Sub-organizations of institutions in computer science journals at the turn of the century

Dalibor Fiala

University of West Bohemia, Department of Computer Science and Engineering

Univerzitní 8, 30614 Plzeň, CZECH REPUBLIC

Phone: +420 377 63 24 29, fax: +420 377 63 24 02, email: dalfia@kiv.zcu.cz

Preprint of: Fiala, D. (2014). Sub-organizations of institutions in computer science journals at the turn of the century. *Malaysian Journal of Library and Information Science*, 19(2), 53-68.

Suborganizations of institutions in computer science journals at the turn of the century

ABSTRACT

Most studies of the research productivity and performance of institutions concentrate on their primary organizations (such as universities) and only few of them deal with suborganizations of institutions (such as schools or departments) on the large scale. In this paper, we bridge this gap in computer science and analyze the metadata on almost 150000 journal articles indexed by Web of Science from the period 1996 – 2005 juxtaposing the first five with the last five years of this epoch. We extract authors' affiliations from the data and aggregate the articles by suborganizations of their authors. Consequently, we produce rankings of computer science departments based on various scientometric indicators and compare the corresponding rankings from the two time periods. We find that, in absolute terms, "IBM Corp.; Thomas J. Watson Research Center" always belongs to the best performers by all criteria whereas in relative terms there is no clear winner. We also examine the complete collaboration and citation networks of departments and visualize the most intense collaborations and citations in both periods. Additionally, we show the most cited departments at the turn of the century of three leading computer science institutions.

Keywords: Institutions; Suborganizations; Production; Citations; Collaboration; Computer science.

INTRODUCTION

Research performance can be measured at many levels – individuals, research groups, departments, institutions, countries, or even continents. Many studies have investigated one or more of them on varying scales. Scientometric research concerning institutions has been usually carried out on the large scale (similarly to countries) involving hundreds of thousands

Preprint of: Fiala, D. (2014). Sub-organizations of institutions in computer science journals at the turn of the century. *Malaysian Journal of Library and Information Science*, 19(2), 53-68.

or even millions of publications while research into departments has not. The reason is not only the much greater size of institutions compared to departments, but also the relative ease of work with institutions in contrast to departments. A common institution (such as a university) is often divided into suborganizations: schools, faculties, departments, divisions, centres, institutes, laboratories, etc. But the hierarchy of suborganizations is not uniform across all institutions – some of them may have a relatively flat structure while others have many suborganizations with their own subunits. Another problem is the inconsistent way authors of research papers state their affiliations. Some researchers prefer indicating their main institution only (we call it the primary organization), others usually state their main institution and their immediate organizational unit with any intermediate suborganizations omitted[1], and yet other scholars like stating their full address and affiliation with all suborganizational units included. In this case, there may be a difficulty with the order of organization and suborganizations in the full affiliation. Sometimes, the complete affiliation is written in a top-down approach (the primary organization followed by a suborganization, followed by its suborganization, etc.), other times the opposite is true. In many cases, the order of suborganizations differs in affiliations that are obviously the same. If there was a standardized way of writing author affiliations in research papers, human annotators of bibliographic databases would certainly make less mistakes in correctly identifying institutions and suborganizations and in unifying or disambiguating author affiliations.

The above issues may be the reason, why large-scale investigations into institutional suborganizations have been avoided to a great extent so far. In this study, we will try to bridge this gap in the field of computer science. We will examine collaboration and citation networks of computer science journal articles with a focus on the suborganizations in the affiliations of

[1] This is also the case of this paper's author's affiliation, where the intermediate suborganization, Faculty of Applied Sciences, is missing.

authors of these articles. In particular, we will attempt to find answers to these research questions: a) what is the composition of author affiliations in terms of their institutions and suborganizations in a set of Web of Science data? b) based on this data set, how do suborganizations perform based on various scientometric indicators and how do the rankings differ in two distinct time periods? And c) what are the most intense collaborations and citations between suborganizations and what is the nature of these collaboration and citation networks? Research question (a) will be explained under section Data and methods, while research questions (b) and (c) will be described under section Results and discussions.

RELATED WORK

As far as bibliometric studies of academic departments are concerned, they are much less numerous than studies on institutions. This and the following paragraph enumerate the most visible research on this topic. García, Rodriguez-Sánchez, and Fdez-Valdivia (2012) compared the research performance of eight economics departments at various universities. The normalized citation impact of 20 Stockholm University's natural science departments was measured by Colliander and Ahlgren (2011). Lee (2010) discusses possible adjustments of research evaluation indicators for institutional suborganizations (called "research units" therein) normalizing for their size and capacity. Ben-David (2010) ranks 11 Israeli economics departments by their citation impact and so does Lazaridis (2010) for 16 Greek university departments in chemistry, chemical engineering, materials science, and physics. Torres-Salinas, Lopez-Cózar, and Jiménez-Contreras (2009) conducted a citation analysis for 50 departments in health sciences of the University of Navarra. Zhou and Leydesdorff (2011) measured the citation impact of 27 departments of the Tsinghua University in Beijing and a citation analysis of 11 departments making up the Cuban National Scientific Research Center was conducted by Arencibia-Jorge, Barrios-Almaguer, Fernández-Hernández, and Carvajal-

Espino (2008). To name some older research: Nederhof, Meijer, Moed, and Van Raan (1993) bibliometrically examined 70 departments of an agricultural university and Nederhof and Noyons (1992) scientometrically compared two Dutch psychology departments with one US and one UK department. Oppenheim (1997) conducted a scientometric analysis of UK departments of anatomy, genetics, and archaeology and Oppenheim (1995) did so for library and information science departments in the UK. Colman, Debra, and Coulthard (1995) evaluated the research performance of 41 British political science departments and a citation analysis of seven British library and information science departments was carried out by Seng and Willet (1995).

Another group of studies concerned with the scientometric evaluation of university departments is based on webometric analyses. Some of the following papers reported a high positive correlation between webometric and standard bibliometric assessments while others did not. Arakaki and Willet (2009) carried out a webometric analysis for 21 library and information science departments in the UK and found no significant correlation between inlinks to the departments and Google Scholar citations to the publications by these departments. An earlier study by Thomas and Willet (2000) came to a similar conclusion by inspecting 17 library and information science departments in the UK and comparing their web-based measures with peer research performance evaluation rankings. On the other hand, Li, Thelwall, Wilkinson, and Musgrove (2005) found significant correlations between webometric indicators and research evaluation measures for 326 chemistry, physics, and biology departments from the UK, Canada, and Australia. The same was done by Tang and Thelwall (2004) for 89 US academic departments in Psychology, Chemistry, and History. Li, Thelwall, Musgrove, and Wilkinson (2003) analyzed the websites of 79 computer departments in the UK and found a strong association between indicators based on inlinks and research productivity and performance ratings. Fiala, Rousselot, and Ježek, K. (2007) carried out a webometric analysis of 80 French computer science departments and laboratories and Fiala, Ježek, and Rousselot (2006) measured the Web impact of 17 Czech computer science departments, but neither study dealt with correlation.

