



# AMiner-mini: A People Search Engine For University

Jingyuan Liu\*, Debing Liu\*, Xingyu Yan\*,  
Li Dong<sup>#</sup>, Ting Zeng<sup>#</sup>, Yutao Zhang\*, and Jie Tang\*

\*Dept. of Com. Sci. and Tech., Tsinghua University

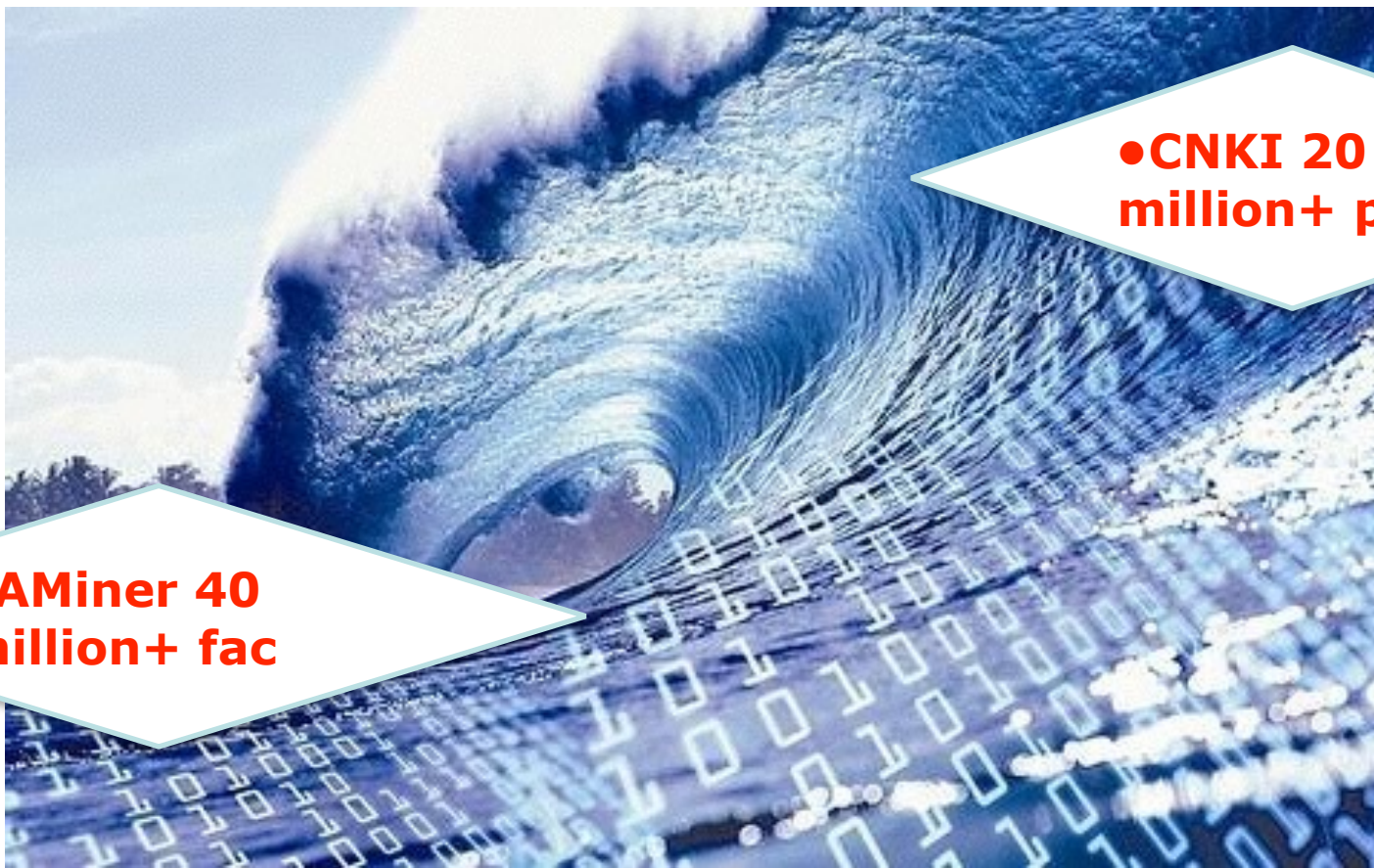
<sup>#</sup>Tsinghua University Library

System website: <http://dlib.lib.tsinghua.edu.cn/>

Paper: <http://keg.cs.tsinghua.edu.cn/jietang/publications/CIKM14-Liu-et-alAminer-mini.pdf>

