Solving Global Grand Challenges with High Performance Data Analytics

David A. Bader

@Prof_DavidBader

http://www.cs.njit.edu/~bader

NJIT
New Jersey Institute of Technology

David A. Bader

Distinguished Professor and
Director, Institute for Data Science

- IEEE Fellow, SIAM Fellow, AAAS Fellow
- Recent Service:
    - White House's National Strategic Computing Initiative (NSCI) panel
    - Computing Research Association Board
    - NSF Advisory Committee on Cyberinfrastructure
    - Council on Competitiveness HPC Advisory Committee
    - IEEE Computer Society Board of Governors
    - IEEE IPDPS Steering Committee
    - Editor-in-Chief, ACM Transactions on Parallel Computing
    - Editor-in-Chief, IEEE Transactions on Parallel and Distributed Systems
- Over $184M of research awards
- 250+ publications, ≥ 11,000 citations, h-index ≥ 61
- National Science Foundation CAREER Award recipient
- Directed: Facebook AI Systems
- Directed: NVIDIA GPU Center of Excellence, NVIDIA AI Lab (NVAIL)
- Directed: Sony-Toshiba-IBM Center for the Cell/B.E. Processor
- Founder: Graph500 List benchmarking “Big Data” platforms
- Recognized as a “RockStar” of High Performance Computing by InsideHPC in 2012 and as HPCwire’s People to Watch in 2012 and 2014.

NJIT
New Jersey Institute of Technology

NJIT

New Jersey Institute
of Technology

NJIT Jumps into Top 100 for 2020 U.S. News College Rankings

http://news.njit.edu/njit-jumps-top-100-2020-us-news-college-rankings

NJIT

The New York Times

SundayReview | OP-ED COLUMNIST

America's Great Working-Class Colleges

David Leonhardt

Jan 18, 2017

An Upward Mobility Top 10

Colleges ranked by percent of students from the bottom fifth of the income scale who end up in the top three-fifths.

1	**New Jersey Institute of Technology**	**85%**
2	Pace University	82%
3	Calif. State – Bakersfield	82%
4	Univ. California – Irvine	81%
5	Calif. Poly – Pomona	81%
6	Xavier of Louisiana	80%
7	SUNY Stony Brook	79%
8	San Jose State	79%
9	CUNY Baruch College	79%
10	Calif. State – Long Beach	78%

Launched in **July 2019**, with inaugural director

David A. Bader

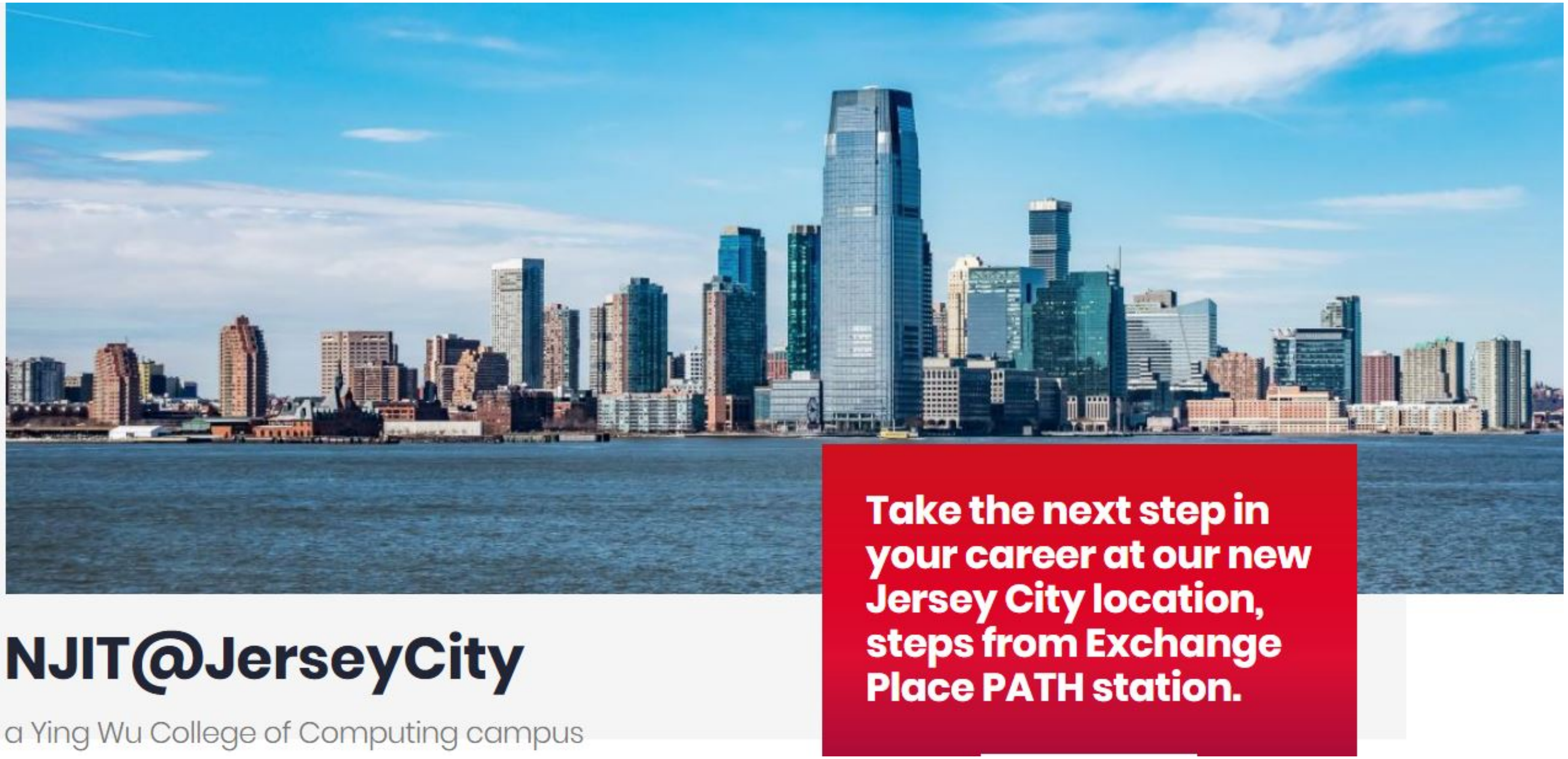
(~35 faculty in current centers)

NJIT Data Science Seminar Series

Wednesday’s 4pm ET

https://njit-institute-for-data-science.eventbrite.com/

Center for Big Data	• Big Data Analytics, Systems, and Tools • Cyberinfrastructure
Cybersecurity Research Center	• Practical encryption • Privacy technologies • Information Assurance
The Structural Analysis of Biomedical Ontologies Center	• Medical Informatics • NIH / National Cancer Institute
FinTech Group	• Financial Services • Insurance Industry
Machine Learning & AI	• Real-world technologies • Industrial partnerships

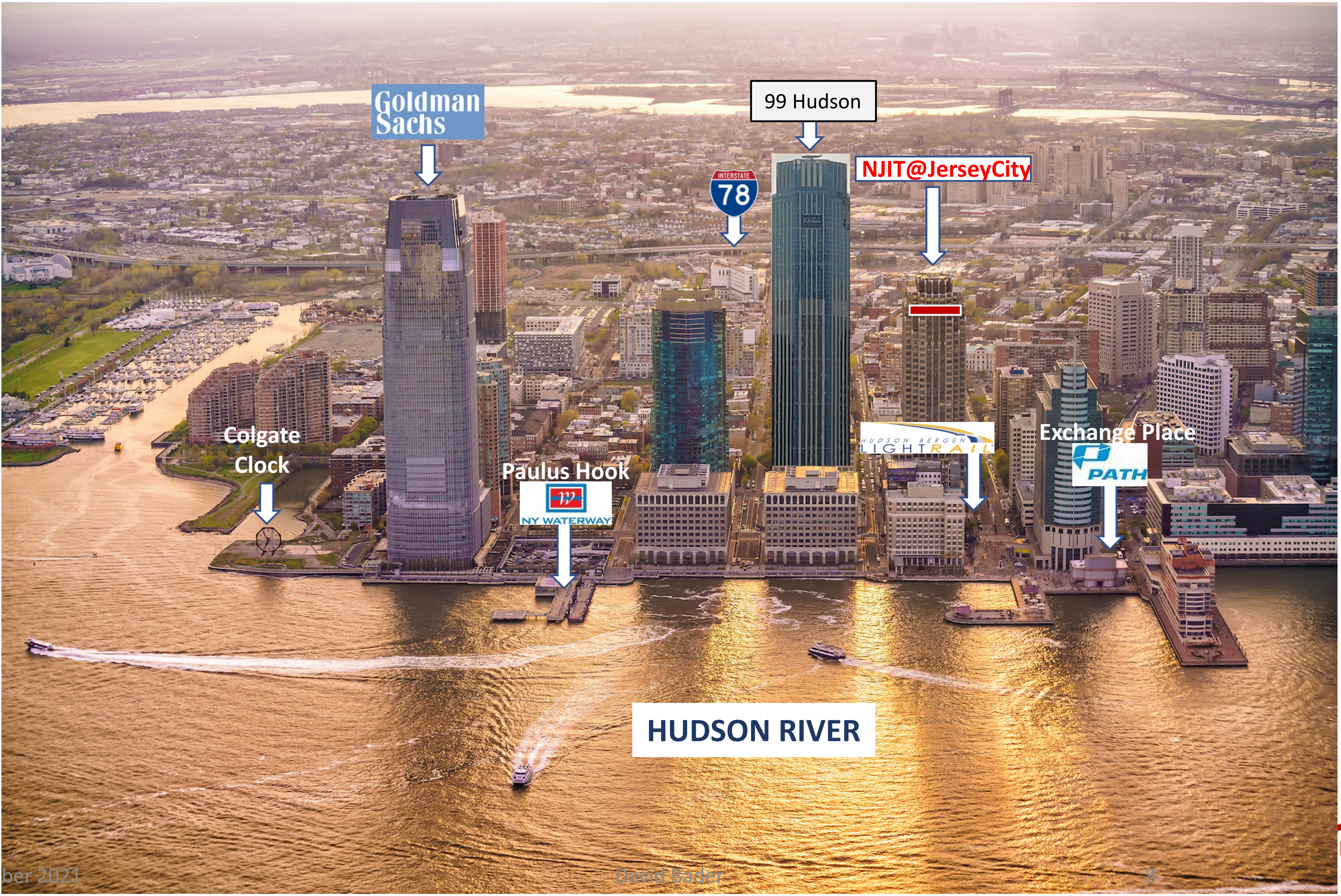


NJIT@JerseyCity

a Ying Wu College of Computing campus

Take the next step in your career at our new Jersey City location, steps from Exchange Place PATH station.