In summary, all of the above studies analyzed hundreds of departments at most and usually (excluding the webometric studies) obtained their bibliometric data for departments by measuring the research productivity and performance of individuals and aggregating the indicators by departments with which the researchers were affiliated. A typical top-down procedure was to compile a list of departments, get names of their researchers (e.g. from departmental websites), compute research evaluation metrics for those researchers, and aggregate the indicators for the departments. In contrast, our bottom-up approach potentially processes tens or hundreds of thousands of departments active in computer science, which is a methodology never described in the available literature.

DATA AND METHODS

In January 2012 we acquired XML data from Thomson Reuters on all articles published from 1996 to 2005 in journals appearing in the 2009 edition of Journals Citation Reports which were classified into the following seven computer science categories: Artificial Intelligence; Cybernetics; Hardware & Architecture; Information Systems; Interdisciplinary Applications; Software Engineering and Theory & Methods. There were 426 journals in those categories in total[2]. Changes in journal names that occurred during the period under study were not taken into account. In this way, we obtained 149347 “core” publications (strictly said, their metadata) and 191447 citations between them. The main difference between this data set and that used by Fiala (2012) is that this one includes only articles (and no other document types such as reviews or letters) and that full records of articles are available, which allows for

[2] A later update of the JCR 2009 database added two more journals which were not part of this study.

authors' affiliations to be extracted. In total, we extracted 362654 author affiliations (addresses), which makes up 2.43 affiliations per paper However, this number is based on all addresses associated with a paper including reprint addresses, which are often duplicates of one's of the researchers address. There was a very small number of addresses (less than 0.3%) that did not contain any (main) organization. (We will refer to the main organizations as institutions.) On the other hand, 77% of addresses had one or more suborganizations. 80% of them had one suborganization, 17% had two suborganizations, and 3% had three suborganizations. There were no addresses with four or more suborganizations. Although 77% of all addresses had one or more suborganizations, the distribution of such addresses in the papers published in different years was not equal. The first two years (1996 and 1997) had a low percentage – only 6.5% and, respectively, 25.3% of addresses in the articles published in those years had some suborganizations. The later years had this share over 80% each. A sample XML record with a researcher's address can be seen in Figure 1.

```xml
<research>
        <rs_address>MIT, Alfred P Sloan Sch Management, Ctr eBusiness, Dept Management
        Sci, Cambridge, MA 02142 USA</rs_address>
        <rs_organization>MIT</rs_organization>
        <rs_suborganizations count="3">
                <rs_suborganization>Alfred P Sloan Sch Management</rs_suborganization>
                <rs_suborganization>Ctr eBusiness</rs_suborganization>
                <rs_suborganization>Dept Management Sci</rs_suborganization>
        </rs_suborganizations>
        <rs_city>Cambridge</rs_city>
        <rs_state>MA</rs_state>
        <rs_country>USA</rs_country>
        <rs_zips count="1">
                <rs_zip location="AP">02142</rs_zip>
        </rs_zips>
</research>
```

Figure 1: An example of a researcher's address XML record

Preprint of: Fiala, D. (2014). Sub-organizations of institutions in computer science journals at the turn of the century. *Malaysian Journal of Library and Information Science*, 19(2), 53-68.

The order in which suborganizations appear in each address determines their “suborganization level”. Typically, an institution is followed by a level-1 suborganization, followed by a level-2 suborganization, followed by a level-3 suborganization at most. An affiliation may thus comprise up to four units, examples of which are shown in Table 1 below. If the order of suborganizations changes, the affiliation is then considered different. Therefore, our present study entirely relies on the order of suborganizations to not change across more occurrences of the same affiliation. Based on experience, however, this is not always the case in Web of Science (or WoS) and is a certain limitation of the study. An attempt to clarify the extent of this problem will be made in the next paragraph.

Table 1: Examples of affiliations with institutions and suborganizations

Institution	Level-1 suborg.	Level-2 suborg.	Level-3 suborg.
MIT			
MIT	Comp Sci Lab		
MIT	Media Lab	Software Agents Grp	
MIT	Alfred P Sloan Sch Management	Ctr Ebusiness	Dept Management Sci

Since most affiliations contained an institution and a (level-1) suborganization, we decided to aggregate affiliations by level-1 suborganizations. In this way, “MIT; Media Lab” (an institution with a level-1 suborganization) and “MIT; Media Lab; Software Agents Grp” (an institution with a level-1 suborganization and a level-2 suborganization) become both “MIT; Media Lab” as an example. Of course, if an institution has no suborganizations at all, the affiliation is not affected by the aggregation (“MIT” just remains “MIT”). For the sake of simplicity, we will refer to the affiliations after this aggregation step as “departments” and this term will also appear in all the text below. Without aggregation, which was possible only after transforming all XML data to a relational database, the citation graph of affiliations we sought to obtain would contain nodes of incomparable sizes – from large universities with thousands of researchers to tiny research groups with a few individuals. As level-2 and level-3 suborganizations appear relatively rarely in affiliations, the aggregation by level-1 suborganizations seems to be the most appropriate, although it is still not optimal due to affiliations with no suborganizations. In total, we got 56400 distinct "true" departments (institutions with no suborganizations not included). As it is impossible to manually check all those department names for real-world duplicates, ambiguities, and other inconsistencies within a reasonable time frame, we closely inspected three random samples of 1000 departments each. But even within the samples it was not feasible to take into account all name changes, variations or translations to unify/disambiguate department names and, therefore, we computed the similarities for all pairs of department names in the sample using a standard algorithm and checked the pairs above a given similarity threshold by hand. As a result, we found less than 1% of department names in each sample to be duplicates. Also, we controlled the order of institution and its suborganization in the names of departments and discovered that, on average, about 1% of names in the samples had (conversely) a suborganization in the first place followed by an institution. Consequently, given the relatively low error rates above and knowing that it is practically impossible to clean all the data, we left the data from Web of Science "as is" except for transforming all names into upper case before comparison.