# Motivation

- Digital Academic Data **Rapid Proliferation**

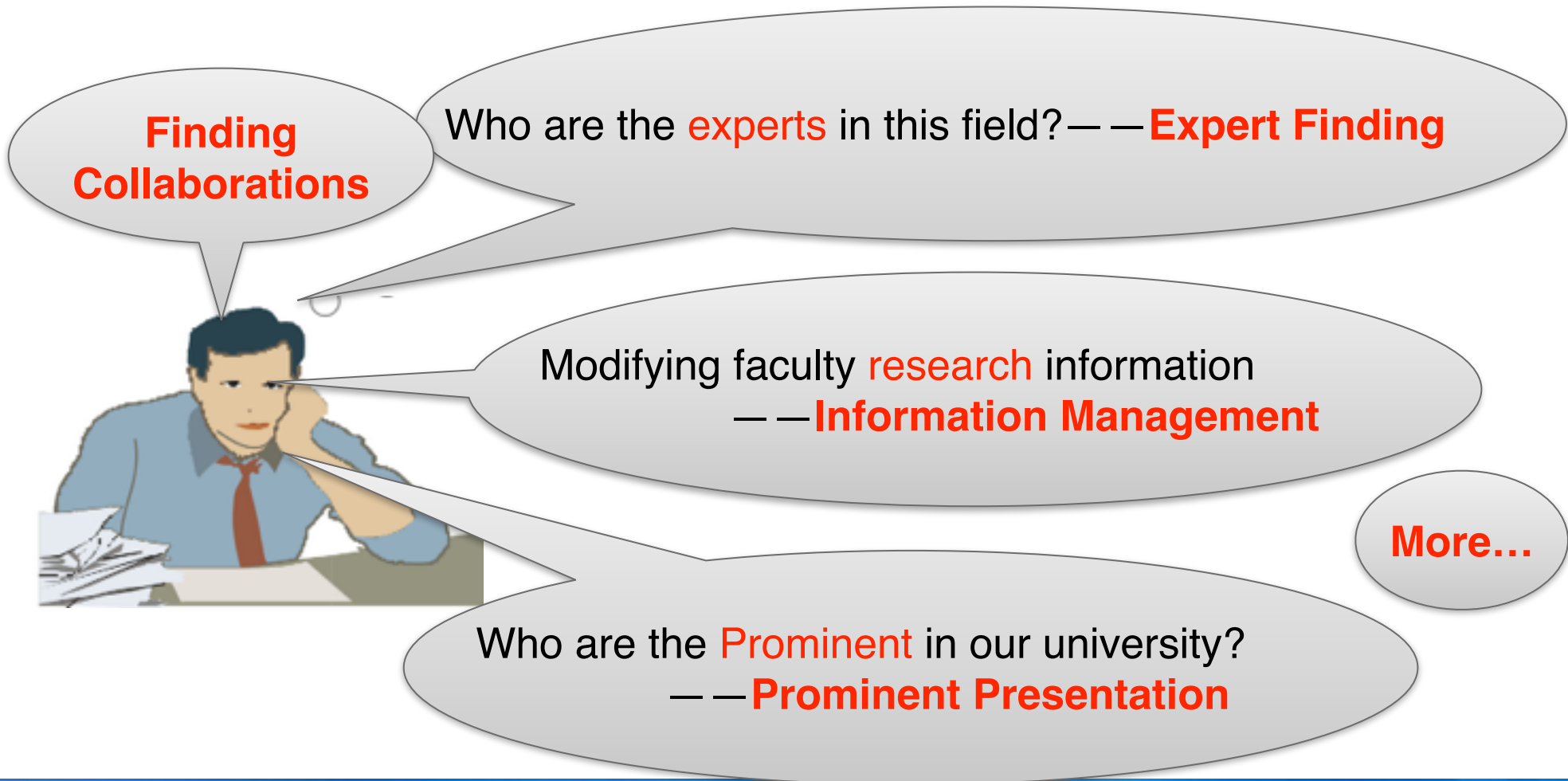


• **AMiner 40  
million+ fac**

• **CNKI 20  
million+ pub**

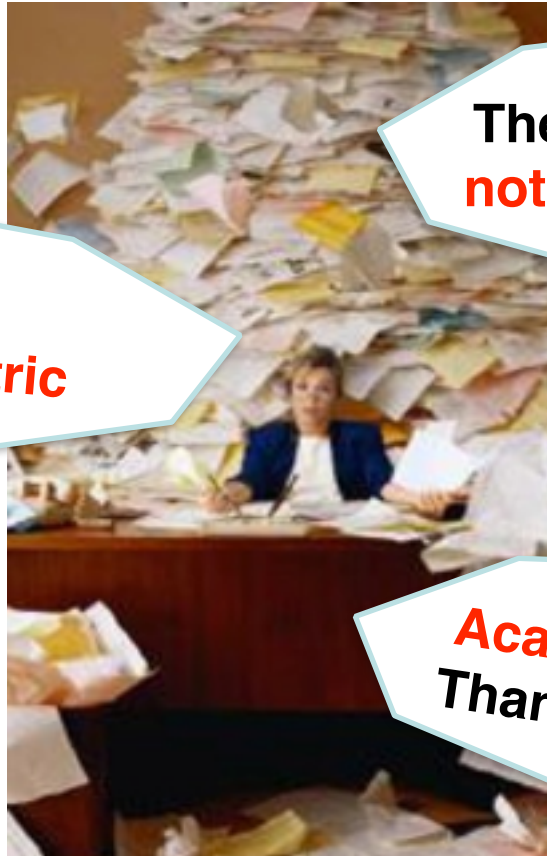
# Motivation

- Satisfying Different User Scenarios



# Motivation

- **People-Centric** rather than Data-Centric



Web Search Trend:  
Data Centric->**People Centric**

The Information need is  
**not only** about Pub

**Academic Search More  
Than Keywords Matching**

# What is AMiner-mini?

- **A People Search Engine for University**
  - **Core Techniques:**
    - **Name Disambiguation**
    - **Academic Search**
  - **System Applications:**
    - **Expert Finding**
    - **Prominent Presentation**
    - **Publication Management**
  - **Distributed Structure:**
    - **Distributed Search**



# System Statistic

- System mainly contains 3 entities:
  - **Faculty:**  
System contains **10918** faculties from **90** department
  - **Papers:**  
System contains **259465** papers range from 1981 to 2014
  - **Course:**  
System contains **10253** courses range from 2001 to 2013

# Academic Search Algorithm

- Modeling Ranking Factors
  - **Relevance:** “**relevance**” between queries and entities
    - Language Model
    - LDA
  - **Importance:** “**important**” and “**influential**”
    - Random Walk
    - Prominent title
  - **Popularity:** “**popular**” entities
    - User feedback
    - Random Serendipity

# Academic Search Algorithm

- Combing Ranking Factors

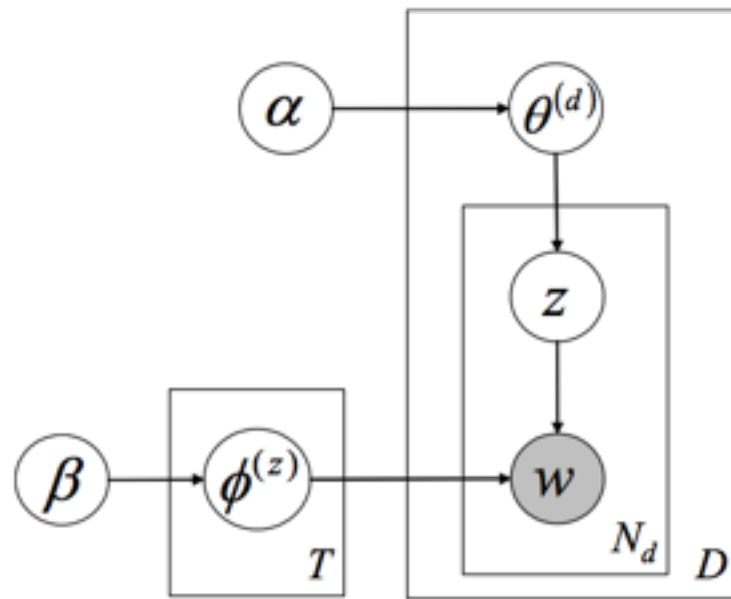
- $Score = \omega_R * Relevance + \omega_I * Importance + \omega_p * Popularity$
- weights are initially **manually** set
- weights are 0.6, 0.2, 0.2 separately