- In Fall 2019, NJIT opened a data science focused satellite campus in Jersey City, NJ, in the heart of the financial services district, and ten minutes from lower Manhattan via the PATH train.
- The flagship offering is a full five-semester part-time M.S. program for data scientists.
- This **M.S. in Data Science** covers basic and advanced methods in statistical inference, machine learning, data visualization, data mining, and big data, all of which are essential skills for a high-performing data scientist.



New Jersey Institute of Technology

Solving real-world challenges

- Urban sustainability
- Healthcare analytics
- Trustworthy, Free and Fair Elections
- Insider threat detection
- Utility infrastructure protection
- Cyberattack defense
- Disease outbreak and epidemic monitoring

NJIT
New Jersey Institute of Technology

High Performance Algorithms for Interactive Data Science at Scale

(PI: Bader) 3/2021 – 2/2022, NSF CCF-2109988

NSF

αρκούδα
massive scale
data science

A real-world challenge in data science is to develop interactive methods for quickly analyzing new and novel data sets that are potentially of massive scale. This award will design and implement fundamental algorithms for high performance computing solutions that enable the interactive large-scale data analysis of massive data sets.

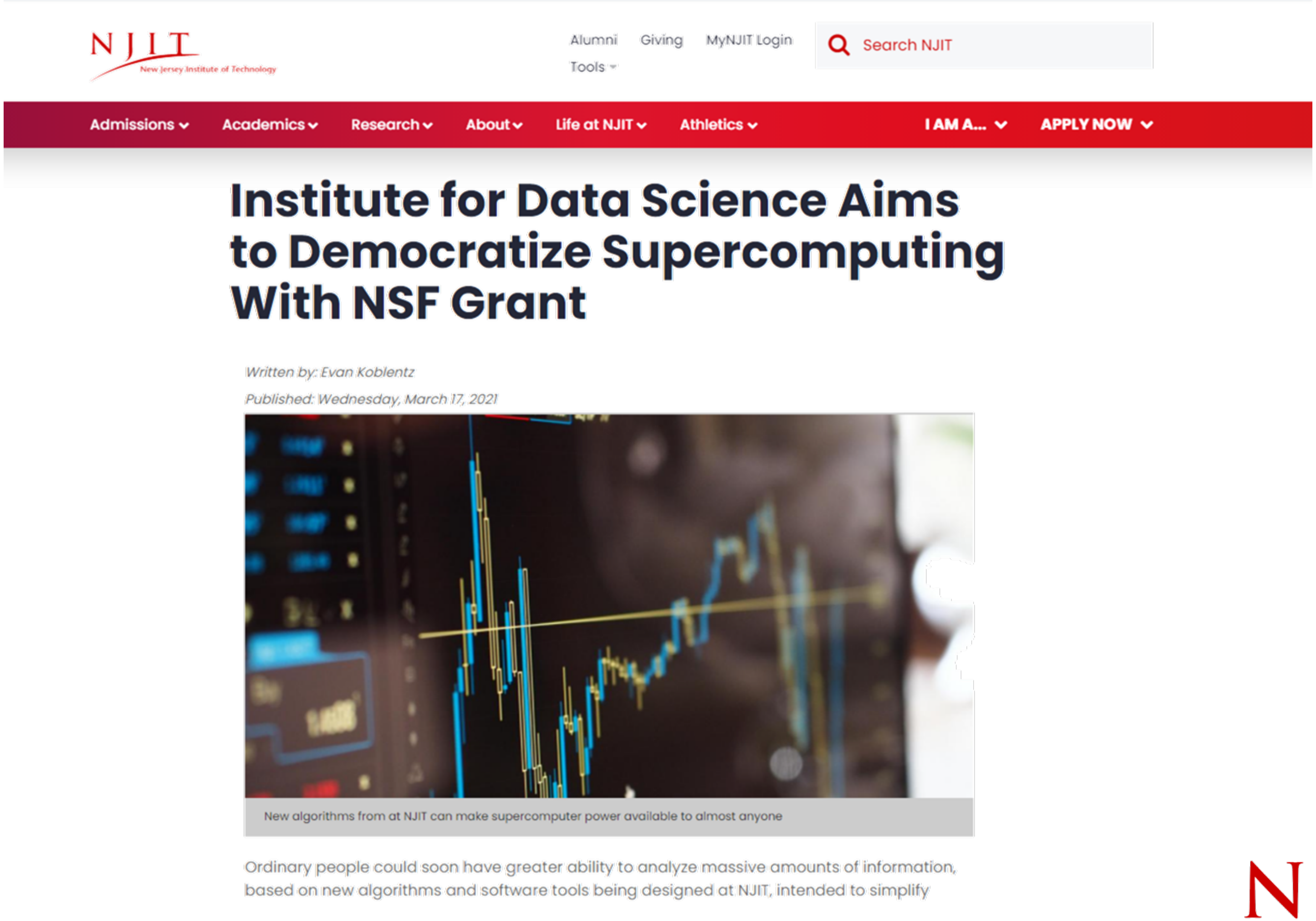
This project focuses on these three important data structures for data analytics:

1) suffix array construction,
2) 'treap' construction, and
3) distributed memory join algorithms,

useful for analyzing large scale strings, implementing random search in large string data sets, and generating new relations, respectively.

To evaluate and show the effectiveness of the proposed algorithms, these algorithms will be implemented in and contribute to an open source NumPy-like software framework that aims to provide productive data discovery tools on massive, dozens-of-terabytes data sets by bringing together the productivity of Python with world-class high performance computing.

NJIT
New Jersey Institute
of Technology



Institute for Data Science Aims to Democratize Supercomputing With NSF Grant

Written by: Evan Koblentz

Published: Wednesday, March 17, 2021

New algorithms from at NJIT can make supercomputer power available to almost anyone

Ordinary people could soon have greater ability to analyze massive amounts of information, based on new algorithms and software tools being designed at NJIT, intended to simplify

https://news.njit.edu/institute-data-science-aims-democratize-supercomputing-nsf-grant

Data Science: Discovery and Innovation

The National Strategic Computing Initiative (NSCI) The NSCI was launched by Executive Order (EO) 13702 in July 2015 to advance U.S. leadership in high performance computing (HPC).

The ability to manipulate data and understand Data Science is becoming increasingly critical to current and future discovery and innovation.

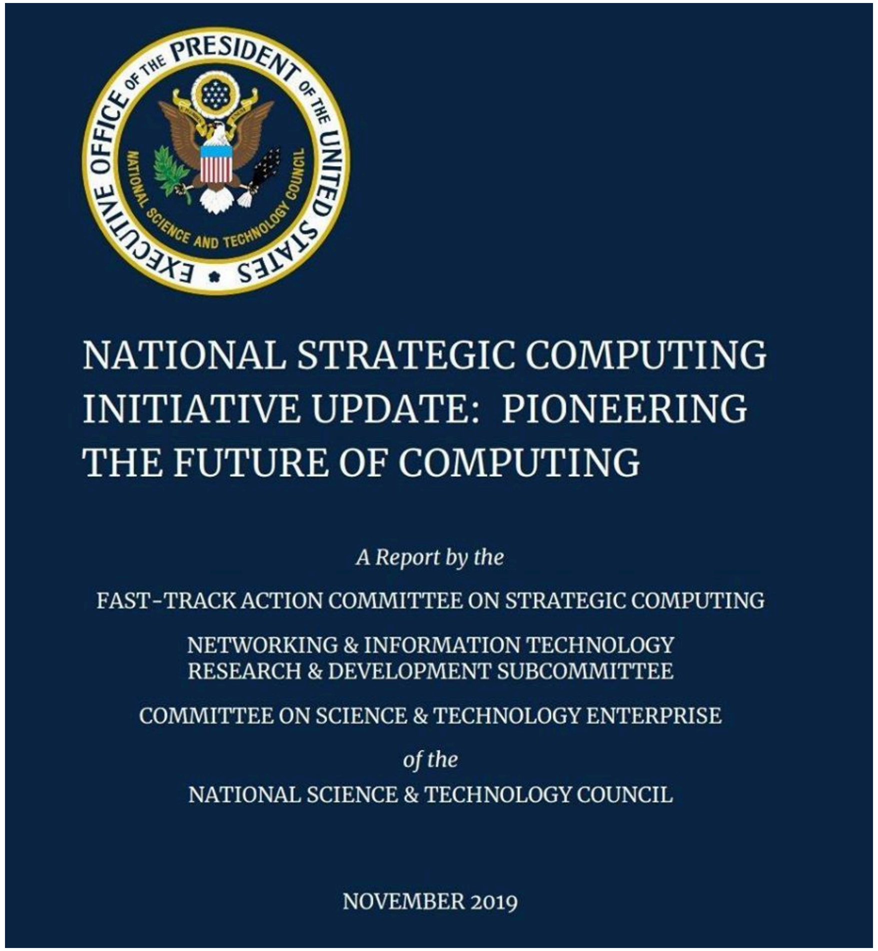
REALIZING THE POTENTIAL OF DATA SCIENCE Final Report from the National Science Foundation Computer and Information Science and Engineering Advisory Committee Data Science Working Group. Francine Berman and Rob Rutenbar, co-Chairs Henrik Christensen, Susan Davidson, Deborah Estrin, Michael Franklin, Brent Hailpern, Margaret Martonosi, Padma Raghavan, Victoria Stodden, Alex Szalay. December 2016

McKinsey predicts that data-driven technologies will bring an additional $300 billion of value to the U.S. health care sector alone, and by 2020, 1.5 million more “data-savvy managers” will be needed to capitalize on the potential of data, “big” and otherwise.