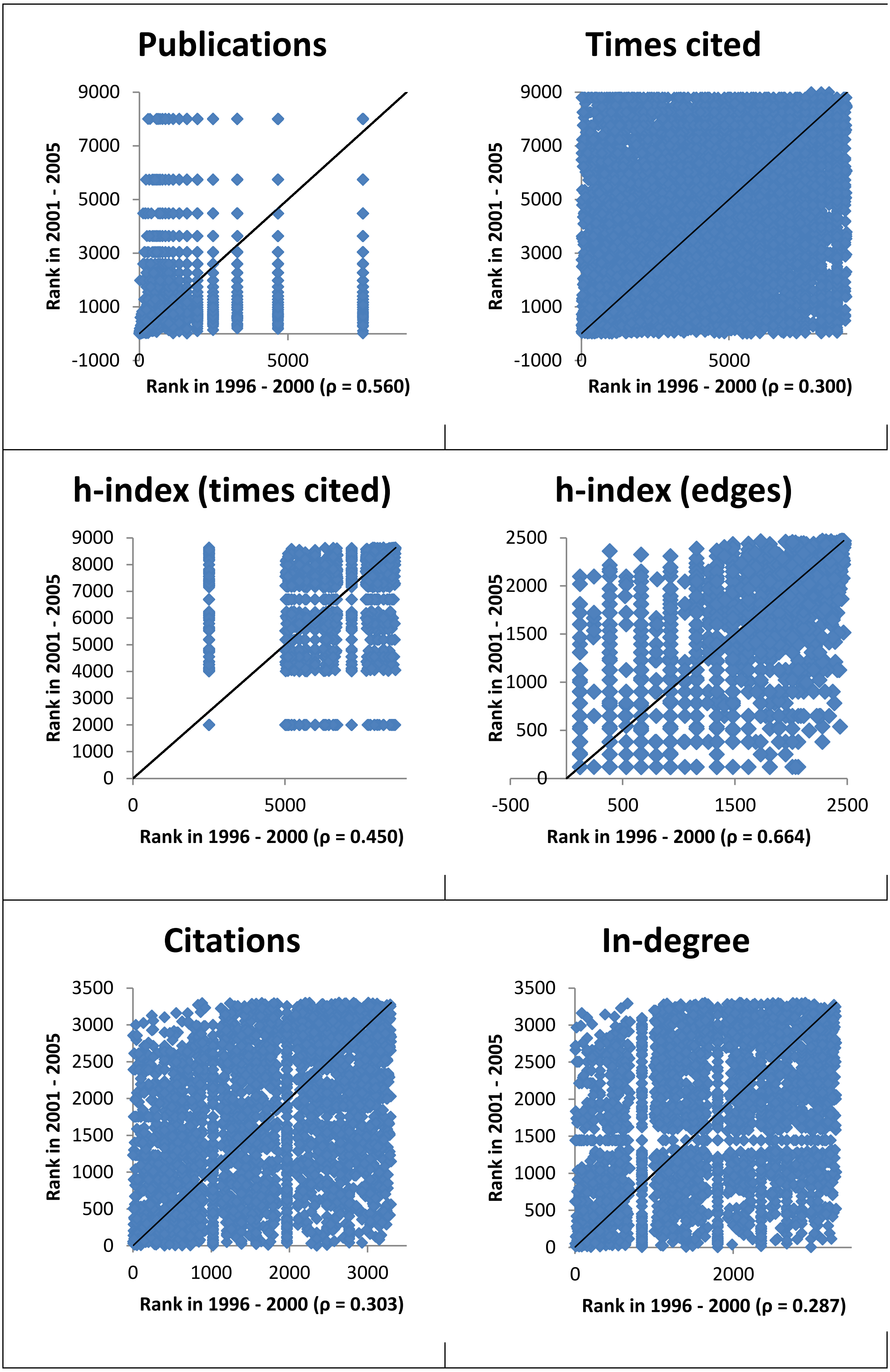
Our main goal was to split the data set into two groups representing computer science journal articles at the end of the 20th century (years 1996 – 2000) and at the beginning of the 21st century (years 2001 – 2005) and to compare the visibility (production), performance, and collaboration of institutional suborganizations in these two periods. In the two article groups, there were 67061 and 82286 publications and 34545 and 48148 citations within them, respectively. For the comparison of production and performance, we chose the following indicators: number of publications, times cited (of publications), h-index (based on times cited or on the citation network), and citations and in-degree within the citation network. As far as collaboration is concerned, the measure used is simply the number of papers jointly published by authors from different departments. The “Times Cited” indicator is adopted from Web of Science as is. For a department, it is the sum of times cited of all publications affiliated with it. In general, these citations can emanate from the whole Web of Science database including from publications outside the core. In contrast, the “Citations” indicator only includes citations from publications within the core of a given year range. Therefore, by definition, “Citations” is always less than or equal to “Times Cited” and “In-degree” (another measure based on the citation network) is always less than or equal to “Citations”. Both “Citations” and “In-degree” depend on the structure of the citation network, but “Citations” may be understood as a “weighted in-degree” whereas “In-degree” itself is unweighted – it simply counts incoming edges for a node in the graph. The directed graph (or citation network) of departments was created from the core set of publications and citations within them using the aggregation step described above and consisted of 17590 nodes (departments) and 76472 edges in 1996 – 2000 and 28967 nodes (departments) and 131383 edges in 2001 – 2005. The sum of weights (i.e. the actual number of citations between departments) was 165883 and 345607 in the respective periods. (Self-citations were discarded after aggregation.) This means that, on average, there were 2.17 and 2.63 citations per edge, 9.43 and 11.93 citations per node, and 4.35 and 4.54 edges per node, which was, respectively, an increase by 21.2%, 26.5%, and 4.4% between the first five years and the last five years of the period under study (1996 – 2005). In other words, the connectivity of departments grew moderately (by about 4%) and the citedness rate increased substantially (by over 20%) between 1996 – 2000 and 2001 – 2005. Typically, a department with many citations and a small in-degree is highly cited from a few “friendly” departments while a department with relatively few citations but a relatively high in-degree may be regarded as having an impact on many others that, however, does not go into depth. In any case, it seems reasonable to study both citations and in-degree as both indicators take into account different aspects of scientific performance.

As for the h-index, which combines both research productivity and impact in a single number, Hirsch (2005) defined it in the following way: if we have a set of publications ordered by the number of times they are cited in descending order, the index h is the largest number h such that there are h publications having at least h citations each. Although it was originally conceived for individual researchers, it can be applied to any entities associated with research papers – also to academic departments. Thus, for instance, a department with an h-index of 20 has published 20 papers at least (productivity) and has received no less than 400 citations (impact). Even if the basic h-index may be amended (e.g. with corrections for multiple coauthorship of papers), it is still a useful scientometric indicator and we employ it as a complement to publication and citation counts in this study. Actually, we use two h-index variants: the first one (h-index – times cited) is based on the times cited count (adopted from WoS) of the core publications aggregated by departments and the second one (h-index – edges) is grounded in the same citation network of departments discussed earlier in relation to citations and in-degree. Again, the indicators produced by the second variant are always less than or equal to the corresponding indicators from the first variant. The first h-index (times cited) represents research performance as a whole whereas the second h-index (edges) is a measure of research performance in the context of the time period and the scientific community under investigation, which we both believe to be important indicators of scientific performance.

