

# Scalable OLAP and Mining of Information Networks\*

Jiawei Han  
Univ. of Illinois at Urbana-Champaign  
hanj@cs.uiuc.edu

Xifeng Yan  
Univ. of California at Santa Barbara  
xyan@cs.ucsb.edu

Philip S. Yu  
University of Illinois at Chicago  
psyu@cs.uic.edu

## ABSTRACT

With the ubiquity of information networks and their broad applications, there have been numerous studies on the construction, online analytical processing, and mining of information networks in multiple disciplines, including social network analysis, World-Wide Web, database systems, data mining, machine learning, and networked communication and information systems. In this tutorial, we present an organized picture on scalable OLAP (online analytical processing) and mining of information networks, with the inclusion of the following topics: (1) an introduction to information networks and information network analysis, (2) general statistical behavior of information networks, (3) mining frequent subgraphs in large graphs and networks, (4) data integration, data cleaning and data validation in information networks, (5) clustering graphs and information networks, (6) classification of graphs and information networks; (7) summarization and simplification of graphs and information networks, (8) OLAP and multidimensional analysis of information networks, (9) evolution of dynamic information networks, and (10) research challenges on OLAP and mining of information networks.

## About the Instructor

**Jiawei Han** is a professor in the Department of Computer Science, University of Illinois at Urbana-Champaign. He has been working on research into data mining, data warehousing, stream data mining, spatial and multimedia data mining, and bio-medical data mining, with over 350 conference and journal publications. He has chaired or served in over 100 program committees of international conferences and workshops. He is serving as Editor-in-Chief of ACM Trans. on Knowledge Discovery from Data, and also served or serv-

\*The work was supported in part by the U.S. National Science Foundation grants IIS-08-42769 and BDI-05-15813, and NASA grant NNX08AC35A. Any opinions, findings, and conclusions expressed here are those of the authors and do not necessarily reflect the views of the funding agencies.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the ACM. To copy otherwise, or to republish, to post on servers or to redistribute to lists, requires a fee and/or special permissions from the publisher, ACM.

*EDBT 2009*, March 24–26, 2009, Saint Petersburg, Russia.  
Copyright 2009 ACM 978-1-60558-422-5/09/0003 ...\$5.00

ing on the editorial boards for Data Mining and Knowledge Discovery, IEEE Trans. on Knowledge and Data Engineering, Journal of Intelligent Information Systems, and Journal of Computer Science and Technology. He is currently serving on the Board of Directors for the Executive Committee of ACM SIGKDD. Jiawei has received the Outstanding Contribution Award at the 2002 Int. Conf. on Data Mining, ACM SIGKDD Innovations Award (2004), and IEEE CS Technical Achievement Award (2005). He is a Fellow of ACM and IEEE. His book “Data Mining: Concepts and Techniques” 2nd ed., (Morgan Kaufmann, 2006), has been popularly used as a textbook worldwide.

**Xifeng Yan** is an assistant professor at the University of California at Santa Barbara and also holds the Venkatesh “Venky” Narayanamurti Chair. He received a PhD degree in Computer Science from the University of Illinois at Urbana-Champaign in 2006. He was a research staff member at the IBM T. J. Watson Research Center between 2006 and 2008. Dr. Yan’s research interests include data mining, databases, and bioinformatics. He has filed 6 patents and published more than 40 papers in refereed journals and conferences. Dr. Yan received the 2007 ACM SIGMOD Doctoral Dissertation Runner-Up Award for his work in graph mining and graph data management.

**Philip S. Yu** is a Professor in the Department of Computer Science at the University of Illinois at Chicago and also holds the Wexler Chair in Information Technology. He was manager of the Software Tools and Techniques group at the IBM Thomas J. Watson Research Center. Dr. Yu is a Fellow of the ACM and the IEEE. He served as the Editor-in-Chief of IEEE Trans. on Knowledge and Data Engineering (2001-2004). He is an associate editor of ACM Trans. on Knowledge Discovery from Data and also ACM Trans. of the Internet Technology. He serves on the steering committee of IEEE Int. Conf. on Data Mining. He was a member of the IEEE Data Engineering steering committee. Dr. Yu received a Research Contributions Award from IEEE Int. Conf. on Data Mining in 2003. His research interests include data mining and database systems. He has published more than 530 papers in refereed journals and conferences. He holds or has applied for more than 300 US patents. Dr. Yu is an IBM Master Inventor.