Manyika, J. et al. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Retrieved from http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation

National Strategic Computing Initiative (NSCI) Update

14 Nov 2019



EXECUTIVE OFFICE OF THE PRESIDENT OF THE UNITED STATES • NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

NATIONAL STRATEGIC COMPUTING INITIATIVE UPDATE: PIONEERING THE FUTURE OF COMPUTING

A Report by the

FAST-TRACK ACTION COMMITTEE ON STRATEGIC COMPUTING

NETWORKING & INFORMATION TECHNOLOGY RESEARCH & DEVELOPMENT SUBCOMMITTEE

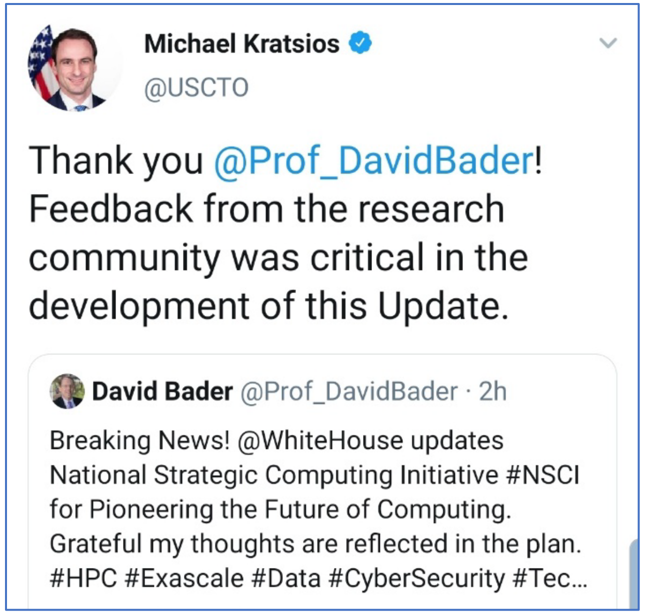
COMMITTEE ON SCIENCE & TECHNOLOGY ENTERPRISE

of the

NATIONAL SCIENCE & TECHNOLOGY COUNCIL

NOVEMBER 2019

In recognition of the fast-changing computing landscape, the updated plan places new emphasis on the following areas as compared to the 2016 plan:



Michael Kratsios @USCTO

Thank you @Prof_DavidBader! Feedback from the research community was critical in the development of this Update.

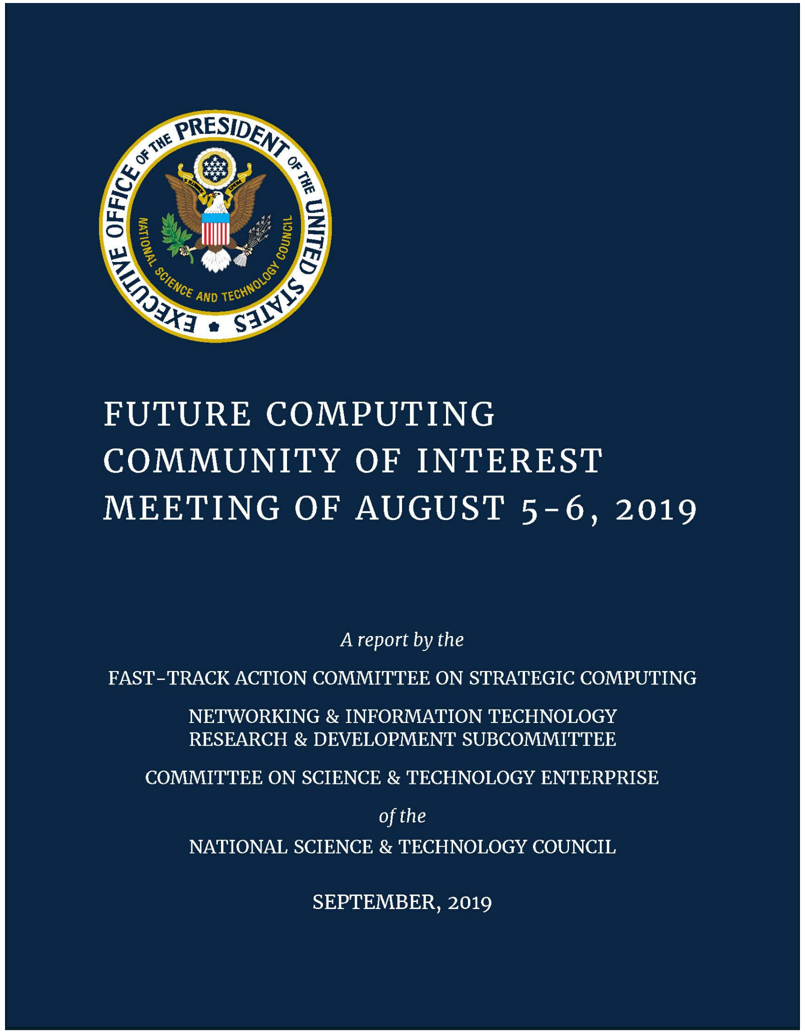
David Bader @Prof_DavidBader · 2h

Breaking News! @WhiteHouse updates National Strategic Computing Initiative #NSCI for Pioneering the Future of Computing. Grateful my thoughts are reflected in the plan. #HPC #Exascale #Data #CyberSecurity #Tec...

- Computing hardware, with a focus on the 10-year horizon and beyond;
- Software infrastructure that will enable effective and sustainable use of new computing;
- Overall infrastructure, from data usage and management to cybersecurity, foundries, and prototypes;
- And the development of new real-world applications, systems, and opportunities for future computing.

Future of Advanced Computing Ecosystem (FACE)

2019-2020


EXECUTIVE OFFICE OF THE PRESIDENT OF THE UNITED STATES
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

FUTURE COMPUTING
COMMUNITY OF INTEREST
MEETING OF AUGUST 5-6, 2019

A report by the

FAST-TRACK ACTION COMMITTEE ON STRATEGIC COMPUTING

NETWORKING & INFORMATION TECHNOLOGY
RESEARCH & DEVELOPMENT SUBCOMMITTEE

COMMITTEE ON SCIENCE & TECHNOLOGY ENTERPRISE

of the

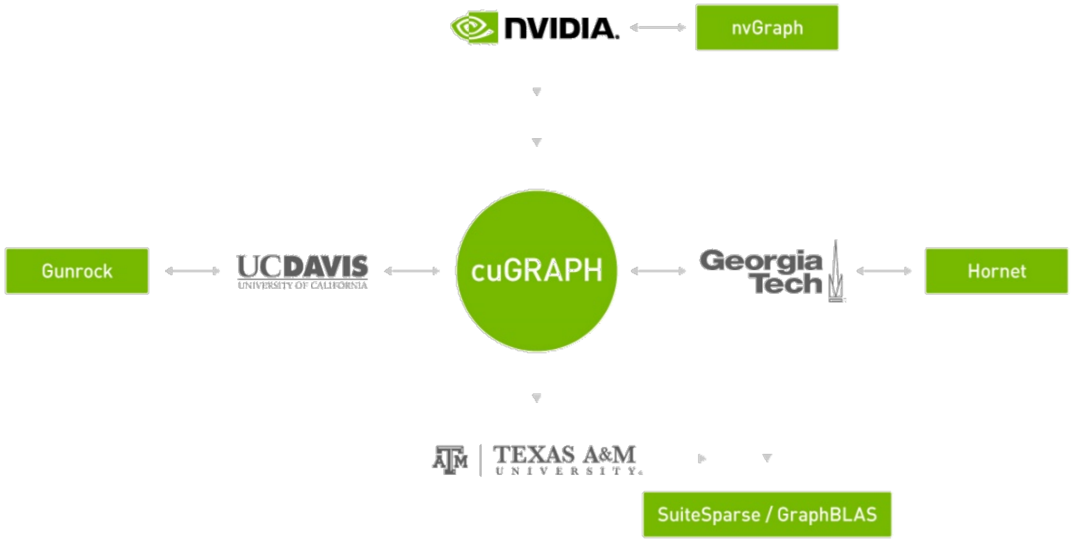
NATIONAL SCIENCE & TECHNOLOGY COUNCIL

SEPTEMBER, 2019

- Community of Interest meeting on “Future Computing,” NITRD High End Computing (HEC) Interagency Working Group (IWG) and the National Strategic Computing Initiative (NSCI) Joint Program Office for Strategic Computing (JPO-SC), NITRD National Coordination Office, Washington, DC, August 5-6, 2019.
- OSTP Convening: Pioneering the Future Advanced Computing Ecosystem, NSTC Subcommittee on the Future Advanced Computing Ecosystem (FACE), White House, Office of Science and Technology Policy (OSTP), National Science and Technology Council (NSTC), virtual, August 17-18, 2020.

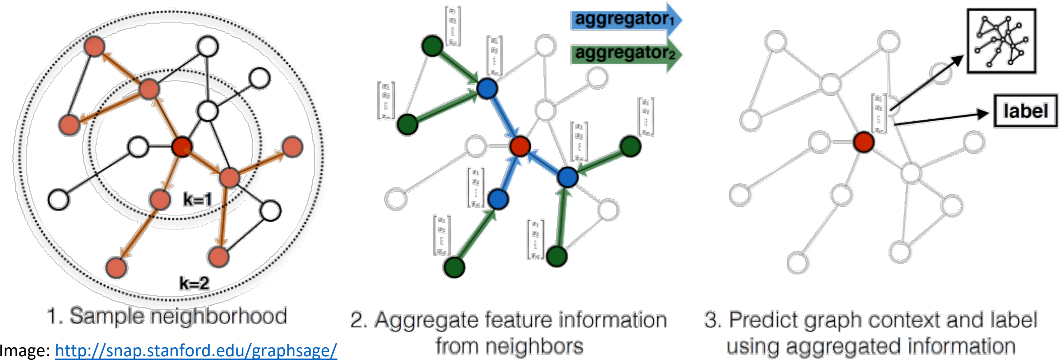
AI Lab (NVAIL) 2019, PI: Bader
Building the Future of Graph Analytics with RAPIDS

“Prof. David Bader and his lab … are leaders in high performance computing algorithms, with a focus on both static and dynamic graph algorithms. With Prof. Bader and his lab, we will work on the design and implementation of scalable graph algorithms and graph primitives for integrating into cuGRAPH, leveraging their Hornet framework.” – Sandra Skaff, NVIDIA, April 2019



2019 Facebook AI Systems Award: Scalable Graph Learning Algorithms

Project Aim: Develop scalable graph learning algorithms and implementations that open the door for learned graph models on massive graphs



1. Sample neighborhood

2. Aggregate feature information from neighbors

3. Predict graph context and label using aggregated information

Image: http://snap.stanford.edu/graphsage/
Inductive Representation Learning on Large Graphs. W.L. Hamilton, R. Ying, and J. Leskovec arXiv:1706.02216 [cs.SI], 2017.