RESULTS AND DISCUSSION

Suborganizations Performances and Rankings: 1996-2000 and 2001-2005

It would be possible to present many charts and tables in the results section with various rankings of departments achieved by different criteria, but we will show the main findings of this study as compressed as possible. In Figure 2, we can see the scatter plots of ranks assigned to individual departments by six scientometric indicators we discussed in the previous section in the periods 1996 – 2000 and 2001 – 2005. The top left chart is a plot of department ranks by the number of publications. The rankings by publications are moderately correlated with the Spearman’s correlation coefficient of 0.560 and include many tied ranks, which we can recognize from the many rows and columns of rank marks. Of course, departments having a rank mark farther from the diagonal have a more differing number of publications in the two periods than departments whose rank mark is closer to the diagonal. (A lower rank means a better position.) The department with the most publications was “IBM Corp; Thomas J Watson Res Ctr” in both periods. The best performer from one period to another was “Nanyang Technol Univ; Sch Comp Engn” represented by the bottom rightmost mark and the lowest performer was “NEC Corp Ltd; C&C Media Res Labs” represented by the top leftmost rank mark. As for the rankings by times cited, they are quite uncorrelated as can be seen from the top right chart of Figure 2. These and any further rankings depicted in Figure 2, however, do not use absolute indicators (times cited, h-index, citations, and in-degree) but relative measures that take into account the number of publications of each department appearing in the specific period. The top performer between 1996 – 2000 and 2001 – 2005 is “Univ Montreal; Cirano” and the lowest performer is “Univ Vigo; Etsi Telecomunicac”. As the total sum of times cited of publications in both periods is almost the same (roughly 1.1 million), the low correlation can only be explained by a real change in the scientific performance of the departments under study in the context of the whole Web of Science database.



Preprint of: Fiala, D. (2014). Sub-organizations of institutions in computer science journals at the turn of the century. *Malaysian Journal of Library and Information Science*, 19(2), 53-68.

Figure 2: Scatter plots of departments' ranks by various indicators in two periods

On the other hand, both h-index scatter plots (based on times cited or citation network in the middle left and middle right charts) show a moderate correlation with the h-index (edges) being more positively correlated (ρ = 0.664) than that grounded in times cited. As opposed to citation counts that consider all publications of a department, the h-index takes into account the top-performing (in terms of citations) publications only as explained earlier. Therefore, it seems that this measure is more robust to changes of the time period under investigation and that the indicator depending on the citation network of core publications is even more resilient than the h-index calculated in the context of the whole Web of Science database (ρ = 0.450). Although both basic h-index rankings contain many tied ranks, this is most markedly demonstrated by the h-index (times cited) rankings, where about 5000 departments share the same position in 1996 – 2000 and about 4000 departments tie the same rank in 2001 – 2005. These are all the departments for which the ratio of h-index (times cited) versus publication count is 1. The top performer in both h-index rankings is "Univ Houston; Dept Geosci" whereas the least performers are "Tech Univ; Moscow Inst Aviat" by the h-index (times cited) and "Pohang Univ Sci & Technol; Dept Chem Engn" by the h-index (edges). Rankings of departments based on citations and in-degree of their publications in the citation network of departments are, again, uncorrelated as we can see on the bottom charts of Figure 2. The inter-period winner is "Rutgers State Univ; Sch Commun Informat & Lib Studies" as far as citations are concerned and "Univ Osaka Prefecture; Dept Ind Engn" as for the in-degree. On the contrary, "Univ Oxford; Dept Mat" and "Okayama Univ Sci; Dept Informat & Comp Engn" are the top losing departments by citations and in-degree, respectively. Even though the number of citations (and edges) in the citation graph approximately doubled between the two time periods, the low correlations (being significant at the 0.01 level two-tailed as all other correlation coefficients in Figure 2) may be easily explained by the fact that the citation networks changed, indicating a change of the scientific performance of the departments under study.

Collaborations and Citations Between Suborganizations

In addition to the various scientometric performance indicators of departments discussed above, we were also interested whether and how collaboration patterns of departments changed in the decade at the turn of the century. In Figure 3, the top 30 most intense collaborations between "computer science" departments in 1996 – 2000 and 2001 – 2005 are visualized using the Cytoscape software by Shannon et al. (2003). Both graphs are undirected with edge weights corresponding to the number of collaborations (i.e. the number of papers published jointly by researchers of both departments) between two departments and node sizes reflecting the number of publications of each department in the specific period. The most intense collaboration in the first time period was between "Univ Maryland; Dept Comp Sci" and "Univ Maryland; Inst Adv Comp Studies", together forming a connected component with other University of Maryland departments ("Dept Elect Engn" collaborating with "Syst Res Inst", "Syst Res Inst" with "Dept Comp Sci", and "Dept Comp Sci" with "Umiacs") and one external department ("Bar Ilan Univ; Dept Math & Comp Sci" collaborating with "Inst Adv Comp Studies"). The second and third most intense collaborations in the same period occurred between "Univ Illinois; Coordinated Sci Lab" and "Univ Illinois; Dept Elect & Comp Engn" and between "MIT; Comp Sci Lab" and "MIT; Dept Math", respectively. While the first collaboration is a self-contained component, the latter is part of a three-member component together wirh "Carnegie Mellon Univ; Sch Comp Sci" collaborating with "MIT; Comp Sci Lab". Other notable three-node components include Yale University intra-institutional collaborations ("Dept Diagnost Radiol" with "Dept Elect Engn" and "Dept Comp Sci" with "Dept Elect Engn") and collaborations of Taiwanese institutions' departments ("Natl Chiao Tung Univ; Dept Comp

& Informat Sci” with “Natl Taiwan Univ Sci & Technol; Dept Elect Engn”) and (“Acad Sinica; Inst Informat Sci” with “Natl Chiao Tung Univ; Dept Comp & Informat Sci”). All in all, about 17% of all 27567 collaborations between departments in 1996 – 2000

