CHARLES UNIVERSITY PRAGUE
FACULTY OF MATHEMATICS AND PHYSICS
Department of Software Engineering
Malostranské nám. 25, 118 00 Prague 1, Czech Republic

Fax: +420 221 914 323
Phone: +420 221 914 265

Prague, August 20, 2007

The Board of Doctoral Study
Faculty of Mathematics and Physics

Report on the PhD Thesis by Václav Petříček

This thesis is a considerable work covering a long-term study of properties of networks in real datasets. During his study the candidate took a part in international teams working in this area and participated there on obtaining the main contributions of his work.

The work focuses on analysis citation networks like CiteSeer and DBLP, e-governmental websites, as well as social networks associated to Flickr and Yahoo!360. Although there is a lot of works devoted to studying structural properties of web sites, this work reports newly about e-government web sites. Another part of the thesis deals with network-informed recommender systems. The candidate provides new algorithms based on knowledge of user' social network. Experiments accompanying all these results use large datasets, which in fact increases their plausibility.

After some necessary background introduced in Chapters 2 and 3, the candidate summarizes recent knowledge associated with a given area. The core of thesis includes Chapter 5-7. Chapter 5 describes two representative databases used for citation analysis, namely DBLP and CiteSeer. Attribute and citation network data analysis are studied in detail, including the proposal of original formal models. This enables a comparison of both services. In Chapter 6 a website navigation analysis in a special domain – e-government – is presented. Metrics proposed in this context may be used to assess websites and they also contribute to improvements of their structure. Algorithms introduced in Chapter 7 use social network knowledge to improve recommendations. Some of them are fused with collaborative filtering. Associated experiments use Flickr and Yahoo!360 data sets. A notion of social network homophily is used in this research. Improvements in recommendations have been high as it is confirmed by the experiments. The thesis concludes with a discussion of contributions.

The thesis uses partly a formal approach, however their main research method is an experiment, which is relevant in this case. The thesis also offers a lot of possibilities for other research, as it is mentioned at the end of Chapter 8.

The results presented in the thesis have been published in proceeding papers of representative international conferences, like e.g. the 14th and 15th Int. Conference on WWW. These papers are also cited internationally in a number of scientific works. Consequently, I recommend that the candidate be awarded the Doctor degree.