# AMiner: Mining Deep Knowledge from Big Scholar Data

Jie Tang
Department of Computer Science and Technology, Tsinghua University
Tsinghua National Laboratory for Information Science and Technology (TNList)
jietang@tsinghua.edu.cn

## ABSTRACT

**AMiner**[1] is the second generation of the ArnetMiner system [6]. We focus on developing author-centric analytic and mining tools for gaining a deep understanding of the large and heterogeneous networks formed by authors, papers, venues, and knowledge concepts. One fundamental goal is how to extract and integrate semantics from different sources. We have developed algorithms to automatically extract researchers' profiles from the Web [5] and resolve the name ambiguity problem [3], and connect different professional networks [9]. We also developed methodologies to incorporate knowledge from the Wikipedia and other sources into the system [7, 2] to bridge the gap between network science and the web mining research. In this talk, I will focus on answering two fundamental questions for author-centric network analysis: who is who? and who are similar to each other?

The system has been in operation since 2006 and has collected more than 100,000,000 author profiles, 100,000,000 publication papers, and 7,800,000 knowledge concepts. It has been widely used for collaboration recommendation [4], similarity analysis [8], and community evolution [1].

## Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data Mining

## Keywords

Academic search; Network analysis; Knowledge graph; Big scholar data

## 1. BIOGRAPHY

**Jie Tang** is an associate professor with the Department of Computer Science and Technology at Tsinghua University, and was also visiting scholar at Cornell University and Hong Kong University of Science and Technology. His interests include social network analysis, data mining, and machine learning. He has published more than 100 journal/conference papers and holds 14 patents. He served as PC Co-Chair of CIKM'16, WSDM'15, ASONAM'15, SocInfo'12, KDD-CUP/Workshop/Local Co-Chair of KDD'11-15, and as the PC member of more than 50 international conferences. He is the founder of AMiner.org (ArnetMiner), which has attracted more than 7 million independent IP accesses from 220 countries/regions in the world. He was honored with the Newton Advanced Scholarship Award, CCF Young Scientist Award, and NSFC Excellent Young Scholar.

## 2. REFERENCES

[1] Y. Sun, J. Tang, J. Han, C. Chen, and M. Gupta. Co-evolution of multi-typed objects in dynamic star networks. *IEEE TKDE*, 26(12):2942–2955, 2014.

[2] J. Tang, Z. Fang, and J. Sun. Incorporating social context and domain knowledge for entity recognition. In *WWW'15*, pages 517–526, 2013.

[3] J. Tang, A. Fong, B. Wang, and J. Zhang. A unified probabilistic framework for name disambiguation in digital library. *IEEE TKDE*, 24(6):975–987, 2012.

[4] J. Tang, S. Wu, J. Sun, and H. Su. Cross-domain collaboration recommendation. In *KDD'12*, pages 1285–1294, 2012.

[5] J. Tang, L. Yao, D. Zhang, and J. Zhang. A combination approach to web user profiling. *ACM TKDD*, 5(1):1–44, 2010.

[6] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su. Arnetminer: Extraction and mining of academic social networks. In *KDD'08*, pages 990–998, 2008.

[7] Z. Wang, J. Li, Z. Wang, and J. Tang. Cross-lingual knowledge linking across wiki knowledge bases. In *WWW'12*, pages 459–468, 2012.

[8] J. Zhang, J. Tang, C. Ma, H. Tong, Y. Jing, and J. Li. Panther: Fast top-k similarity search on large networks. In *KDD'15*, pages 1445–1454, 2015.

[9] Y. Zhang, J. Tang, Z. Yang, J. Pei, and P. Yu. Cosnet: Connecting heterogeneous social networks with local and global consistency. In *KDD'15*, pages 1485–1494, 2015.

[1] http://aminer.org