IEEE/ACM CCGRID 2023

The 23rd IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing

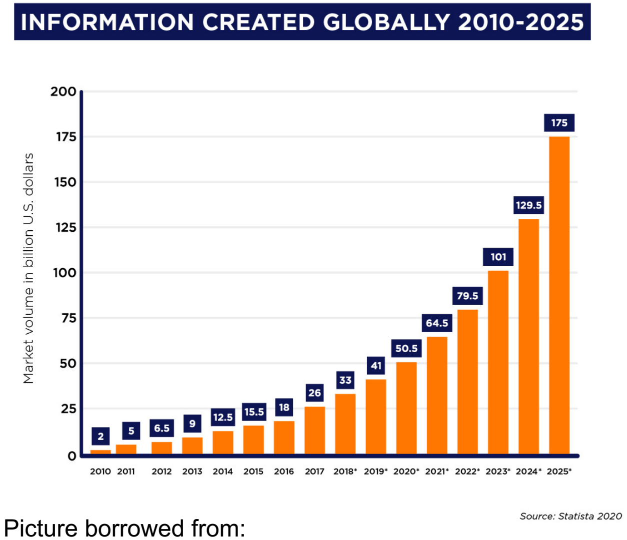
Highly Scalable Large-Scale Asynchronous Graph Processing using Actors

Youssef Elmougy*, Akihiro Hayashi, and Vivek Sarkar
Georgia Institute of Technology, Atlanta GA USA

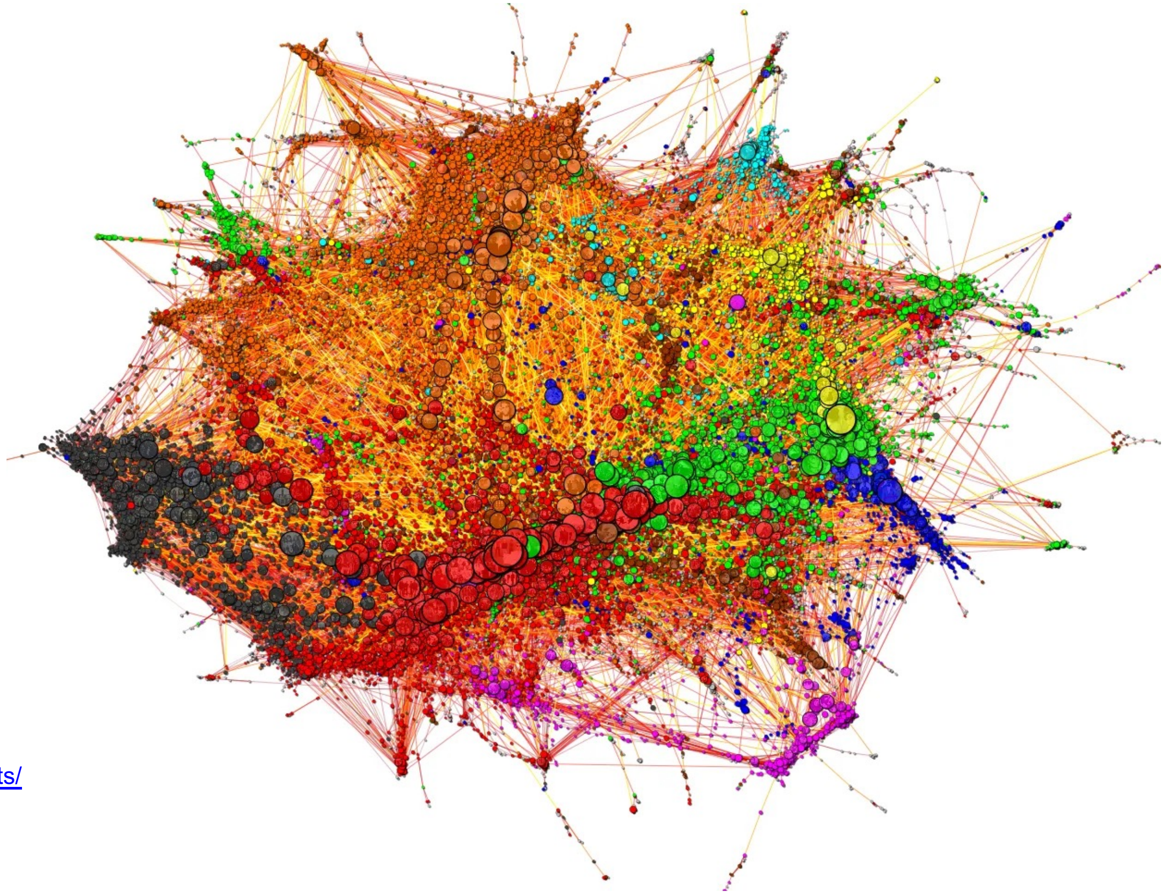
* Corresponding Author and Presenting Author

Georgia Tech

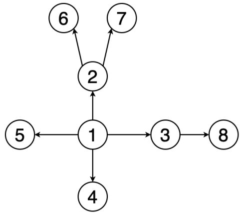
The Growth of Big Data

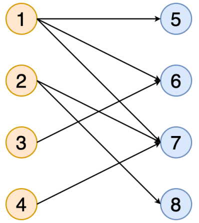


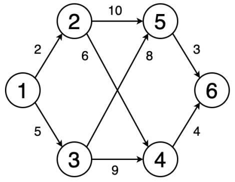
Picture borrowed from:
https://www.iteratorshq.com/blog/big-data-business-impacts/

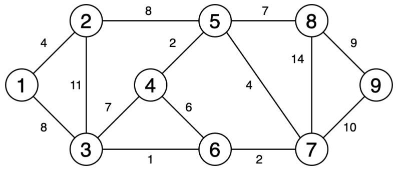


Graph algorithms have become increasingly important for solving problems in many computational domains





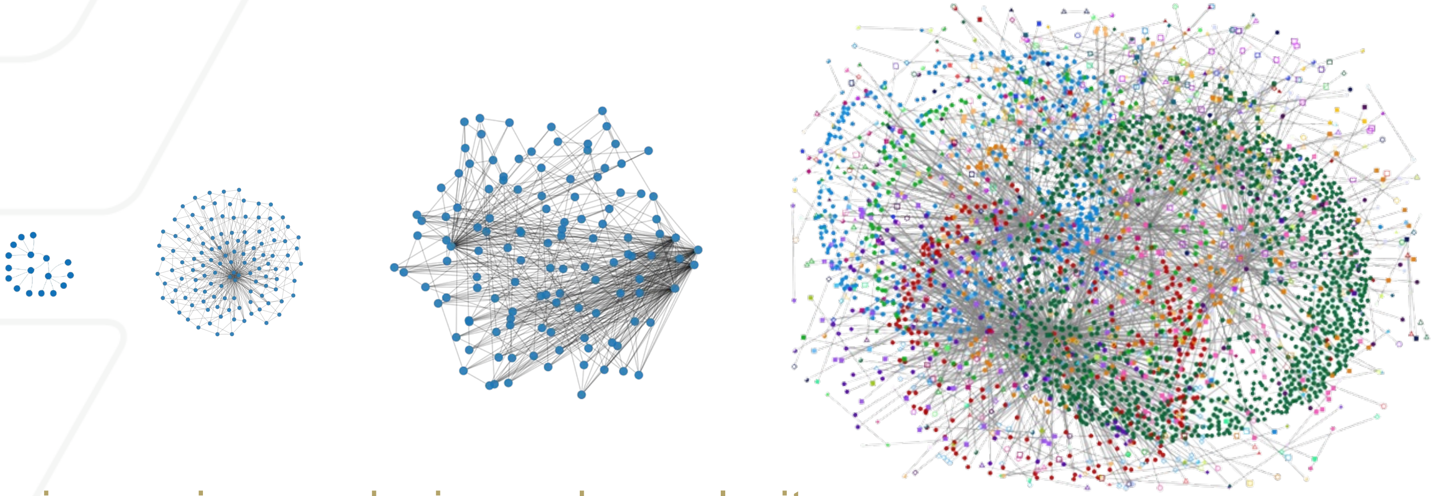


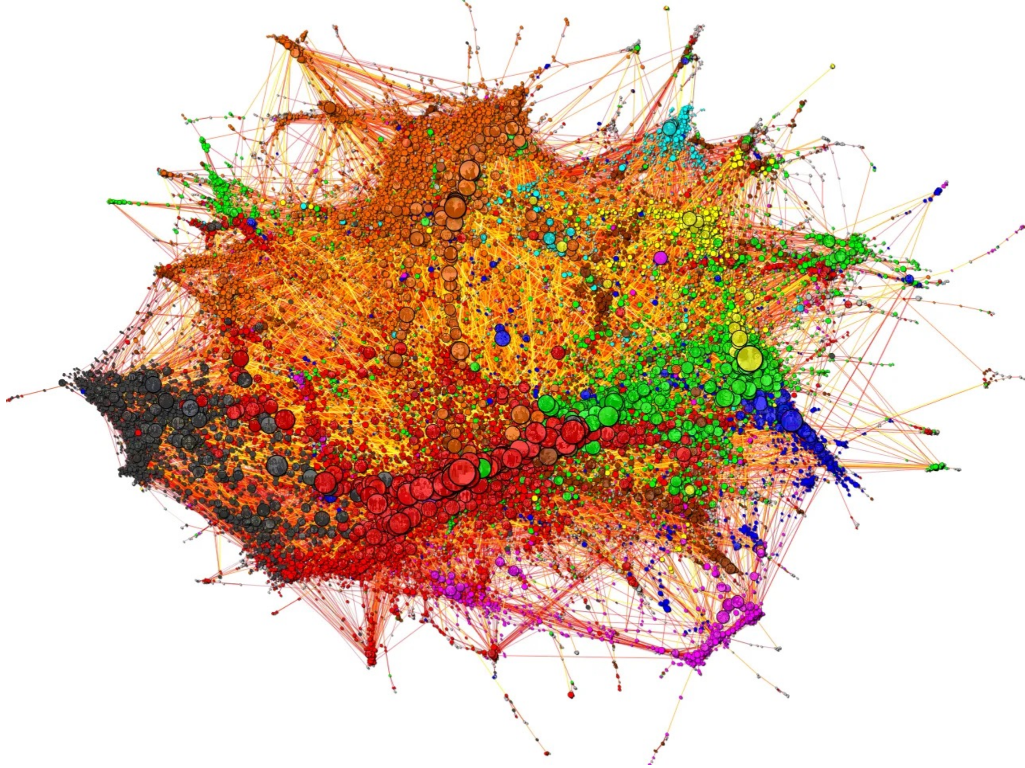


The scale of these graphs present difficulties to their processing and analysis!

