LDBC Graphalytics:
A Benchmark for Large-Scale Graph Analysis on Parallel and Distributed Platforms

Tim Hegeman
Distributed Systems Group, TU Delft

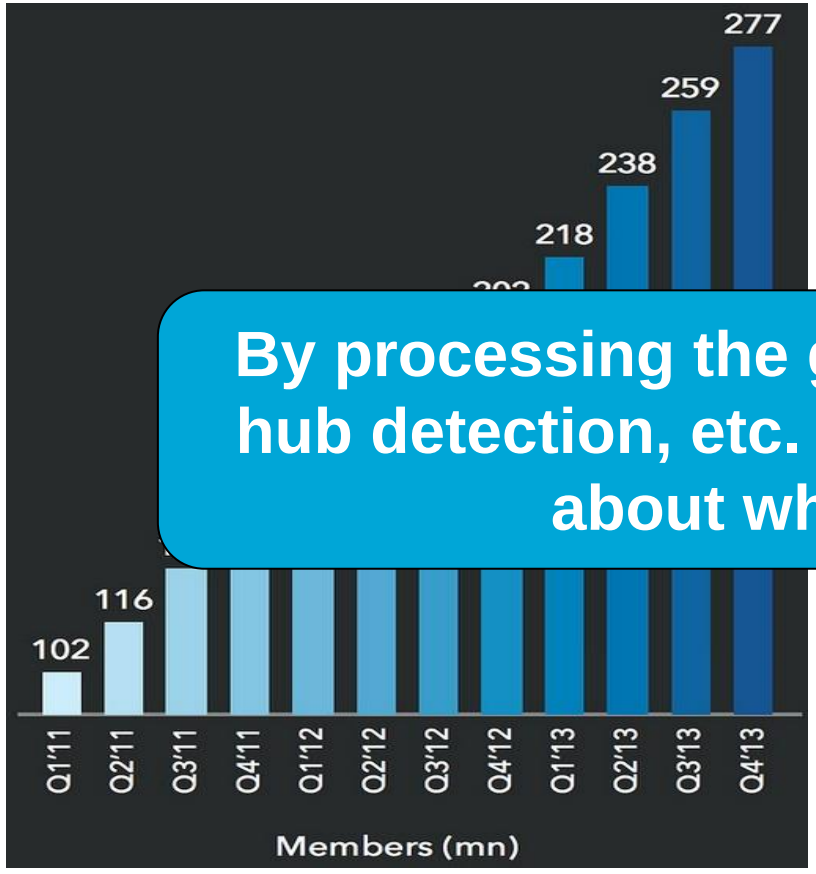
Published in PVLDB vol. 9.13 (2016)
Co-authored by Alexandru Iosup, Wing Lung Ngai, Stijn Heldens, Arnau Prat-Pérez, Thomas Manhardt, Hassan Chafi, Mihai Capotă, Narayanan Sundaram, Michael Anderson, Ilie Gabriel Tănase, Yinglong Xia, Lifeng Nai, Peter Boncz

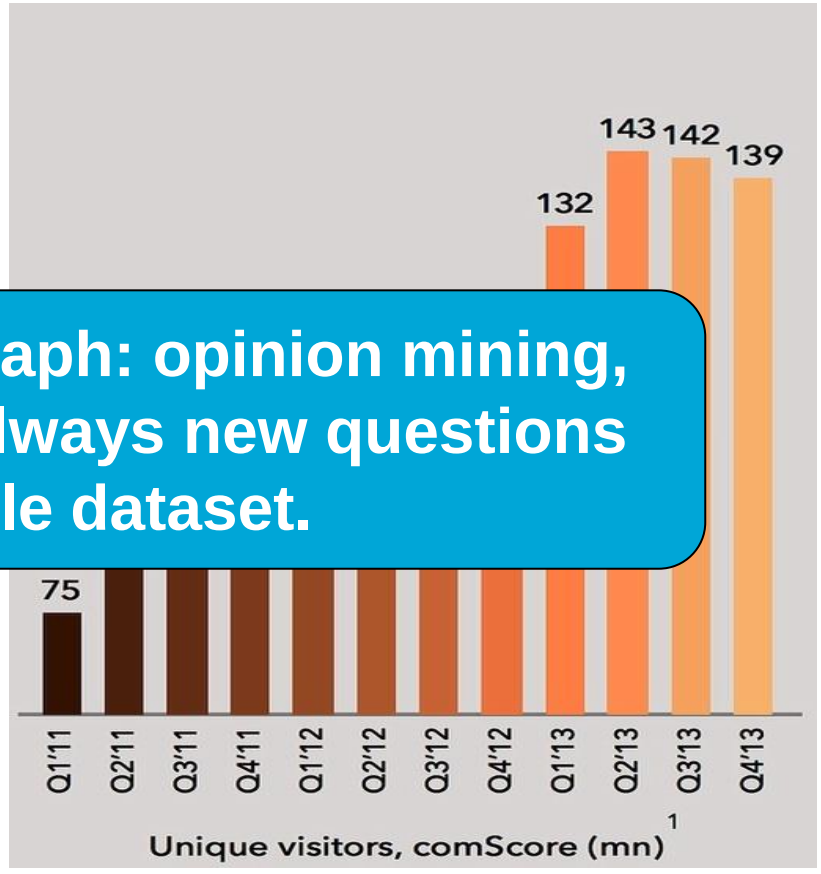
Graphs Are at the Core of Our Society: The LinkedIn Example

