SODA / ALENEX / ANALCO 2017
FINAL PROGRAM & ABSTRACTS

siam
acm Association for Computing Machinery

ACM-SIAM Symposium on Discrete Algorithms
January 16-19, 2017
Hotel Porta Fira
Barcelona, Spain

ANALCO17 Meeting on Analytic Algorithmics and Combinatorics
January 16-17, 2017
Hotel Porta Fira
Barcelona, Spain

Organized by:
Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

UPC

ALENEX17
Meeting on Algorithm Engineering & Experiments
January 17-18, 2017
Hotel Porta Fira
Barcelona, Spain

SODA is jointly sponsored by the ACM Special Interest Group on Algorithms and Computation Theory (SIGACT) and the SIAM Activity Group on Discrete Mathematics

The SIAG on Discrete Mathematics focuses on combinatorics, graph theory, cryptography, discrete optimization, mathematical programming, coding theory, information theory, game theory, and theoretical computer science, including algorithms, complexity, circuit design, robotics, and parallel processing. This activity group provides an opportunity to unify pure discrete mathematics and areas of applied research such as computer science, operations research, combinatorics, and the social sciences. It organizes the SIAM Conference on Discrete Mathematics; co-sponsors, with ACM SIGACT, the annual Symposium on Discrete Algorithms; and sponsors minisymposia at SIAM meetings and conferences. The activity group also runs DM-Net, an electronic forum; publishes an electronic newsletter; and maintains a website and a member directory. Every two years, the activity group also awards the Dénes König Prize to a young researcher for outstanding research in the area of discrete mathematics.

siam
Society for Industrial and Applied Mathematics
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA
Telephone: +1-215-382-9800 Fax: +1-215-386-7999
Conference E-mail: `meetings@siam.org`
Conference Web: www.siam.org/meetings/
Membership and Customer Service: (800) 447-7426 (US & Canada) or +1-215-382-9800 (worldwide)

http://www.siam.org/meetings/da17
http://www.siam.org/meetings/alenex17
http://www.siam.org/meetings/analco17