Georgia Tech

The Scalability Problem





increasing graph size and complexity

- ❑ There is a need for parallel computing resources to meet the computational and memory requirements
- ❑ Existing algorithms and software that perform well for mainstream parallel scientific applications are not necessarily efficient for large-scale graph applications

It is critical to develop lightweight and scalable systems to efficiently process large-scale graphs!

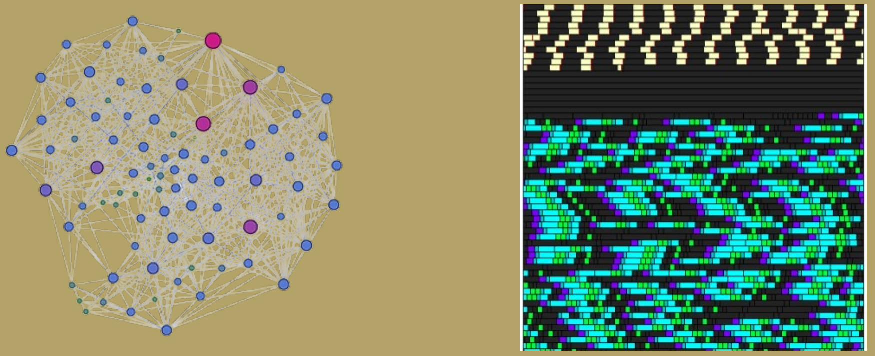
Georgia Tech

The 4 Overarching Challenges in Parallel Graph Processing

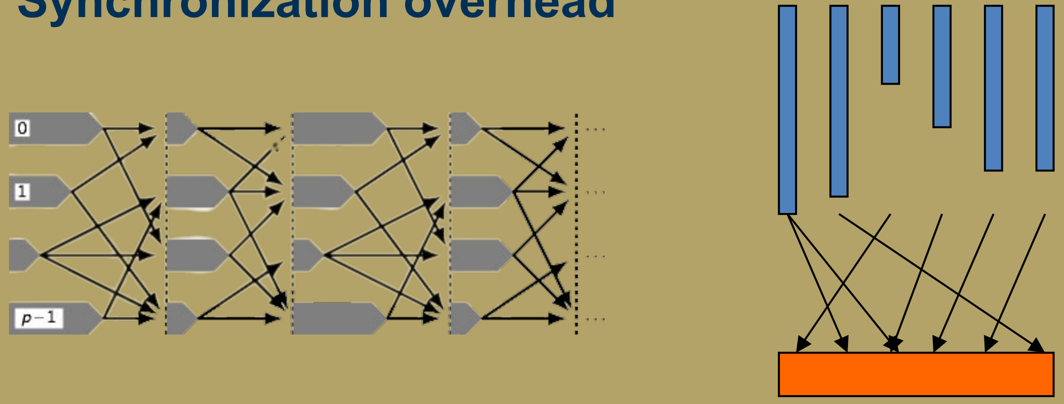
1 Size and scale of the problem

increasing graph size and complexity

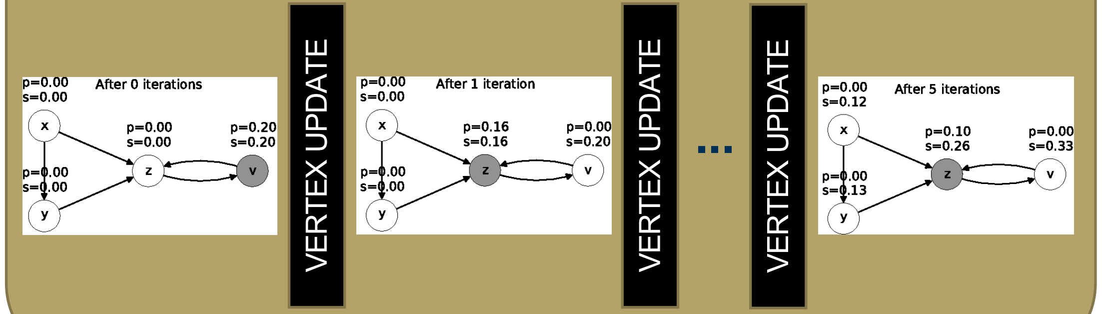
3 Poor locality of memory access



2 Synchronization overhead



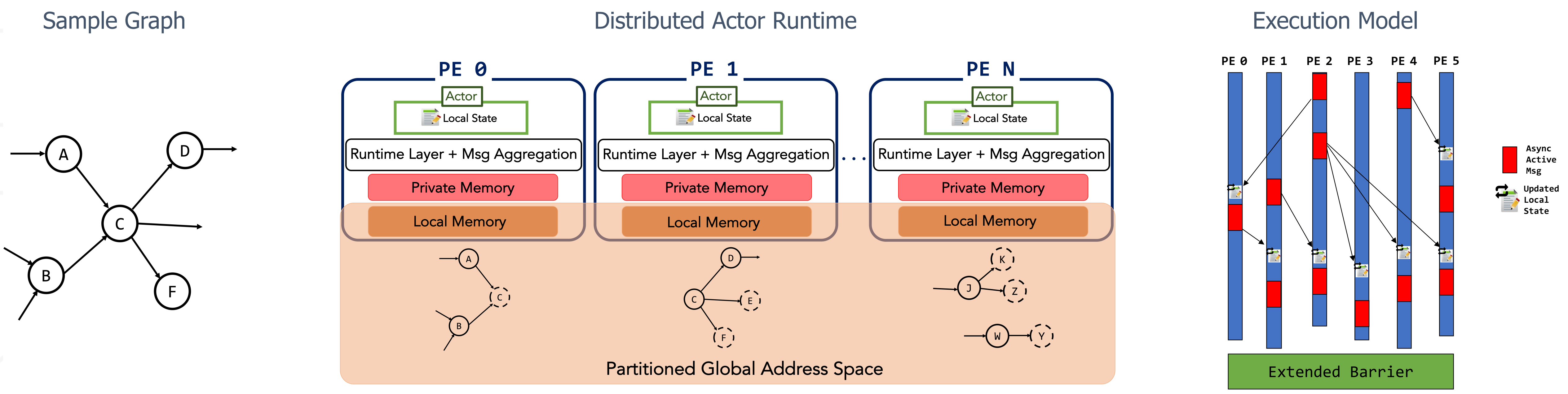
4 Slow convergence



This paper studies the scalability of our novel actor-based programming system to overcome the inherent challenges of large-scale graph processing!

Georgia Tech

Actor-based Scalable Architecture for Graph Processing



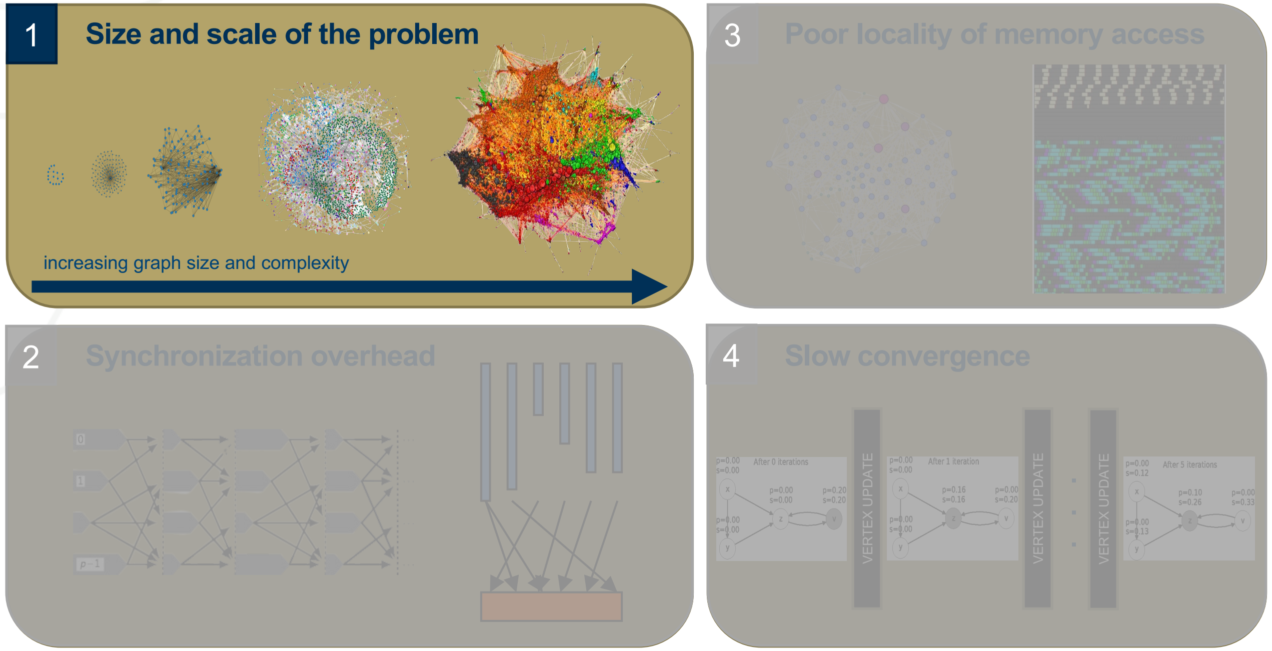
- Presents a lightweight, asynchronous computation model
- Utilizes fine-grained asynchronous actor messages to express point-to-point remote operations
- Treats actors as primitives of computation, where actors are inherently isolated and share no mutable state
- Actors process messages sequentially within its mailbox, thereby avoiding data races and synchronization