Deep Learning (DL) has significantly impacted the tasks of speech recognition, image classification, object detection and recommendation

Complex tasks: self-driving, super-human image recognition, recommendation engines, machine natural language translation, content selection, learning patterns of life

Techniques used in DL: convolutional neural networks (CNNs) → applicable for Euclidean data types and does not apply for Graphs

Solution: embedding graphs into a lower dimensional Euclidean space, generating a regular structure

1. developing a scalable high performance graph learning system based on GCNs algorithms, like GraphSage, by improving the workflow on shared-memory NUMA machines balancing computation between threads, optimizing data movement, and improving memory locality
2. investigate graph learning algorithm: specific decompositions and develop new strategies for graph learning that can inherently scale well while maintaining high accuracy

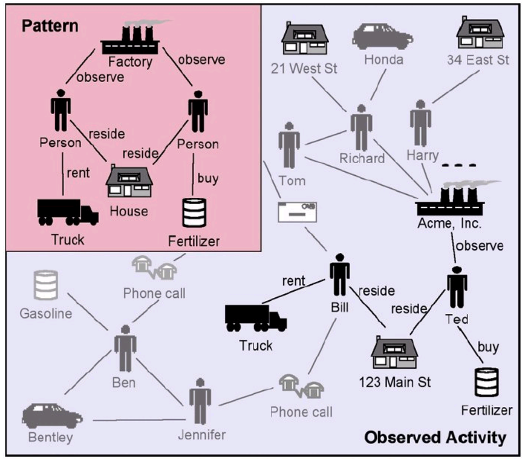
- Explore decomposition results from graph theory, for example forbidden graphs and the Embedding Lemma and determine how to apply such results into the field of graph learning
- Investigate whether these decompositions could assist in a dynamic graph setting

Data-Quad

Unknown

Known Objects Unknown

Patterns



Known

Graph Data Science: Real-world challenges

All involve exascale streaming graphs:

- **Health care** → disease spread, detection and prevention of epidemics/pandemics (e.g. SARS, Avian flu, H1N1 “swine” flu)
- **Massive social networks** → understanding communities, intentions, population dynamics, pandemic spread, transportation and evacuation
- **Intelligence** → business analytics, anomaly detection, security, knowledge discovery from massive data sets
- **Systems Biology** → understanding complex life systems, drug design, microbial research, unravel the mysteries of the HIV virus; understand life, disease,
- **Electric Power Grid** → communication, transportation, energy, water, food supply
- **Modeling and Simulation** → Perform full-scale economic-social-political simulations

REQUIRES PREDICTING / INFLUENCE CHANGE IN REAL-TIME AT SCALE

Graphs are pervasive in large-scale data analysis

- Sources of massive data: peta- and exa-scale simulations, experimental devices, the Internet, scientific applications.
- New challenges for analysis: data sizes, heterogeneity, uncertainty, data quality.

Astrophysics

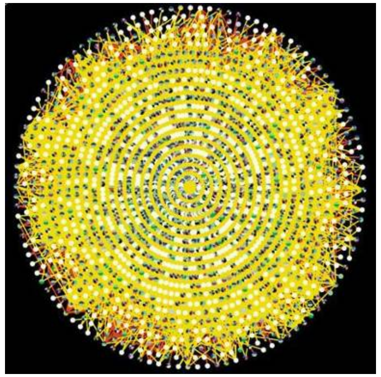
Problem: Outlier detection.
Challenges: massive datasets, temporal variations.
Graph problems: clustering, matching.

Bioinformatics

Problem: Identifying drug target proteins.
Challenges: Data heterogeneity, quality.
Graph problems: centrality, clustering.

Social Informatics

Problem: Discover emergent communities, model spread of information.
Challenges: new analytics routines, uncertainty in data.
Graph problems: clustering, shortest paths, flows.



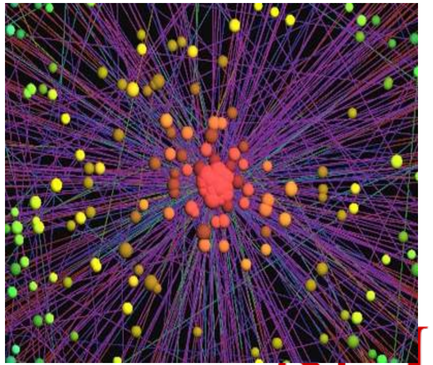
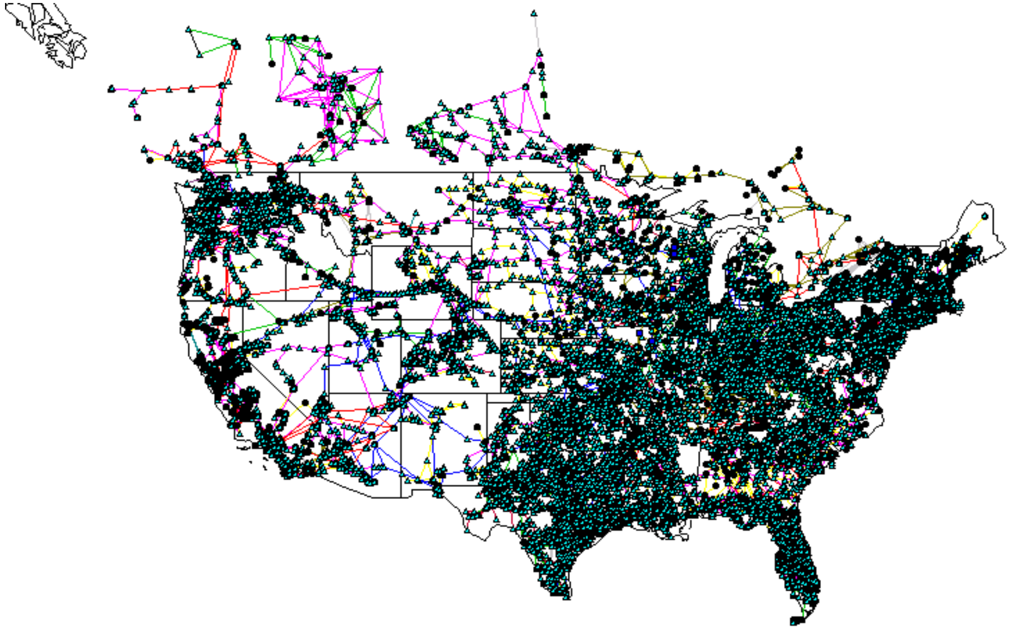
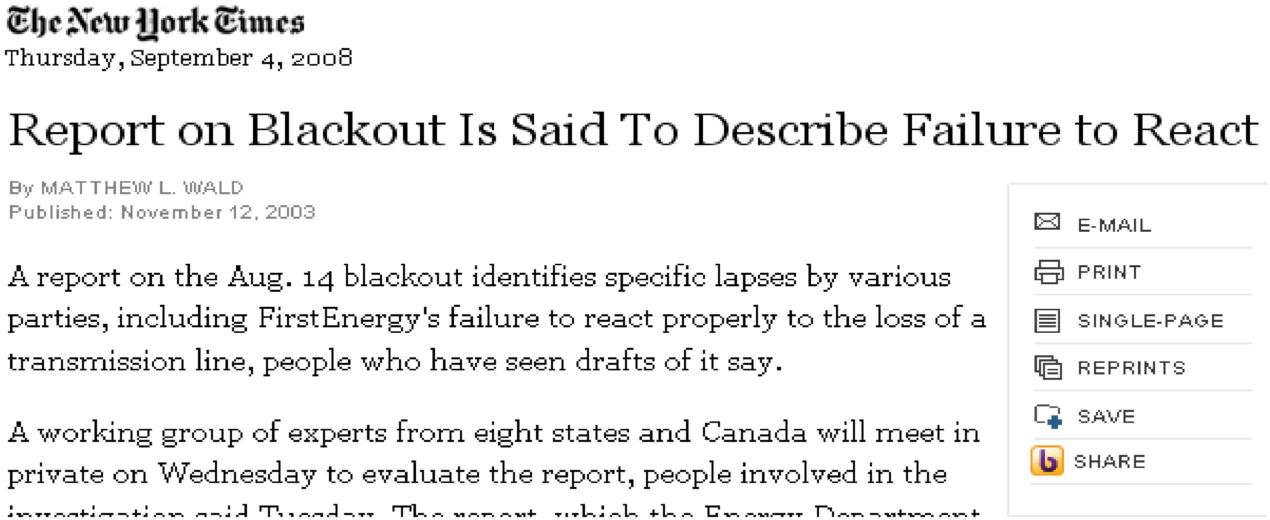


Image sources: (1) http://physics.nmt.edu/images/astro/hst_starfield.jpg
(2,3) www.visualComplexity.com

Massive Data Analytics: Infrastructure

- The U.S. high-voltage transmission grid has >150,000 miles of line.
- Real-time detection of changes and anomalies in the grid is a large-scale problem.
- May mitigate impact of widespread blackouts due to equipment failure or intentional damage.





The New York Times
Thursday, September 4, 2008

Report on Blackout Is Said To Describe Failure to React

By MATTHEW L. WALD
Published: November 12, 2003

A report on the Aug. 14 blackout identifies specific lapses by various parties, including FirstEnergy's failure to react properly to the loss of a transmission line, people who have seen drafts of it say.

