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 $183M of research awards
- 230+ publications, ≥ 8,400 citations, h-index ≥ 54
- National Science Foundation CAREER Award recipient
- Directed: NVIDIA GPU Center of Excellence
- 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 Jumps into Top 100 for 2020 U.S. News College Rankings


9 September 2019
America’s Great Working-Class Colleges

David Leonhardt

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.

<table>
<thead>
<tr>
<th>Rank</th>
<th>College</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New Jersey Institute of Technology</td>
<td>85%</td>
</tr>
<tr>
<td>2</td>
<td>Pace University</td>
<td>82%</td>
</tr>
<tr>
<td>3</td>
<td>Calif. State – Bakersfield</td>
<td>82%</td>
</tr>
<tr>
<td>4</td>
<td>Univ. California – Irvine</td>
<td>81%</td>
</tr>
<tr>
<td>5</td>
<td>Calif. Poly – Pomona</td>
<td>81%</td>
</tr>
<tr>
<td>6</td>
<td>Xavier of Louisiana</td>
<td>80%</td>
</tr>
<tr>
<td>7</td>
<td>SUNY Stony Brook</td>
<td>79%</td>
</tr>
<tr>
<td>8</td>
<td>San Jose State</td>
<td>79%</td>
</tr>
<tr>
<td>9</td>
<td>CUNY Baruch College</td>
<td>79%</td>
</tr>
<tr>
<td>10</td>
<td>Calif. State – Long Beach</td>
<td>78%</td>
</tr>
</tbody>
</table>
Founded in 1880
Awarded 2,500+ Degrees in 2015
2016 Fall enrollment > 11,000 students

**Forbes**
April 10, 2018

**AMERICAN DREAM U**
BEST VALUE COLLEGES WITH THE HIGHEST UPWARD MOBILITY RATES

We looked at the top 100 Forbes Best Value Colleges 2018 to identify the 10 schools that had the most success with upward mobility—the percentage of students from the bottom 20% income distribution who reach the top 20%—and then ranked them by alumni with the highest mid-career salaries.

<table>
<thead>
<tr>
<th>School</th>
<th>Average Mid-Career Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEW JERSEY INSTITUTE OF TECHNOLOGY</td>
<td>$119.7K</td>
</tr>
<tr>
<td>UNIVERSITY OF CALIFORNIA IRVINE</td>
<td>$118.8K</td>
</tr>
<tr>
<td>UCLA</td>
<td>$114.8K</td>
</tr>
<tr>
<td>STONY BROOK UNIVERSITY</td>
<td>$108.8K</td>
</tr>
<tr>
<td>CALIFORNIA STATE POLYTECHNIC UNIVERSITY</td>
<td>$108.3K</td>
</tr>
<tr>
<td>CUNY BARUCH COLLEGE</td>
<td>$103K</td>
</tr>
<tr>
<td>CUNY QUEENS COLLEGE</td>
<td>$95.2K</td>
</tr>
<tr>
<td>CUNY BROOKLYN COLLEGE</td>
<td>$94.8K</td>
</tr>
<tr>
<td>CUNY CITY COLLEGE</td>
<td>$90.4K</td>
</tr>
<tr>
<td>CUNY HUNTER COLLEGE</td>
<td>$87.5K</td>
</tr>
</tbody>
</table>

DATA: The Equal Opportunity Project, Payscale
The Colleges

- Engineering: 52%
- Computing: 25%
- Science and Liberal Arts: 10%
- Management: 7%
- Architecture and Design: 6%
- Honors: 6%
Computing at NJIT

• Largest computing program in the NY/NJ metro area

• All undergraduate students take a capstone project course – working with companies

• Active internships and co-op programs in cooperation with industry, government and others

• Student clubs organize extra-curricula activities, e.g. hackathons and field trips, and provide tutoring
NJIT Joins the Ranks of the Nation’s Elite Research Universities With its New R1 Classification