NOTE: “Actor” and “Selector” will be used interchangeably

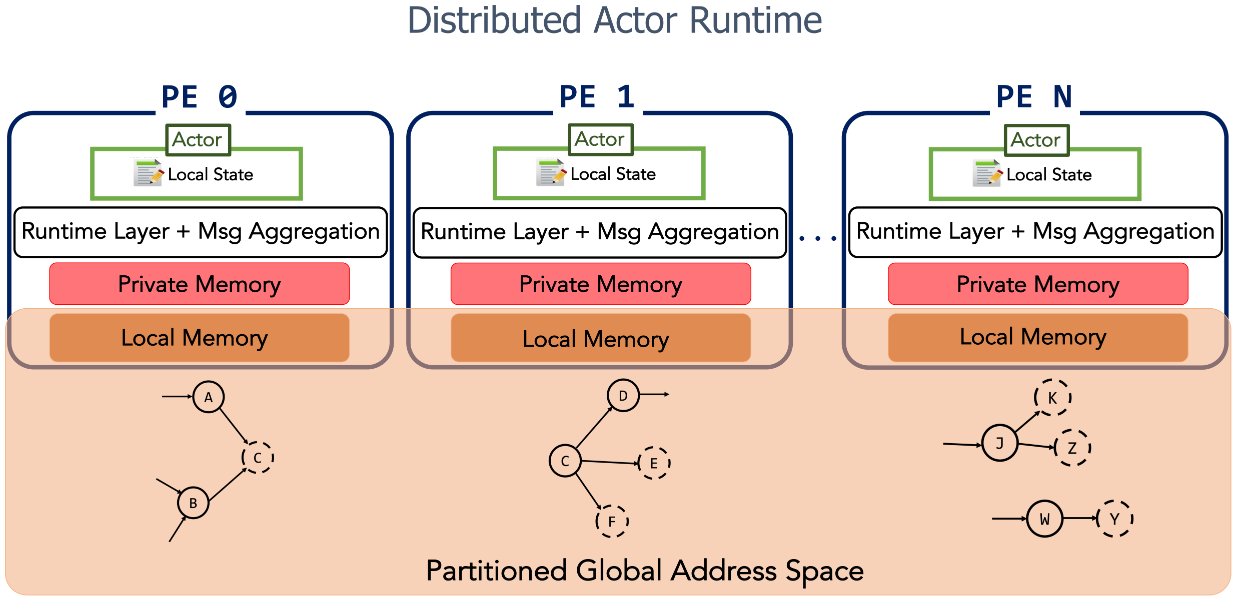
Georgia Tech

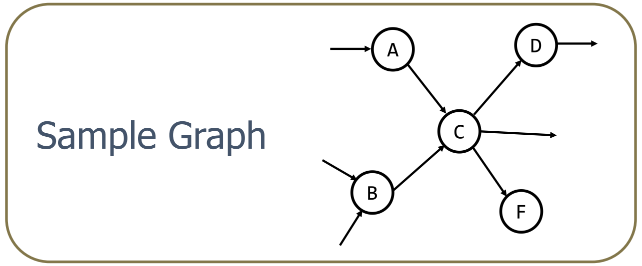
Actor-based Scalable Architecture for Graph Processing

The 4 overarching challenges of parallel graph processing:



SOLUTION



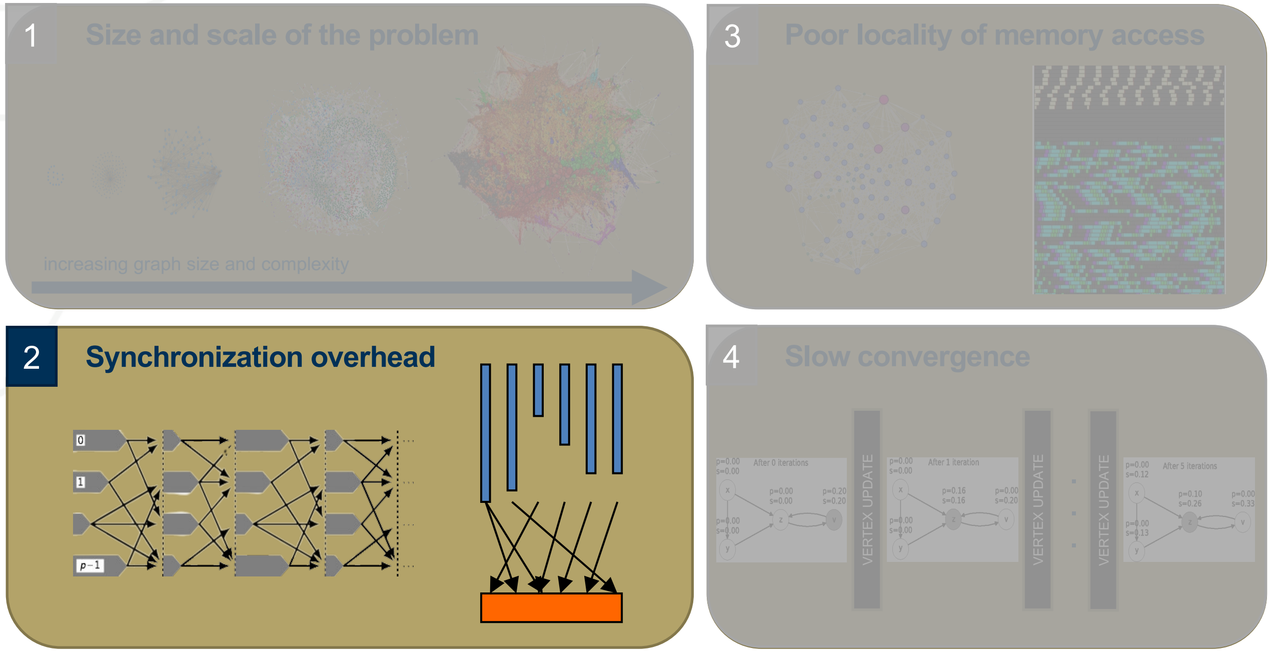


- Large-scale graph can be distributed across multiple PEs, where the local partition per PE is small enough to fit in its local memory

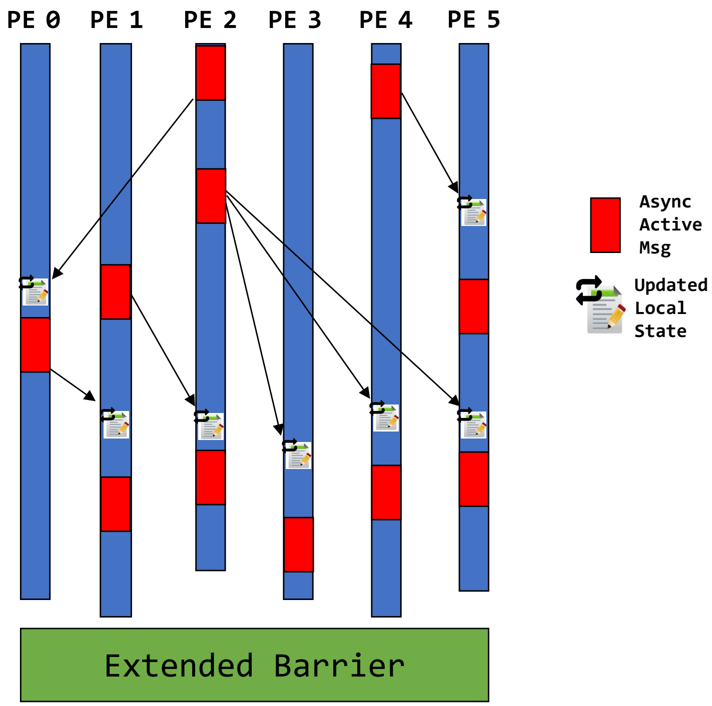
Actor-based Scalable Architecture for Graph Processing

SOLUTION

The 4 overarching challenges of parallel graph processing:



Execution Model

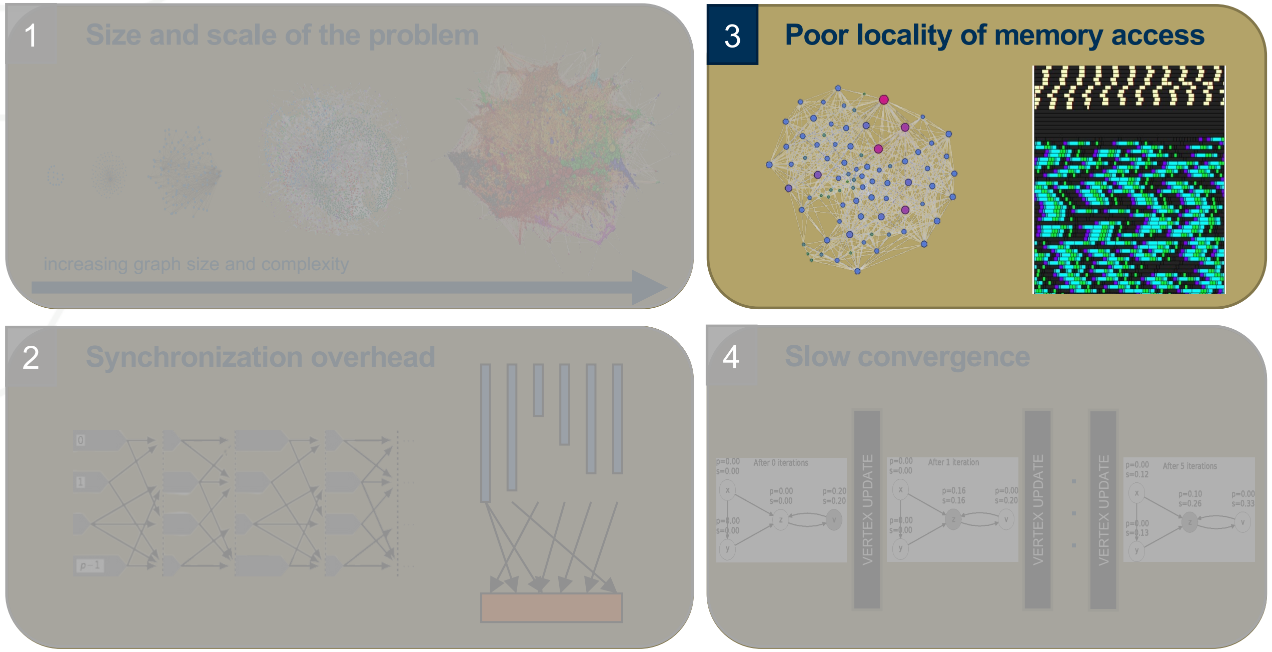


- Fine-grained Asynchronous Bulk-Synchronous (FABS) Parallelism model
- Reduces barriers and time spent idling at barriers, further reducing stall cycles

