Title: Cyberinfrastructure for Scalable Graph Execution

Principal Investigator: Jeanine Cook

Sponsor: National Science Foundation

Collaborators: Thomas Sterling, Louisiana State University and Agricultural and Mechanical College and Andrew Lumsdaine, Indiana University

Summary:
Graphs and graph algorithms have long been a fundamental abstraction in computer science. Important research areas that depend on graph algorithms include areas as diverse as knowledge discovery, bioinformatics, proteomics and genomics, data mining and search, electronic design automation, computer vision, and Internet routing. Although many graph analysis tasks are today performed sequentially, problem sizes continue to grow, necessitating the use of parallel computers. Parallelization of graph algorithms is difficult; the paradigms, software, and hardware that have worked well for developing mainstream parallel scientific applications are not well matched to large-scale graph problems.

The goal of this research is to enable and facilitate efficient execution of dynamic graph applications on current terascale and petascale systems, as well as on future exascale systems. Effectively addressing the shortcomings of existing approaches requires a complete re-thinking of the entire system stack. Accordingly, we use a comprehensive and integrated hardware and software system for scalable graph execution. The hardware will be co-designed with the software and will include the following important capabilities:

- *graph library* containing key algorithms, common graph execution kernels, and a user API specifically targeted at graph-based problems and scalable graph-based computations
- *new parallel programming model* to express fine-grained parallelism and expose sufficient and meaningful opportunities for scalable performance
- *graph execution platform* that realizes the new parallel programming model and provides all necessary system functionality to enable scalable performance
- *new graph-optimized core architecture* developed to support efficient parallel graph analysis constructs that cannot be found on today's commercial or HPC processors

Work will be informed by the needs of real application efforts with which researchers have ongoing collaborations. This research will enable and fundamentally improve scalability and user productivity for graph-based applications. The research focuses on the programming challenges of constructing correct, reliable, and optimal graph application programs while achieving efficient (with respect to time and energy), scalable and reliable execution on terascale, petascale, and exascale systems. This project will provide a fundamental, new understanding of the basics of large-scale graph processing, how it differs from conventional processing, and how to build scalable systems to efficiently solve large-scale graph problems. In particular, objectives for this work include characterization of processing overheads and the manners in which they limit graph processing scalability, performance models and data models that properly capture graph algorithms and associated processing overheads, the co-design process for developing graph-specific hardware, and experimental verification with a full proof-of-concept graph execution environment.