David S. Johnson (1945–2016)
In Memoriam

General Information

CONTENTS

SODA COMMITTEES

Program Committee Chair

Philip Klein
Brown University, USA

Program Committee

Pankaj Agarwal
Duke University, USA

Yair Bartal
Hebrew University, Israel

Gerth Stølting Brodal
Aarhus University, Denmark

Sébastien Bubeck
Microsoft Research, USA

Anne Driemel
Eindhoven University of Technology, Netherlands

David Eppstein
University of California-Irvine, USA

Jeff Erickson
University of Illinois at Urbana-Champaign, USA

Rong Ge
Duke University, USA

MohammadTaghi Hajiaghayi
University of Maryland, USA

Thomas P. Hayes
University of New Mexico, USA

Dorit Hochbaum
University of California-Berkeley, USA

Ken-ichi Kawarabayashi
National Institute of Informatics, Japan

Jochen Koenemann
University of Waterloo, Canada

Ravishankar Krishnaswamy
Microsoft Research, India

Reut Levi
Max Planck Institute for Informatics, Germany

Dániel Marx
Hungarian Academy of Sciences, Hungary

Nicole Megow
Technical University of Munich, Germany

Ruta Mehta
University of Illinois at Urbana-Champaign, USA

Vahab Mirrokni
Google, USA

Shay Mozes
The Interdisciplinary Center Herzliya, Israel

Kamesh Munagala
Duke University, USA

Jelani Nelson
Harvard University, USA

Lorenzo Orecchia
Boston University, USA

Alessandro Panconesi
Sapienza University of Rome, Italy

Andrea Pietracaprina
University of Padova, Italy

Ely Porat
Bar-Ilan University, Israel

Vijaya Ramachandran
University of Texas at Austin, USA

Adi Rosén
University Paris Diderot, France

Maria Serna
Universitat Politècnica de Catalunya, Spain

Aaron Sidford
Stanford University, USA

Mikkel Thorup
University of Copenhagen, Denmark

Dorothea Wagner
Karlsruhe Institute of Technology, Germany

Roger Wattenhofer
ETH Zurich, Switzerland

Matt Weinberg
Princeton University, USA

Virginia Vassilevska Williams
Stanford University, USA

Peter Winkler
Dartmouth College, USA

Grigory Yaroslavtsev
University of Pennsylvania, USA

Ke Yi
Hong Kong University of Science and Technology, China

Neal Young
University of California-Riverside, USA

Steering Committee

Pavol Hell
Simon Fraser University, Canada

Daniel Král'
University of Warwick, United Kingdom

Dana Randall
Georgia Institute of Technology, USA

Cliff Stein (chair)
Columbia University, USA

Shang-Hua Teng
University of Southern California, USA

ALENEX COMMITTEES

Program Committee Co-Chairs

Sándor Fekete
Technische Universität Braunschweig, Germany

Vijaya Ramachandran
University of Texas at Austin, USA

Program Committee

Deepak Ajwani
Nokia Bell Labs, Ireland

Hannah Bast
University of Freiburg, Germany

Rezaul Chowdhury
Stony Brook University, USA

Cid de Souza
University of Campinas, Brazil

Irene Finocchi
Università di Roma "La Sapienza", Italy

Martin Nöllenburg
TU Vienna, Austria

Valentin Polishchuk
Linköping University, Sweden

Rajeev Raman
University of Leicester, United Kingdom

Baruch Schieber
IBM Research, USA

Subhash Suri
UC Santa Barbara, USA

Sergei Vassilvitskii
Google Inc., USA

Carola Wenk
Tulane University, USA

Renato Werneck
Amazon, USA

Christos Zaroliagis
CTI and University of Patras, Greece

Steering Committee

David Applegate
AT&T Labs Research, USA

Lars Arge
Aarhus University, Denmark

David Bader
Georgia Institute of Technology, USA

Guy Blelloch
Carnegie Mellon University, USA

Ulrik Brandes
Universität Konstanz, Germany

Gerth Stølting Brodal
Aarhus University, Denmark

David Eppstein
University of California-Irvine, USA

Irene Finocchi
Università di Roma "La Sapienza", Italy

Michael Goodrich (chair)
University of California-Irvine, USA

John Hershberger
Mentor Graphics, USA

Matthias Müller-Hannemann
Martin-Luther-University Halle-Wittenberg, Germany

Petra Mutzel
Technische Universität Dortmund, Germany

Cliff Stein
Columbia University, USA

Renato Werneck
Amazon, USA

ANALCO COMMITTEES

Program Committee Co-Chairs

Conrado Martínez
Universitat Politècnica de Catalunya, Spain

Mark D. Ward
Purdue University, USA

Program Committee

Olivier Bodini
Université Paris Nord, France

Mireille Bousquet-Mélou
Université de Bordeaux, France

Elie de Panafieu
Bell Labs France, France

Luc Devroye
McGill University, Canada

Martin Dietzfelbinger
Technische Universität Ilmenau, Germany

Bernhard Gittenberger
Technische Universität Wien, Austria

Cecilia Holmgren
Uppsala Universitet, Sweden

Mihyun Kang
Technische Universität Graz, Austria

Marni Mishna
Simon Fraser University, Canada

Cyril Nicaud
Université Paris Est, France

Marc Noy
Universitat Politècnica de Catalunya, Spain

Dana Randall
Georgia Institute of Technology, USA

Steering Committee

Michael Drmota
Technische Universität Wien, Austria

James Allen (Jim) Fill
Johns Hopkins University, USA

Hsien-Kuei Hwang
Academia Sinica, Taiwan

Conrado Martínez
Universitat Politècnica de Catalunya, Spain

Markus Nebel
University of Bielefeld, Germany

Robert Sedgewick (chair)
Princeton University, USA

Wojciech Szpankowski
Purdue University, USA

Mark D. Ward
Purdue University, USA

LOCAL ORGANIZING COMMITTEE

Josep Díaz
Universitat Politècnica de Catalunya, Spain

Amalia Duch
Universitat Politècnica de Catalunya, Spain

Conrado Martínez (chair)
Universitat Politècnica de Catalunya, Spain

Marc Noy
Universitat Politècnica de Catalunya, Spain

Rodrigo Silveira
Universitat Politècnica de Catalunya, Spain

Gabriel Valiente
Universitat Politècnica de Catalunya, Spain

Special thanks go to Mrs. Olga Núñez and Mrs. Yolanda López from Activa Congresos for their invaluable help to organize the conference. Many thanks go also to the SIAM Staff, in particular, to Mrs. Shelby Asen, Dr. William Kolata, Mrs. Kristin O'Neill, Mrs. Linda Thiel and Mrs. Kirsten Wilden.

Registration Desk

The registration desk is located in the Las Arenas Foyer on the first floor. It is open during the following hours:

Sunday, January 15
5:00 PM – 8:00 PM

Monday, January 16
8:00 AM – 5:00 PM

Tuesday, January 17
9:00 AM – 5:00 PM

Wednesday, January 18
9:00 AM – 12:30 PM

Hotel Information

Hotel Porta Fira
Plaza Europa, 45
08908 L'Hospitalet de Llobregat
(Barcelona)
Spain
http://hotelbarcelonaportafira.com/

Hotel Telephone Number

To reach an attendee or leave a message, call +34 932 973 500. If the attendee is a hotel guest, the hotel operator can connect you with the attendee's room.

Hotel Check-in and Check-out Times

Check-in time is 03:00 PM and check-out time is 12:00 PM.

Child Care

Please visit Barcelona Babysitter http://www.bcnbabysitter.com/ for babysitting service information. Barcelona Babysitter offers services in 5+ languages, and is a member of "Barcelona Turismo" (Barcelona Tourism Consortium, an organization operating under official agreement with the Municipal Council and the Official Chamber of Commerce).

