

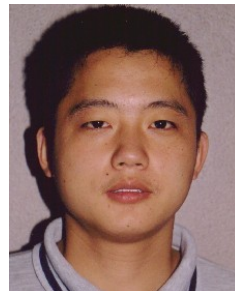
DATA MINING WITH GRAPHS AND MATRICES



Chris Ding



Fei Wang



Tao Li

Chris Ding (Department of Computer Science and Engineering, University of Texas at Arlington, USA)

Fei Wang (Department of Statistical Science, Cornell University, USA)

Tao Li (Florida International University, USA)

Abstract:

The field of data mining and machine learning increasingly adapt methods and algorithms from advanced matrix computations, graph theory and optimization. Prominent examples are spectral clustering, non-negative matrix factorization, Principal Component Analysis (PCA) and Singular Value Decomposition (SVD), graph-Laplacian based semi-supervised learning, diffusion process, etc. Graph and matrix-based methods are rapidly becoming popular and significant in information and knowledge management for the following reasons: (1) Graph and matrix based methods are amenable to vigorous analysis and benefits from the well established knowledge in matrix computations, graph theory and optimization; (2) Compared to probabilistic and information theoretic approaches, graph and matrix based methods are fast, easy to understand and implement; and (3) Graph and matrix based methods are especially suitable for parallel and distributed-memory computers to solve large scale challenging problems such as searching and extracting patterns from the entire Web.

This tutorial will present recent advances in algorithms and methods using graphs and matrices for modeling and analyzing massive, high-dimensional, and nonlinear-structured data. One main goal of the tutorial is to consolidate the recent ideas on information and knowledge management using graphs and matrices. We will summarize some open problems contained in this field and

propose some future trends. We also wish to attract practitioners who seek novel ideas for applications.