AMiner: Toward Understanding 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

In this talk, I present a novel academic search and mining system, **AMiner**[1], the second generation of the ArnetMiner system [9]. Different from traditional academic search systems that focus on document (paper) search, AMiner aims to provide a systematic modeling approach to gain a deep understanding of the large and heterogeneous networks formed by authors, papers they have published, and venues in which they were published. The system extracts researchers' profiles automatically from the Web [7] and integrates them with published papers after name disambiguation [3]. It has collected a large scholar dataset, with more than 130,000,000 researcher profiles and 100,000,000 papers from multiple publication databases. We have also developed an approach named COS-NET [12] to connect AMiner with several professional social networks, such as LinkedIn and VideoLectures, which significantly enriches the scholar metadata. Based on our integrated big scholar data, we devised a unified topic modeling approach to modeling the different entities (authors, papers, venues) simultaneously and providing a topic-level expertise search by leveraging the modeling results [8]. In addition, AMiner offers a set of *researcher-centered* functions, including social influence analysis [5], influence visualization [1], collaboration recommendation [6], relationship mining [4, 10], similarity analysis [11], and community evolution [2]. The system has been in operation since 2006 and has attracted more than 7,000,000 independent IP accesses from over 200 countries/regions.

Categories and Subject Descriptors

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

Keywords

Academic search; Recommendation; Integration; Social influence

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] L. Shi, H. Tong, J. Tang, and C. Lin. Vegas: Visual influence graph summarization on citation networks. *IEEE TKDE*, 2015.

[2] 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.

[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, T. Lou, J. Kleinberg, and S. Wu. Transfer link prediction across heterogeneous social networks. *ACM TOIS*, 2015.

[5] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *KDD'09*, pages 807–816, 2009.

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

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

[8] J. Tang, J. Zhang, R. Jin, Z. Yang, K. Cai, L. Zhang, and Z. Su. Topic level expertise search over heterogeneous networks. *Machine Learning Journal*, 82(2):211–237, 2011.

[9] 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.

[10] C. Wang, J. Han, Y. Jia, J. Tang, D. Zhang, Y. Yu, and J. Guo. Mining advisor-advisee relationships from research publication networks. In *KDD'10*, pages 203–212, 2010.

[11] 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.

[12] 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