# Academic Search Algorithm

- Statistic Topic Model
  - Using LDA to extract **hidden topics** from textural materials



# Academic Search Algorithm

- Search Experiment Result

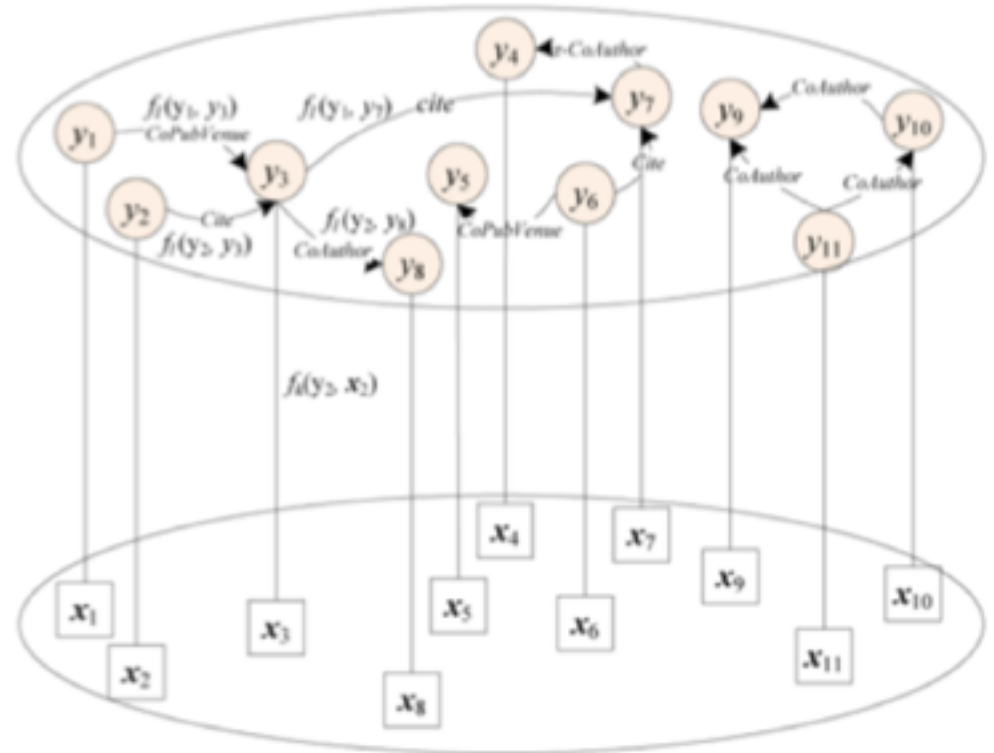
Search Methods	P@5	P@10	MAP
<b>0.3 LDA + 0.7 LM</b>	<b>0.876</b>	0.8	<b>0.912</b>
0.2 LDA + 0.8 LM	0.864	<b>0.81</b>	0.89
0.0 LDA + 1.0 LM	0.872	0.77	0.79
Lucene (TFIDF)	0.773	0.726	0.73

- Obviously **outperforms** baseline (TF-IDF)
- best combination weights: 0.3 LDA + 0.7 LM

# Name Disambiguation Methodology

- Probabilistic HMRF Framework

- Using a Probabilistic HMRF Framework to **cluster ambiguity** papers and courses



# Name Disambiguation Methodology

- Active Learning Strategy
  - Using **active learning** strategy to form three-phases disambiguation framework



# System Applications

- Expert Finding
  - Implement expert finding via **academic search** algorithm
  - Search for faculty, pub, course **simultaneously**

The screenshot displays a web application interface for a search system. At the top, there is a navigation bar with links: 校园搜索 (Campus Search), 院系概况 (Department Overview), 杰出人才 (Outstanding Talents), 高级搜索 (Advanced Search), and 系统介绍 (System Introduction). A search bar is located below the navigation bar, with the text '社会网络' (Social Network) entered and a '搜索' (Search) button.

