

Grigory Yaroslavtsev, <http://grigory.us>

U.S. permanent resident.

3130 Luddy Hall, Bloomington, IN, 47405

E-mail: grigory@grigory.us

INTERESTS Algorithms for big data analysis, machine learning and data science, data privacy.

2016– **Indiana University**, Bloomington, IN.

Assistant Professor of Computer Science. Founding director of the Center for Algorithms and Machine Learning (CAML) <http://caml.indiana.edu>.

2014–2016 **University of Pennsylvania**, Philadelphia, PA.

Postdoctoral Fellow at the **Warren Center for Network and Data Sciences**, hosted by the departments of Computer and Information Sciences and Statistics at the Wharton Business School. Mentors: Michael Kearns (CIS) and Elchanan Mossel (Stat).

2013–2014 **Brown University ICERM**, Providence, RI.

Institute Postdoctoral Fellow. Mentor: Philip Klein.

2010–2013 **Pennsylvania State University**, State College, PA

Ph.D., Thesis: “Efficient Combinatorial Techniques in Sparsification, Summarization and Testing of Large Datasets.” Advisor: Sofya Raskhodnikova.

2008–2010 **Academic University of the Russian Academy of Sciences**, St. Petersburg, Russia

M.S. in Applied Mathematics and Physics.

2004–2008 **St. Petersburg State Polytechnic University**, St. Petersburg, Russia

B.S. in Physics and Technology.

RESEARCH **Microsoft Research, Redmond**, May 2013 – August 2013.

INTERNSHIPS Theory group, mentored by Konstantin Makarychev.

Microsoft Research, Silicon Valley, August 2012 – October 2012.

Theory group, mentored by Alexandr Andoni.

IBM Research, Almaden, May 2012 – July 2012.

Theory group, mentored by David P. Woodruff.

AT&T Labs – Research, May 2011 — August 2011.

Database theory group, mentored by Graham Cormode, Cecilia M. Procopiuc, Divesh Srivastava and Howard Karloff.

AWARDS AND **Facebook Faculty Research Award**, 2017.

ACHIEVEMENTS **Warren Center Postdoctoral Fellowship**, University of Pennsylvania, 2014 — 2016.

Institute Postdoctoral Fellowship, Brown University, ICERM, 2013 — 2014.

Best Graduate Research Assistant at Computer Science and Engineering Department, Penn State, 2012.

TopCoder Open Algorithm Competition Finalist (Top 24 worldwide), 2010.

College of Engineering Fellowship, Penn State, 2010 — 2013.

University Graduate Fellowship, Penn State, 2010 — 2011.

Yandex personal research grant, Academic University of the RAS, 2009 — 2010.

FUNDING

- **Facebook Faculty Research Award.** Award amount: \$35K.
- **NSF CRII Award:** “Algorithms for Noise-Tolerant Function Testing with Applications to Deep Learning”, 2017–2019. Sole PI, award amount: \$175K.

CONFERENCE PAPERS Authors listed in alphabetical order unless otherwise specified:

