Gospel - High-performance graph analytics

Alberto Parravicini
2019-05-29
High-Performance Graph Analytics

Graphs are a gold mine of information
- Social Networks
- Financial transactions
- Recommender Systems

Real graphs are enormous
- Facebook: 2B users, Wikipedia: 200M links

We need high-performance and scalable ways to process graphs
Introducing Gospel

- High-performance heterogeneous architectures for graph analytics
- Make graph processing faster and readily available to researchers and industry

Quick examples:
1. Single-GPU PageRank on Wikipedia in 0.2 sec
2. Real-time Entity Linking with >80% accuracy
The Gospel Graph

GPU Algorithm Acceleration
- PageRank, Graph Visits, Embeddings

Framework/DSL extension
- Green-Marl by Oracle Labs

Embeddings & ML apps
- Entity Linking

ORACLE Labs
Approximating PageRank on GPU

- The original workhorse of Google’s search
- Computation time is a bottleneck, even with GPUs
- No need for 100% accuracy
  - The ranking is what matters!
- Leverage approximate computing
  - Low precision arithmetic, loop perforation, numerical tricks
PageRank on graphs larger than GPU memory

- Most real-world graphs are larger than GPU memory (e.g. the web!)
- Adapt PR to work in these cases
  - Graph partitioning
  - Data compression
  - Double buffering and pipelining
- These techniques can be adapted to many algorithms, and to a multi-GPU scenario
Accelerating embeddings primitives on GPU

- Vertex embeddings are key to perform machine learning on graphs
- Most algorithms have the same primitives: random walks, vertex sampling, etc...
  - Often done on CPU, and results sent to GPU
- Our direction:
  - Do everything on GPU, using modified Breadth-First Visit
Fast Entity Linking via Graph Embeddings (1/2)

● **Entity Linking (EL):** connect Named Entities to unique identities (e.g. Wikipedia Page)

“The Indiana Pacers and Miami Heat [...] meet at Miami's American Airlines Arena”

en.wikipedia.org/wiki/Indiana_Pacers  en.wikipedia.org/wiki/Miami

en.wikipedia.org/wiki/Miami_Heat  ../wiki/American_Airlines_Arena

● Lots of **applications:** search engines, recommender systems, chat bots
Fast Entity Linking via Graph Embeddings (2/2)

- The first EL algorithm to leverage graph embeddings
- SoA results (>80% accuracy) with real-time latency (30 names/sec)
Automatic GPU code generation from graph DSL

- Writing high-performance GPU graph algorithms is difficult
- We can extend Green-Marl, a graph DSL developed by Oracle Labs
  - Graph computation as linear algebra kernels
  - We leverage GraphBlast, GPU library from the authors of Gunrock
  - 2x-10x speedup w.r.t. 56-threads CPU
The Gospel Folks

Alberto Parravicini
Francesco Sgherzi
Elisa Tardini
Nicolò Scipione
Ivan Montalbano
Rolando Brondolin
Davide Bartolini
Rhicheek Patra
Marco Santambrogio

2019-05-29, Google, Mountain View
alberto.parravicini@polimi.it
• Approximating PageRank on GPU
• PageRank on graphs larger than GPU memory
• Accelerating embeddings primitives on GPU
• Fast Entity Linking via Graph Embeddings
• Automatic GPU code generation from Graph DSL

Thank you!
Gospel - High-performance graph analytics

Alberto Parravicini
2019-05-29
alberto.parravicini@polimi.it