Massive Graphs: Big Compute meets Big Data (SIAM AN12 Minisymposium)
Minneapolis, MN
9-13 July 2012

Massive Graphs: The Way Forward

Aydin Buluç, Lawrence Berkeley National Laboratory, USA

Large graph analytics have become an increasingly important in a wide variety of application areas such as internet search, bioinformatics, social media, and cybersecurity. Massive graphs push the state of the art in both big compute and big data. This presentation will collect and present the outstanding questions in this field that have been raised over the course of the mini-symposium.

- Dynamic graphs need attention too
- Need linear or sublinear algorithms
- sampling?
- however, if you depend on sampling, why do you have all that data in the first place?
- Partitioning != Community Detection
- Pragmatic concerns are important
- constants are important, Big-O ignores them
- integration with existing systems very important
- Programming model limits applicability
- Linear Algebraic approach great for problems it fits, not so much for others
- Active Pebbles can be a new model
- Some applications that don't fit BSP
- We're defining only one small layer (graph algorithms) of a full solution (solving some problem) and we must be mindful of what's involved in the rest.
- how to abstract the problem (fit user's problem into our abstractions)
- we are often making assumptions/approximations that may not be true
- Hardware heterogeneity is making a comeback
- GPU, specialized accelerators, XMT
- Graph500 and MPI/BSP friendly, but implementations still use custom versions
- Power usage is important.
- Industry (heavy industry, eg power, oil, etc) are not aware of graph theory
- Computer scientists are getting better at computation/software, but not as good at theory
- Method comparisons are hard: no good norm that says "this is good"
- Need good data generation for development
- API design is important.
- it needs to train the user on underlying theory
- should guide user to an efficient solution.