The main content area is divided into three sections:

- "社会网络" 的相关教师 (Related Teachers of "Social Network")**: This section displays a list of teachers. The first teacher shown is 唐杰 (Tang Jie), a professor at the Department of Computer Science and Technology, Tsinghua University. The second teacher is 曾晓宏 (Zeng Xiaohong), a professor at the School of Information Science and Technology, Tsinghua University. The third teacher is 罗家德 (Luo Jiade), a professor and Ph.D. supervisor at the School of Social Sciences, Tsinghua University. The fourth teacher is 刘惠芬 (Liu Hui fen), a professor at the School of Journalism and Communication, Tsinghua University. The fifth teacher is 任平原 (Ren Pingyuan), a professor at the School of Information Science and Technology, Tsinghua University.
- "社会网络" 的相关课程 (Related Courses of "Social Network")**: This section displays a list of courses. The first course is 计算机网络及应用 (Computer Network and Application) by 李洪强 (Li Hongqiang). The second course is 计算机网络技术 (Computer Network Technology) by 丁勇广 (Ding Yongguang). The third course is 计算机网络技术与实践 (Computer Network Technology and Practice) by 邓在星 (Deng Zaixing). The fourth course is 网络控制 (Network Control) by 李小明 (Li Xiaoming). The fifth course is 网络优化 (Network Optimization) by 谢金良 (Xie Jinliang). The sixth course is 计算机网络 (Computer Network) by 李洪强 (Li Hongqiang). The seventh course is 网络性能 (Network Performance) by 曾白华 (Zeng Baihua).
- "社会网络" 的相关论文 (Related Papers of "Social Network")**: This section displays a list of papers. The first paper is 社会网络对外包治理和美化的影响 (Impact of Social Network on Outsourcing Governance and Beautification) by 刘智勇 (Liu Zhiyong), 刘智坤 (Liu Zhikun), and 黄景明 (Huang Jingming). The second paper is 科学学与科学技术管理 - 2009 (Science and Technology Management - 2009) by 黄景明 (Huang Jingming). The third paper is 竞争环境、法律环境和治理者社会网络治理 (Competitive Environment, Legal Environment and Governance of Social Network Governance) by 黄景明 (Huang Jingming). The fourth paper is 科学学与科学技术管理 - 2009 (Science and Technology Management - 2009) by 黄景明 (Huang Jingming). The fifth paper is 社会网络与战略匹配及其对企业绩效影响的实证研究 (Empirical Study on the Impact of Social Network and Strategic Fit on Enterprise Performance) by 邵晓博 (Shao Xiaobo), 李智坤 (Li Zhikun), and 黄景明 (Huang Jingming). The sixth paper is 科学学与科学技术管理 - 2011 (Science and Technology Management - 2011) by 黄景明 (Huang Jingming).

- Publication Management

- 

**郭宇林**

教授、博导、特聘研究员

✉ [guoy@hangzhou.edu.cn](mailto:guoy@hangzhou.edu.cn)

姓名: 郭宇林

**论文发表分布曲线:**

**近期关键词:**

10 carbon nanotubes

catheter spin polarization

nanotubular helixes

nanotubular semiconductors

carbon nanotubes

carbon nanotubes

carbon nanotubes

carbon nanotubes

carbon nanotubes

carbon nanotubes

carbon nanotubes

**郭宇林的论文: 289**

201 Carbon effects the inverse ability of HCT116 cells by regulating "TGF- $\beta$ "

Zhang M., Lu Y., Fang M., Bai X., Zhao W., Tang Z., Gu B., Li Z., Lu Y.

American Journal of Pathology - 2019

65%

1 近邻效应

近邻效应: 近邻效应

近邻效应: 近邻效应

65%

202 Interfacial thermal conductance of partially wrapped carbon nanotubes: Linear scaling and exponential decay

Chen X., Xu Y., Zou X., Gu B. L., Qian W.

Physical Review B - Condensed Matter and Materials Physics - 2019

65%

203 Mixed ferromagnetic semiconductor Li<sub>2</sub>NiP<sub>2</sub> with denuded charge and spin doping

Deng Z., Zhao X., Gu B., Han W., Zhu J. L., Wang S. C., Li X., Li B. G., Yu H. J., Sato T., Fronzen B., Li J., Zhang J., Wang T., Yang F. L., Kishino S., Yamane Y. J., Jin C.

Physical Review B - Condensed Matter and Materials Physics - 2019

65%

204 Topological insulators in transition-metal intercalated graphene: The role of interfaces in significantly increasing the spin-orbit gap

Li Y., Tang P., Chen P., Wu Z., Gu B. L., Fang X., Zhang S. B., Qian W.