A working group of experts from eight states and Canada will meet in private on Wednesday to evaluate the report, people involved in the investigation said Tuesday. The report, which the Energy Department

Network Analysis for Intelligence and Surveillance

- [Krebs ’04] Post 9/11 Terrorist Network Analysis from public domain information
- Plot masterminds correctly identified from interaction patterns: centrality

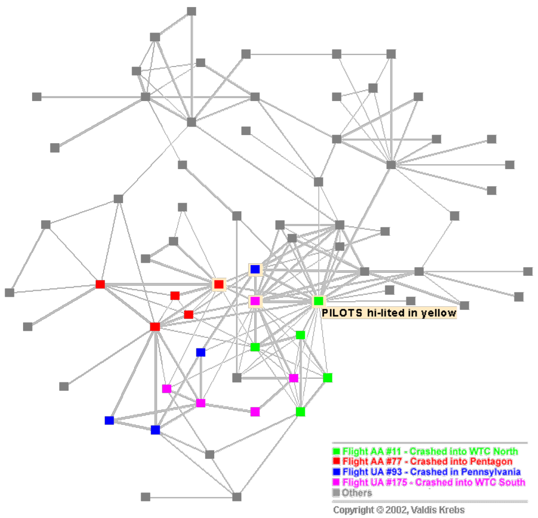

PILOTS hi-lited in yellow
🟩 Flight AA #11 - Crashed into WTC North
🟥 Flight AA #77 - Crashed into Pentagon
🟦 Flight UA #93 - Crashed in Pennsylvania
🟪 Flight UA #175 - Crashed into WTC South
⬛ Others
Copyright © 2002, Valdis Krebs

Image Source: http://www.orgnet.com/hijackers.html

- A global view of entities is often more insightful
- Detect anomalous activities by exact/approximate graph matching

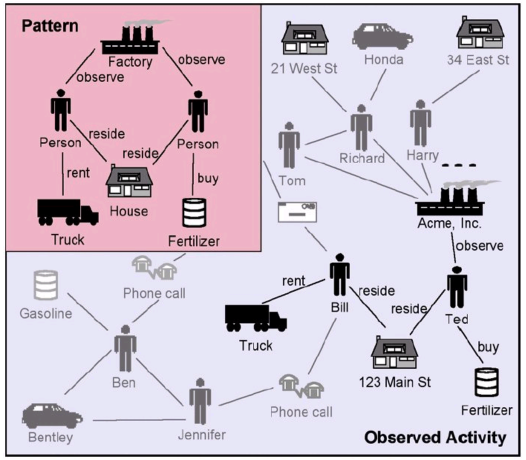
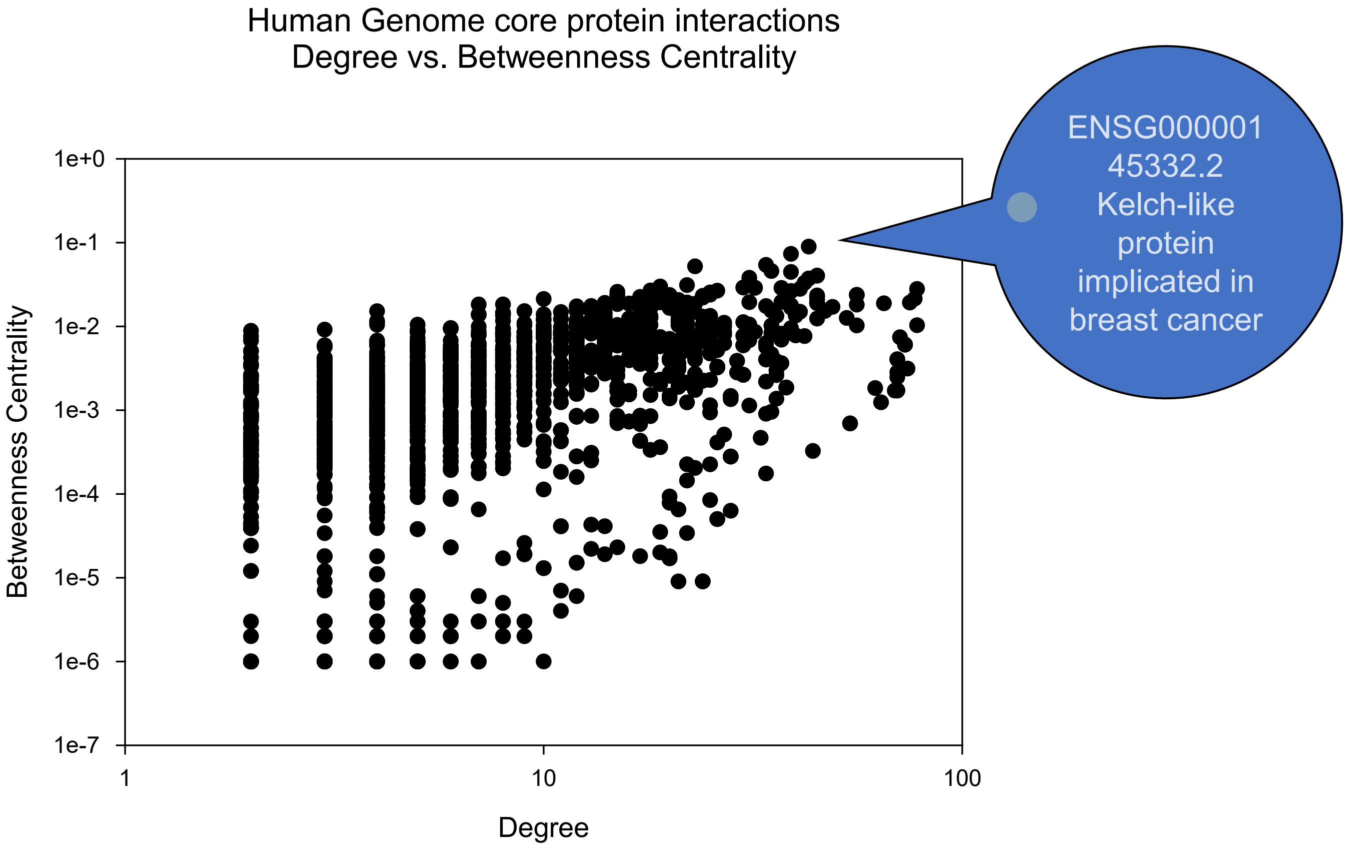


Image Source: T. Coffman, S. Greenblatt, S. Marcus, Graph-based technologies for intelligence analysis, CACM, 47 (3, March 2004): pp 45-47

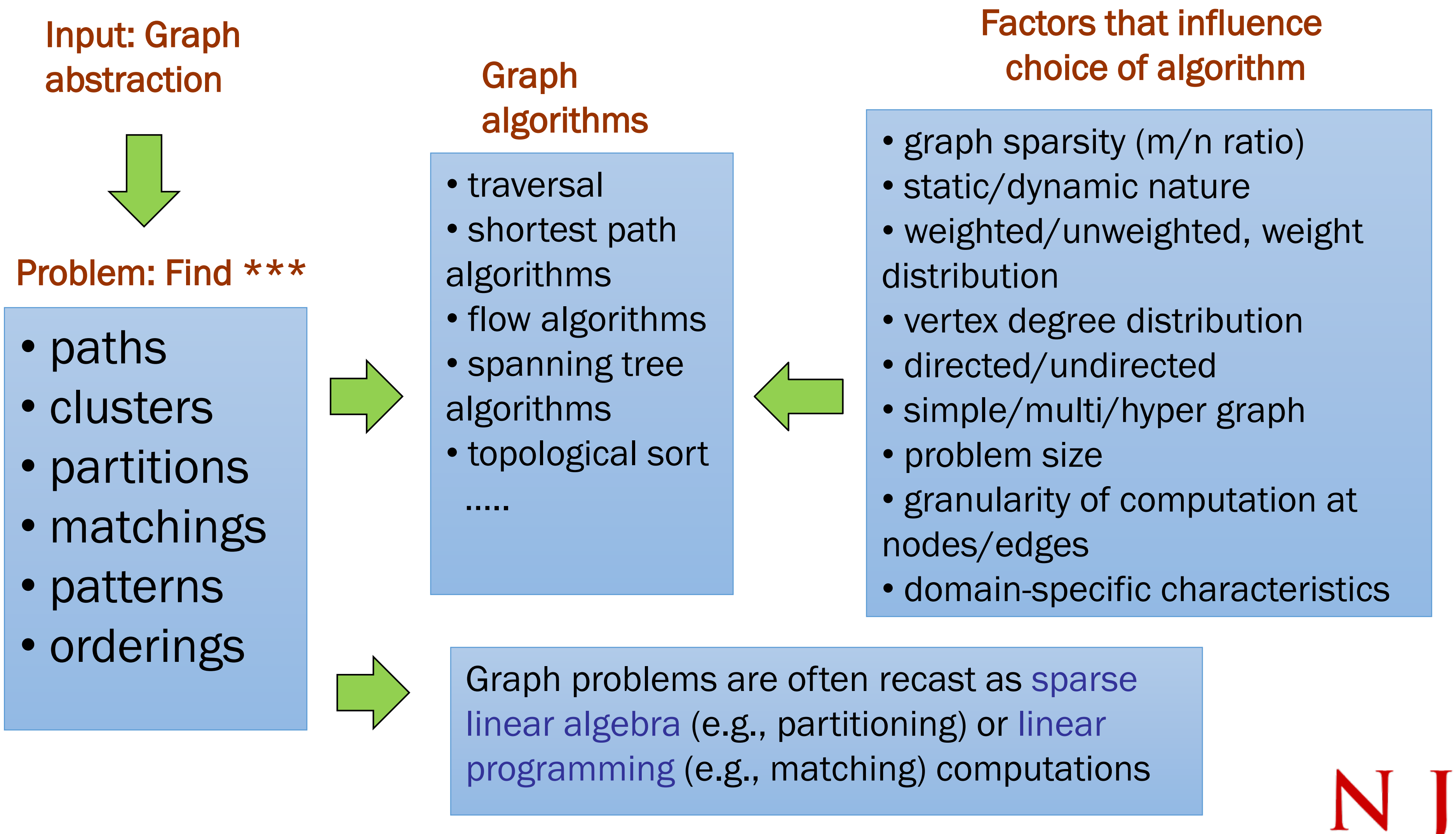
Massive Data Analytics: Public Health

- CDC/national-scale surveillance of public health
- Cancer genomics and drug design
  - Computed Betweenness Centrality of Human Proteome



New Jersey Institute of Technology

Characterizing Graph-theoretic computations



Input: Graph abstraction

Problem: Find ***

- paths
- clusters
- partitions
- matchings
- patterns
- orderings

Graph algorithms

- traversal
- shortest path algorithms
- flow algorithms
- spanning tree algorithms
- topological sort

.....