PAPERS

- “Linear Sketching over \mathbb{F}_2 ”, with S. Kannan, E. Mossel and S. Sanyal.
CCC 2018 (33rd Conference on Computational Complexity).
 - “Tight Bounds on Linear Sketches of Approximate Matchings”, with S. Assadi, S. Khanna and Y. Li.
SODA 2016 (27th Annual ACM-SIAM Symposium on Discrete Algorithms).
 - “Amplification of One-Way Information Complexity via Codes and Noise Sensitivity”, with M. Molinaro and D. Woodruff.
ICALP 2015 (42nd International Colloquium on Automata, Languages and Programming).
 - “Near Optimal LP Rounding Algorithm for Correlation Clustering on Complete and Complete k-partite Graphs”, with S. Chawla, K. Makarychev and T. Schramm.
STOC 2015 (47th ACM Symposium on the Theory of Computing).
 - “Certifying Equality with Limited Interaction”, with J. Brody, A. Chakrabarti, R. Kondapally and D. Woodruff.
RANDOM 2014 (18th International Workshop on Randomization and Computation).
 - “Beyond Set Disjointness: The Communication Complexity of Finding the Intersection”, with J. Brody, A. Chakrabarti, R. Kondapally and D. Woodruff.
PODC 2014 (33rd ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing).
 - “Parallel Algorithms for Geometric Graph Problems”, with A. Andoni, K. Onak and A. Nikolov.
STOC 2014 (46th ACM Symposium on the Theory of Computing).
 - “ L_p -testing”, with P. Berman and S. Raskhodnikova.
STOC 2014 (46th ACM Symposium on the Theory of Computing).
 - “Lower Bounds for Testing Properties of Functions over Hypergrid Domains”, with E. Blais and S. Raskhodnikova.
CCC 2014 (29th IEEE Conference on Computational Complexity).
 - ¹“Accurate and Efficient Private Release of Datacubes and Contingency Tables”. G. Yaroslavtsev, G. Cormode, C. Procopiuc and D. Srivastava.
ICDE 2013 (29th IEEE International Conference on Data Engineering).
 - “Beating the Direct Sum Theorem in Communication Complexity with Implications for Sketching”, with Marco Molinaro and David Woodruff.
SODA 2013 (24th Annual ACM-SIAM Symposium on Discrete Algorithms).
- Invited to a special issue of “Algorithmica” on “Information Complexity and Applications”.**

¹This is the only paper with non-alphabetical ordering of authors

- “Learning Pseudo-Boolean k-DNF and Submodular Functions”, with S. Raskhodnikova. SODA 2013 (24th Annual ACM-SIAM Symposium on Discrete Algorithms).
- “Primal-dual algorithms for Node-Weighted Network Design in Planar Graphs”, with P. Berman. APPROX 2012 (15th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems).
- “Private Analysis of Graph Structure”, with V. Karwa, S. Raskhodnikova and A. Smith. VLDB 2011 (37th International Conference on Very Large Data Bases), Research track.
- “Improved Approximation for the Directed Spanner Problem”, with P. Berman, A. Bhattacharyya, K. Makarychev and S. Raskhodnikova. ICALP 2011 (38th International Colloquium on Automata, Languages and Programming).
Runner-up for the Best Paper Award, invited to a special issue of a journal “Information and Computation”.
- “Steiner Transitive-Closure Spanners of Low-Dimensional Posets”, with P. Berman, A. Bhattacharyya, E. Grigorescu, S. Raskhodnikova and D. Woodruff. ICALP 2011 (38th International Colloquium on Automata, Languages and Programming).
- “Finding Efficient Circuits using SAT-solvers”, with A. Kojevnikov and A. Kulikov. SAT 2009 (12th International Conference on Theory and Applications of Satisfiability Testing).

JOURNAL
PAPERS

Authors listed in alphabetical order:

- “Private Algorithms for the Protected in Social Network Search”, with M. Kearns, A. Roth and S. Wu.
PNAS (Proceedings of the National Academy of Sciences), via direct submission, 2016.
- “Certifying Equality with Limited Interaction”, with J. Brody, A. Chakrabarti, R. Kondapally and D. Woodruff.
Algorithmica, special issue on “Information Complexity and Applications”, to appear.
- “Private Analysis of Graph Structure”, with V. Karwa, S. Raskhodnikova and A. Smith. ACM Transactions on Database Systems, 2014.
- “Steiner Transitive-Closure Spanners of Low-Dimensional Posets”, with P. Berman, A. Bhattacharyya, E. Grigorescu, S. Raskhodnikova and D. Woodruff. Combinatorica, 2014.
- “Approximation Algorithms for Spanner Problems and Directed Steiner Forest”, with P. Berman, A. Bhattacharyya, K. Makarychev and S. Raskhodnikova. Information and Computation, special issue for ICALP’11, 2012.
- “New upper bounds on the Boolean Circuit Complexity of Symmetric Functions”, with E. Demenkov, A. Kojevnikov and A. Kulikov. Information Processing Letters, 2010.

