The State of Graph Processing

APIs, Libraries, Benchmarks, and Programming Languages

Samuel Pollard

December 14, 2016

1 APIs and Libraries

- **Pregel.** API; developed by Google; distributed; vertex centric, Bulk-Synchronous Parallel model; inspired many other platforms such as Giraph and GPS; original paper; 2010.
- **GPS (Graph Processing System).** API; developed by Stanford; distributed; vertex centric, Bulk-Synchronous Parallel model; open source; similar to Pregel but with dynamic graph repartitioning and other enhancements. original paper; website; First appeared 2013; apprears inactive.
- **GraphX** Library.

2 Programming Languages

- **GP (Graph Programs).** Nondeterministic; serial; original paper: [6]; website: `https://www.cs.york.ac.uk/plasma/wiki/index.php?title=GP_(Graph_Programs)`; appears to be more theoretical and used for program verification; still active.
- **Gremlin.** functional, data flow; distributed; a way to interact with graph databases; [7]; website: `http://tinkerpop.apache.org/gremlin.html`.

3 Benchmarks

- **Graphalytics.** CPU and GPU; supports GraphMat, PowerGraph, GraphBIG, Giraph, GraphX, Neo4j, and MapReduce; [2]; still active; so far I can only get PowerGraph and GraphBIG running; website: `http://graphalytics.ewi.tudelft.nl`.
- **GAP (Graph Algorithm Platform).** CPU; shared Memory (OpenMP); `http://gap.cs.berkeley.edu/benchmark.html`; last active October 2016 on Github. [1]
- **GraphBIG.** CPU and GPU; shared Memory and CUDA; last active February 2016.
- **Lonestar.** CPU and GPU; shared memory and CUDA; part of Galois. First appeared 2011; last update appears to be in 2015 though I am in recent (December 2016) contact with someone working on the project.

4 Dynamic Graphs

- **STINGER.** Data structure and library; Georgia Institute of Technology and various national laboratories; streaming model; parallel or serial, distributed or shared; original paper; website; First appeared 2009; still active on Github.
- **GraphJet.** Java library; parallel but single-machine; original purpose was for real time recommendations from Twitter; `https://github.com/twitter/GraphJet`; [8].
- **GraphStream.** Java library; I'm not sure but it appears to be serial; `http://graphstream-project.org/`; [4] first active 2007; appears still active.
- **Kineograph.** API; parallel and distributed; [3]; 2012; appears to be inactive but influenced GraphX, GraphChi, and PowerGraph.
- **DynoGraph.** Performance Analysis; `https://github.com/sirpoovey/DynoGraph` [5]; connected components, PageRank, BFS; right now just a repo and a poster at SC'16.

5 Visualization

There are many graph visualization tools out there, most notably Graphviz (and its associated DOT file format). Likewise, databases such as Neo4j have their own visualization tools. This is concerned only with visualizations which are scalable to a large number of vertices and edges.

- **Gephi.** website

References

[1] Scott Beamer, Krste Asanovic, and David A. Patterson. The GAP benchmark suite. *CoRR*, abs/1508.03619, 2015.

[2] Mihai Capotă, Tim Hegeman, Alexandru Iosup, Arnau Prat-Pérez, Orri Erling, and Peter Boncz. Graphalytics: A big data benchmark for graph-processing platforms. In *Proceedings of the GRADES'15*, GRADES'15, pages 7:1–7:6, New York, NY, USA, 2015. ACM.

[3] Raymond Cheng, Ji Hong, Aapo Kyrola, Youshan Miao, Xuetian Weng, Ming Wu, Fan Yang, Lidong Zhou, Feng Zhao, and Enhong Chen. Kineograph: Taking the pulse of a fast-changing and connected world. In *Proceedings of the 7th ACM European Conference on Computer Systems*, EuroSys '12, pages 85–98, New York, NY, USA, 2012. ACM.

[4] Antoine Dutot, Frédéric Guinand, Damien Olivier, and Yoann Pigné. GraphStream: A Tool for bridging the gap between Complex Systems and Dynamic Graphs. In *Emergent Properties in Natural and Artificial Complex Systems. Satellite Conference within the 4th European Conference on Complex Systems*, EECS '07, Dresden, Germany, October 2007.

[5] Eric Hein and Tom Conte. Dynograph: Benchmarking dynamic graph analytics. SC '16, 2016.

[6] Detlef Plump. The graph programming language GP. In *Proceedings of the 3rd International Conference on Algebraic Informatics*, CAI '09, pages 99–122, Berlin, Heidelberg, 2009. Springer-Verlag.

[7] Marko A. Rodriguez. The gremlin graph traversal machine and language. *CoRR*, abs/1508.03843, 2015.

[8] Aneesh Sharma, Jerry Jiang, Praveen Bommannavar, Brian Larson, and Jimmy Lin. Graphjet: Real-time content recommendations at twitter. In *Proceedings of the VLDB Endowment.*