Factors that influence choice of algorithm

- graph sparsity (m/n ratio)
- static/dynamic nature
- weighted/unweighted, weight distribution
- vertex degree distribution
- directed/undirected
- simple/multi/hyper graph
- problem size
- granularity of computation at nodes/edges
- domain-specific characteristics

Graph problems are often recast as sparse linear algebra (e.g., partitioning) or linear programming (e.g., matching) computations

Streaming Analytics move us from reporting the news to predictive analytics

Traditional HPC

- Great for "static" data sets.
- Massive scalability at the cost of programmability.
- Great for dense problems.
  - Sparse problems typically underutilize the system.

Streaming Analytics

- Requires specialized analytics and data structures.
- Rapidly changing data.
- Low data re-usage.
  - Focused on memory operations and not FLOPS.

Graph Data Science

- Are there new graph techniques? Do they scale? Can the computational systems (algorithms, machines) handle massive networks with billions to trillions of items? Can the techniques tolerate noisy data, massive data, streaming data, etc. …
- **Communities may overlap, exhibit different properties and sizes, and be driven by different models**
  - **Detect communities (static or emerging)**
  - **Identify important individuals**
  - **Detect anomalous behavior**
  - **Given a community, find a representative member of the community**
  - **Given a set of individuals, find the best community that includes them**
  - **Find congestion, weak points, anomalies, surprises, …**

Massive Streaming Graph Analytics

Analysts

(A, B, t1, poke)

(A, C, t2, msg)

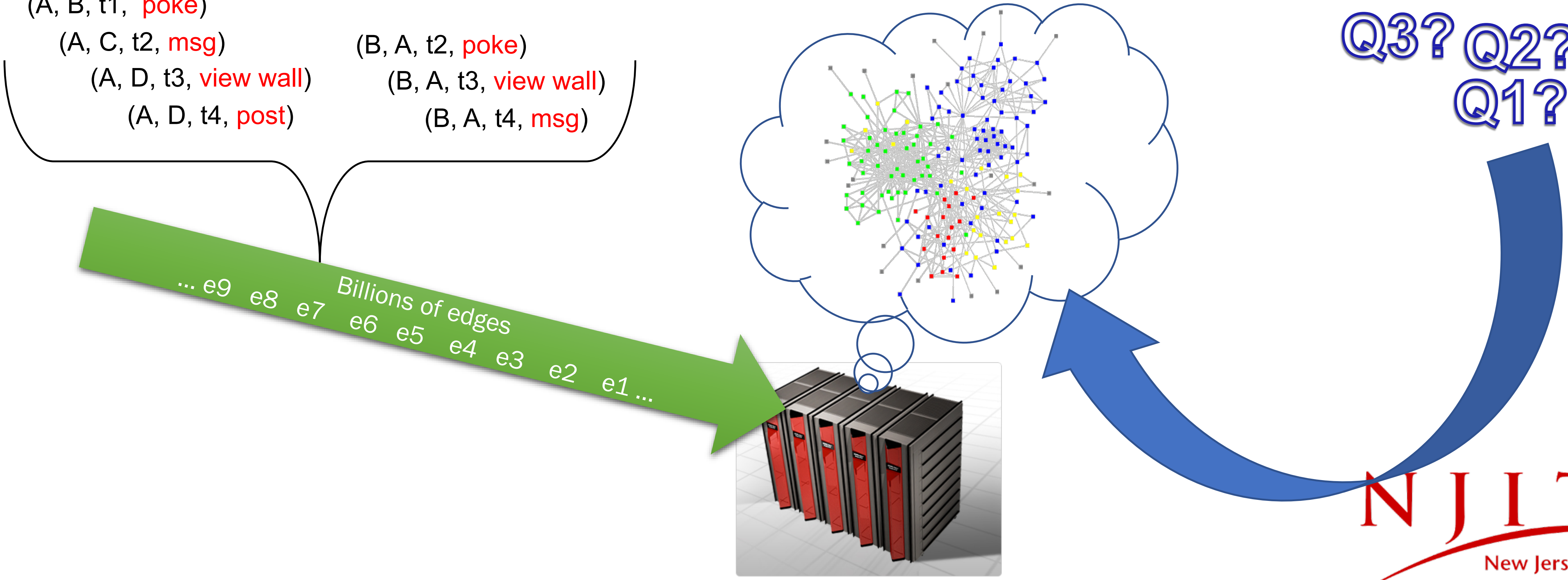
(A, D, t3, view wall)

(A, D, t4, post)

(B, A, t2, poke)

(B, A, t3, view wall)

(B, A, t4, msg)

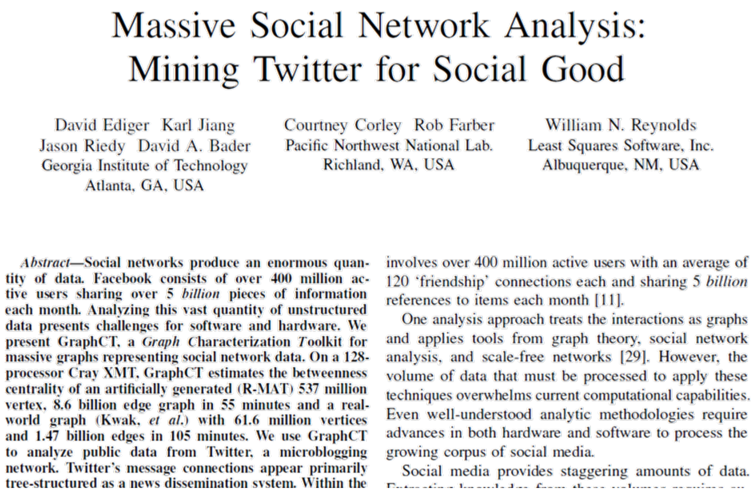


Hierarchy of Interesting Analytics

- **Extend single-shot graph queries to include time.**
  - Are there *s*-*t* paths between time T_1 and T_2?
  - What are the important vertices at time *T*?
- **Use persistent queries to monitor properties.**
  - Does the path between *s* and *t* shorten drastically?
  - Is some vertex suddenly very central?
- **Extend persistent queries to fully dynamic properties.**
  - Does a small community stay independent rather than merge with larger groups?
  - When does a vertex jump between communities?
- **New types of queries, new challenges...**

Mining Twitter for Social Good

ICPP 2010



Massive Social Network Analysis: Mining Twitter for Social Good

David Ediger Karl Jiang
Jason Riedy David A. Bader
Georgia Institute of Technology
Atlanta, GA, USA

Courtney Corley Rob Farber
Pacific Northwest National Lab.
Richland, WA, USA

William N. Reynolds
Least Squares Software, Inc.
Albuquerque, NM, USA

Abstract—Social networks produce an enormous quantity of data. Facebook consists of over 400 million active users sharing over 5 *billion* pieces of information each month. Analyzing this vast quantity of unstructured data presents challenges for software and hardware. We present GraphCT, a *Graph* Characterization *Toolkit* for massive graphs representing social network data. On a 128-processor Cray XMT, GraphCT estimates the betweenness centrality of an artificially generated (R-MAT) 537 million vertex, 8.6 billion edge graph in 55 minutes and a real-world graph (Kwak, *et al.*) with 61.6 million vertices and 1.47 billion edges in 105 minutes. We use GraphCT to analyze public data from Twitter, a microblogging network. Twitter's message connections appear primarily tree-structured as a news dissemination system. Within the

involves over 400 million active users with an average of 120 'friendship' connections each and sharing 5 *billion* references to items each month [11].

One analysis approach treats the interactions as graphs and applies tools from graph theory, social network analysis, and scale-free networks [29]. However, the volume of data that must be processed to apply these techniques overwhelms current computational capabilities. Even well-understood analytic methodologies require advances in both hardware and software to process the growing corpus of social media.

Social media provides staggering amounts of data.

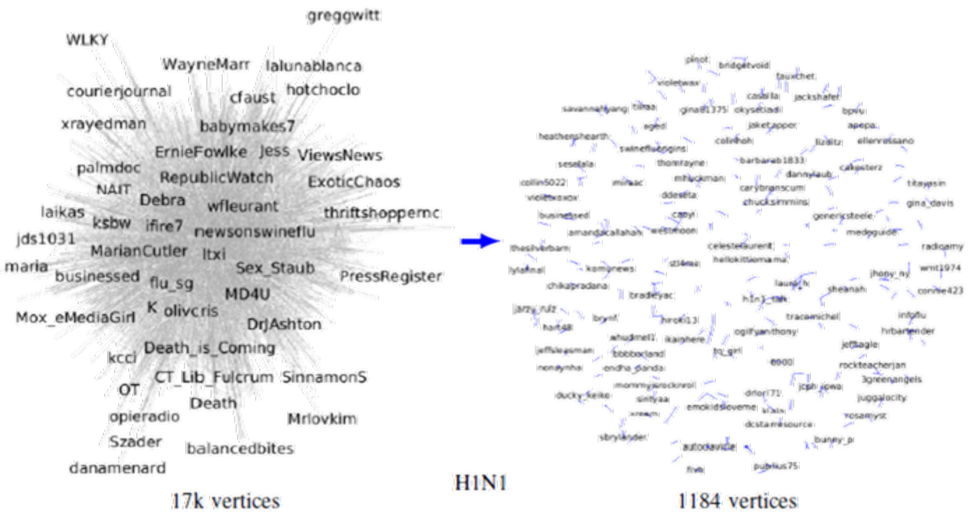


Fig. 3. Subcommunity filtering on Twitter data sets

TOP 15 USERS BY BETWEENNESS CENTRALITY

Rank	Data Set	
	H1N1	atlflood
1	@CDCFlu	@ajc
2	@addthis	@driveafastercar
3	@Official_PAX	@ATLCheap
4	@FluGov	@TWCi
5	@nytimes	@HelloNorthGA
6	@tweetmeme	@11AliveNews
7	@mercola	@WSB_TV
8	@CNN	@shaunking
9	@backstreetboys	@Carl
10	@EllieSmith_x	@SpaceyG
11	@TIME	@ATLINtownPaper
12	@CDCemergency	@TJsDJs
13	@CDC_eHealth	@ATLien
14	@perezhilton	@MarshallRamsey
15	@billmaher	@Kanye