PREPRINTS Co-authors listed in alphabetical order:

- “Approximate \mathbb{F}_2 -Sketching of Valuation Functions”. In submission to ICALP’18.
- “Massively Parallel Algorithms for Single-Linkage Clustering under ℓ_p -Distances”, with Adithya Vadapalli. In submission to ICML’18. [ArXiv:1710.01431](#)
- “Going for Speed: Sublinear Algorithms for Dense r-CSPs”. [ArXiv:1407.7887](#)
- “Online Algorithms for Machine Minimization”, with N. Devanur, K. Makarychev and D. Panigrahi. [ArXiv:1403.0486](#)

TEACHING “**Data Structures (honors)**”

- Indiana University, Bloomington, CSCI-H343, Spring 2018.

“**Applied Algorithms**”

- Indiana University, Bloomington, CSCI-B505, Fall 2017.

“**Foundations of Data Science**”

- Indiana University, Bloomington, CSCI-B609, Fall 2016 and 2017.

“**Algorithms for Big Data**”

- University of Pennsylvania, CIS 700, Fall 2015.

“**Computational Learning Theory**”

- University of Pennsylvania, CIS 625, Spring 2015 (co-teaching with Michael Kearns).

“**Sublinear Algorithms for Big Data**”

- University of Buenos Aires, Argentina. 15-hour crash course. July – August 2014.

Tutorial “**Algorithms for MapReduce and Beyond**” (with Sergei Vassilvitskii, Google)

- 24th International Conference on Information and Knowledge Management (CIKM 2015), Melbourne, Australia.

Guest lecturer at undergraduate classes:

- CIS 399, “Foundations of Data Science”, University of Pennsylvania, Spring 2016.
- CMPSC 464, “Introduction to the Theory of Computing”, Pennsylvania State University, Fall 2010.

Extracurricular education for high-school students:

- Prepared training contests for the United States team in International Olympiad in Informatics 2011.
- Co-founder and coordinator of St. Petersburg network of extracurricular education in informatics for high-school students (<http://spbtc.ru>) (2009-2010).

SERVICE Internal service at Indiana University:

- Head of the undergraduate and graduate research award committee (’16–)
- Graduate admission committee member (’16–’17).

Service to federal funding agencies:

- Panelist for grant proposals for NSF CCF core programs (2018, 2017).
- Reviewer for Israeli Science Foundation grant proposals.

Program committees:

- 21st International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX’18).
- 5th Workshop on Algorithms and Systems for MapReduce and Beyond (BeyondMR’18) at SIGMOD/PODS’18.
- 23rd International Computing and Combinatorics Conference (COCOON’17).

- **28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'17).**
- 24th Annual European Symposium on Algorithms (ESA'16), Design and Analysis Track.
- 41st International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM'15), Foundations of Computer Science Track.

Organizer:

- **Linear Sketching as a Tool for Everything**
1-day workshop at IEEE FOCS, Oct 2017.
<http://caml.indiana.edu/linear-sketching.focs.html>
- **67th Midwest Theory Day**
2-day workshop at Indiana University, Bloomington, Apr 2017.
<http://caml.indiana.edu/mtd.html>
- **Big Data through the Lens of Sublinear Algorithms**
2-day workshop at Rutgers University, DIMACS, Aug 2015.
<http://grigory.us/mpc-workshop-dimacs.html>
- **Algorithmic Frontiers of Modern Massively Parallel Computation**
1-day workshop at ACM FCRC/STOC, Jun 2015.
<http://grigory.us/mpc-workshop-fcrc.html>
- **Sublinear Algorithms and Big Data Day** at Brown University, ICERM, Apr 2014.
<http://grigory.us/big-data-day.html>
- **Theory Seminar** at the University of Pennsylvania, Computer and Information Sciences Department (2014 – 2016). <http://theory.cis.upenn.edu/seminar/>
- **Theory Seminar** at Brown CS Department and Brown University, ICERM (2013 – 2014). <http://grigory.us/theory-seminar-brown-spring14.html>

REVIEWING

Theory conferences:

- STOC (ACM Symposium on the Theory of Computing): 2018, 2017, 2016, 2015.
- FOCS (IEEE Symposium on Foundations of Computer Science): 2017, 2015, 2014, 2013, 2012.
- SODA (ACM Symposium on Discrete Algorithms): 2018, 2016, 2013, 2012.
- ICALP (International Colloquium on Automata, Languages and Programming): 2018, 2017, 2015, 2014, 2013.
- CCC (Conference on Computational Complexity): 2018, 2016.
- ITCS (Innovations in Theoretical Computer Science): 2018.
- RANDOM (Workshop on Randomization and Computation): 2017, 2015, 2014.
- APPROX (Workshop on Approximation Algorithms for Combinatorial Optimization Problems): 2012.
- ESA (European Symposium on Algorithms): 2015.
- MFCS (Symposium on Mathematical Foundations of Computer Science): 2013, 2010.

