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

Dalibor Fiala
Department of Computer Science and Engineering,
University of West Bohemia,
Univerzitní 8, 30614 Plzeň, CZECH REPUBLIC
e-mail: dalfia@kiv.zcu.cz

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 sub-organizations 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 150,000 journal articles indexed by the 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 sub-organizations 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 performer. 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; Sub-organizations; Research productivity; 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 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 sub-organizations: such as schools, faculties, departments, divisions, centres, institutes, and laboratories. But the hierarchy of sub-organizations is not uniform across all institutions – some of them may have a relatively flat structure while others have many sub-organizations with their own subunits. Another problem is the inconsistent way authors of research papers state their affiliations. Some authors 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 sub-organizations omitted1, and yet other

1 This is also the case of this paper’s author’s affiliation, where the intermediate sub-organization, Faculty of Applied Sciences, is missing.
Fiala, D.

scholars prefer stating their full address and affiliation with all sub-organizational units included. In this case, there may be a difficulty with the order of organization and sub-organizations in the full affiliation. Sometimes, the complete affiliation is written in a top-down approach (the primary organization followed by a sub-organization, followed by its sub-organization, etc.), other times the opposite is true. In many cases, the order of sub-organizations 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 sub-organizations and in unifying or disambiguating author affiliations.

The above issues may be the reason why large-scale investigations into institutional sub-organizations have been avoided to a great extent so far. In this study, we attempt 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 sub-organizations in the affiliations of authors of these articles. In particular, we attempt to find answers to these research questions:

(a) what is the composition of author affiliations in terms of their institutions and sub-organizations in a set of Web of Science data?

(b) based on this data set, how do sub-organizations 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 sub-organizations and what is the nature of these collaboration and citation networks?

Research question (a) is addressed under section Materials and Methods, while research questions (b) and (c) are 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, Rodríguez-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 Collinder and Ahlgren (2011). Lee (2010) discussed possible adjustments of research evaluation indicators for institutional sub-organizations (called “research units” therein) normalizing for their size and capacity. Ben-David (2010) ranked 11 Israeli economics departments by their citation impact and so did Lazaridis (2010) for 16 Greek university departments in chemistry, chemical engineering, materials science, and physics. Torres-Salinas, López-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 Arenobia-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 et al. (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 et al. (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 (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 that has never been described in the available literature.

MATERIALS AND METHOD

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 (JCR) 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. Changes in journal names that occurred during the period under study were not taken into account. In this way, we obtained 149,347 “core” publications (strictly said, their metadata) and 191,447 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

---

2 A later update of the JCR 2009 database added two more journals which were not part of this study.
are available, which allows for authors’ affiliations to be extracted. In total, we extracted 362,654 author affiliations (addresses), which make 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 of the researchers’ address. There were 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 sub-organizations; 80% of them had one sub-organization; 17% had two sub-organizations; and 3% had three sub-organizations. There were no addresses with four or more sub-organizations. Although 77% of all addresses had one or more sub-organizations, 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% respectively, and 25.3% of the addresses in the articles published in those years had some sub-organizations. The later years had this share over 80% each. A sample XML record with a researcher’s address is presented 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

The order in which sub-organizations appear in each address determines their “sub-organization level”. Typically, an institution is followed by a level-1 sub-organization, followed by a level-2 sub-organization, followed by a level-3 sub-organization at most. An affiliation may thus comprise up to four units, examples of which are shown in Table 1 below. If the order of sub-organizations changes, the affiliation is then considered different. Therefore, our present study entirely relies on the order of sub-organizations 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 is made in the next paragraph.
Suborganizations of Institutions in Computer Science Journals

<table>
<thead>
<tr>
<th>Institution</th>
<th>Level-1 suborg.</th>
<th>Level-2 suborg.</th>
<th>Level-3 suborg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIT</td>
<td>Comp Sci Lab</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIT</td>
<td>Media Lab</td>
<td>Software Agents Grp</td>
<td></td>
</tr>
<tr>
<td>MIT</td>
<td>Alfred P Sloan Sch Management Ctr Ebusiness</td>
<td>Dept Management Sci</td>
<td></td>
</tr>
</tbody>
</table>

Since most affiliations contained an institution and a (level-1) sub-organization, we decided to aggregate affiliations by level-1 sub-organizations. In this way, “MIT; Media Lab” (an institution with a level-1 sub-organization) and “MIT; Media Lab; Software Agents Grp” (an institution with a level-1 sub-organization and a level-2 sub-organization) become both “MIT; Media Lab” as an example. Of course, if an institution has no sub-organizations 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 sub-organizations appear relatively rarely in affiliations, the aggregation by level-1 sub-organizations seems to be the most appropriate, although it is still not optimal due to affiliations with no sub-organizations. In total, we got 56,400 distinct “true” departments (institutions with no sub-organizations not included). As it is impractical 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 sub-organization in the names of departments and discovered that, on average, about 1% of names in the samples had (conversely) a sub-organization 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 WoS “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 sub-organizations in these two periods. In the two article groups, there were 67,061 and 82,286 publications and 34,545 and 48,148 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 WoS as it 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 WoS 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 17,590 nodes (departments) and 76,472 edges in 1996 – 2000 and 28,967 nodes (departments) and 131,383 edges in 2001 – 2005. The sum of weights (i.e. the actual number of citations between departments) was 165,883 and 345,607 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


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. Figure 2 presents scatter plots of ranks assigned to individual departments by six scientometric indicators in the periods 1996 – 2000 and 2001
– 2005. discussed in the previous section. 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.

