Graphitti:
Parallelization and Distribution for Large-Scale Graph Processing

Marcin Biczak, Yong Guo, Alexandru Iosup, Nefeli Papapetrou Lampraki, Ate Penders, Ana Lucia Varbanescu (TUDelft)
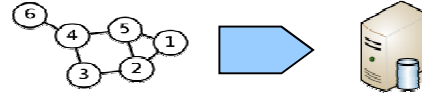
" Graph processing at any scale. "

Why do we need graph processing?
• **Understand, Manage, Operate**
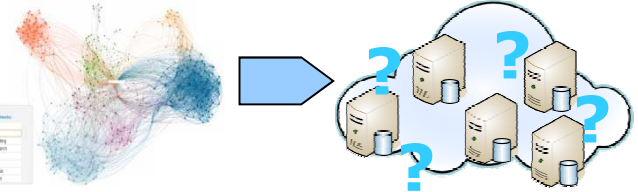• Social Networks, Gaming, Bioinformatics, Logistics, Maps

The problem

• **Now**

• **Future**

Graphitti, our solution

Use all parallelism in the infrastructure, at all scales. It's that simple.

• **Standard data format**
  - Share traces with the community
  - Use extensions for specific modeling aspects
  - Text-based, easy to parse for custom tasks
  - Additional SQL-compatible data format (GWF-SQLite)

• **Standard API and multiple back-ends**
  - Provide ready-to-use tools to the community
  - Ease-of-use

• **Infrastructure-, Algorithm-, and Data-dependent**
  - Towards best performance and minimal effort
  - Core computer science research: scheduling,
    monitoring, benchmarking, workload modeling, etc.
  - Co-design: data+algorithm+machine

Applications

• PageRank, other reputation
• Game Analytics
• Community Management
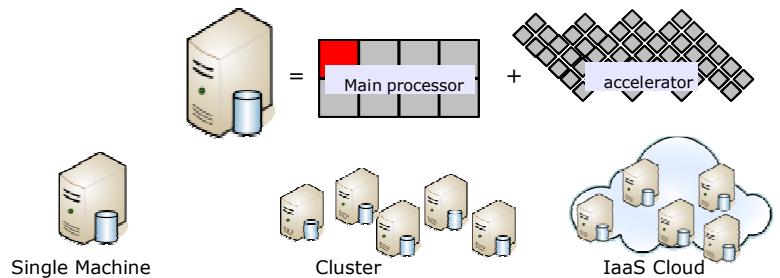• Person management
• Link-formation prediction

Algorithms

• Traversals: BFS, DFS, etc.
• All-source shortest path
• Logistics: Maxflow, etc.
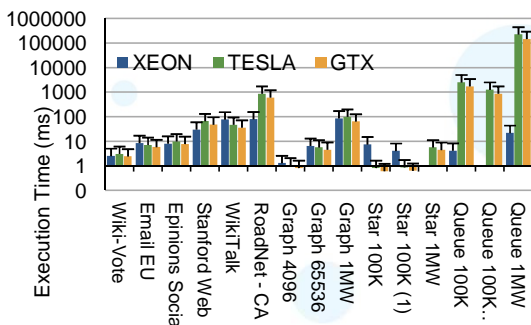• Community detection
• Betweenness centrality

Common Data Format

Common API

Multiple Back-ends

Single Machine   Cluster   IaaS Cloud

Name	Period	Size (GB)	Node (M)	Edge (M)	Category
KGS	2002/02-2009/03	2	0.8	27.4	Chess Game
FICS	1997/11-2011/09	168	0.4	144.2	Chess Game
BBO	2009/11-2009/12	10	3.9	12.9	Card Game
XFire	2008/05-2011/12	58	7.7	34.7	OMGN
Dota League	2006/07-2011/03	23	0.1	3.0	RTS
DotaLicious	2010/04-2012/02	1	0.1	0.6	RTS
Dota Garena	2009/09-2010/05	1	0.3	0.1	RTS
WoWAH	2006/01-2009/10	3	0.1	N/A	MMORPG

Timeline

Mar 2011 — Dec — Mar 2012 — Dec — Mar 2014

idea   group   online   Graphitti working

A.L.Varbanescu@tudelft.nl

TUDelft

Technische Universiteit Delft