Machine learning and learning theory conferences:

- NIPS (Neural Information Processing Systems): 2017, 2016.
- ICML (International Conference on Machine Learning): 2018.
- COLT (Conference on Learning Theory): 2016.
- ICLR (International Conference on Learning Representations): 2018.
- ALT (Conference on Algorithmic Learning Theory): 2014.

Databases and large-scale data processing conferences:

- PODS (ACM Symposium on Principles of Database Systems): 2018, 2017, 2016.
- SPAA (ACM Symposium on Parallelism in Algorithms and Architectures): 2017.
- VLDB (Conference on Very Large Databases): 2012.
- CIKM (ACM International Conference on Information and Knowledge Management): 2014.

Journals:

- SICOMP (SIAM Journal on Computing)
- CSUR (ACM Computing Surveys)
- I&C (Information and Computation)
- TKDE (IEEE Transactions on Knowledge and Data Engineering)
- ToC (Theory of Computing)
- RSA (Random Structures and Algorithms)
- Algorithmica

EXTERNAL
TALKS

- Massively Parallel Algorithms and Hardness for Single-Linkage Clustering under ℓ_p -Distances
 - Stanford University, CA. Theory Seminar. May 2018.
 - University of Warwick, UK. Workshop on Data Summarization. March 2018.
- Linear Sketching for Functions over the Boolean Hypercube
 - 68th Midwest Theory Workshop. Toyota Technological Institute, Chicago, IL. April 2018.
 - IEEE FOCS, Berkeley, CA. Workshop “Linear Sketching as a Tool for Everything”. October 2017.
- Clustering on Clusters 2049: Massively Parallel Algorithms for Clustering Graphs and Vectors
 - Facebook, Menlo Park, CA. Tech Talk. October 2017.
- Computational and Communication Complexity in Massively Parallel Computing
 - ITMO University, St. Petersburg, Russia. Departmental Colloquium. June 2017.
 - Higher School of Economics, Moscow, Russia. Workshop on Complexity of Computation, Communication, Descriptions and Proofs. June 2017.
- Clustering on Clusters: Massively Parallel Algorithms for Clustering Graphs and Vectors
 - Facebook, Menlo Park, CA. Tech Talk. February 2017.
- Linear Sketching over \mathbb{GF}_2
 - St. Petersburg Department of Steklov Institute of Mathematics, St. Petersburg, Russia. Theory Seminar. May 2017.
 - Moscow State University, Moscow, Russia. Kolmogorov Seminar, May 2017.
 - Banff Workshop on Communication Complexity and Applications II, Banff, Canada. March 2017.
 - Columbia University, New York, NY. Theory Seminar. November 2016.
 - University of Pennsylvania, Philadelphia, PA. Theory Seminar. October 2016.
 - University of Utah, Salt Lake City, UT. Theory Seminar, September 2016.
 - University of Illinois, Urbana, IL. Theory Seminar, August 2016.
 - Microsoft Research, Redmond, WA. Theory Seminar, June 2016.
- What’s New in “The Big Data Theory”?
 - College of William and Mary, Williamsburg, VA. Departmental Colloquium, February 2016.
 - Georgetown University, Washington, DC. Departmental Colloquium, February 2016.
 - Indiana University, Bloomington, IN. Departmental Colloquium, February 2016.
 - University of Colorado, Boulder, CO. Departmental Colloquium, February 2016.
 - Boston University, Boston, MA. Departmental Colloquium, February 2016.
 - Drexel University, Philadelphia, PA. Departmental Colloquium, March 2016.
- Fast Fourier Sparsity Testing over the Boolean Hypercube
 - University of Wisconsin, Madison. Theory Seminar. August 2015.
- Near Optimal LP Rounding for Correlation Clustering
 - Cornell University, Ithaca, NY. Theory Seminar. May 2015.
 - MIT, Boston, MA. Algorithms and Complexity Seminar. April 2015.
 - Microsoft Research, Redmond, WA. Theory Seminar. March 2015.