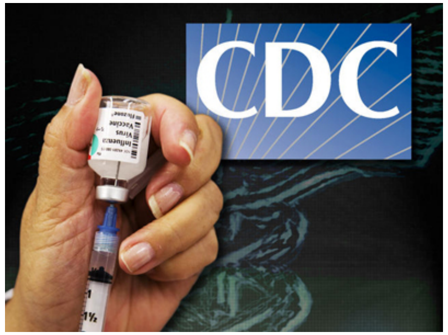


Image credit: bioethicsinstitute.org

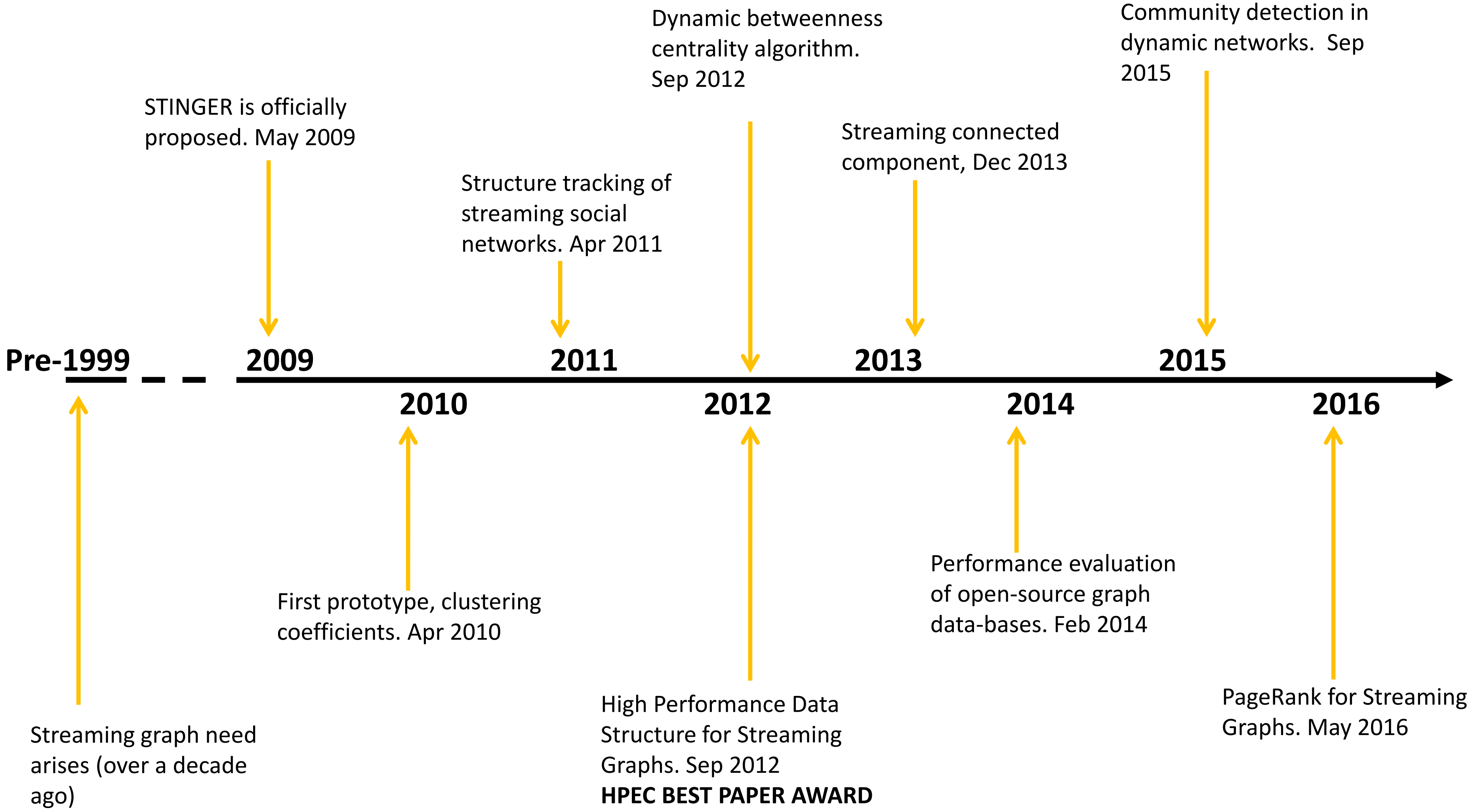
Pacific Northwest NATIONAL LABORATORY

NJIT New Jersey Institute of Technology

STING Initiative: Focusing on Globally Significant Grand Challenges

- Many globally-significant grand challenges can be modeled by **Spatio-Temporal Interaction Networks and Graphs** (or "STING").
- Emerging real-world graph problems include:
  - Detecting community structure in large social networks
  - Defending the nation against cyber-based attacks
  - Discovering insider threats (e.g. Ft. Hood shooter, WikiLeaks)
  - Improving the resilience of the electric power grid
  - Detecting and preventing disease in human populations.
- Unlike traditional applications in computational science and engineering, solving these problems at scale often raises new research challenges due to:
  - Sparsity and the lack of locality in the massive data
  - Design of parallel algorithms for massive, streaming data analytics
  - The need for new exascale supercomputers that are energy-efficient, resilient, and easy-to-program

STINGER – Time Frame



Pre-1999

Streaming graph need arises (over a decade ago)

2009

STINGER is officially proposed. May 2009

2010

First prototype, clustering coefficients. Apr 2010

2011

Structure tracking of streaming social networks. Apr 2011

2012

Dynamic betweenness centrality algorithm. Sep 2012

High Performance Data Structure for Streaming Graphs. Sep 2012
HPEC BEST PAPER AWARD