Written By: Tracey Raizman
Cyber Innovations for Solving Global Grand Challenges

LANL Roadrunner with IBM Cell B.E. Top500 No. 1 system from June 2008 to June 2009

Intel HIVE processor (2019)

IBM Watson with POWER7/8, won Jeopardy in Feb 2010

NVIDIA GPUs used in 127 of Top500 systems, incl. top 2 (in USA), and fastest in Europe and Japan. (Nov. 2018)

Cray XMT with ThreadStorm proc. Massively Multithreaded Architecture

IBM BlueGene/Q. Record breaking performance over 10PF sustained on science apps
Data Science: Discovery and Innovation

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

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.

“Advances in communications and the democratization of other technologies have also generated an ability to create and share vast and exponentially growing amounts of information farther and faster than ever before. This abundance of data provides significant opportunities for the IC, including new avenues for collection and the potential for greater insight, but it also challenges the IC’s ability to collect, process, evaluate, and analyze such enormous volumes of data quickly enough to provide relevant and useful insight to its customers.”

→ “Develop and maintain capabilities to acquire and evaluate data to obtain a deep understanding of the global political, diplomatic, military, economic, security, and informational environment.”
Exascale Streaming Data Analytics: Real-world challenges

All involve analyzing massive streaming complex networks:

- **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

Ex: discovered minimal changes in O(billions)-size complex network that could hide or reveal top influencers in the community

**Sample queries:**

- **Allegiance switching:** identify entities that switch communities.
- **Community structure:** identify the genesis and dissipation of communities
- **Phase change:** identify significant change in the network structure

**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](http://physics.nmt.edu/images/astro/hst_starfield.jpg)  
(2,3) [www.visualComplexity.com](http://www.visualComplexity.com)
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

• A global view of entities is often more insightful

• Detect anomalous activities by exact/approximate graph matching

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

Characterizing Graph-theoretic computations

**Input: Graph abstraction**
- paths
- clusters
- partitions
- matchings
- patterns
- orderings

**Problem: Find ***

**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.
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)

Billions of edges... e9 e8 e7 e6 e5 e4 e3 e2 e1...
Mining Twitter for Social Good

ICPP 2010

Massive Social Network Analysis: Mining Twitter for Social Good

David Ediger, Karl Jiang, Jason Ricely, David A. Bader, and William N. Reynolds
Georgia Institute of Technology
Atlanta, GA, 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 X1M, GraphCT estimates the betweenness centrality of an artificially generated R-MAT 557 million vertex, 86 billion edge graph in 25 minutes and a real-world graph (Kwak, et al.) with 614 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 avg 120 ‘friendship’ connections each and sharing 5 references to items each month [11].

One analysis approach treats the interactions as 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 advances in both hardware and software to process growing corpus of social media.

Social media provides staggering amounts of data: 400 million tweets per minute; 340 million videos per day; over 3000 tweets per second; 300 million photos per day; and, 17 billion photos on Instagram. In comparison, the raw data from a major cell phone network provider contains over 800 million users with 4000 million text messages, 2500 million calls each day. These numbers may be skewed for the social media model.

Thus, researchers are now looking beyond the traditional social network model to explain the social network data. The question is which model makes the most sense for the underlying data, and how social network data relates to social media. Understanding these models is critical to designing and implementing effective social media analysis and analytics.