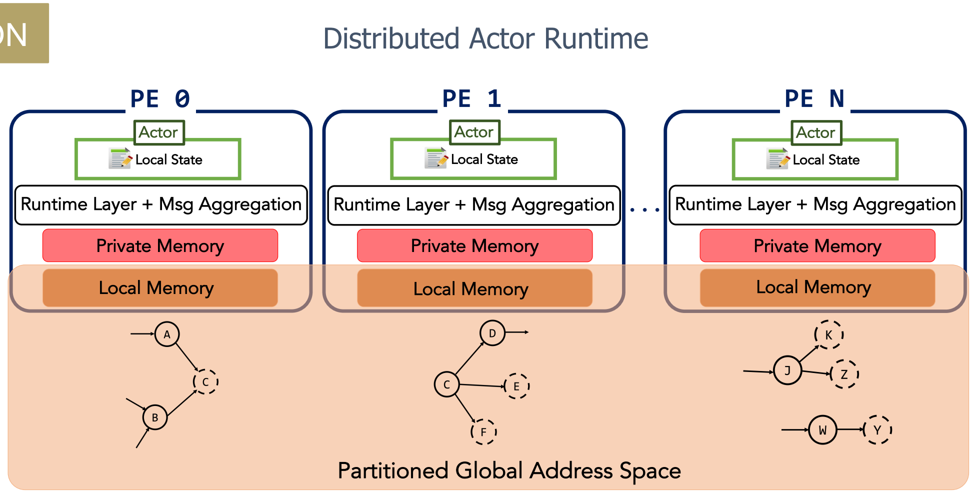
Georgia Tech

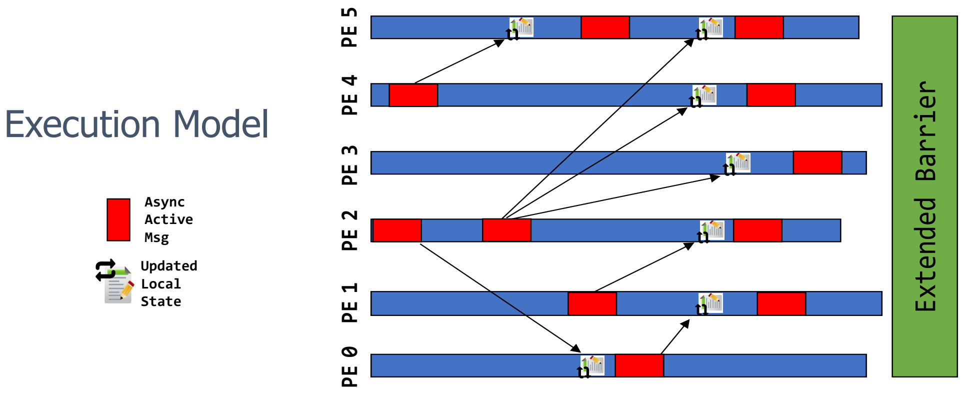
Actor-based Scalable Architecture for Graph Processing

The 4 overarching challenges of parallel graph processing:



SOLUTION



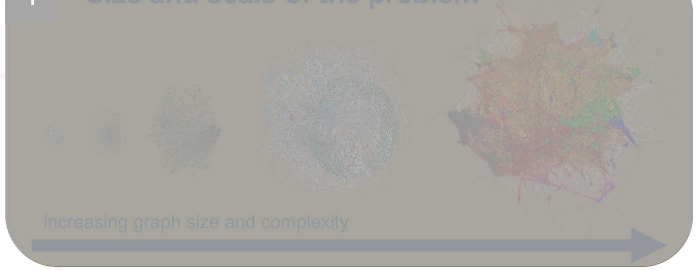


- Well suited for irregular applications due to the Partitioned Global Address Space (PGAS)
- Expresses point-to-point remote operations as short, one-sided fine-grained async messages
- Message aggregation allows for low overhead and high network utilization
- Computation is migrated to where the data is located (moving compute to data)

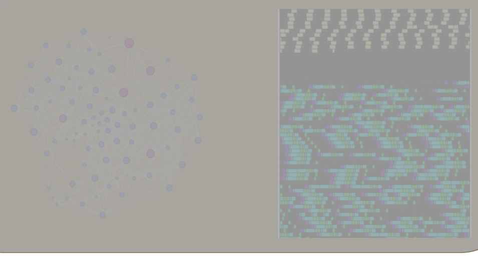
Actor-based Scalable Architecture for Graph Processing

The 4 overarching challenges of parallel graph processing:

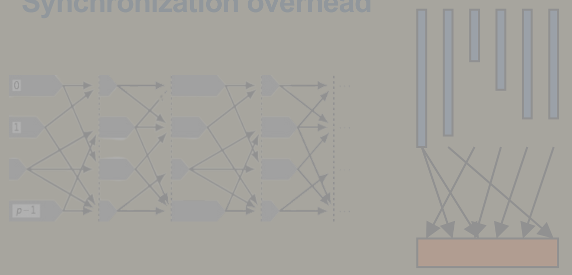
1 Size and scale of the problem



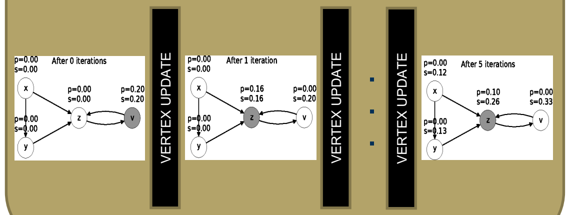
3 Poor locality of memory access



2 Synchronization overhead

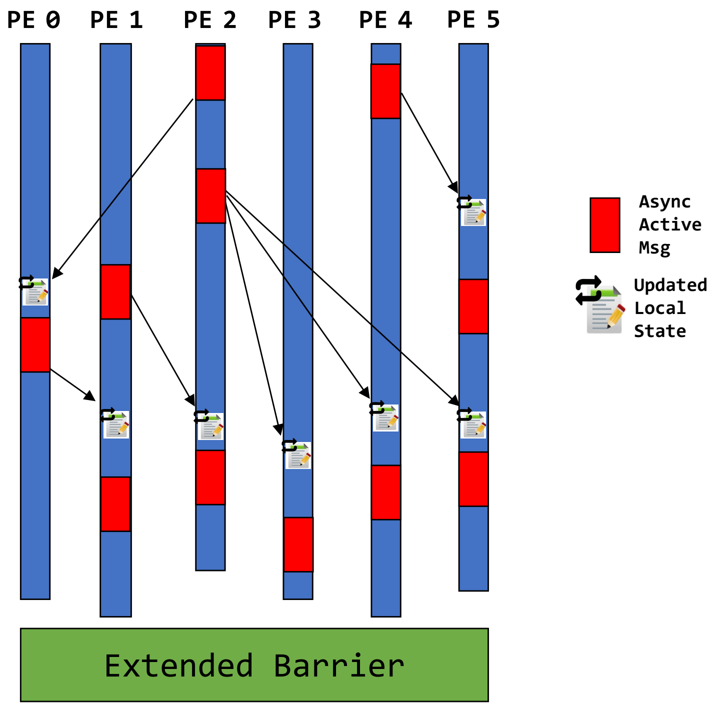


4 **Slow convergence**



SOLUTION

Execution Model



- As actor messages are executed, the local state of the actor is updated, allowing the next received actor message to utilize the updated state within the same super-step
- Updates the neighboring vertices with most recent values within the same iteration