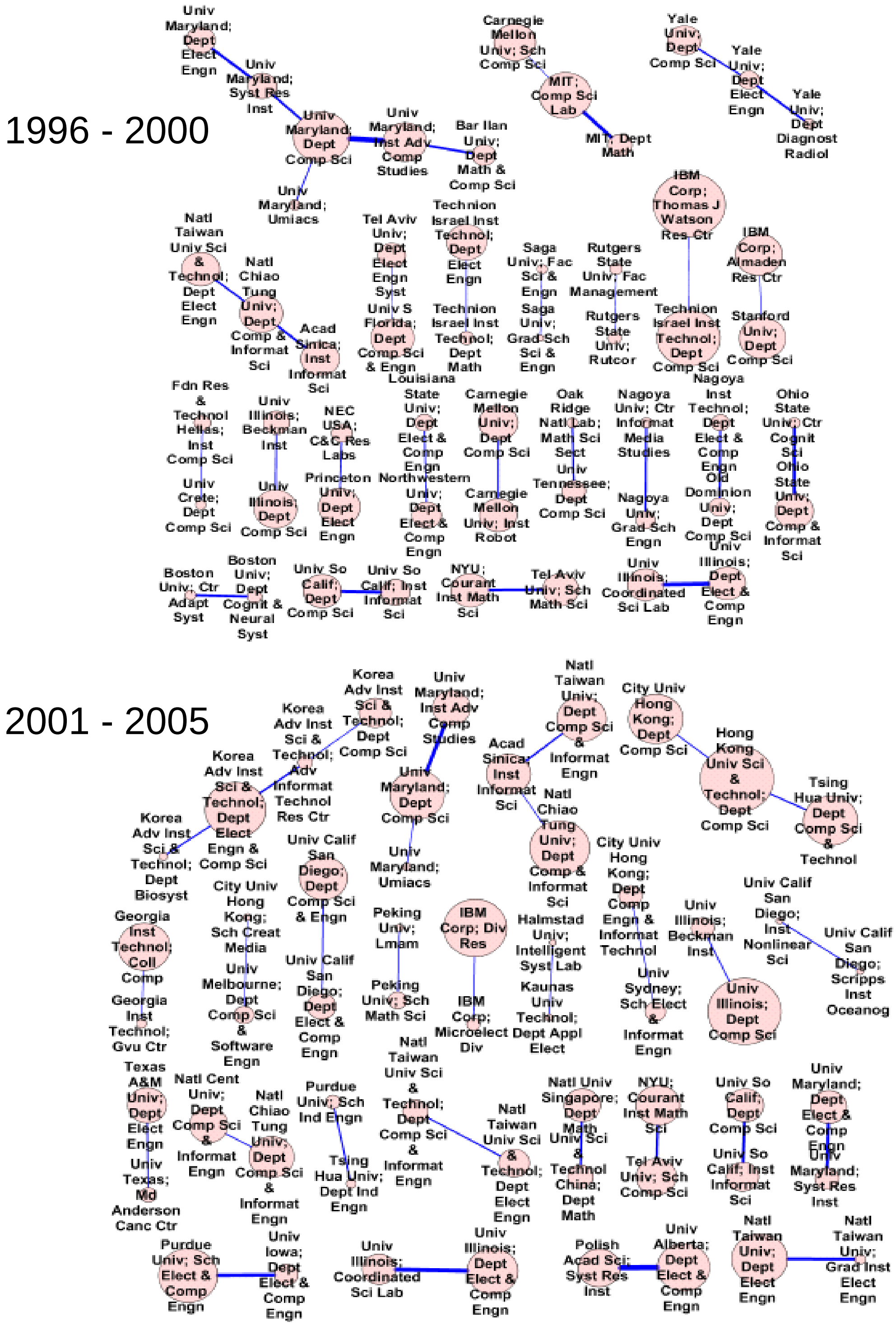


Figure 3: Top 30 most intense collaborations in two periods

Preprint of: Fiala, D. (2014). Sub-organizations of institutions in computer science journals at the turn of the century. *Malaysian Journal of Library and Information Science*, 19(2), 53-68.

were intra-institutional. As for the years 2001 – 2005, the three most frequent collaborations are “Polish Acad Sci; Syst Res Inst” with “Univ Alberta; Dept Elect & Comp Engn”, “Univ Illinois; Coordinated Sci Lab” with “Univ Illinois; Dept Elect & Comp Engn”, and “Univ Maryland; Dept Comp Sci” with “Univ Maryland; Inst Adv Comp Studies”. While the last two collaborations are also frequent in the preceding five years, the first collaboration (between a Polish and a Canadian department) is quite surprising because none of the departments is particularly well known in the context of computer science. By more closely inspecting the joint publications of the two departments, however, we found that they were coauthored by “Pedrycz, W”, a researcher affiliated with both departments. In addition to the Taiwanese computer science departments, which collaborate similarly to the preceding time period, there emerged new components of Asian departments collaborating with each other. In particular, there is a group of four “Korea Adv Inst Sci & Technol” departments with collaborations between “Adv Informat Technol Res Ctr” and “Dept Elect Engn & Comp Sci”, “Dept Biosyst” and “Dept Elect Engn & Comp Sci”, and “Adv Informat Technol Res Ctr” and “Dept Comp Sci” and a group of three Chinese departments with collaborations between “Hong Kong Univ Sci & Technol; Dept Comp Sci” and “Tsing Hua Univ; Dept Comp Sci & Technol” and between “City Univ Hong Kong; Dept Comp Sci” and “Hong Kong Univ Sci & Technol; Dept Comp Sci”. The proportion of intra-institutional collaborations remained quite stable in comparison to the previous period – a little less than 17% of all 60848 collaborations between departments in 2001 – 2005 were intra-institutional again.

As far as citations between individual departments in both epochs are concerned, Figure 4 visualizes the top 30 most intense citations in 1996 – 2000 and in 2001 – 2005. In the directed graphs, node sizes correspond to the sum of times cited of a department’s publications (computed from the whole of the Web of Science database) and edge weights reflect the number of citations from one department to another. The three most intense citations between departments in 1996 – 2000 are from “Lucent; Opt Networking Grp” to “Lucent Technol; Opt Networking Grp”, from “Lucent; Opt Networking Grp” to “AT&T Bell Labs; Lightwave Syst Res Dept”, and from “Riken; Brain Sci Inst” to “Riken; Frontier Res Program”. Since the first citation is very likely a self-citation, which should normally be discarded, the third “true” most intense citation is from “IBM Corp; Div Res” to “IBM Japan Ltd;