Top 15 Users by Betweenness Centrality

<table>
<thead>
<tr>
<th>Rank</th>
<th>HIN1</th>
<th>Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>@CDCflu</td>
<td>@ajc</td>
</tr>
<tr>
<td>2</td>
<td>@addthis</td>
<td>@driveafaster</td>
</tr>
<tr>
<td>3</td>
<td>@Official_PAX</td>
<td>@ATLCheap</td>
</tr>
<tr>
<td>4</td>
<td>@FluGov</td>
<td>@TWCi</td>
</tr>
<tr>
<td>5</td>
<td>@nytimes</td>
<td>@HelloNorthGA</td>
</tr>
<tr>
<td>6</td>
<td>@tweetmeme</td>
<td>@11AliveNews</td>
</tr>
<tr>
<td>7</td>
<td>@morcola</td>
<td>@WSB_TV</td>
</tr>
<tr>
<td>8</td>
<td>@CNN</td>
<td>@shaunking</td>
</tr>
<tr>
<td>9</td>
<td>@backstreetboys</td>
<td>@Carl</td>
</tr>
<tr>
<td>10</td>
<td>@EllieSmith_x</td>
<td>@SpaceyG</td>
</tr>
<tr>
<td>11</td>
<td>@TIME</td>
<td>@ATLINtownPA</td>
</tr>
<tr>
<td>12</td>
<td>@CDCEmergency</td>
<td>@TJsDJs</td>
</tr>
<tr>
<td>13</td>
<td>@CDC_eHealth</td>
<td>@ATLien</td>
</tr>
<tr>
<td>14</td>
<td>@perezhilton</td>
<td>@MarshallRamsey</td>
</tr>
<tr>
<td>15</td>
<td>@billmaher</td>
<td>@Kanye</td>
</tr>
</tbody>
</table>

Top 15 Users by Betweenness Centrality

Image credit: bioethicsinstitute.org
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

- STINGER is officially proposed. May 2009

2009

- Streaming graph need arises (over a decade ago)
- 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
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.

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.

AND…
STING: High-level architecture

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

• Massive-Scale Streaming Analytics require
  • Simple programming model
    • Simple API.
    • CSR-like in concept.
    • STINGER has a lot more under the hood.
  • Extremely fast updates
    • Millions of updates per second.
    • These must not be bottlenecks for updating an analytic.
    • STINGER offers these

• STINGER has major performance benefits
  • Thousands of times faster than static graph computation.
  • Hundreds of thousands of updates per second for numerous analytics.
  • Real-time monitoring of underlying network.
• Today homeland security knowledge is captured in reports
  • These are written in a natural language, can be multiple pages, include references, footnotes, figures, etc.

• Analysts manually discover and re-discover relationships among the unclassified security reports and open data
  • Searching and filtering help uncover relationships, but all complex queries must be done by hand and manually

• Graphs are a natural and important structure for connecting this knowledge
Approach

• We developed a knowledge graph for the Department of Homeland Security, with a focus on special events, working with DHS/FBI/NYPD

• This knowledge graph is built on top of:
  • Reports
  • Spreadsheets
  • Raw datasets

• Enterprise-class performance with STINGER
Using the Knowledge Graph

After the graph has been created, relationships between reports and events can be maintained.

Analysts are able to use this graph to quickly triage new events / information and link together otherwise hard-to-connect data.

Furthermore, advanced and predictive analytics can be run to gain further insight.
Data Sources

Several sources contribute to homeland security knowledge graph

Focus on three datasets:

- A collection of reports provided by partners
- The Global Terrorism Database (GTD) from the University of Maryland
- The Homegrown Violent Extremists (HVE) list from the University of Nebraska Omaha
Ingesting Unstructured Text

Vertices: Proper nouns were extracted and given labels of *Media*, *Events*, *Groups*, *Individuals*, and *Locations*

Properties: Attacks and threats were extracted and placed into the *AttackType* label (e.g. explosive devices or vehicle attacks)

Edges were created when a report indicated a connection between the vertices
On November 28, 2016, a terrorist vehicle-ramming and stabbing attack occurred at 9:52 a.m. EST at Ohio State University (OSU)'s Watts Hall in Columbus, Ohio. The attacker, Somali refugee Abdul Razak Ali Artan, was shot and killed by the first responding OSU police officer, and 13 people were hospitalized for injuries.