Sources: Vincenzo Cosenza, The State of LinkedIn, http://vincos.it/the-state-of-linkedin/
via Christopher Penn, http://www.shiftcomm.com/2014/02/state-linkedin-social-media-dark-horse/

TU Delft Delft University of Technology
VU VRIJE UNIVERSITEIT AMSTERDAM

LinkedIn's Service Analysis



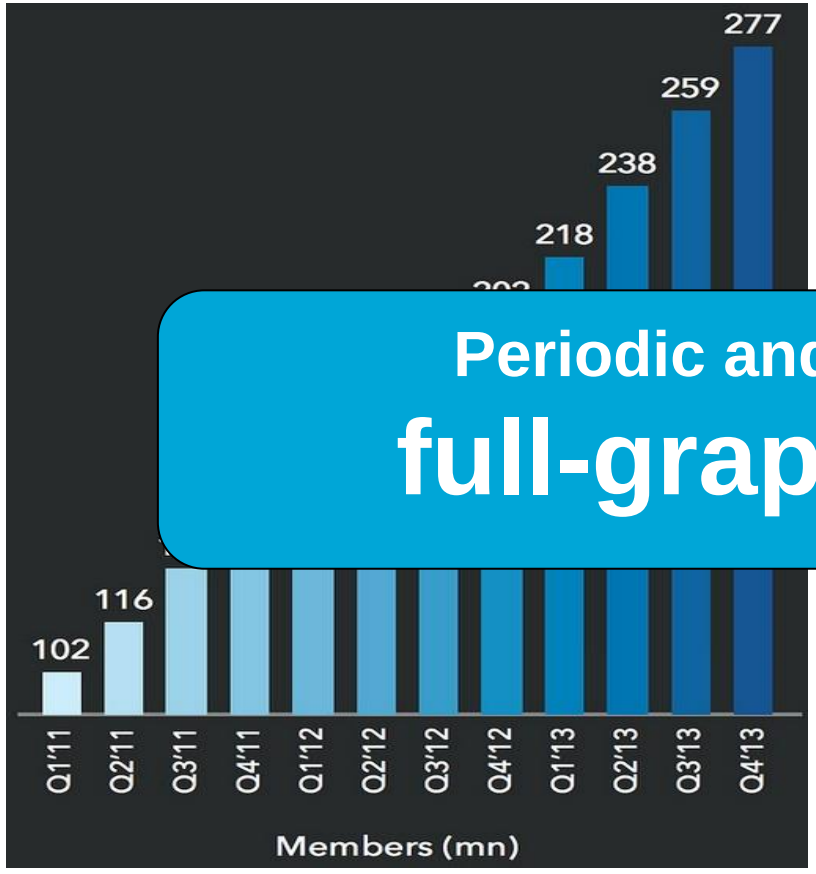


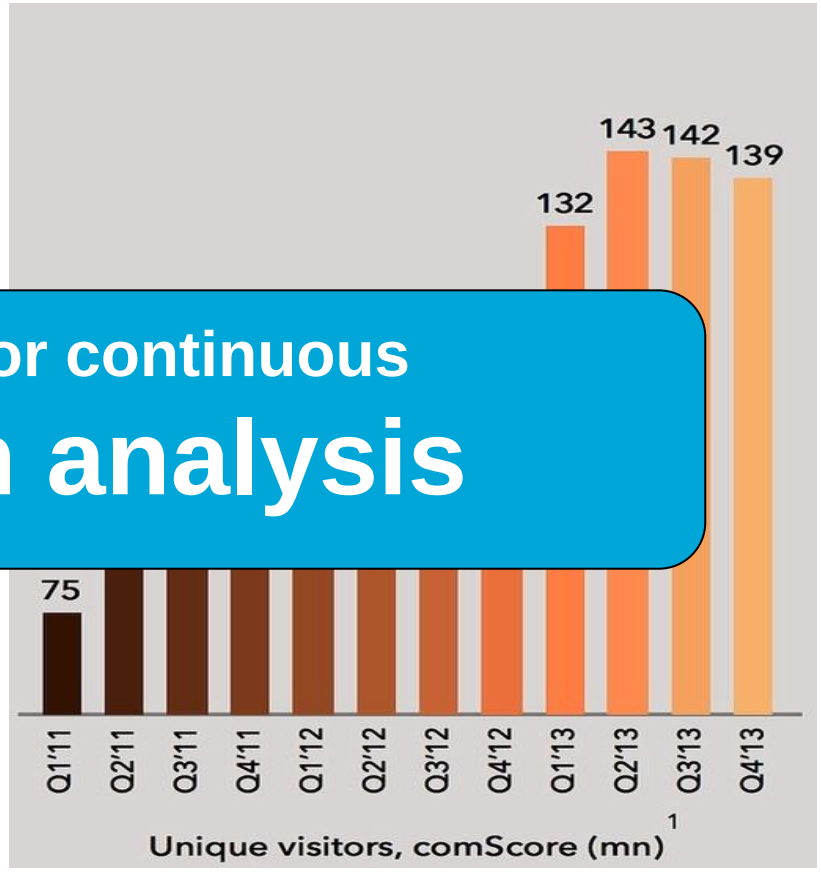
By processing the graph: opinion mining, hub detection, etc. Always new questions about whole dataset.

Sources: Vincenzo Cosenza, The State of LinkedIn, http://vincos.it/the-state-of-linkedin/