- Google Research, NYC. Google Tech Talk. February 2015.
- Rutgers University, New Brunswick, NJ. Theory Seminar. January 2015.
- Carnegie Mellon University, Pittsburgh, PA. Theory Lunch. January 2015.
- Pennsylvania State University, State College, PA. CSE Departmental Colloquium. January 2015.
- Beyond Set Disjointness: The Communication Complexity of Finding the Intersection
 - MIT, Boston, MA. Theory of Distributed Systems Seminar. May 2014.
- “The Big Data Theory” and Randomized Algorithms
 - Georgia Tech, Atlanta, GA. ARC Colloquium. March 2014.
- Approximating Graph Problems: The Old and The New
 - Yahoo! Research, NYC. February 2014.
 - MIT, Boston, MA. Algorithms and Complexity Seminar. February 2014.
 - Toyota Technological Institute, Chicago IL. Theory Seminar. February 2014.
- L_p -Testing
 - University of Pennsylvania, Statistics Student Seminar, November 2014.
 - Columbia University, New York, NY. Theory Seminar. October 2014.
 - Microsoft Research, Redmond. Theory Lunch. January 2014.
 - Harvard University, Boston MA. Theory Seminar. November 2013.
 - Brown University, Providence RI. Theory Seminar. November 2013.
 - IBM Almaden Research Center, San Jose, CA. October 2013.
- Property Testing and Communication Complexity
 - MIT, Boston, MA. Algorithms and Complexity Seminar. September 2013.
- Beating the Direct Sum in Communication Complexity with Implications for Sketching.
 - Aarhus University, Denmark. Theory Seminar. May 2013.
 - MIT, Boston, MA. Algorithms and Complexity Seminar. December 2012.
 - Princeton University, Princeton, NJ. Theory lunch. November 2012.
- Parallel Algorithms for Geometric Problems
 - ISMP’15 (22nd International Symposium on Mathematical Programming). July 2015.
 - Johns Hopkins University, Baltimore, MD. Theory Seminar. November 2014.
 - University of Maryland, College Park, MD. Capital Area Theory Seminar. October 2014.
 - University of Pennsylvania, Philadelphia, PA. Theory Seminar. August 2014.
 - University of Massachusetts, Amherst. Theory Seminar. May 2014.
 - Google Research, NYC. Google Tech Talk. March 2014.
 - Sandia Labs, Livermore, CA. March 2014.
 - Stanford University, Stanford, CA. Theory Seminar. March 2014.
 - Microsoft Research SVC, Mountain View, CA. Lab Meeting. October 2012.
- Learning and Testing Submodular Functions.
 - Microsoft Research, Redmond. Theory Seminar. June 2013.
 - University of Melbourne, Theory Seminar, April 2013.
 - UCLA, Los Angeles, LA. Theory Seminar. February 2013.
 - Weizmann Institute of Science, Rehovot, Israel. Theory Seminar. December 2012.
 - Harvard University, Boston, MA. Theory Seminar. December 2012.
 - Carnegie-Mellon University, Pittsburgh, PA. Theory Lunch, December 2012.
 - Carnegie-Mellon University, Pittsburgh, PA. Operations Research Seminar. December 2012.
 - IBM T.J. Watson Research Center, Yorktown Heights, NY. Integer Programming Seminar. November 2012.
 - Columbia University, New York, NY. Theory Seminar. October 2012.
 - Microsoft Research SVC, Mountain View, CA. Theory Seminar. October 2012.
 - IBM Almaden Research Center, San Jose, CA. Theory Seminar. May 2012.
- Advances in Directed Spanners.
 - University of Sydney, Theory Seminar, April 2013.

