ISTeC

Colorado State University

The Information Science & Technology Center

ISTeC.ColoState.edu

Colorado State University's
Information Science and Technology Center (ISTeC)
presents two lectures by

Dr. David Bader

Georgia Institute of Technology
Professor, School of Computational
Science and Engineering

ISTeC Distinguished Lecture

in conjunction with the
Electrical and Computer Engineering Department and Computer Science Department Seminar Series

"Petascale Computing for Computational Biology and Genomics"

Monday, October 4, 2010
Reception: 10:30 a.m.
Lecture: 11:00 – 12:00 noon
Location: Lory Student Center Room 205

◆

Joint Electrical and Computer Engineering Department and Computer Science Department Special Seminar

sponsored by ISTeC

"Massive-scale Analysis of Streaming Social Networks"

Friday, October 1, 2010
Lecture: 9:00 – 10:00 a.m.
Location: Computer Science Room 425

ABSTRACTS

“Petascale Computing for Computational Biology and Genomics”

Graph theoretic problems are representative of fundamental kernels in traditional and emerging computational sciences such as chemistry, biology, and medicine, as well as applications in national security. Yet they pose serious challenges for parallel machines due to non-contiguous, concurrent accesses to global data structures with low degrees of locality. Few parallel graph algorithms outperform their best sequential implementation due to long memory latencies and high synchronization costs. In this talk, we consider several graph theoretic kernels for connectivity and centrality and discuss how the features of petascale architectures will affect algorithm development, ease of programming, performance, and scalability. Our large-scale graph algorithms are applied to real-world problems in phylogenetic reconstruction of evolutionary histories, inference of gene function in protein interaction networks, and cancer research.

“Massive-scale Analysis of Streaming Social Networks”

Emerging real-world graph problems include detecting community structure in large social networks, improving the resilience of the electric power grid, and detecting and preventing disease in human populations. Unlike traditional applications in computational science and engineering, solving these problems at scale often raises new challenges because of sparsity and the lack of locality in the data, the need for additional research on scalable algorithms and development of frameworks for solving these problems on high performance computers, and the need for improved models that also capture the noise and bias inherent in the torrential data streams. The explosion of real-world graph data poses a substantial challenge: How can we analyze constantly changing graphs with billions of vertices? Our approach leverages the Cray XMT's fine-grained parallelism and flat memory model to scale to massive graphs. On the Cray XMT, our static graph characterization package GraphCT summarizes such massive graphs, and our ongoing STINGER streaming work updates clustering coefficients on massive graphs at a rate of tens of thousands of updates per second.

SPEAKER BIOGRAPHY

David A. Bader (http://www.cc.gatech.edu/~bader/) is a Full Professor in the School of Computational Science and Engineering, College of Computing, at Georgia Institute of Technology. He received his Ph.D. in 1996 from The University of Maryland, was awarded a National Science Foundation (NSF) Postdoctoral Research Associateship in Experimental Computer Science. He is an NSF CAREER Award recipient, an investigator on several NSF and NIH awards, was a distinguished speaker in the IEEE Computer Society Distinguished Visitors Program, and a member of the IBM PERCS team for the DARPA High Productivity Computing Systems program. Bader currently serves on the NVIDIA-Cray team under the DARPA Ubiquitous High Performance Computing program. Dr. Bader directed the Sony-Toshiba-IBM Center of Competence for the Cell Broadband Engine Processor. Dr. Bader also serves on the Research Advisory Council for Internet2, the Steering Committees of the IPDPS and HiPC conferences, and is the General Chair of IPDPS 2010 and Chair of SIAM PP12. He is an associate editor for several high impact publications including the ACM Journal of Experimental Algorithmics (JEA), IEEE DSOnline, and Parallel Computing, has been an associate editor for the IEEE Transactions on Parallel and Distributed Systems (TPDS). He is an IEEE Fellow and a Member of the ACM. Dr. Bader's interests are at the intersection of high-performance computing and computational biology and genomics. He has co-chaired the IEEE International Workshop on High-Performance Computational Biology (HiCOMB), co-organized the NSF Workshop on Petascale Computing in the Biological Sciences, written several book chapters, and co-edited special issues of the Journal of Parallel and Distributed Computing (JPDC) and IEEE TPDS on high-performance computational biology. He has co-authored over 100 articles in peer-reviewed journals and conferences, and his main areas of research are in parallel algorithms, combinatorial optimization, and computational biology and genomics.

To arrange a meeting with the speaker, please contact MaryAnn Stroub at (970) 491-2708
or mstroub@engr.colostate.edu.

ISTeC (Information Science and Technology Center) is a university-wide organization for promoting, facilitating, and enhancing CSU’s research, education, and outreach activities pertaining to the design and innovative application of computer, communication, and information systems. For more information please see ISTeC.ColoState.edu.