(source: Wikipedia)
Ingested Data

- 29 Joint Threat Assessments reports were parsed
  - Provided by DHS
  - Recent (2016 to 2018) and relevant to the problem
- 114 vertices, 163 edges
Ingesting the Global Terrorism Database / HVE

• Database is provided in the form of a spreadsheets with structured coding

• Converted the spreadsheets to CSV files, and then ingested primarily using `LOAD CSV`

• Pre-processing performed with Python scripts

Images from http://www.start.umd.edu/gtd/
Graph Schema and Ingestion

- Schema is focused around database rows
  - This creates a nearly bipartite graph from rows and columns
- Other connections are then created through queries
- **325K vertices, 2.7M edges**
Running Analytics on the Homeland Security Knowledge Graph

Louvain running on US events, visualized in Gephi

Node Importance

<table>
<thead>
<tr>
<th>Id</th>
<th>eigencentrality</th>
<th>pageranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1</td>
<td>0.024539</td>
</tr>
<tr>
<td>6</td>
<td>0.998713</td>
<td>0.024502</td>
</tr>
<tr>
<td>5</td>
<td>0.99301</td>
<td>0.024298</td>
</tr>
<tr>
<td>8</td>
<td>0.846495</td>
<td>0.020458</td>
</tr>
<tr>
<td>7</td>
<td>0.829617</td>
<td>0.019987</td>
</tr>
<tr>
<td>28</td>
<td>0.529627</td>
<td>0.012339</td>
</tr>
<tr>
<td>22</td>
<td>0.525788</td>
<td>0.012185</td>
</tr>
<tr>
<td>16</td>
<td>0.479301</td>
<td>0.011143</td>
</tr>
<tr>
<td>32</td>
<td>0.306425</td>
<td>0.007603</td>
</tr>
<tr>
<td>39</td>
<td>0.301179</td>
<td>0.007432</td>
</tr>
<tr>
<td>67</td>
<td>0.289466</td>
<td>0.006846</td>
</tr>
<tr>
<td>18</td>
<td>0.219309</td>
<td>0.005264</td>
</tr>
<tr>
<td>27</td>
<td>0.198742</td>
<td>0.005009</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eigencentrality</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.198742</td>
<td>Unknown</td>
</tr>
<tr>
<td>0.073040</td>
<td>Anti-Abortion extremists</td>
</tr>
<tr>
<td>0.063670</td>
<td>Left-Wing Militants</td>
</tr>
<tr>
<td>0.047164</td>
<td>FALN</td>
</tr>
<tr>
<td>0.032517</td>
<td>New World Liberation Front</td>
</tr>
<tr>
<td>0.030430</td>
<td>Black Nationalists</td>
</tr>
<tr>
<td>0.028223</td>
<td>White extremists</td>
</tr>
<tr>
<td>0.027993</td>
<td>Animal Liberation Front</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

8 August 2019  David A. Bader
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

• **Jason Riedy**, Research Scientist, (Georgia Tech)
• **Oded Green**, Research Scientist, (Georgia Tech)
• Current Graduate Students (Georgia Tech):
  • Xiaojing An
  • James Fox
  • Kasimir Gabert
  • Euna Kim
• Recent Bader Alumni:
  • Dr. Eisha Nathan (Lawrence Livermore National Lab)
  • Dr. Vipin Sachdeva (IBM)
  • Dr. Anita Zakrzewska (Lawrence Livermore National Lab)
  • Dr. Lluis Miquel Munguia (Google)
  • Prof. Kamesh Madduri (Penn State)
  • Dr. David Ediger (GTRI)
  • Dr. James Fairbanks (GTRI)
  • Dr. Seunghwa Kang (Pacific Northwest National Lab)
Contribution

• Developed new algorithm for local community detection that incrementally updates results
• Outputs results similar to static re-computation
  • Average recall and precision 0.80-0.99 and 0.59-0.98 across real graphs
  • Recall and precision do not decrease over time
  • Can start with no data and build community
• Faster than re-computing with static algorithm
  • Dynamic algorithm most beneficial for small batch sizes
  • Reaches up to two orders of magnitude dynamic speedup
Bader, Related Recent Publications (2005-2009)


- Karl Jiang, David Ediger, and David A. Bader. “Generalizing k-Betweenness Centrality Using Short Paths and a Parallel Multithreaded Implementation.” The 38th International Conference on Parallel Processing (ICPP), Vienna, Austria, September 2009.