Physical Review B - Condensed Matter and Materials Physics - 2019

65%

205 Field effect tunable spin loss in ultra-thin films of magnetically doped nanotubular helices

Zhen L., Tang P., Gu B. L., Qian W.

Physical Review Letters - 2019

65%



# System Applications

- Prominent Presentation

- Present **prominent** faculties with honored title



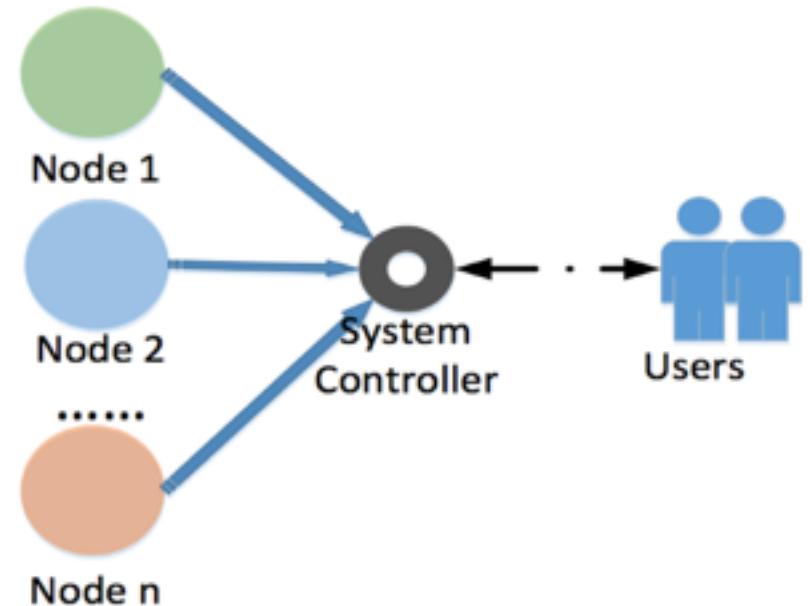
# System Applications

- PersonInfo Presentation
  - Research interest
- Academic social network
- Research Trend
- Research Topics



# Distributed Structure

- **Intra- and Inter-** university level academic services
  - work as single node
  - connect via web server
- **Distributed Search**
  - system controller
  - rerank search result



# Deploy your AMiner-mini

- System is cooperated with THU lib
- System is an ongoing project, THU version:
  - <http://dlib.lib.tsinghua.edu.cn/>
- We plan to build open-source project, find us:
  - [git@github.com:toothacher17/AMiner-mini.git](https://github.com/toothacher17/AMiner-mini)
- We are willing to help deploy your own AMiner-mini, contact us:
  - <http://keg.cs.tsinghua.edu.cn/jietang/>
- The system is developed under J2EE Tapestry Structure

# Reference

- J. Tang, A.C.M. Fong, B. Wang, and J. Zhang. A Unified Probabilistic Framework for Name Disambiguation in digital library. In *TKDE*, Volume 24, Issue 6, Pages 975-987, 2012
- K. Balog, Y. Fang, M. de Rijke, P. Serdyukov and L. Si. Expertise Retrieval. In *FTIR*, Volume 6, 2012
- J. Tang, J. Zhang, R. Jin, Z. Yang, K. Cai, L. Zhang, and Z. Su. Topic Level Expertise Search over Heterogeneous Networks. In *Machine Learning Journal*, Volume 82, Issue 2, Pages 211-237, 2011
- R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval (2<sup>nd</sup> Edition)*. China Machine Press, 2010
- J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang and Z. Su. ArnetMiner: Extraction and Mining of Academic Social Network. In *KDD'08*, pages 990-998, 2008.
- A. Ferreira, M. Gnocalves, and A. Laender. A Brief Survey of Automatic Methods for Author Name Disambiguation. In *SIGMOD'12*, 2012
- T. Joachims, L. Granka, H. Hembrooke, F. Radlinski, and G. Gay. Evaluating the Accuracy of Implicit Feedback from Clicks and Query Reformulations in Web Search. In *TIS*, Volume 25, 2007
- G. Coulouris, J. Dollimore, and T. Kindberg. *Distributed systems: Concepts and Design (5<sup>th</sup> Edition)*. China Machine Press, 2011.
- M. Ge, C. Delgado-Battenfeld, and D. Jannach. Beyond accuracy: Evaluating recommender systems by coverage and serendipity. In *RecSys'10*, 2010

That is all!