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 (\(\rho = 0.664\)) 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 (\(\rho = 0.450\)). 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 can be seen on the bottom charts of Figure 2. The inter-period performer 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 least performing 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.
Figure 2: Scatter Plots of Departments’ Ranks by Various Indicators in Two Periods

- **Publications**
  - Rank in 2001-2005 vs. Rank in 1996-2000 ($\rho = 0.560$)

- **Times cited**
  - Rank in 2001-2005 vs. Rank in 1996-2000 ($\rho = 0.300$)

- **h-index (times cited)**
  - Rank in 2001-2005 vs. Rank in 1996-2000 ($\rho = 0.450$)

- **h-index (edges)**
  - Rank in 2001-2005 vs. Rank in 1996-2000 ($\rho = 0.664$)

- **Citations**
  - Rank in 2001-2005 vs. Rank in 1996-2000 ($\rho = 0.303$)

- **In-degree**
  - Rank in 2001-2005 vs. Rank in 1996-2000 ($\rho = 0.287$)
Collaborations and Citations between Sub-organizations

In addition to the various scientometric performance indicators of departments discussed earlier, we were also interested to know 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 with “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 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 60,848 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 WoS 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; 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 department frequently citing each other, which is a phenomenon commented on earlier. Regarding the proportion of intramural 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, and characteristic path length, 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).
### Table 2: Top 20 Departments of Three Leading Universities in Computer Science