2013

Streaming connected component, Dec 2013

2014

Performance evaluation of open-source graph data-bases. Feb 2014

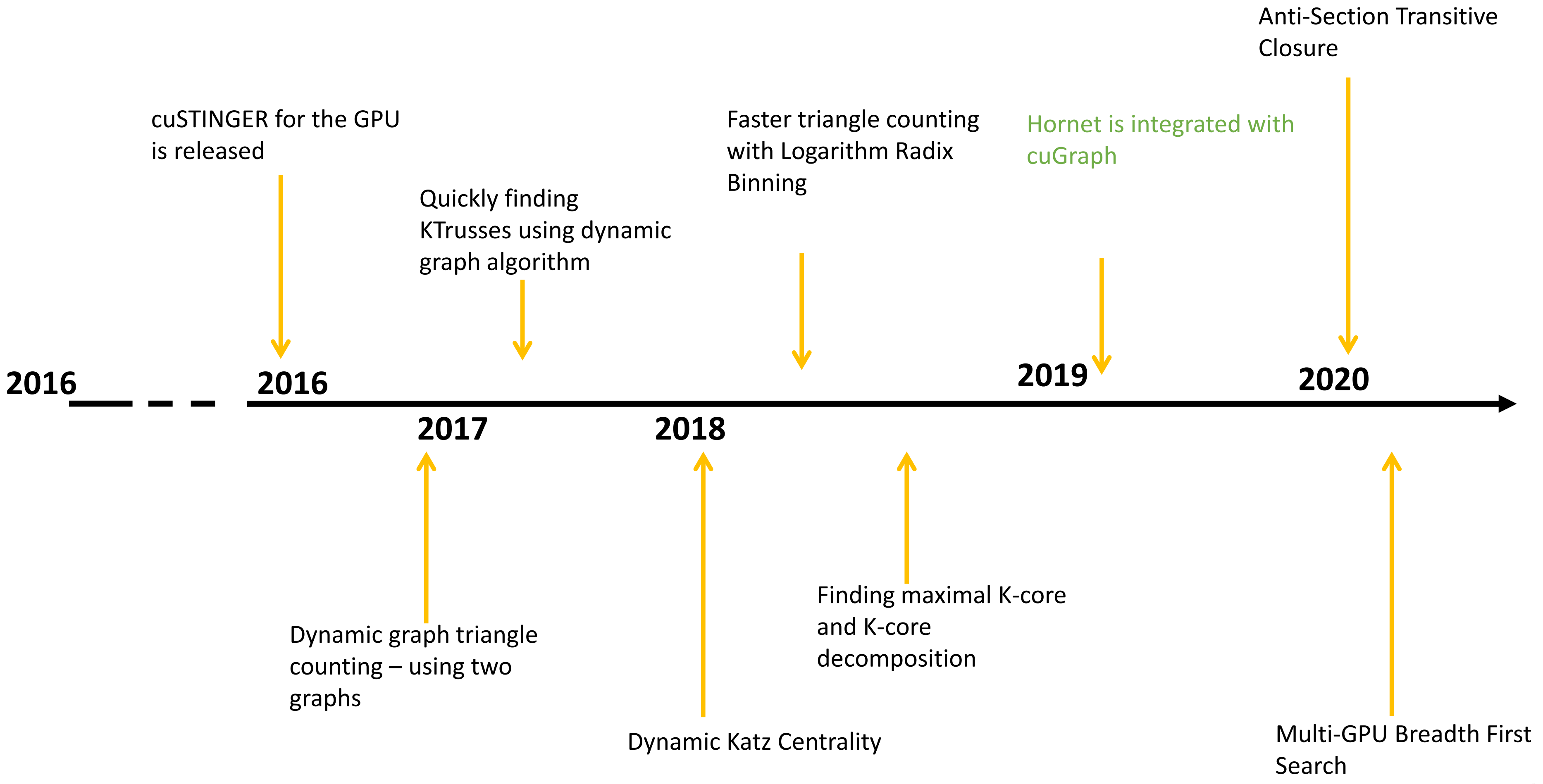
2015

Community detection in dynamic networks. Sep 2015

2016

PageRank for Streaming Graphs. May 2016

Hornet (GPU only) – Time Frame



2016

2016

cuSTINGER for the GPU is released

2017

Dynamic graph triangle counting – using two graphs

Quickly finding KTrusses using dynamic graph algorithm

2018

Dynamic Katz Centrality

Faster triangle counting with Logarithm Radix Binning

Finding maximal K-core and K-core decomposition

2019

Hornet is integrated with cuGraph

2020

Anti-Section Transitive Closure

Multi-GPU Breadth First Search

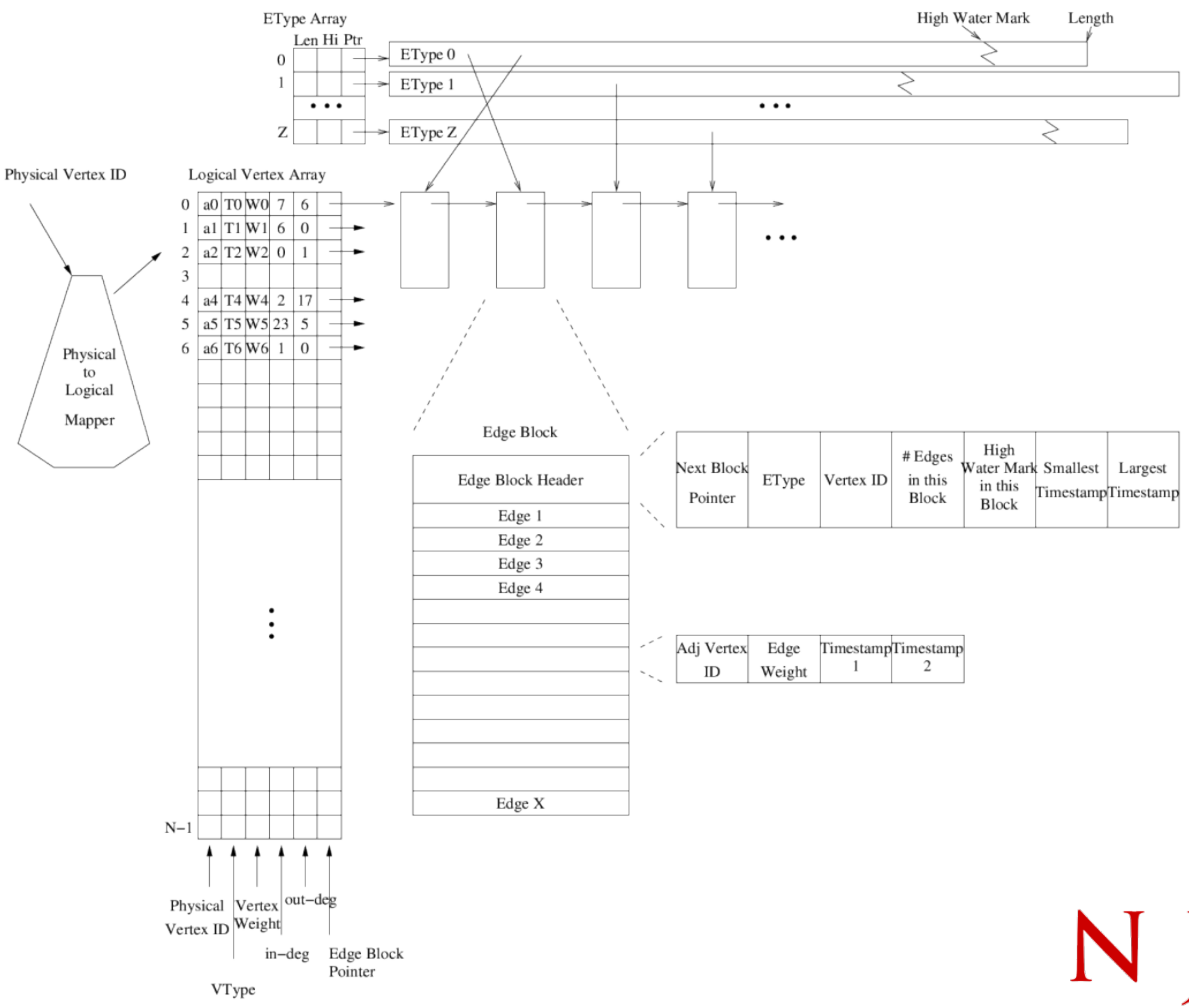
STING Extensible Representation (STINGER)

Design goals

- Enable algorithm designers to implement dynamic graph algorithms with ease.
- Portable semantics for various platforms
- Good performance for all types of graph problems and algorithms - static and dynamic.
- Assumes globally addressable memory access
- Support multiple, parallel readers and a single writer
  - One server manages the graph data structures
  - Multiple analytics run in background with read-only permissions.

STING Extensible Representation (STINGER)

- Semi-dense edge list blocks with free space
- Compactly stores timestamps, types, weights
- Maps from application IDs to storage IDs
- Deletion by negating IDs, separate compaction



STINGER as an analysis package

http://www.stingergraph.com/

Anything that a static graph package can do (and a whole lot more):

Parallel agglomerative clustering:
Find clusters that are optimized for a user-defined edge scoring function.

K-core Extraction:
Extract additional communities and filter noisy high-degree vertices.

Classic breadth-first search:
Performs a parallel breadth-first search of the graph starting at a given source vertex to find shortest paths.

Parallel connected components:
Finds the connected components in a static network.

AND…

Streaming edge insertions and deletions:
New edge insertions, updates, and deletions in batches or individually. Optimized to update at rates of over 3 million edges per second on graphs of one billion edges.

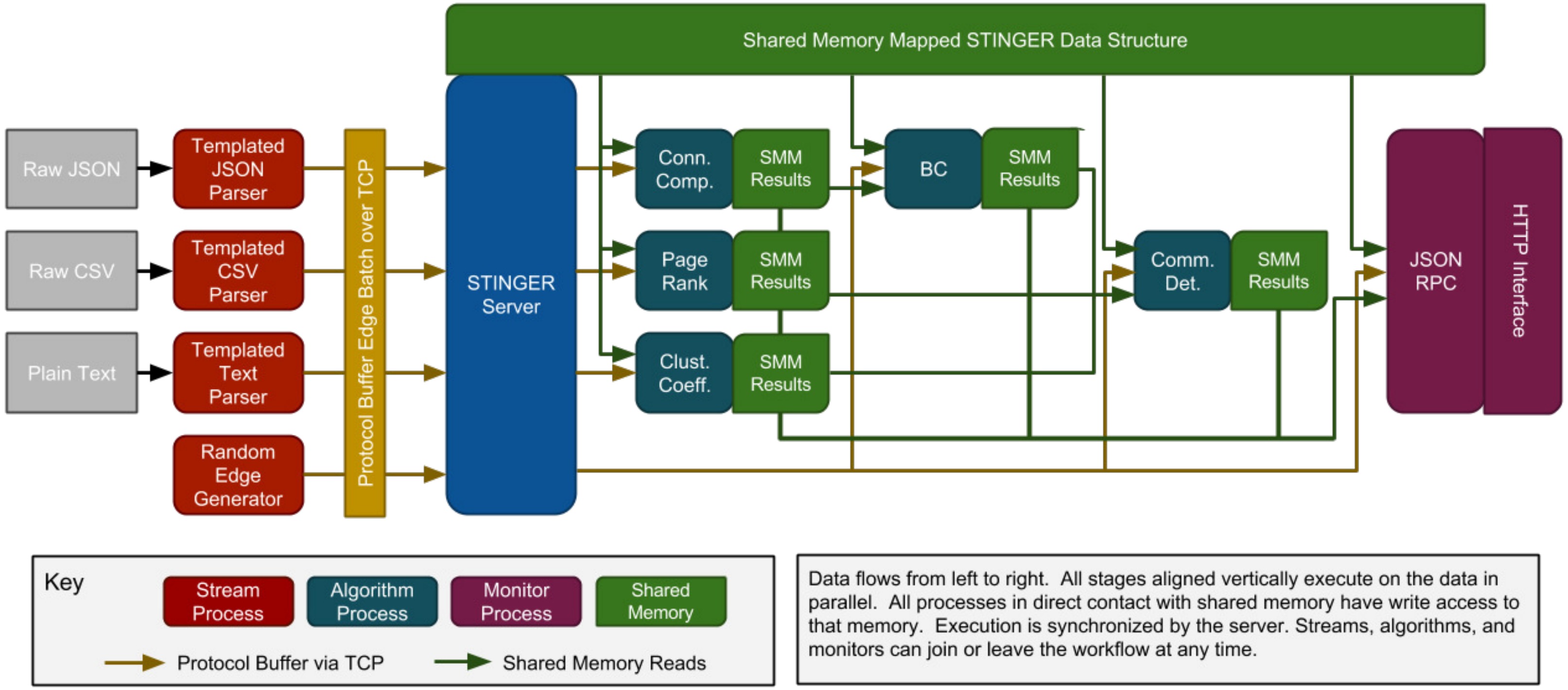
Streaming clustering coefficients:
Tracks the local and global clustering coefficients of a graph.

Streaming connected components:
Real time tracking of the connected components.

Streaming Betweenness Centrality:
Find the key points within information flows and structural vulnerabilities.

Streaming community detection:
Track and update the community structures within the graph as they change.

STING: High-level architecture



Data flows from left to right. All stages aligned vertically execute on the data in parallel. All processes in direct contact with shared memory have write access to that memory. Execution is synchronized by the server. Streams, algorithms, and monitors can join or leave the workflow at any time.

- Server: Graph storage, kernel orchestration
- OpenMP + sufficiently POSIX-ish
- Multiple processes for resilience

Why not existing technologies?

- Traditional SQL databases
  - Not structured to do any meaningful graph queries with any level of efficiency or timeliness
- Graph databases - mostly on-disk
  - Distributed disk can keep up with storing / indexing, but is simply too slow at random graph access to process on as the graph updates
- Hadoop and HDFS-based projects
  - Not really the right programming model for many structural queries over the entire graph, random access performance is poor
- Smaller graph libraries, processing tools
  - Can't scale, can't process dynamic graphs, frequently leads to impossible visualization attempts

Conclusions

- Solving massive-scale analytics will require new
  - High-performance computing platforms
  - Streaming algorithms
  - Energy-efficient implementations
- Mapping applications to high performance architectures may yield performance improvements of six or more orders of magnitude.
- Solving real-world challenges such as:
  - Urban sustainability
  - Healthcare analytics
  - Trustworthy, Free and Fair Elections
  - Insider threat detection
  - Utility infrastructure protection
  - Cyberattack defense
  - Disease outbreak and epidemic monitoring

Acknowledgments

- Dr. Zhihui Du, NJIT
- Oliver Alvarado Rodriguez, NJIT
- Oded Green, (NVIDIA)
- Recent Bader Alumni:
  - **Dr. Eisha Nathan** (Lawrence Livermore National Lab)
  - **Dr. Vipin Sachdeva** (IBM)
  - **Dr. Anita Zakrzewska** (Trovares)
  - **Dr. Lluis Miquel Munguia** (Google)
  - **Prof. Kamesh Madduri** (Penn State)
  - **Dr. David Ediger** (GTRI)
  - **Prof. James Fairbanks** (University of Florida)
  - **Dr. Seunghwa Kang** (NVIDIA)