via Christopher Penn, http://www.shiftcomm.com/2014/02/state-linkedin-social-media-dark-horse/

TU Delft Delft University of Technology
VU VRIJE UNIVERSITEIT AMSTERDAM

LinkedIn's Service Analysis





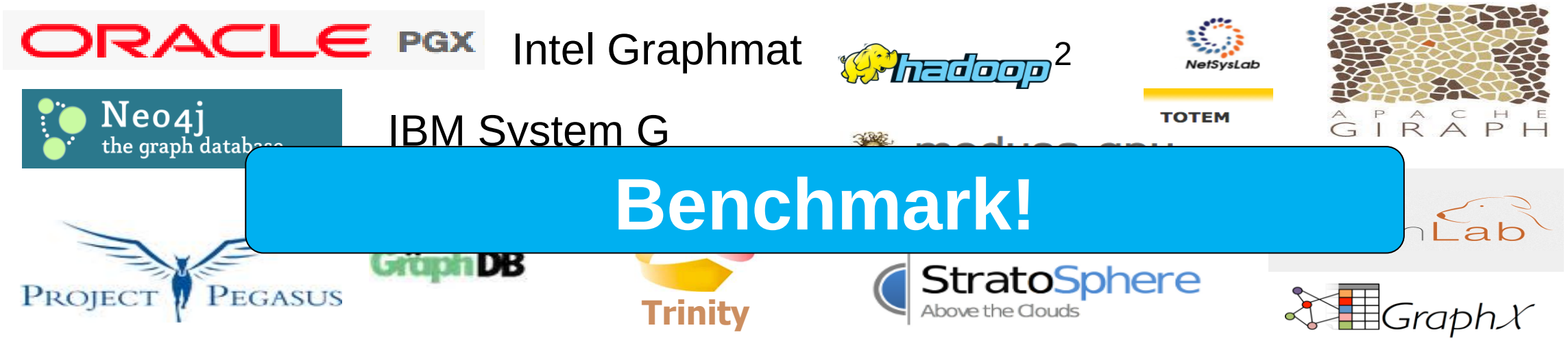
Periodic and/or continuous

full-graph analysis

Sources: Vincenzo Cosenza, The State of LinkedIn, http://vincos.it/the-state-of-linkedin/
via Christopher Penn, http://www.shiftcomm.com/2014/02/state-linkedin-social-media-dark-horse/

Graph Processing Platforms

Graph Processing Platforms



What Is the Performance of Graph Processing Platforms?

