How do Computer Science Professors at the UIUC Collaborate with Each Other?

Liang Tao & Ted Stelling
Outline

• Introduction
• Data Collection
• Collaboration Network
• Topic Model
• Search Engine
• Conclusion
Introduction

• Some popular academic search engines
  • Google Scholar
  • Microsoft Academic Search
  • ArnetMiner

• Unsupported features
  • Overview of whole collaboration network
  • Find out the collaboration pattern from scholar perspective
Terminology

• Collaboration
  • On research papers at conferences or publication of these papers in academic journals

• Dataset: DBLP
  • Specific for Computer Science field
  • 2.6 million publications
  • 1.4 million researchers
  • 25,000 journal volumes & 24,000 conferences
Data Collection

- Offline Collection
- Online Collection
  - Author / Publication / Coauthor
  - Java SAX Parser
- Data Analysis
  - NodeXL
Basic Statistics

- 61 Faculty
- 5019 Scholars
- 6289 Publications

Top professors with the highest number of coauthors

<table>
<thead>
<tr>
<th>Name</th>
<th>Coauthor</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jiawei Han</td>
<td>581</td>
<td>665</td>
</tr>
<tr>
<td>William D. Gropp</td>
<td>308</td>
<td>215</td>
</tr>
<tr>
<td>Marc Snir</td>
<td>286</td>
<td>120</td>
</tr>
<tr>
<td>Tarek F. Abdelzaher</td>
<td>277</td>
<td>243</td>
</tr>
<tr>
<td>Klara Nahrstedt</td>
<td>264</td>
<td>360</td>
</tr>
<tr>
<td>Chengxiang Zhai</td>
<td>237</td>
<td>239</td>
</tr>
<tr>
<td>Roy H. Campbell</td>
<td>227</td>
<td>236</td>
</tr>
<tr>
<td>David A. Padua</td>
<td>226</td>
<td>177</td>
</tr>
<tr>
<td>Laxmikant V. Kale</td>
<td>193</td>
<td>225</td>
</tr>
<tr>
<td>Carl A. Gunter</td>
<td>178</td>
<td>123</td>
</tr>
</tbody>
</table>
The Collaboration Network

Isolated Group

Algorithms for finding group by clustering

Topic Model Analysis

• Topic model analysis based on each papers title.
• Compare topics identified with UIUC topics
• Compare topics identified with collaboration clusters

• Architecture, Compilers and Parallel Computing
• Systems and Networking
• Theory and Algorithms
• Artificial Intelligence
• Programming Languages, Formal Methods, and Software Engineering
• Database and Information Systems
• Graphics, Visualization, and HCI
• Scientific Computing
• Bioinformatics and Computational Biology
Search Engine

Tool to validate findings

• Individual Professor Profiles
• Publication Search
• Degree of Collaboration
Conclusions & Limitations

• Collaboration can be determined by authorship analysis.
• Examining a collection of documents by authors' collaboration provides a different view of similarity.
• Expansion is limited by the availability of the author information.
Possible Uses

• Students would be able to identify schools with concentrations in their area of interest
• Researchers could identify others to possibly collaborate with on new research
• If expanded, identify experts in a particular field when needed
Future Work

- Develop a process to make researchers names unique
- Use the number of links to determine the degree of collaboration
- Adjust the degree of collaboration by the position of the author on the paper
- Expand beyond the University of Illinois faculty
- Expand beyond the Computer Science field
Questions?