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stanford Univ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Dept Comp Sci</td>
<td>3839</td>
<td>5139</td>
</tr>
<tr>
<td>2 Dept Elect Engn</td>
<td>1394</td>
<td>1605</td>
</tr>
<tr>
<td>3 Ctr Turbulence Res</td>
<td>908</td>
<td>1410</td>
</tr>
<tr>
<td>4 Stanford Linear Accelerator Ctr</td>
<td>640</td>
<td>584</td>
</tr>
<tr>
<td>5 Dept Psychol</td>
<td>582</td>
<td>482</td>
</tr>
<tr>
<td>6 Comp Syst Lab</td>
<td>489</td>
<td>Sch Med</td>
</tr>
<tr>
<td>7 Dept Stat</td>
<td>456</td>
<td>Stanford Med Informat</td>
</tr>
<tr>
<td>8 Sch Med</td>
<td>404</td>
<td>Dept Math</td>
</tr>
<tr>
<td>9 Dept Petr Engn</td>
<td>260</td>
<td>Informatt Lab</td>
</tr>
<tr>
<td>10 Psychol &amp; Neurosci Dept</td>
<td>192</td>
<td>Sci Comp &amp; Computat Math Program</td>
</tr>
<tr>
<td>11 Dept Engr Econ Syst &amp; Operat Res</td>
<td>126</td>
<td>Grad Sch Business</td>
</tr>
<tr>
<td>12 Dept Math</td>
<td>125</td>
<td>Sci Comp Computat Math Program</td>
</tr>
<tr>
<td>13 Dept Radiol</td>
<td>118</td>
<td>Dept Management Sci &amp; Engn</td>
</tr>
<tr>
<td>14 Gates Comp Sci</td>
<td>114</td>
<td>Ctr Turbulence Res</td>
</tr>
<tr>
<td>15 Dept Commun</td>
<td>113</td>
<td>Dept Civil &amp; Environm Engn</td>
</tr>
<tr>
<td>16 Dept Management Sci &amp; Engn</td>
<td>110</td>
<td>Dept Neurosurg</td>
</tr>
<tr>
<td>17 Dept Mech Engn</td>
<td>110</td>
<td>Environm Fluid Mech Lab</td>
</tr>
<tr>
<td>18 Persuas Technol Lab</td>
<td>103</td>
<td>Dept Radiol</td>
</tr>
<tr>
<td>19 Dept Chem Engn</td>
<td>88</td>
<td>Ctr Study Language &amp; Informat</td>
</tr>
<tr>
<td>20 Dept Hlth Res &amp; Policy</td>
<td>87</td>
<td>Dept Elect Engn &amp; Comp Sci</td>
</tr>
<tr>
<td><strong>MIT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Media Lab</td>
<td>3380</td>
<td>Dept Elect Engn &amp; Comp Sci</td>
</tr>
<tr>
<td>2 Comp Sci Lab</td>
<td>3198</td>
<td>Artificial Intelligence Lab</td>
</tr>
<tr>
<td>3 Artificial Intelligence Lab</td>
<td>2831</td>
<td>Comp Sci Lab</td>
</tr>
<tr>
<td>4 Dept Brain &amp; Cognit Sci</td>
<td>1367</td>
<td>Media Lab</td>
</tr>
<tr>
<td>5 Ctr Biol &amp; Computat Learning</td>
<td>1319</td>
<td>Informatt &amp; Decis Syst Lab</td>
</tr>
<tr>
<td>6 Dept Elect Engn &amp; Comp Sci</td>
<td>913</td>
<td>Comp Sci &amp; Artificial Intelligence Lab</td>
</tr>
<tr>
<td>7 Dept Math</td>
<td>796</td>
<td>Dept Mech Engn</td>
</tr>
<tr>
<td>8 Lincoln Lab</td>
<td>727</td>
<td>Dept Math</td>
</tr>
<tr>
<td>9 Informat &amp; Decis Syst Lab</td>
<td>509</td>
<td>Dept Brain &amp; Cognit Sci</td>
</tr>
<tr>
<td>10 Dept Mech Engn</td>
<td>497</td>
<td>Alfred P Sloan Sch Management</td>
</tr>
<tr>
<td>11 Elect Res Lab</td>
<td>465</td>
<td>CSAIL</td>
</tr>
<tr>
<td>12 Alfred P Sloan Sch Management</td>
<td>450</td>
<td>Elect Res Lab</td>
</tr>
<tr>
<td>13 Dept Ocean Engn</td>
<td>352</td>
<td>Whitehead Inst</td>
</tr>
<tr>
<td>14 Vis &amp; Modeling Media Lab</td>
<td>333</td>
<td>Brain Sci Dept</td>
</tr>
<tr>
<td>15 Dept Biol</td>
<td>212</td>
<td>Ctr Informat Syst Res</td>
</tr>
<tr>
<td>16 AI Lab</td>
<td>179</td>
<td>Dept Aeronaut &amp; Astronaut</td>
</tr>
<tr>
<td>17 Dept Phys</td>
<td>163</td>
<td>Ctr Biol &amp; Computat Learning</td>
</tr>
<tr>
<td>18 Dept Elect Engn</td>
<td>161</td>
<td>Dept Engn Mech</td>
</tr>
<tr>
<td>19 Ctr Operat Res</td>
<td>142</td>
<td>Lincoln Lab</td>
</tr>
<tr>
<td>20 Dept Aeronaut &amp; Astronaut</td>
<td>123</td>
<td>Sloan Sch Management</td>
</tr>
<tr>
<td><strong>Univ Calif Berkeley</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Dept Elect Engn &amp; Comp Sci</td>
<td>4145</td>
<td>Dept Elect Engn &amp; Comp Sci</td>
</tr>
<tr>
<td>2 Div Comp Sci</td>
<td>3209</td>
<td>Div Comp Sci</td>
</tr>
<tr>
<td>3 Lawrence Berkeley Lab</td>
<td>2232</td>
<td>Dept Stat</td>
</tr>
<tr>
<td>4 Elect Engn &amp; Comp Sci Div</td>
<td>1741</td>
<td>Lawrence Berkeley Lab</td>
</tr>
<tr>
<td>5 Dept Math</td>
<td>1229</td>
<td>Comp Sci Div</td>
</tr>
<tr>
<td>6 Dept Stat</td>
<td>971</td>
<td>Dept Comp Sci</td>
</tr>
<tr>
<td>7 Dept EECS</td>
<td>487</td>
<td>Lawrence Berkeley Natl Lab</td>
</tr>
<tr>
<td>8 Dept Mech Engn</td>
<td>362</td>
<td>Dept EECS</td>
</tr>
<tr>
<td>9 Dept Phys</td>
<td>361</td>
<td>Dept Ind Engn &amp; Operat Res</td>
</tr>
<tr>
<td>10 Sch Informat Syst</td>
<td>333</td>
<td>Dept Math</td>
</tr>
</tbody>
</table>

---
CONCLUSIONS AND FUTURE WORK

Research papers are written by authors who are affiliated with institutions and, very often, with sub-organizations 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 them are concerned with sub-organizations of institutions (such as schools or departments) on the large scale. In the present study, we conducted a profound analysis of WoS 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 sub-organizations 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 as follows:

• We analyzed 36,2654 authors’ addresses from 149,347 articles and attempted to aggregate the articles by level-1 sub-organizations (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 sub-organizations. Though, this required level of detail has systematically appeared in WoS 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 performer. 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.

The limitation of this study is that it relies on the presence of sub-organizations in authors’ affiliations and on the consistent way the order and names of sub-organizations in affiliations are treated by WoS. The first problem results in institutions (with no sub-organizations) 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 well known fact 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


