

**Presents the Fall 2012 EECS Seminar Series**

**Dr. Rong Jin**  
Michigan State University

**“Crowdclustering with Sparse Pairwise Labels: A Matrix Completion Approach”**  
Friday, October 19, 2012 • 2:00 p.m. • HEC 117

**ABSTRACT**

Crowdsourcing utilizes human ability by distributing tasks to a large number of workers. It is especially suitable for data clustering because it measures similarity between objects based on manual annotations, capturing the human perception of similarity among objects. This is in contrast to most clustering algorithms that face the challenge of finding an appropriate similarity measure for the given dataset. The main challenge of crowdclustering arises from the noisy nature of human annotations. In this talk, I will describe our work on crowdclustering that addresses the noisy annotation problem by the technique of matrix completion. I will show, both theoretically and empirically, that the proposed approach works well even when the majority of manual annotations are noisy. I will also describe our recent efforts for crowdclustering in a more challenging scenario, where most objects do not receive manual annotations. We extend our matrix completion framework by incorporating the techniques of semi-supervised clustering and robust distance metric learning.

**BIOGRAPHY**

Dr. Jin is an associate professor in the Department of Computer Science and Engineering at Michigan State University. His research is focused on statistical machine learning and its application to information retrieval. He has worked on a variety of machine learning algorithms/theories and their application to a wide range of applications, including information retrieval, collaborative filtering, document clustering, and visual object recognition. He has published over 180 conference and journal articles on related topics. Dr. Jin holds a Ph.D. in Computer Science from Carnegie Mellon University in 2003. He received the NSF Career Award in 2006, and the best student paper of COLT in 2012.