"Big Data Research: Methods, Systems, and Applications"

Abstract: We are in the era of “Big Data”, which is loosely defined as the application of data driven approaches to solve problems arising in a wide-range of domains in science, engineering, government, and business. Big Data holds the promise of allowing us to tackle problems at a scale, complexity, and fidelity that was previously impossible, enables us to achieve a deep understanding about the world around us, and revolutionize every aspect of our daily life.

In this talk, I present an overview of some recent work in my laboratory that spans various aspects of “Big Data” research including development of new algorithms, runtime systems, and applications of data analysis methods to emerging areas. On the algorithms side, the talk will focus on methods for nearest-neighbor recommender systems, on methods for analyzing dynamic relational networks towards finding patterns of relational co-evolution, on methods for partitioning and clustering networks on multi-core architectures, and on parallel methods for sparse tensor decomposition. On the systems side, the talk will focus on our work in developing runtime systems to allow the automated out-of-core execution of distributed memory message-passing programs, which provides a framework for solving very large problems on moderate size clusters and still achieve high-levels of computational performance. Finally, on the application side, the talk will present our work on employing “Big Data” approaches to analyze higher education data towards addressing issues related to academic pathways, effective pedagogy, retention, and persistence.

Bio: George Karypis is an ADC Chair of Digital Technology and Professor at the Department of Computer Science & Engineering at the University of Minnesota, Twin Cities. His research interests spans the areas of data mining, high performance computing, information retrieval, collaborative filtering, bioinformatics, cheminformatics, and scientific computing. His research has resulted in the development of software libraries for serial and parallel graph partitioning (METIS and ParMETIS), hypergraph partitioning (hMETIS), for parallel Cholesky factorization (PSPASES), for collaborative filtering-based recommendation algorithms (SUGGEST), clustering high dimensional datasets (CLUTO), finding frequent patterns in diverse datasets (PAFI), and for protein secondary structure prediction (YASSPP). He has coauthored over 250 papers on these topics and two books (“Introduction to Protein Structure Prediction: Methods and Algorithms” (Wiley, 2010) and “Introduction to Parallel Computing” (Publ. Addison Wesley, 2003, 2nd edition)). In addition, he is serving on the program committees of many conferences and workshops on these topics, and on the editorial boards of the IEEE Transactions on Knowledge and Data Engineering, ACM Transactions on Knowledge Discovery from Data, Data Mining and Knowledge Discovery, Social Network Analysis and Data Mining Journal, International Journal of Data Mining and Bioinformatics, the journal on Current Proteomics, Advances in Bioinformatics, and Biomedicine and Biotechnology.

Host: Prof. Jakub Szefer, Dept. of Electrical Engineering, Yale University