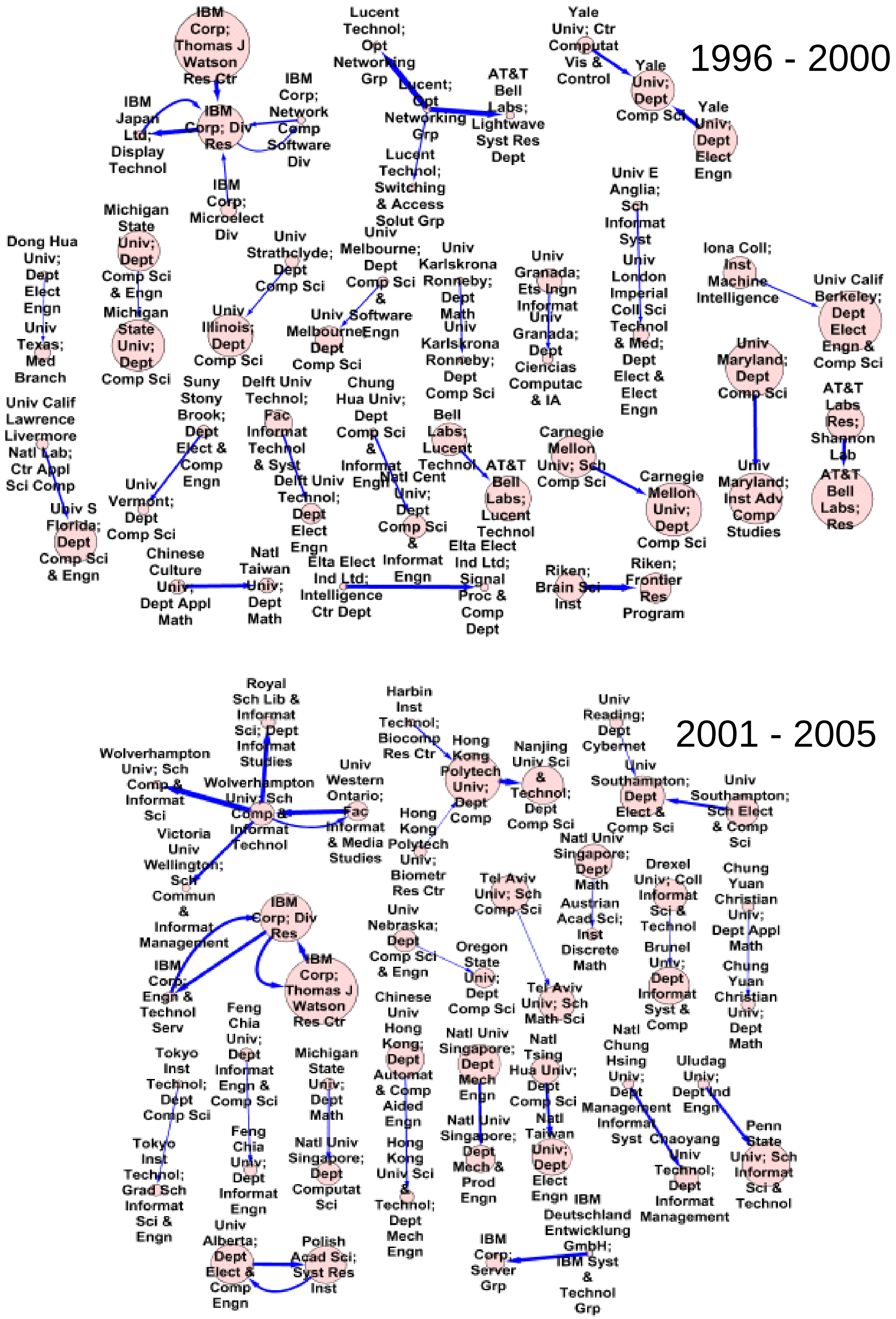


Figure 4: Top 30 most intense citations in two periods

Preprint of: Fiala, D. (2014). Sub-organizations of institutions in computer science journals at the turn of the century. *Malaysian Journal of Library and Information Science*, 19(2), 53-68.

Display Technol”. There is an apparent citation community consisting of four “IBM Corp” departments (“Div Res”, “Thomas J Watson Res Ctr”, “Network Comp Software Div”, and “Microelect Div”) and one “IBM Japan Ltd” department (“Display Technol”). Within this community, “IBM Corp; Div Res” is central – it is cited by all other departments and it cites two other departments whereas “IBM Corp; Thomas J Watson Res Ctr” is the department that is most cited from outside of the IBM component as indicated by the great node size. Another striking feature of the top chart in Figure 4 is the community of three “Yale Univ” departments, in which “Dept Comp Sci” and “Dept Elect Engn” are comparable as for times cited, but only “Dept Comp Sci” attracts citations from the other two. A remarkable change in the citation graph of “computer science” departments occurred in 2001 – 2005 when there emerged a community in which “Wolverhampton Univ; Sch Comp & Informat Technol” often cited “Wolverhampton Univ; Sch Comp & Informat Sci” (possibly a self-citation), “Royal Sch Lib & Informat Sci; Dept Informat Studies”, “Victoria Univ Wellington; Sch Commun & Informat Management”, and “Univ Western Ontario; Fac Informat & Media Studies” in decreasing order of citations and was itself heavily cited by “Univ Western Ontario; Fac Informat & Media Studies”. There also appeared a community of four Chinese departments with “Hong Kong Polytech Univ; Dept Comp” cited by “Harbin Inst Technol; Biocomp Res Ctr” and by “Hong Kong Polytech Univ; Biometr Res Ctr” and citing “Nanjing Univ Sci & Technol; Dept Comp Sci”. The IBM community became smaller with “IBM Corp; Engn & Technol Serv” replacing “IBM Japan Ltd; Display Technol”, “IBM Corp; Network Comp Software Div”, and “IBM Corp; Microelect Div”. However, “IBM Corp; Div Res” remained the central department and “IBM Corp; Thomas J Watson Res Ctr” was still the department most cited from outside. However, it was not the most cited department (in terms of the absolute sum of times cited) as such – it ranked second after “Carnegie Mellon Univ; Inst Robot” in 1996 – 2000 and also second after “Univ Calif Berkeley; Dept Elect Engn & Comp Sci” in 2001 - 2005. Even in this citation graph here we can see a Polish and a Canadian depart ment frequently citing each other, which is a phenomenon commented on earlier. Regarding the proportion of intra-institutional citations between departments, it decreased slightly from 11% in 1996 – 2000 to 8% in 2001 – 2005 out of 56201 and 252287 total citations between “true” departments, respectively.

Although a detailed analysis of the topology of the collaboration and citation networks of departments was not the aim of this study, we computed the main descriptive parameters such as the clustering coefficient, number of connected components, network diameter and radius, characteristic path length, etc. for all networks, but we did not find any striking differences between the corresponding networks in two different epochs. Thus, to conclude the section on the main results achieved, we present in Table 2 the rankings by times cited of the best departments of the three leading universities in computer science according to the Academic Ranking of World Universities 2012 (www.arwu.org). These universities are “Stanford Univ”, “MIT”, and “Univ Calif Berkeley”. We can see that while there is one strong department at Stanford (“Dept Comp Sci”) and at Berkeley (“Dept Elect Engn & Comp Sci”) that maintained and even strengthened their positions at the turn of the century (most notably “Dept Comp Sci” at Stanford), there are several comparably well performing departments at MIT that earn research impact to their institution in computer science journals with “Dept Elect Engn & Comp Sci” replacing “Media Lab” as the top department at the beginning of the 21st century. We can expect that also other leading universities in computer science belong to one of the two groups – either having one top department (like Stanford or Berkeley) or a couple of very good departments (like MIT).

