaclav Petricek has looked at three very different, yet related problems and produced interesting results in each.

Recommender systems are a very useful tool to both assist users (e.g. reduce information overload) and web service providers (e.g. improve sales). However, when a new user joins a site, the performance of a recommender system is often poor due to the lack of knowledge of the new user. In this thesis, a solution is proposed for the case where a user is already part of a social network. This information can be used to enhance the performance of a recommender system during this initial phase. Comprehensive experimental results on two datasets and a variety of recommender algorithms support this work.

Governments throughout the world are investing large amounts of money on e-government services. The scale of this investment is such that it is important to verify that the money is well-spent. A variety of qualitative assessments are performed annually to provide rankings of e-government services. However, it would be very useful if a more quantitative analysis could be performed. Vaclav Petricek has considered this problem and studied the correspondence between a variety of network

(graph) structures associated with an e-government website and its navigability. A significant user study was then performed to verify the conclusions.

Citation networks are increasingly being used as part of assessments of the quality of scientific research. Most databases are manually constructed. However, two very popular systems, GoogleScholar and CiteSeer, are constructed automatically based on crawls of the web and analysis of the collected documents. Vacalav Petricek compared two databases, CiteSeer and a manually constructed database, DBLP, to see whether there an automatically generated citation network exhibited any bias. This analysis revealed that this is indeed the case. Specifically, papers with fewer than three authors are under-reported in CiteSeer. A model was then proposed to explain this phenomenon.

In all these contributions, the research is characterized by very strong experimentation together with theory.

New scientific results and importance

The scientific results of this work are described above. The results of this work have been published in a number of prestigious, peer-reviewed international conferences, including the 9th European Conf. on research and Advanced technology for Digital Libraries (EDCL'05), and the World Wide Web Conf. (WWW'05 and WWW'06)

Applications of results

Recommender systems are extensively used on the Web. Improvements to these systems will help both users and web-service providers.

Accurate quantitative performance analysis of e-government websites will help ensure that government funds are well-spent.

An accurate understanding of autonomously constructed citations networks will permit a more accurate assessment of research outputs.

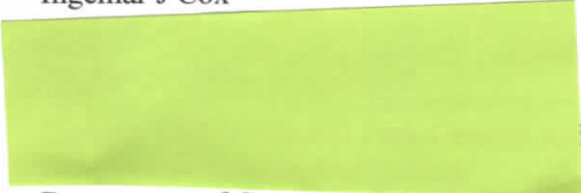
Future work

A number of avenues for future work have been identified in the thesis. These are interesting questions which I look forward to reading the results of.

Recommendation

I strongly recommend Vaclav Petricek to be awarded a Ph.D.

Ingemar J Cox



Department of Computer Science &
Department of Electronic and Electrical Engineering