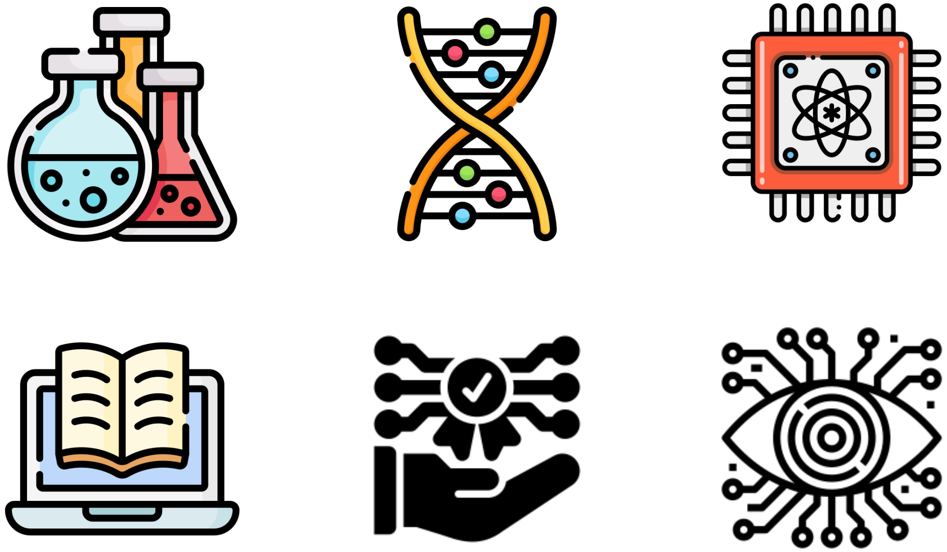
Georgia Tech

Showing the System's Extreme Scalability with PageRank & Jaccard Index

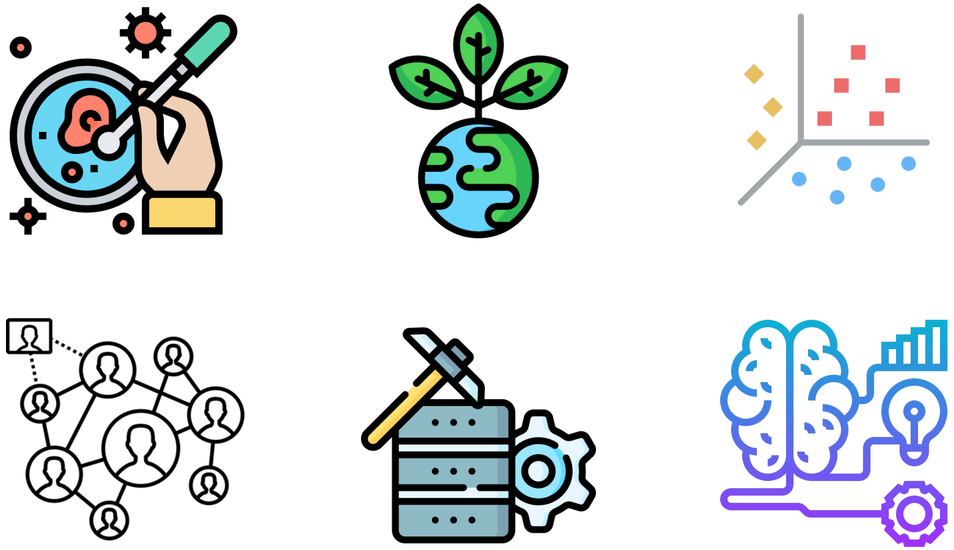
We focus on PageRank and Jaccard Index due to two reasons:

1. They show iterative vs non-iterative application types
2. They have been applied to many real-world problems with social impact

PageRank

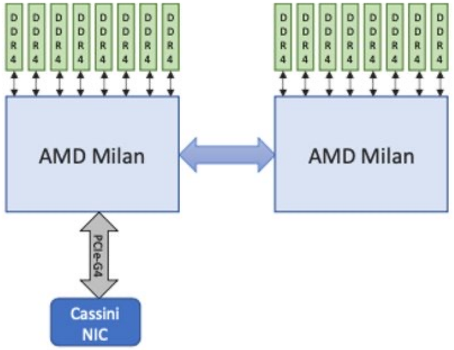


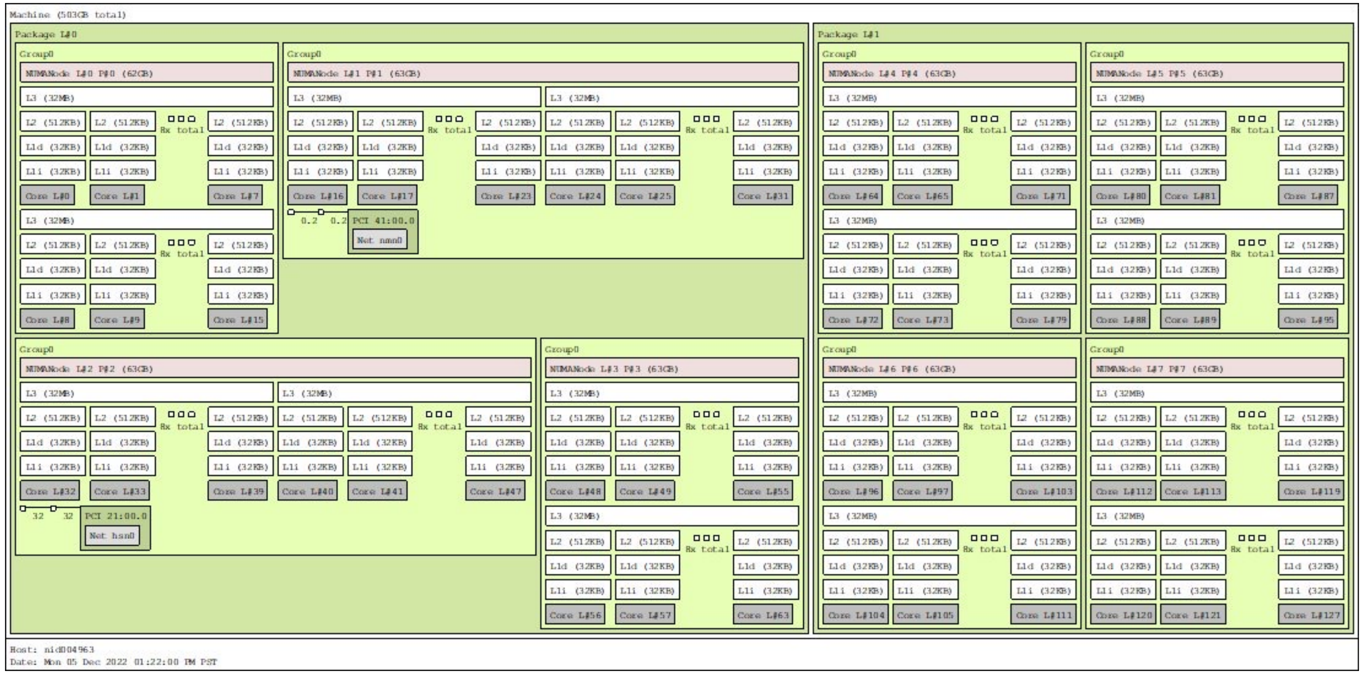
Jaccard Index



Experimental Setup and Architecture

- **Metrics**: execution time (in seconds) and traversed edges per second (TEPS)
- Experiments conducted on the CPU nodes of the **Perlmutter supercomputer** at the National Energy Research Scientific Computing Center (NERSC)
  - 2x AMD EPYC 7763 (Milan) CPUs
  - 64 physical cores per CPU
  - 512 GB memory
  - 1x HPE Cray Slingshot Interconnect
- Results for **different dimensions of scalability** are presented



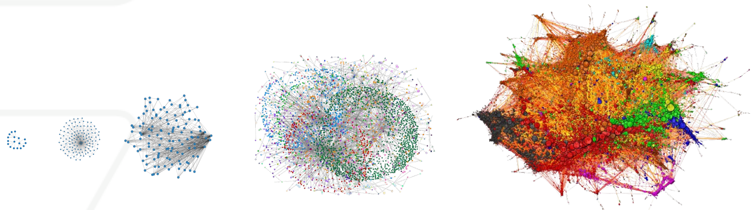


Picture borrowed from: https://docs.nersc.gov/systems/perlmutter/architecture/

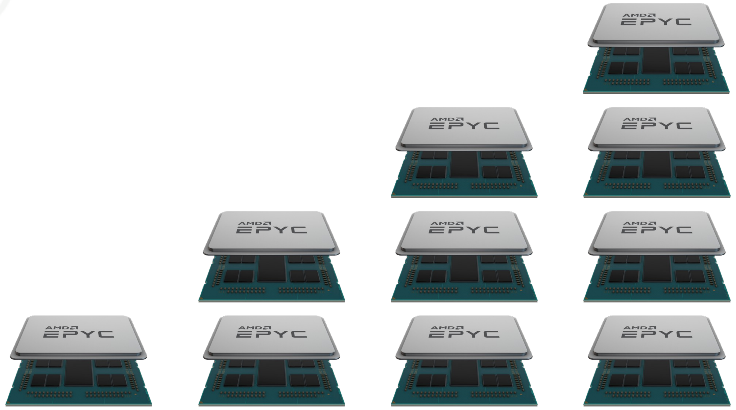
Dimensions of Scalability

Scaling performance is shown using three experiment types:

(1) SCALE1

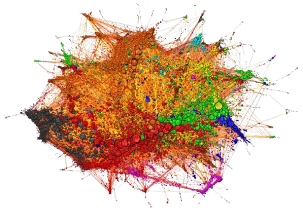


increasing graph size (WEAK SCALING, 5000 vertices per core)

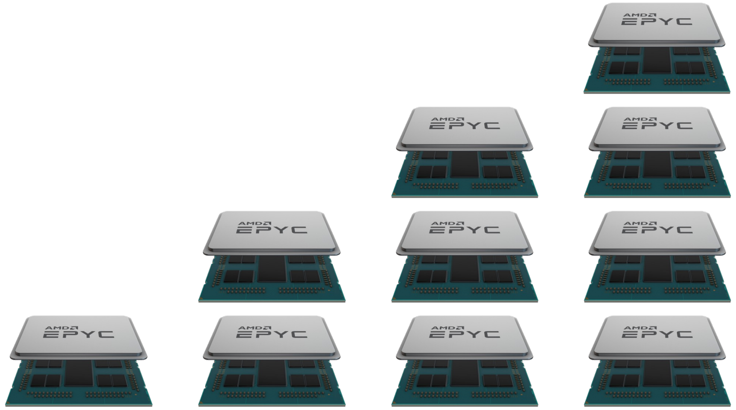


increasing #nodes (SCALE-OUT, from 256 to 65,536 cores
from 1 to 1024 nodes)

(2) SCALE2

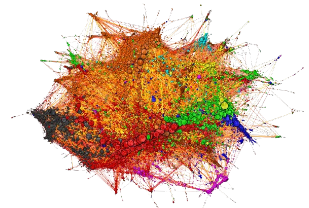


constant graph size
(STRONG SCALING, 10.2M vertices & 696.4M edges)