Metrics Diversity

Graph Diversity

Algorithm Diversity

- Graph500
  - Single application (BFS), Single class of synthetic datasets. @ISC16: future diversification.
- Few existing platform-centric comparative studies
  - Prove the superiority of a given system, limited set of metrics
- GreenGraph500, GraphBench, XGDBench
  - Issues with representativeness, systems covered, metrics, …

What Is the Performance of Graph Processing Platforms?

Metrics Diversity

Graph Diversity

Algorithm Diversity

Graphalytics = comprehensive benchmarking suite for graph processing across many platforms

Graphalytics, in a nutshell

- An LDBC benchmark
- Advanced benchmarking harness
- Many classes of algorithms used in practice
- Diverse real and synthetic datasets
- Diverse set of experiments representative for practice
- Renewal process to keep the workload relevant
- Extended toolset for bottleneck analysis
- Enables comparison of many platforms, community-driven and industrial

http://ldbcouncil.org/ldbc-graphalytics

TU Delft Delft University of Technology
VU VRIJE UNIVERSITEIT AMSTERDAM

Results: Experimental Setup (1)

Graphalytics has been implemented for 3 community-driven platforms (Giraph, GraphX, PowerGraph) and 3 industry-driven platforms (PGX, GraphMat, OpenG).

PGX

GraphMat

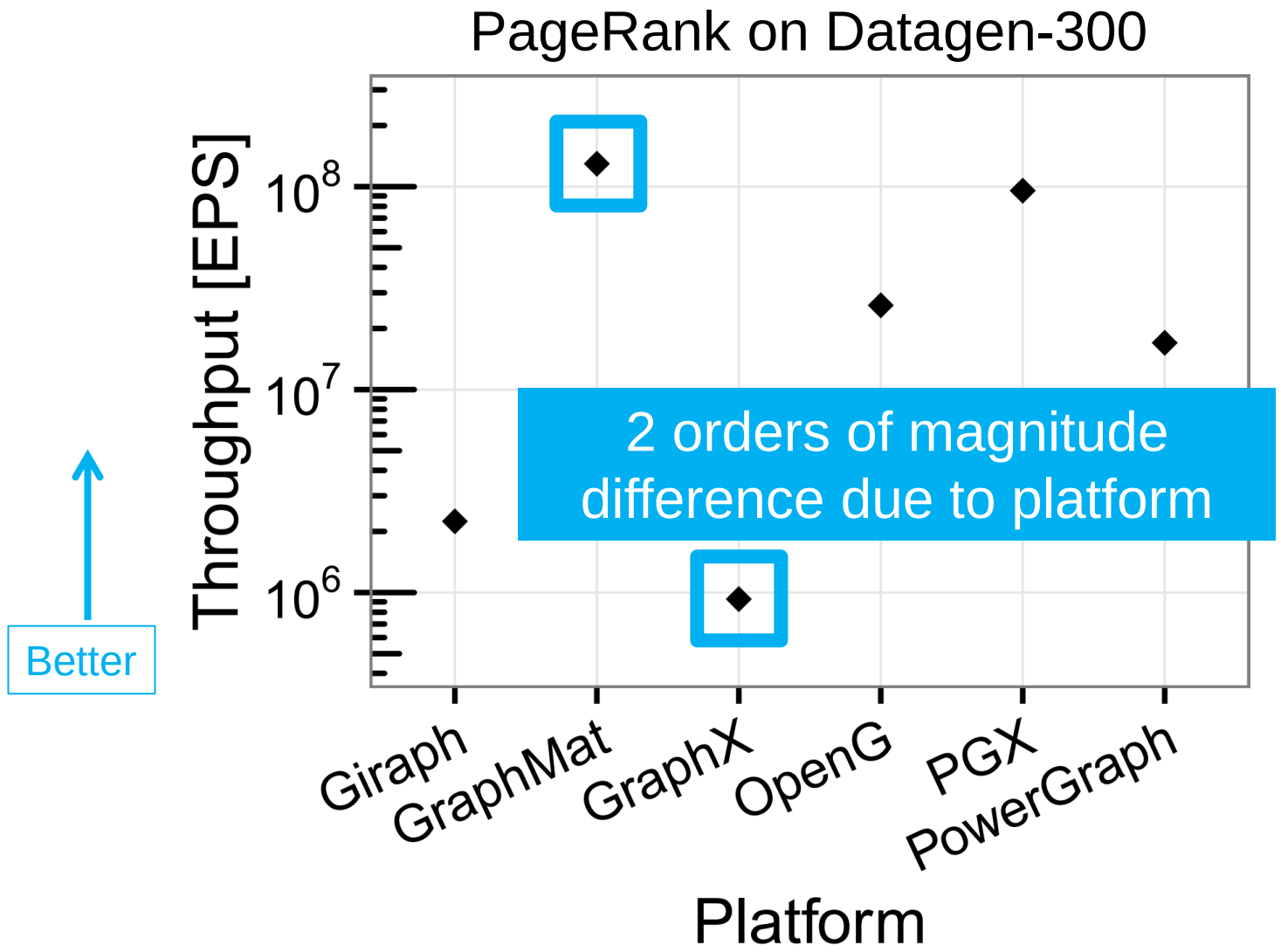
OpenG

Results: Experimental Setup (2)

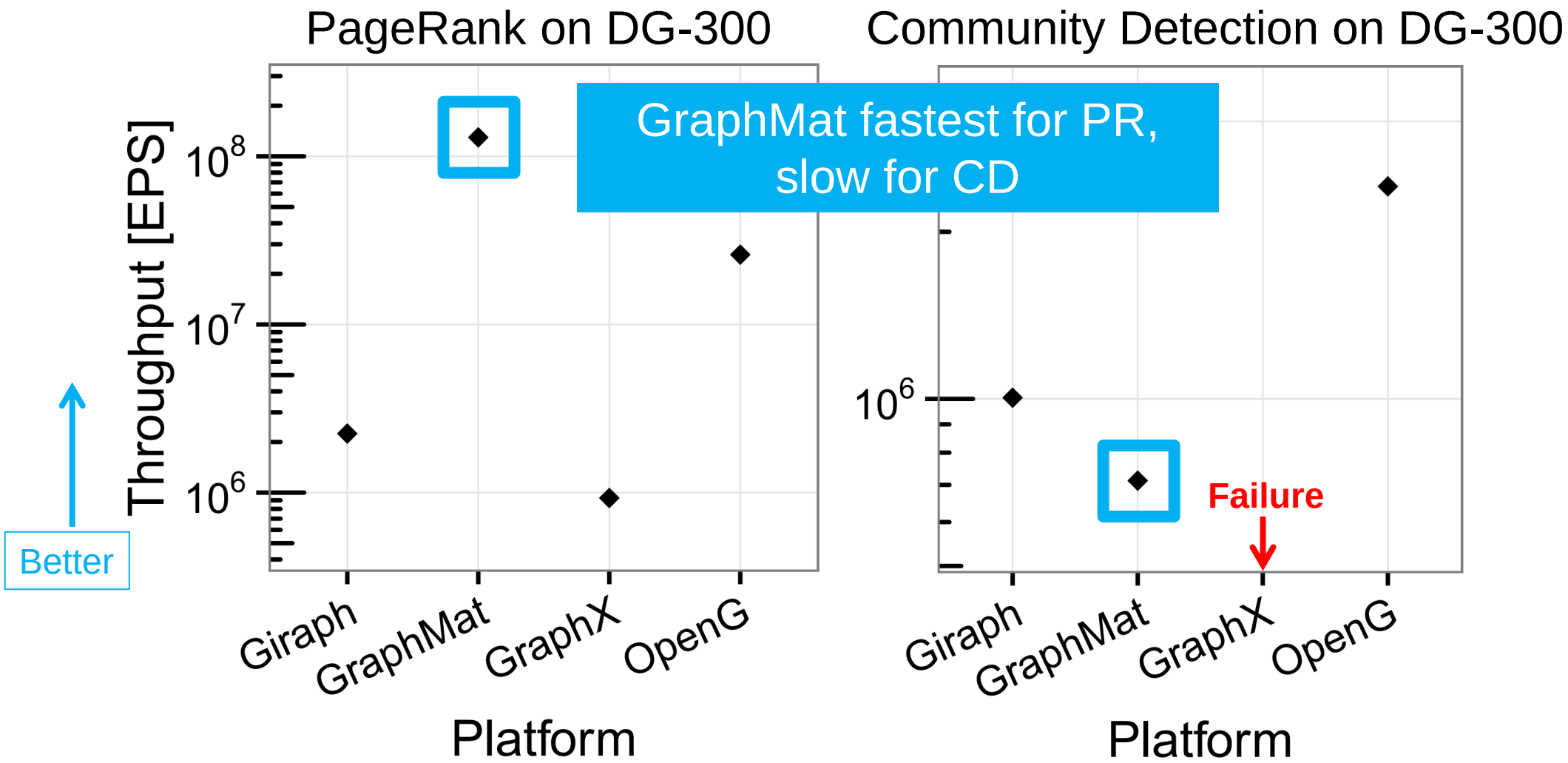
All experiments were performed by TU Delft on DAS-5 (Distributed ASCI Supercomputer, the Dutch national supercomputer for Computer Science research).

