AMiner: Toward Understanding Big Scholar Data

Jie Tang
Department of Computer Science and Technology, Tsinghua University
Tsinhua 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 \cite{AMiner2010}. 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 \cite{AMiner2010} and integrates them with published papers after name disambiguation \cite{Shi2010}. 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 COSNET \cite{Zhang2014} 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 \cite{Tang2015}. In addition, AMiner offers a set of researcher-centered functions, including social influence analysis \cite{Tang2013}, influence visualization \cite{Tang2013}, collaboration recommendation \cite{Tang2014b}, relationship mining \cite{Tang2013a, Tang2014a}, similarity analysis \cite{Tang2014}, and community evolution \cite{Tang2012}. 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.

Acknowledgements. The work is supported by the National High-tech R&D Program (2014AA015103), National Basic Research Program of China (2014CB340506, 2012CB316006), Natural Science Foundation of China (61222212), and a research fund supported by Huawei Inc.

2. REFERENCES
\begin{thebibliography}{12}
\end{thebibliography}

\footnotetext[1]{{http://aminer.org}}