Performance of Graph and Biological Analytics on the IBM Cell Broadband Engine Processor

David A. Bader
Tan M. Tran
Georgia Institute of Technology
**Background and Intent**

- To develop analytical schemes to identify similarities between sequences of symbols to assist computational biologists
  - Optimize for the IBM Cell Broadband Engine Processor
  - Find common genes and closely matches between sequences
- Five kernels
- **Methods:**
  - Local alignment: Water Smith dynamic programming (Kernel 1)
  - Searching for similarities (Kernels 2 and 3)
  - Global alignment (Kernel 4)
  - Multiple alignment: Center Star method (Kernel 5)
- Each kernel operates on either the original sequences, the results of the previous kernel, or both
- To be entirely integer and character based
Background and Intent

- To develop a scalable synthetic compact application that consists of four kernels requiring irregular access to a large, directed, weight multi-graph
  - Optimize for the IBM Cell Broadband Engine Processor
  - Untimed Scalable Data Generator generates the graph as tuples of vertex pairs and corresponding weights.

- Methods:
  - Cache-friendly adjacency lists (Kernel 1)
  - Parallel scan to classify larger sets of graph (Kernel 2)
  - Parallel BFS (Kernel 3)
  - Dijkstra Single-Source Shortest Path and Betweenness Centrality algorithms (Kernel 4)

- David A. Bader and Kamesh Madduri designed and implemented the first parallel Betweenness Centrality algorithm on symmetric multiprocessors.

- To be entirely integer and character based