- Carnegie-Mellon University, Theory Lunch, November 2011.
- University of Maryland, Capital Area Theory Seminar, November 2011.
- Private Analysis of Graph Structure
 - AT&T Labs – Research, Florham Park, NJ. August 2011.
- Improved Approximation for the Directed Spanner Problem
 - AT&T Labs – Research, Florham Park, NJ. Mathematics Research Colloquium and Informal Seminar. June 2011.
 - Moscow State University. Combinatorial Optimization Seminar. May 2011.
 - IBM T.J. Watson Research Center, Yorktown Heights, NY. Integer Programming Seminar. April 2011.
 - St. Petersburg Institute of Fine Mechanics and Optics. Theory Seminar. December 2010.
- Linear bounds on circuit complexity and feebly one-way permutations
 - Pennsylvania State University, State College, PA. Theory Seminar. April 2010.

MEDIA
COVERAGE

For paper “Private Algorithms for the Protected in Social Network Search”:

- PBS Newshour “[The secret things you give away through your phone metadata](#)”
- Schneier on Security “[Research on Balancing Privacy with Surveillance.](#)”
- Association of American Universities “[Penn Researchers Balance Privacy and Security in Network Analysis.](#)”
- ACM Tech News / The Daily Pennsylvanian: “[Penn Professor’s Computer Algorithm Could Fight Terrorism While Protecting Privacy.](#)”
- Quartz: “[There may be a way to allow mass surveillance and preserve our privacy at the same time.](#)”
- Pacific Standard: “[Searching Private Data, and Ensuring It Stays Private.](#)”
- Wired (German): “[Ein neuer bewachungs-Algorithmus soll in Social Media nur auf Terroristen zielen.](#)”
- Vice Motherboard: “[Algorithms Claim to Hunt Terrorists While Protecting the Privacy of Others.](#)”
- The Naked Scientists Podcast: “[National Security Algorithm.](#)”

VISITS AND
CONSULTING

- Facebook, Menlo Park, CA. Consultant, Spring’17–. (Host: Sergey Pupyrev)
- Google Research, NYC. Weekly visitor in Fall’14–Spring’15. (Host: Silvio Lattanzi)
- Microsoft Research, Redmond, WA. 03/08/15–03/14/15, 01/08/14–01/12/14. (Host: Konstantin Makarychev)
- IBM T.J. Watson Research Center, Yorktown Heights, NY. 04/19/11–04/21/11, 11/13/12–11/15/12. (Hosts: Konstantin Makarychev, Vishwanath Nagarajan)
- AT&T Labs – Research, Florham Park, NJ. 11/18/11–11/25/11. (Host: Howard Karloff)
- Weizmann Institute of Science, Rehovot, Israel. 12/27/12–01/04/13. (Host: Robert Krauthgamer)
- University of Melbourne, Australia. 04/12/13–04/20/13. (Host: Anthony Wirth)
- Aarhus University, Denmark. 05/17/13–05/25/13. (Host: Joshua Brody)

STUDENTS
MENTORED

At Indiana University:

- Dmitrii Avdyukhin (Ph.D.), 2017–.
- Nikolai Karpov (Ph.D.), 2017.
- Adithya Vadapalli (Ph.D.), 2016–.

Ph.D. students mentored while a postdoc:

- Sepehr Assadi (University of Pennsylvania), joint paper in SODA’16.
- Yang Li (University of Pennsylvania), joint paper in SODA’16.

- Steven Wu (University of Pennsylvania), joint paper in PNAS'16.
- Tselil Schramm (UC Berkeley), joint paper in STOC'15.
- Eli Fox-Eppstein (Brown University).
- David Meierfrankenfeld (Brown Univeristy).

PATENTS

- “A Communication and Message-Efficient Protocol for Computing the Intersection Between Different Sets of Data”, with David P. Woodruff. U.S. patent #9438704. IBM Almaden Research Center, San Jose, CA.

MEMBERSHIPS ACM SIGACT, ACM SIAM, IEEE

TECH SKILLS C/C++, STL, Java, Windows/Linux, \LaTeX , CPLEX/Gurobi/AMPL, MapReduce/Hadoop.