CONCLUSIONS AND FUTURE WORK

Research papers are written by authors who are affiliated with institutions and, very often, with suborganizations within these institutions such as schools, laboratories, divisions, groups, or departments. Most studies of the research performance and collaboration of institutions concentrate on their primary organizations, i.e. on the institutions as such, and only few of

Table 2: Top 20 departments of three leading universities in computer science

	1996 – 2000		2001 - 2005	
		Stanford Univ		
1	Dept Comp Sci	3839	Dept Comp Sci	5139
2	Dept Elect Engn	1394	Dept Elect Engn	1605
3	Ctr Turbulence Res	908	Dept Stat	1410
4	Stanford Linear Accelerator Ctr	640	Dept Mech Engn	584
5	Dept Psychol	582	Comp Syst Lab	482
6	Comp Syst Lab	489	Sch Med	381
7	Dept Stat	456	Stanford Med Informat	381
8	Sch Med	404	Dept Math	314
9	Dept Petr Engn	260	Informat Syst Lab	279
10	Psychol & Neurosci Dept	192	Sci Comp & Computat Math Program	272
11	Dept Engn Econ Syst & Operat Res	126	Grad Sch Business	255
12	Dept Math	125	Sci Comp Computat Math Program	247
13	Dept Radiol	118	Dept Management Sci & Engn	246
14	Gates Comp Sci	114	Ctr Turbulence Res	236
15	Dept Commun	113	Dept Civil & Environm Engn	232
16	Dept Management Sci & Engn	110	Dept Neurosurg	219
17	Dept Mech Engn	110	Environm Fluid Mech Lab	188
18	Persuas Technol Lab	103	Dept Radiol	185
19	Dept Chem Engn	88	Ctr Study Language & Informat	163
20	Dept Hlth Res & Policy	87	Dept Elect Engn & Comp Sci	160
		MIT		
1	Media Lab	3380	Dept Elect Engn & Comp Sci	3231
2	Comp Sci Lab	3198	Artificial Intelligence Lab	2319
3	Artificial Intelligence Lab	2831	Comp Sci Lab	2309
4	Dept Brain & Cognit Sci	1367	Media Lab	2134
5	Ctr Biol & Computat Learning	1319	Informat & Decis Syst Lab	1761
6	Dept Elect Engn & Comp Sci	913	Comp Sci & Artificial Intelligence Lab	1422
7	Dept Math	796	Dept Mech Engn	803
8	Lincoln Lab	727	Dept Math	775
9	Informat & Decis Syst Lab	509	Dept Brain & Cognit Sci	758
10	Dept Mech Engn	497	Alfred P Sloan Sch Management	408
11	Elect Res Lab	465	CSAIL	328
12	Alfred P Sloan Sch Management	450	Elect Res Lab	295
13	Dept Ocean Engn	352	Whitehead Inst	273
14	Vis & Modeling Media Lab	333	Brain Sci Dept	257
15	Dept Biol	212	Ctr Informat Syst Res	206
16	AI Lab	179	Dept Aeronaut & Astronaut	200
17	Dept Phys	163	Ctr Biol & Computat Learning	199
18	Dept Elect Engn	161	Dept Engn Mech	198
19	Ctr Operat Res	142	Lincoln Lab	191
20	Dept Aeronaut & Astronaut	123	Sloan Sch Management	190
		Univ Calif Berkeley		
1	Dept Elect Engn & Comp Sci	4145	Dept Elect Engn & Comp Sci	7292
2	Div Comp Sci	3209	Div Comp Sci	3787
3	Lawrence Berkeley Lab	2232	Dept Stat	3416
4	Elect Engn & Comp Sci Div	1741	Lawrence Berkeley Lab	960
5	Dept Math	1229	Comp Sci Div	637
6	Dept Stat	971	Dept Comp Sci	578
7	Dept EECS	487	Lawrence Berkeley Natl Lab	540
8	Dept Mech Engn	362	Dept EECS	527
9	Dept Phys	361	Dept Ind Engn & Operat Res	468
10	Sch Informat Syst	333	Dept Math	448
11	Neurol Unit	305	BISC	377
12	Int Comp Sci Inst	291	Elect Res Lab	354
13	Lawrence Berkeley Natl Lab	263	Dept Mech Engn	272
14	CS Div	230	Space Sci Lab	250
15	Telerobot Unit	223	Elect Engn & Comp Sci Div	212
16	Dept Comp Sci	222	Berkeley Sensor & Actuator Ctr	185
17	Elect Res Lab	190	Dept Elect & Comp Syst Engn	178
18	Dept Civil & Environm Engn	181	Grad Sch	144
19	Berkeley Wireless Res Ctr	135	EECS Comp Sci Div	133

20	Dept Mol & Cell Biol	104	Sch Informat Management & Syst	118

them are concerned with suborganizations of institutions (such as schools or departments) on the large scale. In the present study, we conducted a profound analysis of Web of Science metadata on computer science journal articles published in the period 1996 – 2005 and juxtaposed the first five with the last five years of this epoch with the aim of finding collaboration and citation patterns of computer science institutional suborganizations as viewed from the perspective of high-impact computer science journals at the turn of the century. The main contributions of the present study are the following:

- We analyzed 362654 authors' addresses from 149347 articles and attempted to aggregate the articles by level-1 suborganizations (we call them departments) of their authors' affiliations.
- We split the data set into two groups of articles published in 1996 – 2000 and in 2001 – 2005, computed several well-known scientometric indicators such as citations, in-degree, or h-index for the departments and compared the research production and (relative) impact of departments in the two distinct time periods. We also calculated the correlation of the rankings.
- We created complete collaboration and citation graphs of departments as they looked like in 1996 – 2000 and in 2001 – 2005 and identified the most intense collaborations and citations between departments.

Based on the methodology described above, we obtained the following main results:

- We found out that 0.3% of all affiliations did not comprise any primary organization, but 77% of all affiliations included one or more suborganizations. Though, this required level of detail has systematically appeared in Web of Science records only since 1998.
- In absolute terms, "IBM Corp; Thomas J Watson Res Ctr" always belongs to the best performers by all the criteria we examined in both periods. However, in relative terms

(scientometric indicators proportional to the number of publications), there is no clear winner. The least correlations (and, therefore, the greatest change in the rankings based on the two periods) were determined in in-degree, citations, and times cited.