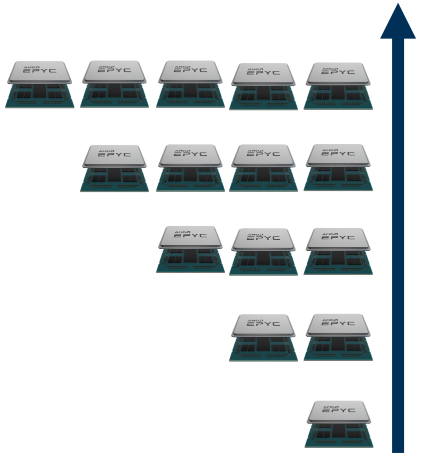


increasing #nodes (SCALE-OUT, from 256 to 65,536 cores
from 1 to 1024 nodes)

(3) SCALE3



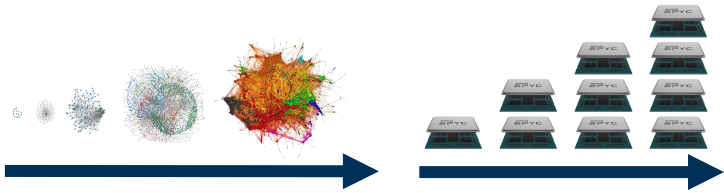
constant graph size
(STRONG SCALING, 10.2M vertices & 696.4M edges)



increasing #cores per node (SCALE-UP, from 1 core/node to 128 cores/node)