Environment: 1 machine (64GB, 2x8 cores)

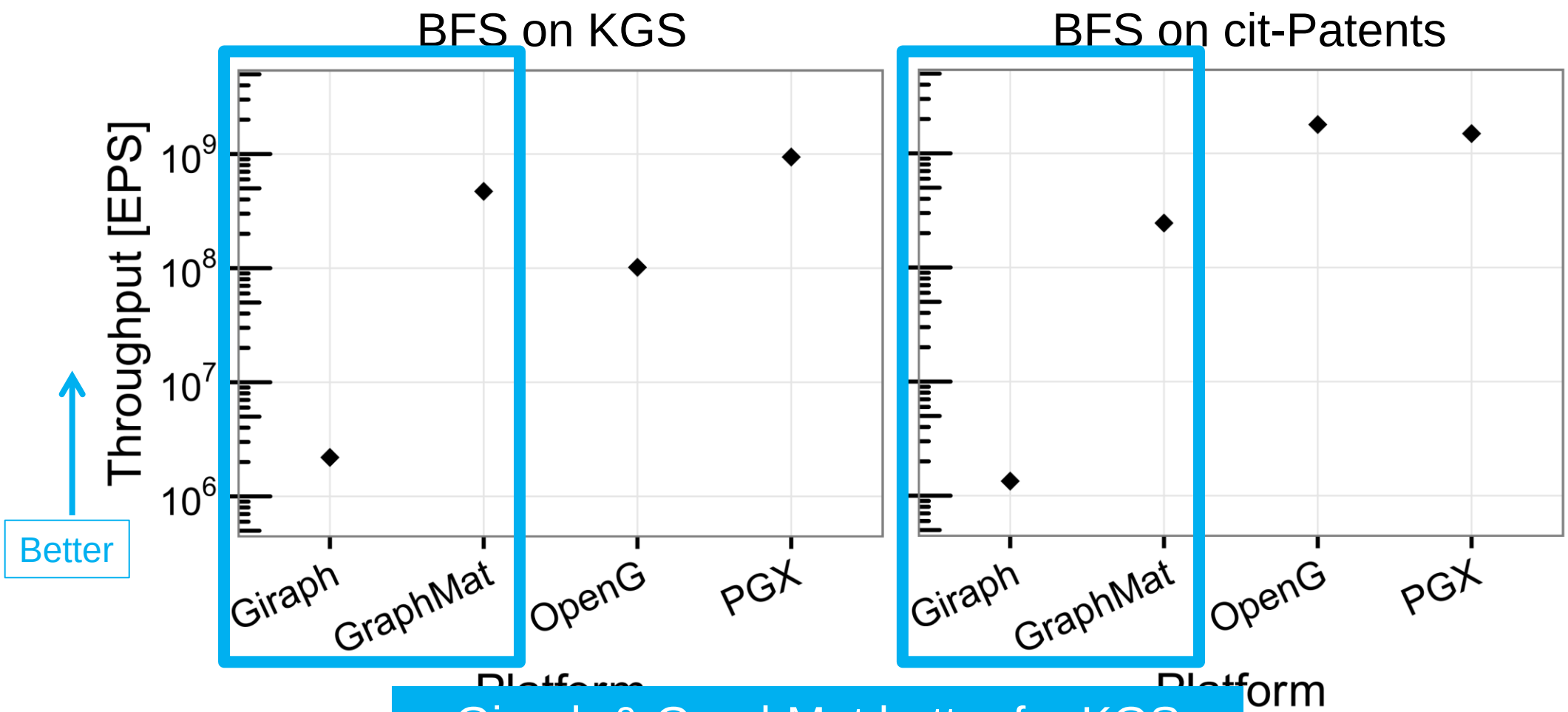
The Platform Has Large Impact



The Algorithm Has Large Impact

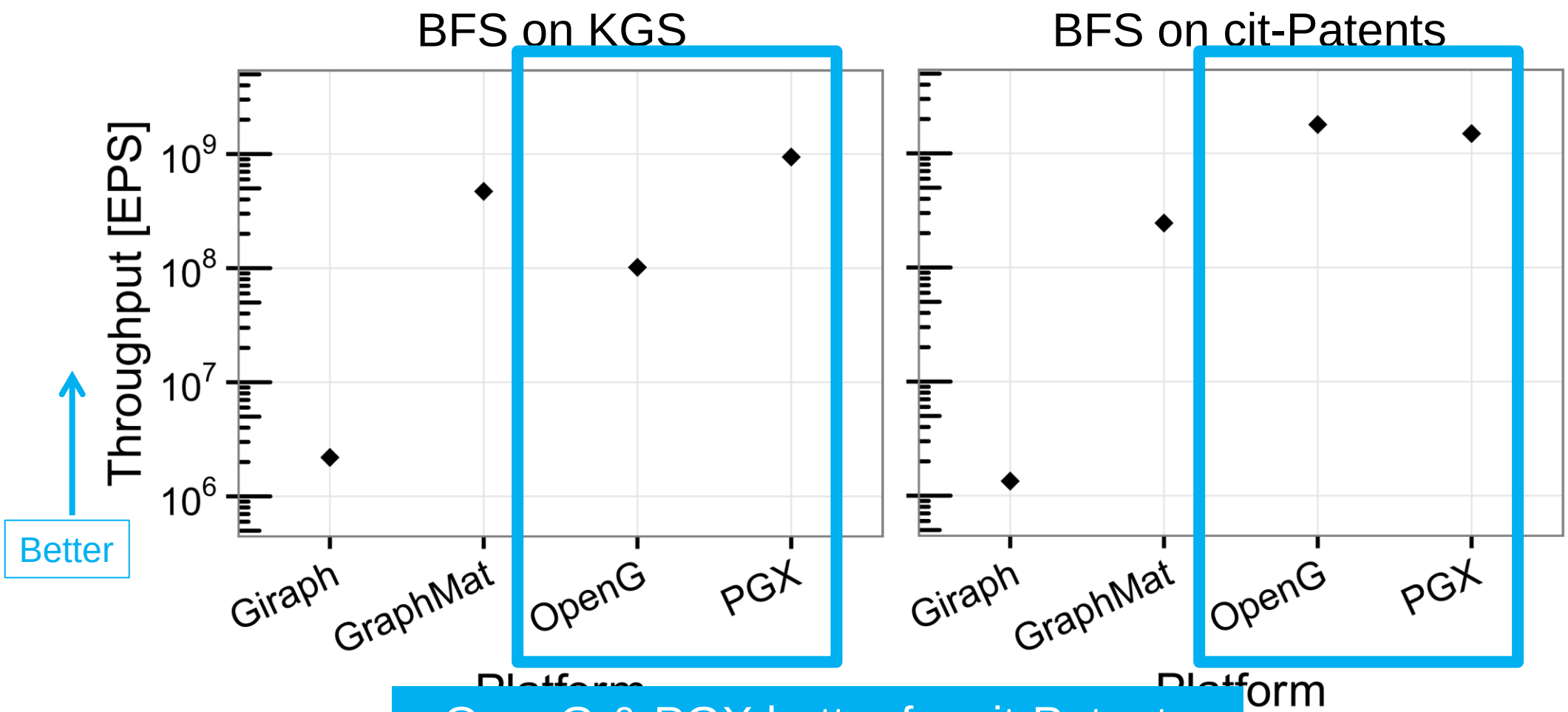


The Dataset Has Large Impact



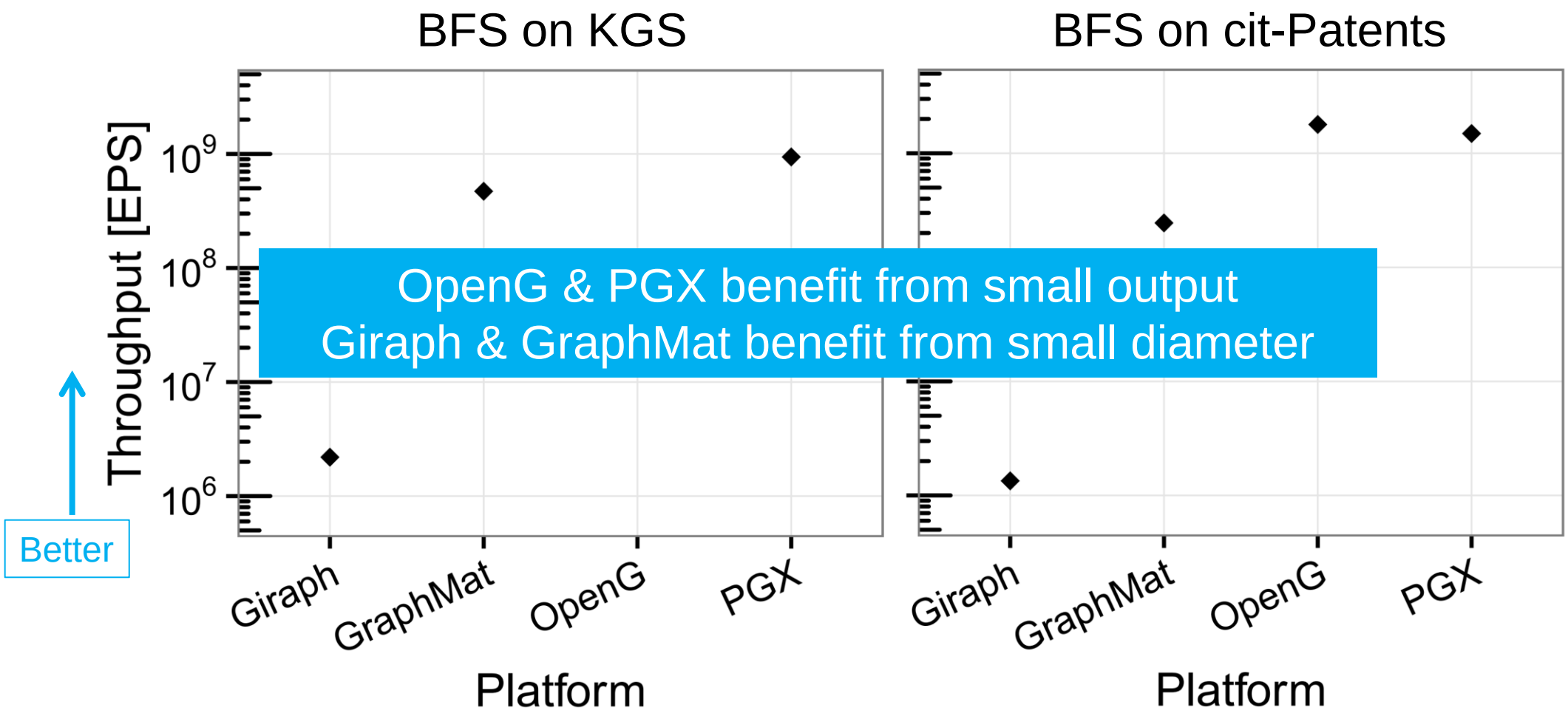
Giraph & GraphMat better for KGS

The Dataset Has Large Impact



OpenG & PGX better for cit-Patents

The Dataset Has Large Impact



Lessons learned

Performance of graph processing is a non-trivial function of
(Platform, Algorithm, Dataset, …), the P-A-D triangle

Understanding performance requires in-depth analysis
We are building tools for manual/automated bottleneck analysis

All current platforms can also have drawbacks
Ease-of-use/programmability of a platform is very important
Significant knowledge required to tune a system

Ongoing work

Finalizing version 1.0 of the benchmark specification using feedback from the community and our partners

Launching an online archive for sharing Graphalytics results

Organizing a global competition for graph-processing systems

Take-Home Message

Graphalytics: comprehensive benchmark and framework for comparing graph processing platforms

The P-A-D triangle:
Performance = f (Platform, Algorithm, Dataset, …)

Global competition and public archive of benchmark results are coming

Use Graphalytics to test your graph processing platform:
http://ldbcouncil.org/ldbc-graphalytics

LDBC Graphalytics:
A Benchmark for Large-Scale Graph Analysis on Parallel and Distributed Platforms

Tim Hegeman
Distributed Systems Group, TU Delft

Published in PVLDB vol. 9.13 (2016)
Co-authored by Alexandru Iosup, Wing Lung Ngai, Stijn Heldens, Arnau Prat-Pérez, Thomas Manhardt, Hassan Chafi, Mihai Capotă, Narayanan Sundaram, Michael Anderson, Ilie Gabriel Tănase, Yinglong Xia, Lifeng Nai, Peter Boncz