- Even though the absolute numbers of nodes and edges in the collaboration and citation graphs of departments grew substantially between 1996 – 2000 and 2001 – 2005, the basic topological features of the corresponding networks remained very similar. We found that about 17% of collaborations and roughly 10% of citations between departments were intra-institutional in both periods. In any case, the later period was marked by a visible rise of Asian computer science departments in both productivity and influence.

A certain limitation of this study is that it relies on the presence of suborganizations in authors' affiliations and on the consistent way the order and names of suborganizations in affiliations are treated by Web of Science. The first problem results in institutions (with no suborganizations) being compared to departments in some cases and the second issue may cause some departments being underestimated due to name variations. Although we have shown that only about 1% of department names may be duplicates, both problems currently exist. Another limitation results from the fact that it is well known that computer science research is often published in conference proceedings rather than in journals. Even if, in our experience, many research results from influential computer science conferences are also published in journals later on (in an extended form), a natural extension of the present study would be including (more recent) data on computer science publications from journals as well as from conferences to get a more complete picture of the computer science research landscape. Alternatively, a different scientific domain might be chosen and the results (i.e. the dominating departments) could be compared to those from computer science.

Acknowledgements

This work was supported by the European Regional Development Fund (ERDF), project "NTIS - New Technologies for Information Society", European Centre of Excellence, CZ.1.05/1.1.00/02.0090.

REFERENCES

Arakaki, M. and Willett, P. 2009. Webometric analysis of departments of librarianship and information science: A follow-up study. *Journal of Information Science*, Vol. 35, no. 2: 143-152.

Arencibia-Jorge, R., Barrios-Almaguer, I., Fernández-Hernández, S. and Carvajal-Espino, R. 2008. Applying successive H indices in the institutional evaluation: A case study. *Journal of the American Society for Information Science and Technology*, Vol. 59, no. 1: 155-157.

Ben-David, D. 2010. Ranking Israel's economists. *Scientometrics*, Vol. 82, no. 2: 351-364.

Colliander, C. and Ahlgren, P. 2011. The effects and their stability of field normalization baseline on relative performance with respect to citation impact: A case study of 20 natural science departments. *Journal of Informetrics*, Vol. 5, no. 1: 101-113.

Colman, A.M., Debra, D. and Coulthard, B. 1995. A bibliometric evaluation of the research performance of British university politics departments: Publications in leading journals. *Scientometrics*, Vol. 32, no. 1: 49-66.

Fiala, D. 2012. Time-aware PageRank for bibliographic networks. *Journal of Informetrics*, Vol. 6, no. 3, 370-388.

Fiala, D., Ježek, K. and Rousselot, F. 2006. Finding Authoritative Researchers on Academic Web Sites. *Proceedings of World Academy of Science Engineering and Technology*, Vol. 17: 74-79.

Fiala, D., Rousselot, F. and Ježek, K. 2007. Ranking algorithms for Web sites: Finding Authoritative Academic Web Sites and Researchers. In J. Filipe, J. Cordeiro, B. Encarnaçao and V. Pedrosa, eds. *Proceedings of the 3rd International Conference on Web Information Systems and Technologies*. Barcelona: INSTICC: 372-375.

García, J.A., Rodriguez-Sánchez, R. and Fdez-Valdivia, J. 2012. A comparison of top economics departments in the US and EU on the basis of the multidimensional prestige of influential articles in 2010. *Scientometrics*, Vol. 93, no. 3: 681-698

Hirsch, J.E. 2005. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 102, no. 46: 16569-16572.

Lazaridis, T. 2010. Ranking university departments using the mean h-index. *Scientometrics*, Vol. 82, no. 2: 211-216.

Lee, G.J. 2010. Assessing publication performance of research units: Extensions through operational research and economic techniques. *Scientometrics*, Vol. 84, no. 3: 717-734.

Li, X., Thelwall, M., Musgrove, P. and Wilkinson, D. 2003. The relationship between the WIFs or inlinks of computer science departments in UK and their RAE ratings or research productivities in 2001. *Scientometrics*, Vol. 57, no. 2: 239-255.

Li, X., Thelwall, M., Wilkinson, D. and Musgrove, P. 2005. National and international university departmental web site interlinking. Part 1: Validation of departmental link analysis. *Scientometrics*, Vol. 64, no. 2: 151-185.

Nederhof, A.J., Meijer, R.F., Moed, H.F. and Van Raan, A.F.J. 1993. Research performance indicators for university departments: A study of an agricultural university. *Scientometrics*, Vol. 27, no. 2: 157-178.

Nederhof, A.J. and Noyons, E.C.M. 1992. Assessment of the international standing of university departments' research: A comparison of bibliometric methods. *Scientometrics*, Vol. 24, no. 3: 393-404.

Oppenheim, C. 1995. The correlation between citation counts and the 1992 Research Assessment Exercise ratings for British library and information science university departments. *Journal of Documentation*, Vol. 51, no. 1: 18-27.

Oppenheim, C. 1997. The correlation between citation counts and the 1992 Research Assessment Exercise ratings for British research in genetics, anatomy and archaeology. *Journal of Documentation*, Vol. 53, no. 5: 477-487.

Seng, L.B. and Willett, P. 1995. The citedness of publications by United Kingdom library schools. *Journal of Information Science*, Vol. 21, no. 1: 68-71.

Shannon, P., Markiel, A., Ozier, O., Baliga N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, Vol. 13, no. 11: 2498-2504.

Tang, R. and Thelwall, M. 2004. Patterns of national and international Web inlinks to US academic departments: An analysis of disciplinary variations. *Scientometrics*, Vol. 60, no. 3: 475-485.

Thomas, O. and Willett, P. 2000. Webometric analysis of departments of librarianship and information science. *Journal of Information Science*, Vol. 26, no. 6 : 421-428.

Torres-Salinas, D., Lopez-Cózar, E.D. and Jiménez-Contreras, E. 2009. Ranking of departments and researchers within a university using two different databases: Web of Science versus Scopus. *Scientometrics*, Vol. 80, no. 3: 761-774.

Zhou, P. and Leydesdorff, L. 2011. Fractional counting of citations in research evaluation: A cross- and interdisciplinary assessment of the Tsinghua University in Beijing. *Journal of Informetrics*, Vol. 5, no. 3, 360-368.

Preprint of: Fiala, D. (2014). Sub-organizations of institutions in computer science journals at the turn of the century. *Malaysian Journal of Library and Information Science*, 19(2), 53-68.