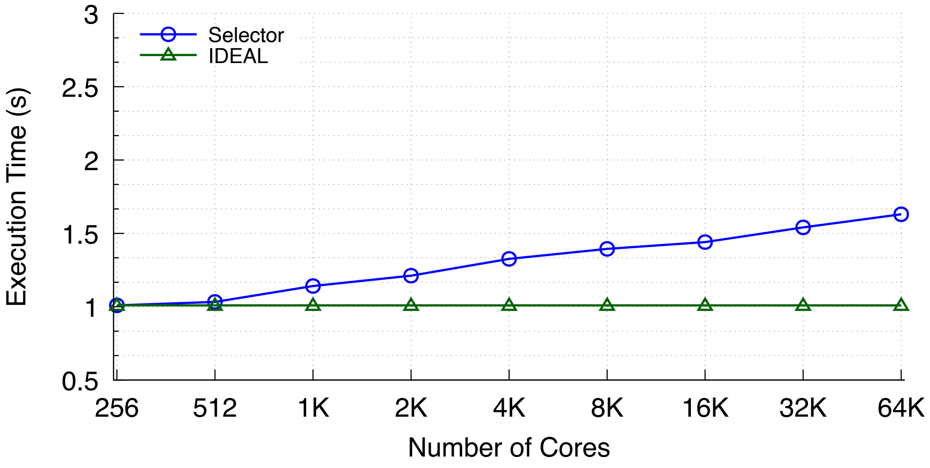
Georgia Tech

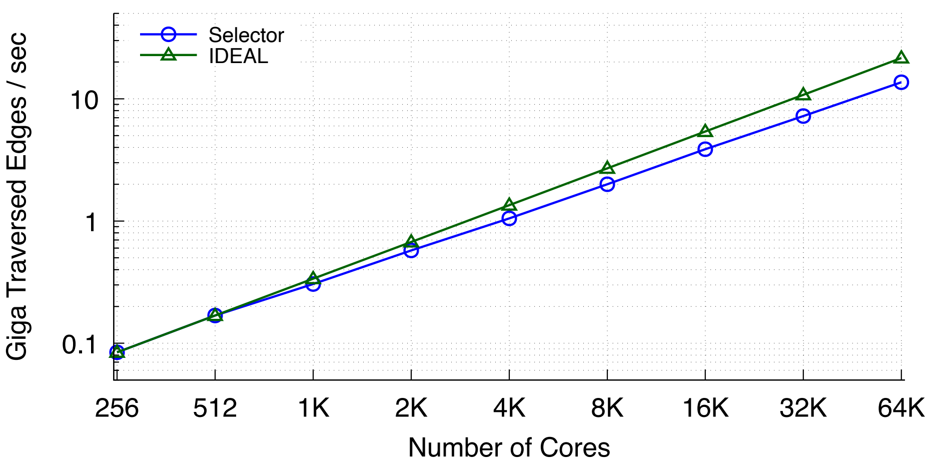
Performance Results: SCALE1



PageRank

13B TEPS

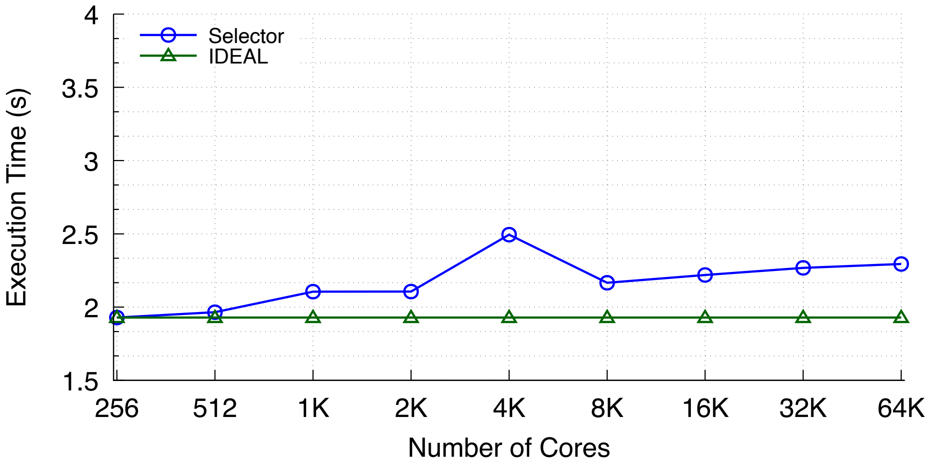


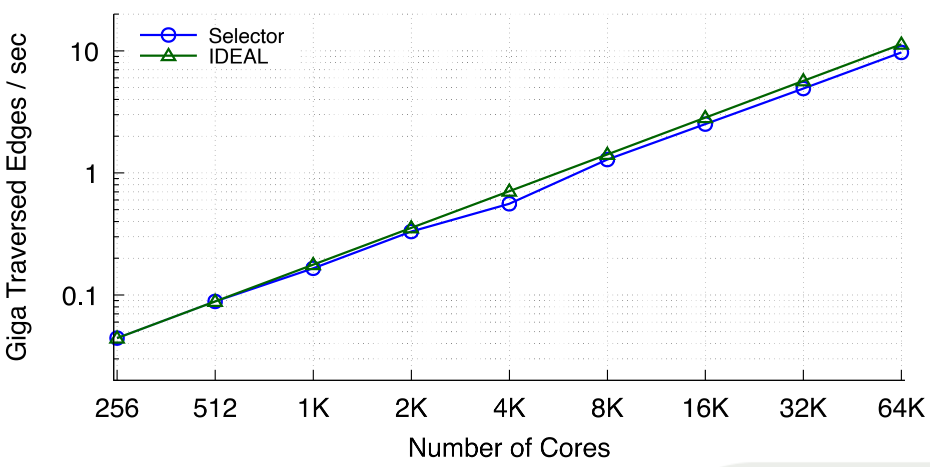


Jaccard Index

85.7% parallel efficiency

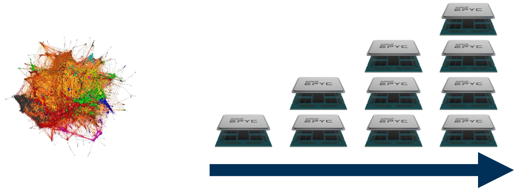
10B TEPS



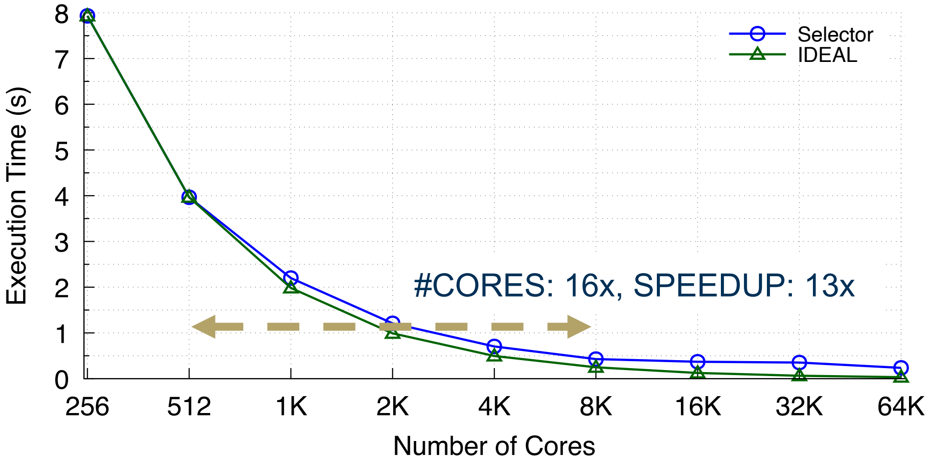


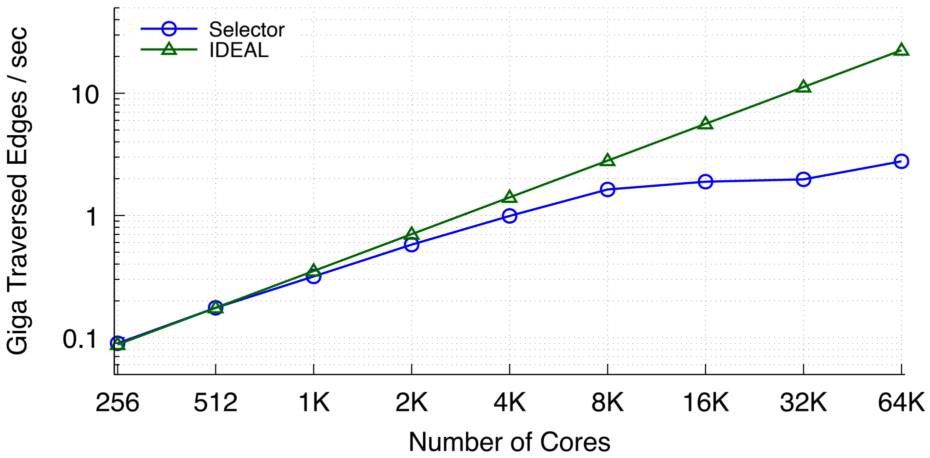
Georgia Tech

Performance Results: SCALE2



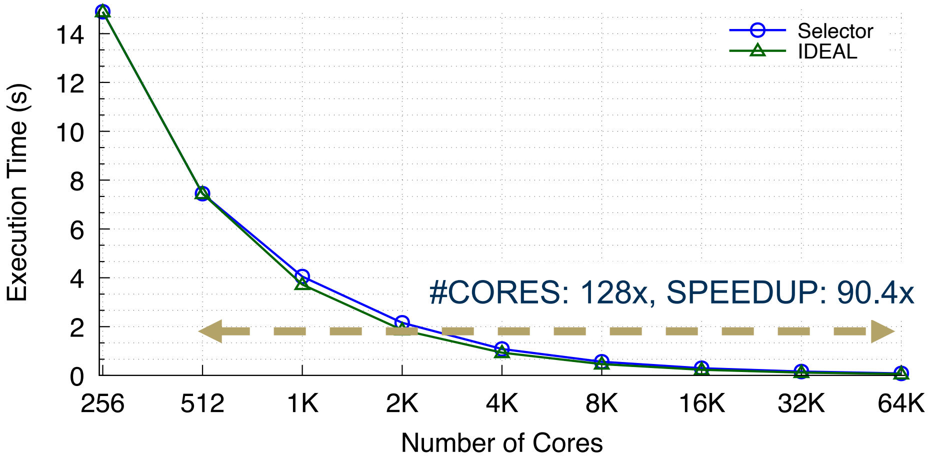
PageRank

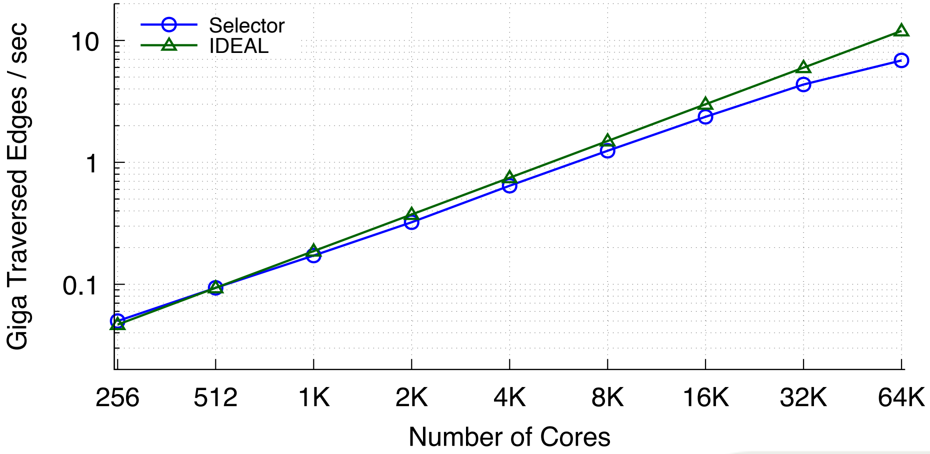




Jaccard Index

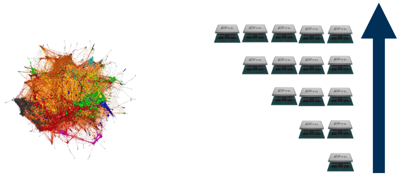
70.6% parallel efficiency





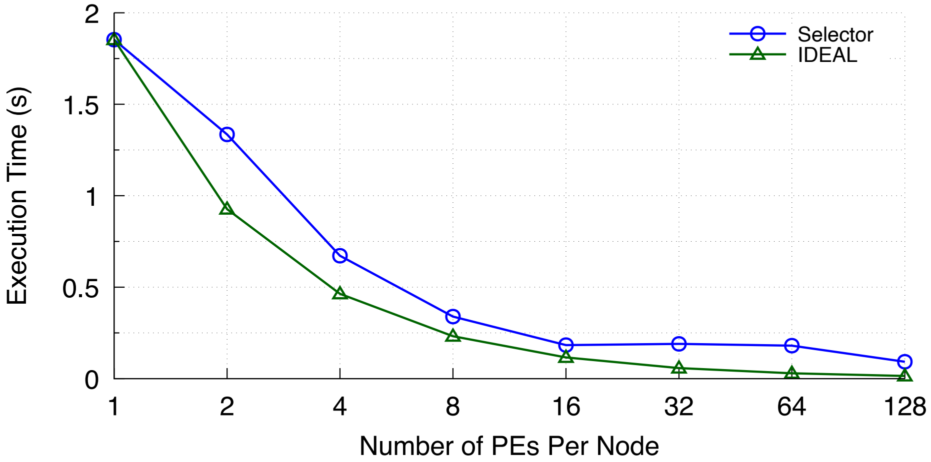
Georgia Tech

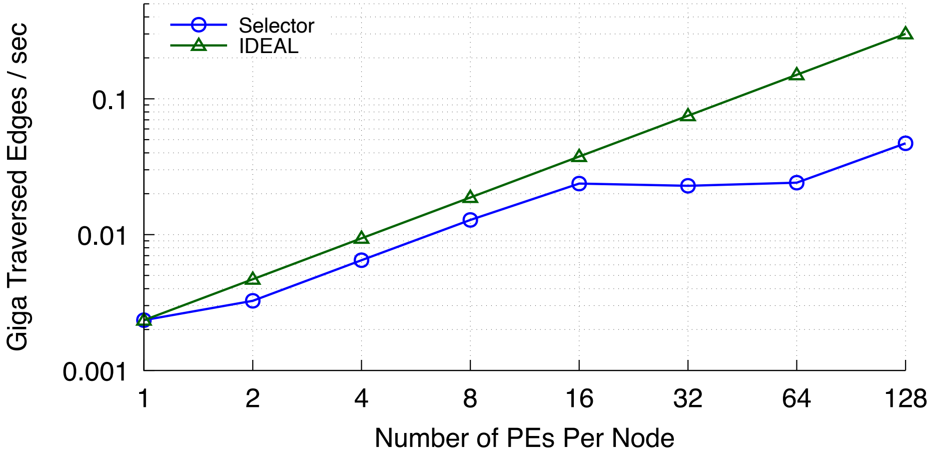
Performance Results: SCALE3



PageRank

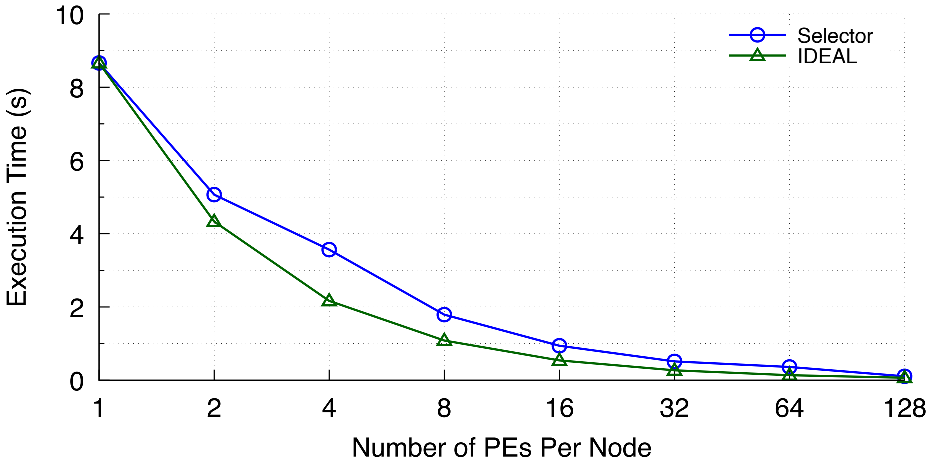
20.1x ↑

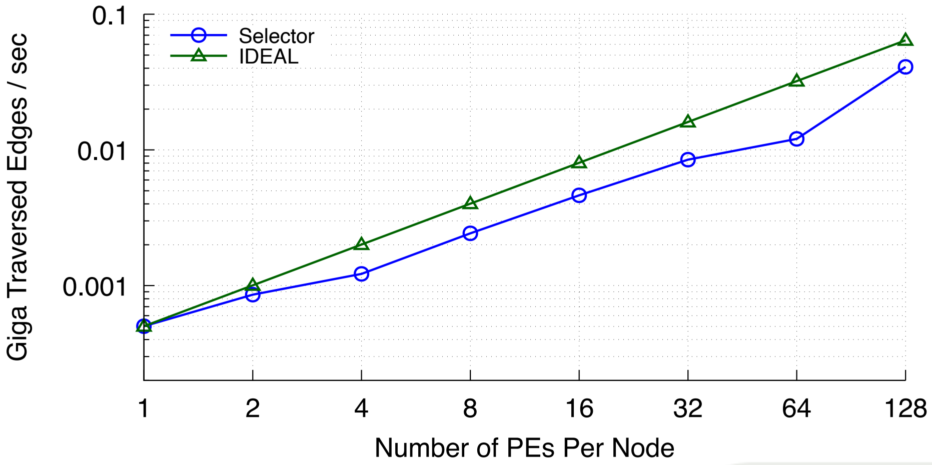




Jaccard Index

81.5x ↑





Georgia Tech

Contrasting to Related Approaches

- We contrast with respect to **remote atomics performance** and **graph application performance**
- Related approaches: OpenSHMEM, UPC, MPI3-RMA, YGM

Remote Atomics

Communication System	Non-Blocking	Read / Get	Update / Set
OpenSHMEM (cray-shmem 7.7.19)	N	40.06	N.A.
OpenSHMEM NBI (cray-shmem 7.7.19)	Y	4.79	4.99
UPC (Berkley-UPC 2022.5.0)	N	30.37	30.03
UPC NBI (Berkley-UPC 2022.5.0)	Y	20.58	N.A.
MPI3-RMA (OpenMPI 4.1.2)	N	25.44	142.04
MPI3-RMA (cray-mpich 7.7.19)	Y	9.67	59.47
YGM	Y	N.A.	> 600
Actors (cray-shmem 7.7.19)	Y	2.70	0.66

