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

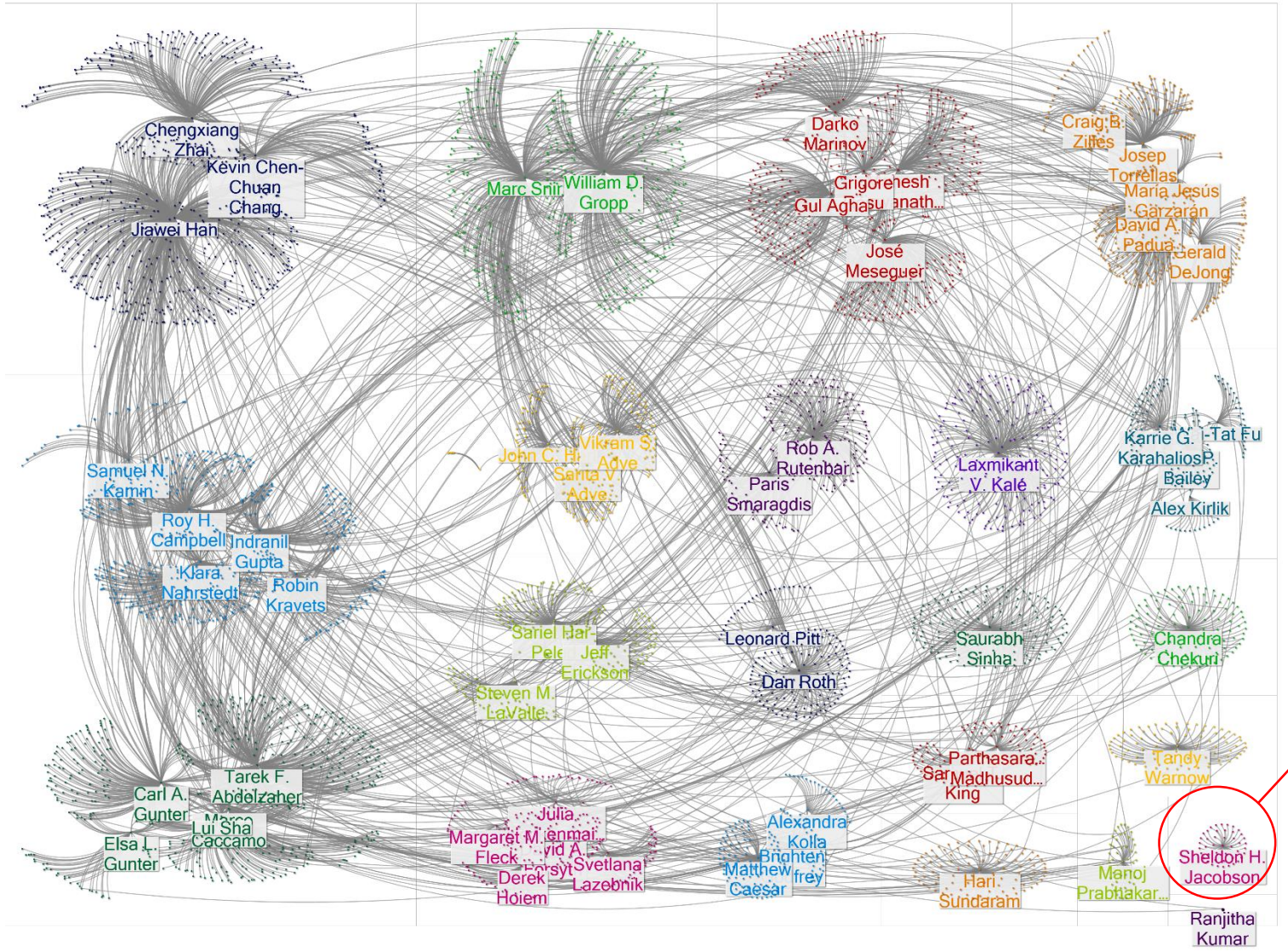
Basic Statistics

- 61 Faculty
- 5019 Scholars
- 6289 Publications

Top professors with the highest number of coauthors

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

The Collaboration Network

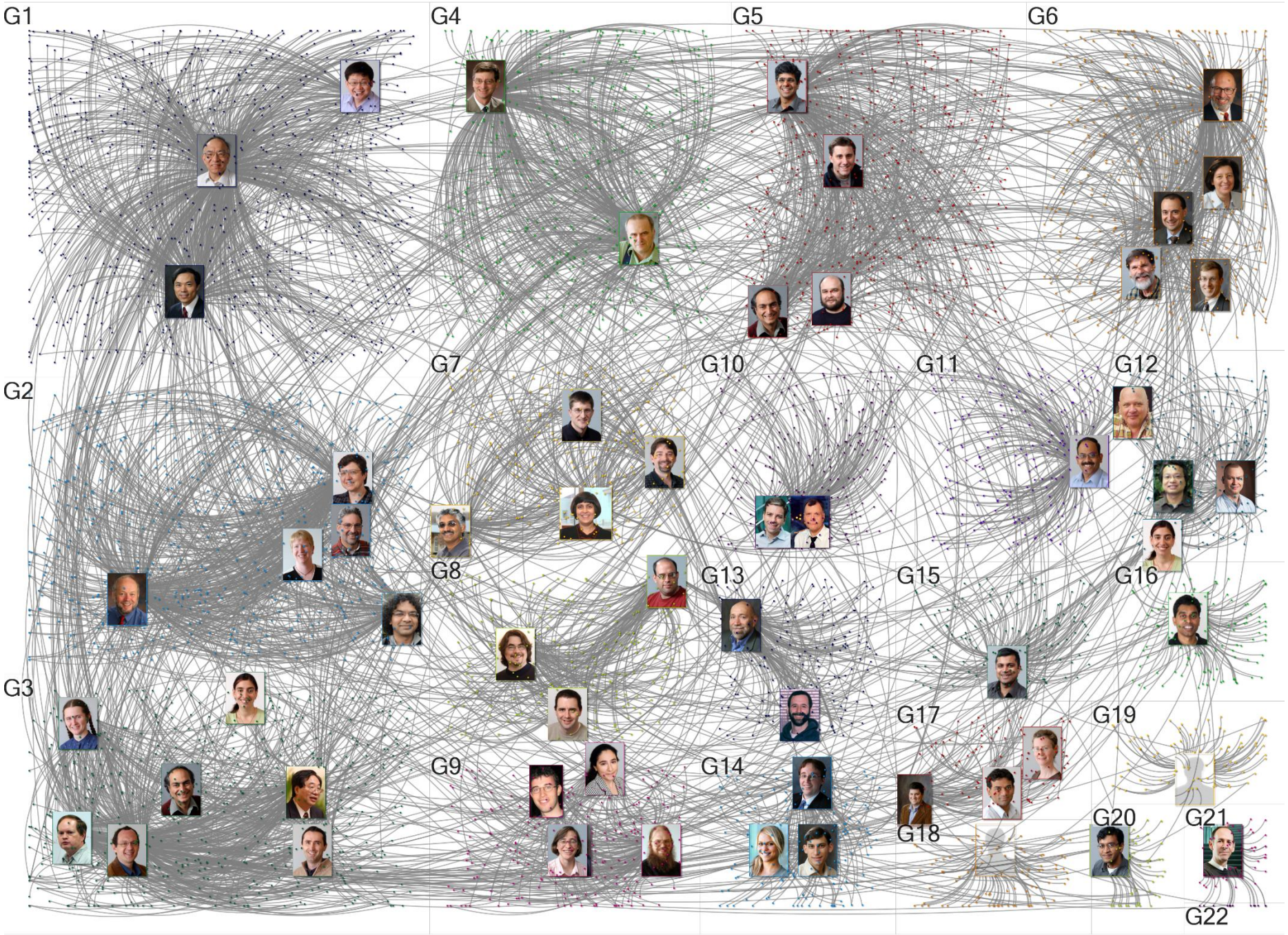


with NodeXL (http://nodexl.codeplex.com) from the Social Media Research Foundation (http://www.smrfoundation.org)

Isolated Group

Algorithms for finding group by clustering

Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks. Phys. Rev. E, 70:066111, Dec 2004.



eated with NodeXL (http://nodexl.codeplex.com) from the Social Media Research Foundation (http://www.smrfoundation.org)

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 determine 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?