Massive-scale Streaming Analytics:
Models, Parallelism, & Real-world Applications

Keynote Talk

David A. Bader
School of Computational Science and Engineering
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332 USA

ABSTRACT

Emerging real-world graph problems include: detecting and preventing disease in human populations; revealing community structure in large social networks; and improving the resilience of the electric power grid. Unlike traditional applications in computational science and engineering, solving these social problems at scale often raises new challenges because of the sparsity and lack of locality in the data, the need for research on scalable algorithms and development of frameworks for solving these real-world problems on high performance computers, and for improved models that capture the noise and bias inherent in the torrential data streams. Highlighting this keynote talk, Bader will discuss the opportunities and challenges in massive data-intensive computing for applications in social sciences, physical sciences, and engineering.

Focusing on parallel algorithm design and implementation, Bader formalizes a practical model for graph analysis on streaming data. In this model, a massive graph undergoes changes from an input stream of edge insertions and removals. The model supports concurrent updating of the graph while algorithms execute concurrently on the dynamic data structure. The talk introduces a concept of validity: an algorithm is valid if the output is correct for a graph consisting of the initial graph with some subset of concurrent changes. Practical examples of this model are given for valid implementations of breadth first search, connected components, PageRank, and triangle counting, all useful graph kernels in real-world applications. This is joint work with E. Jason Riedy and Chunxing Yin.

KEYWORDS

Graph Algorithms; Parallel Computing.

ACKNOWLEDGMENTS