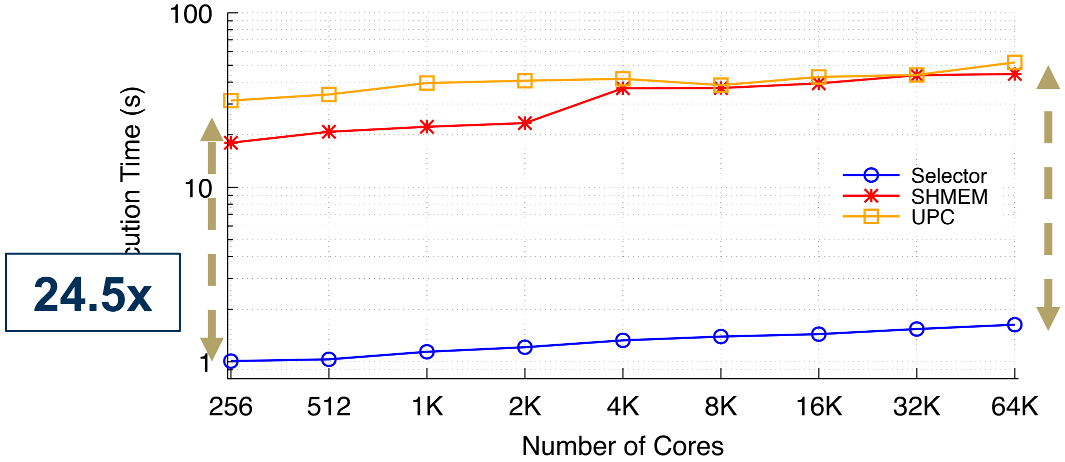
1.8x

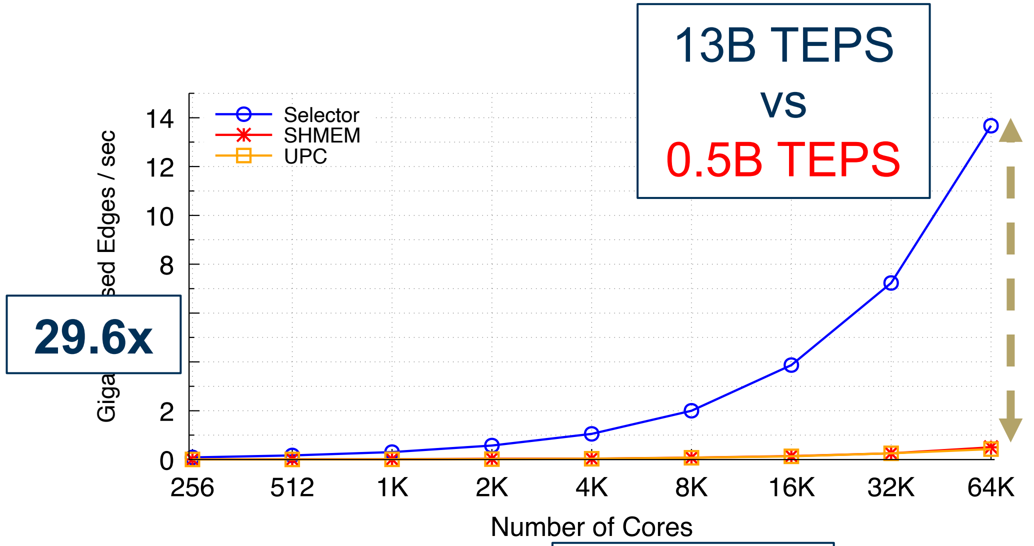
7.6x

Contrasting to Related Approaches

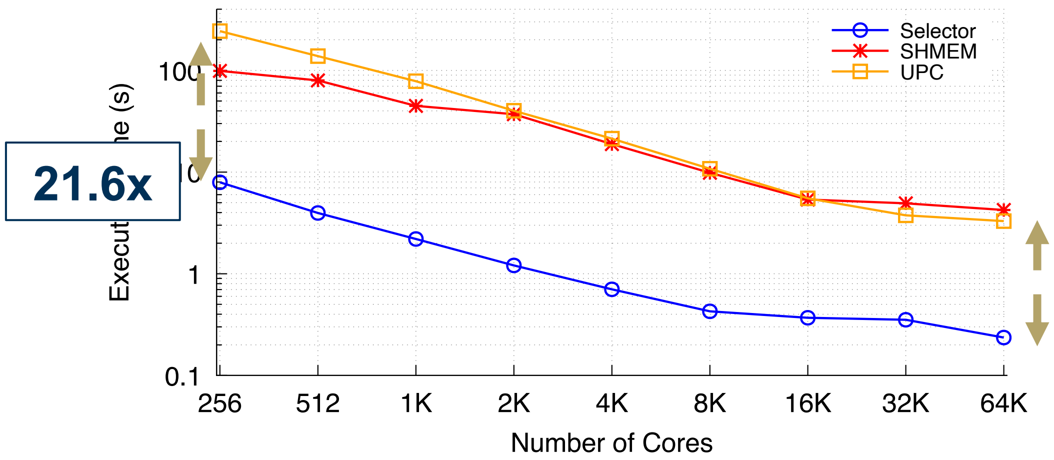
PageRank

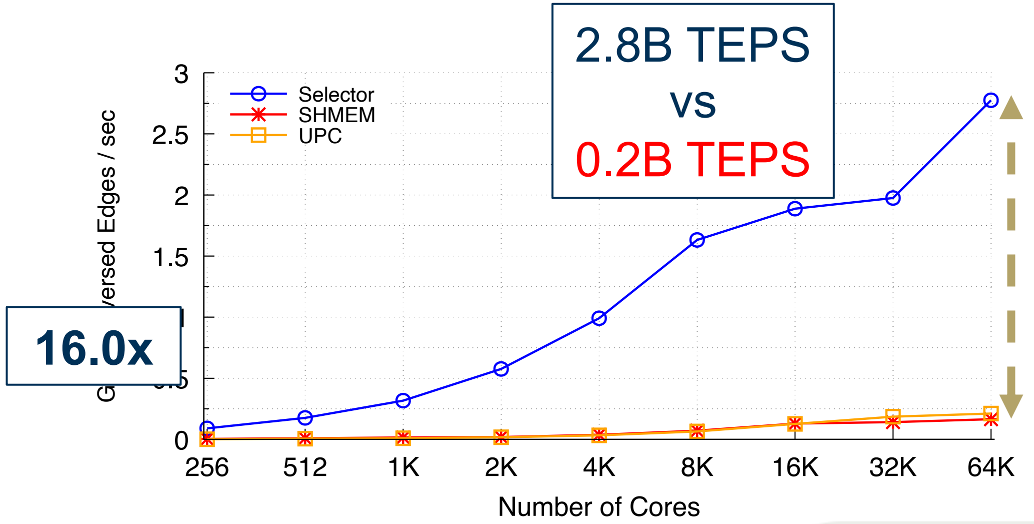
Weak Scaling





Strong Scaling





Georgia Tech

Impact of the Solution

- The actor-based system has shown **scalability**, **productivity**, and **performance**
- The extensibility of this system can front four impacts:

1. The system can be used on graphs of higher scale as well as systems of higher scaled-up/-out hardware resources
2. The system can be extended to other large-scale (iterative/non-iterative) graph applications
3. The system can be applied to structured, regular applications
4. The system can be expanded and compared to other related PGAS and non-PGAS approaches

You can try this at home… Just visit hclib-actor.com !

Bulk Synchronous Parallel - HC

hclib-actor.com/background/bsp/

HClib-Actor Documentation

Search

hclib_actor

0 1

HClib-Actor Documentation

Home

Background

Theory

Bulk Synchronous Parallel

Partitioned Global Address Space

Actor Model

Practice

OpenSHMEM

Bale

Summary

spmat

libgetput

Habanero-C Library (HClib)

Getting Started

Containers

Docker

Singularity

Clusters/Supercomputers

NERSC/ORNL/PACE

Writing HClib-Actor Programs

Bulk Synchronous Parallel

What is the bulk synchronous parallel model?

The Bulk Synchronous Parallel (BSP) model is one of the most popular parallel computation models.

The model consists of:

- A set of processor-memory pairs.
- A communication network that delivers messages in a point-to-point manner.
- Efficient barrier synchronization for all or a subset of the processes.

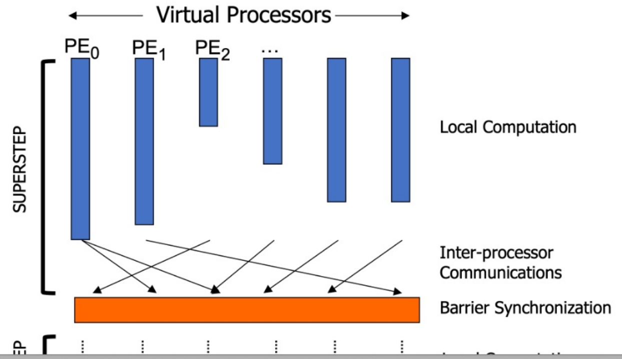


Table of contents

What is the bulk synchronous parallel model?

Single Program Multiple Data (SPMD) Programming

Further Readings

ACKNOWLEDGEMENT

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Advanced Graphical Intelligence Logical Computing Environment (AGILE) research program, under Army Research Office (ARO) contract number W911NF22C0083. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government.

Georgia Tech

IEEE/ACM CCGRID 2023

The 23rd IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing

Highly Scalable Large-Scale Asynchronous Graph Processing using Actors

Youssef Elmougy, Akihiro Hayashi, and Vivek Sarkar
Georgia Institute of Technology, Atlanta GA USA

Thank you for your attention!

Georgia Tech