# COMPUTER SCIENCE SEMINAR

## *Computational Phenotyping using Tensor Factorization and Tensor Network*

### Dr. Jimeng Sun
### Georgia Institute of Technology

**Abstract:** Computational phenotyping is the process of converting heterogeneous electronic health records (EHRs) into meaningful clinical concepts (phenotypes). Tensor factorization has been shown as a successful unsupervised approach for discovering phenotypes. However, tensor methods have some major limitations for phenotyping:
1) unable to incorporate existing medical knowledge;
2) fail to handle high-order tensors (e.g., order ¿ 5) .
We will talk about two of our recent developments in addressing these challenges:

First, we proposed Rubik, a constrained non-negative tensor factorization and completion method for phenotyping. Rubik incorporates 1) guidance constraints to align with existing medical knowledge, and 2) pairwise constraints for obtaining distinct, non-overlapping phenotypes. Rubik also has built-in tensor completion that can significantly alleviate the impact of noisy and missing data. We evaluate Rubik on two large EHR datasets. Our results show that Rubik can discover more meaningful and distinct phenotypes than the baselines.

Second, we extended a theoretical framework called tensor networks for analyzing high-order tensors. We developed an efficient sparse hierarchical Tucker model (Sparse H-Tucker) for finding interpretable tree-structured factorizations from sparse high-order tensor. Sparse H-Tucker scales nearly linearly in the number of non-zero tensor elements. We applied Sparse H-Tucker on a real EHR dataset for learning a disease hierarchy. The resulting tree structure provides an interpretable disease hierarchy, which is confirmed by a clinical expert.

Bio Jimeng Sun is an Associate Professor of School of Computational Science and Engineering at College of Computing in Georgia Institute of Technology. Prior to joining Georgia Tech, he was a research staff member at IBM TJ Watson Research Center. His research focuses on health analytics using electronic health records and data mining, especially in designing novel tensor analysis and similarity learning methods and developing large-scale predictive modeling systems. He has published over 80 papers, filed over 20 patents (5 granted). He has received ICDM best research paper award in 2008, SDM best research paper award in 2007, and KDD Dissertation runner-up award in 2008. Dr. Sun received his B.S. and M.Phil. in Computer Science from Hong Kong University of Science and Technology in 2002 and 2003, and PhD in Computer Science from Carnegie Mellon University in 2007.

### Friday, April 22, 2016, 3:00 pm
### Mathematics and Science Center: W301

MATHEMATICS AND COMPUTER SCIENCE
EMORY UNIVERSITY