Bader, Related Recent Publications (2010-2011)

• David Ediger, Karl Jiang, E. Jason Riedy, and David A. Bader. “Massive Streaming Data Analytics: A Case Study with Clustering Coefficients,” Fourth Workshop in Multithreaded Architectures and Applications (MTAAP), Atlanta, GA, April 2010.

• Seunghwa Kang, David A. Bader. “Large Scale Complex Network Analysis using the Hybrid Combination of a MapReduce cluster and a Highly Multithreaded System,” Fourth Workshop in Multithreaded Architectures and Applications (MTAAP), Atlanta, GA, April 2010.


• Virat Agarwal, Fabrizio Petrini, Davide Pasetto and David A. Bader. “Scalable Graph Exploration on Multicore Processors,” The 22nd IEEE and ACM Supercomputing Conference (SC10), New Orleans, LA, November 2010.


Bader, Related Recent Publications (2012)


Bader, Related Recent Publications (2013)


- David A. Bader, Henning Meyerhenke, Peter Sanders, and Dorothea Wagner (eds.), Graph Partitioning and Graph Clustering, American Mathematical Society, 2013.

- E. Jason Riedy, Henning Meyerhenke, David Ediger and David A. Bader, "Parallel Community Detection for Massive Graphs," in David A. Bader, Henning Meyerhenke, Peter Sanders, and Dorothea Wagner (eds.), Graph Partitioning and Graph Clustering, American Mathematical Society, Chapter 14, pages 207-222, 2013.


- J. Riedy and D.A. Bader, "Multithreaded Community Monitoring for Massive Streaming Graph Data," 7th Workshop on Multithreaded Architectures and Applications (MTAAP), Boston, MA, May 24, 2013.

- D. Ediger and D.A. Bader, "Investigating Graph Algorithms in the BSP Model on the Cray XMT," 7th Workshop on Multithreaded Architectures and Applications (MTAAP), Boston, MA, May 24, 2013.

- O. Green and D.A. Bader, "Faster Betweenness Centrality Based on Data Structure Experimentation," International Conference on Computational Science (ICCS), Barcelona, Spain, June 5-7, 2013.

- Z. Yin, J. Tang, S. Schaeffer, and D.A. Bader, "Streaming Breakpoint Graph Analytics for Accelerating and Parallelizing the Computation of DCJ Median of Three Genomes," International Conference on Computational Science (ICCS), Barcelona, Spain, June 5-7, 2013.

- T. Senator, D.A. Bader, et al., "Detecting Insider Threats in a Real Corporate Database of Computer Usage Activities," 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Chicago, IL, August 11-14, 2013.


- A. Zakrzewska and D.A. Bader, "Measuring the Sensitivity of Graph Metrics to Missing Data," 10th International Conference on Parallel Processing and Applied Mathematics (PPAM), Warsaw, Poland, September 8-11, 2013.


- A. McLaughlin and D.A. Bader, "Revisiting Edge and Node Parallelism for Dynamic GPU Graph Analytics," 8th Workshop on Multithreaded Architectures and Applications (MTAAP), held in conjunction with The IEEE International Parallel and Distributed Processing Symposium (IPDPS 2014), Phoenix, AZ, May 23, 2014.


Bader, Related Recent Publications (2016-2017)


- Anita Zakrzewska and David A. Bader, “Aging Data in Dynamic Graphs: A Comparative Study,” 2nd International Workshop on Dynamics in Networks (DyNo), held in conjunction with IEEE/ACM International Conference on Advances in Social Networks Analysis and Modeling (ASONAM), San Francisco, CA, August 18, 2016.


Bader, Related Recent Publications (2018)