Corporate Members and Affiliates

SIAM corporate members provide their employees with knowledge about, access to, and contacts in the applied mathematics and computational sciences community through their membership benefits. Corporate membership is more than just a bundle of tangible products and services; it is an expression of support for SIAM and its programs. SIAM is pleased to acknowledge its corporate members and sponsors. In recognition of their support, non-member attendees who are employed by the following organizations are entitled to the SIAM member registration rate.

Corporate/Institutional Members

The Aerospace Corporation
Air Force Office of Scientific Research
Amazon
Aramco Services Company
Bechtel Marine Propulsion Laboratory
The Boeing Company
CEA/DAM
Department of National Defence (DND/CSEC)
DSTO-Defence Science and Technology Organisation
Hewlett-Packard
Huawei FRC French R&D Center
IBM Corporation
IDA Center for Communications Research, La Jolla
IDA Center for Communications Research, Princeton
Institute for Defense Analyses, Center for Computing Sciences
Lawrence Berkeley National Laboratory
Lockheed Martin
Los Alamos National Laboratory
Max-Planck-Institute for Dynamics of Complex Technical Systems
Mentor Graphics
National Institute of Standards and Technology (NIST)
National Security Agency (DIRNSA)
Oak Ridge National Laboratory (managed by UT-Battelle for the Department of Energy)
Sandia National Laboratories
Schlumberger-Doll Research
United States Department of Energy
U.S. Army Corps of Engineers, Engineer Research and Development Center
US Naval Research Labs

List current December 2016.

Leading the applied mathematics community: Join SIAM and save!

Join your peers in supporting the premier professional society for applied mathematicians and computational scientists. SIAM members receive subscriptions to SIAM Review, SIAM News and SIAM Unwrapped, and enjoy substantial discounts on SIAM books, journal subscriptions, and conference registrations.

Free Student Memberships are available to students who attend an institution that is an Academic Member of SIAM, are members of Student Chapters of SIAM, or are nominated by a Regular Member of SIAM.

Join online today at www.siam.org/joinsiam to join online or contact SIAM Customer Service:

Telephone: +1-215-382-9800 (worldwide); or 800- 447-7426 (U.S. and Canada only)
Fax: +1-215-386-7999
E-mail: membership@siam.org
Postal mail:
Society for Industrial and Applied Mathematics
3600 Market Street, 6th floor
Philadelphia, PA 19104-2688 USA

Standard Audio/Visual Set-Up in Meeting Rooms

All lecture rooms are equipped with one screen, one data projector (beamer) and a Windows laptop. We strongly encourage speakers to produce PDF files for their presentations and upload them through the CloudSlides server (details were provided by email to the speakers; check with the registration desk if necessary). Each laptop will synchronize the PDF files uploaded by the speakers of the corresponding sessions, thus very little time will be needed to switch from a presentation to the next.

If you need special hardware or software (other than a PDF reader like Adobe Acrobat) you will have to bring your own laptop. The data projectors support VGA connections only. Presenters requiring an HDMI or alternate connection must provide their own adaptor. Please try the connection to the beamer, screen resolution, etc. before your presentation if you have to use your own computer. Bear in mind that any unnecessary delay to set-up your computer for the presentation will consume part of your time slot. If you plan to use your own computer for the presentation, it is helpful that you tell the organizers and the chair of your session about it in advance. If you have questions regarding availability of equipment in the meeting room of your presentation, please send email to `soda2017@cs.upc.edu` or contact a member of the local organization at the registration desk.

Internet Access

There is free wireless Internet access in all the public areas of the hotel, as well as in all the guest rooms. Details for connection to the hotel wireless network will be provided to participants while they register for the conference, or during check-in as hotel guests.

Whova: SODA 2017 Mobile App

Download the Whova app (http://whova.io/portal/asda_201701) from Google Play or the App Store in your smartphone, signup a free account with the same email address that you provied to register for SODA/ALENEX/ANALCO and *voilà!* enjoy an on-line full agenda of the event, with access to the abstracts, create your personal agenda choosing which talks and sessions you'd like to attend, get reminders for events in your personal agenda, add your notes to talks and other events, network and share contact info with other participants, make comments, share pictures of the conference, tweet them or post them in your Facebook wall in a simple way from the app, . . .

If you already have a Whova account you shall see that you have access to the SODA/ALENEX/ANALCO 2017 agenda. If you registered for SODA/ALENEX/ANALCO with an email address that is different from the email address of your Whova account, search for SODA within the app and introduce the invitation code sent by email to you to join the event.

Exhibits

Table top displays will be situated in the Las Arenas Foyer for book, journal and other exhibit materials.
Exhibitors:

- Princeton University Press
- SIAM
- Susquehanna International Group

Registration Fee Includes

- Admission to all technical sessions
- ALENEX/ANALCO Business Meeting (Monday, January 16)
- Morning and evening coffee breaks daily
- Lunches daily
- Proceedings of SODA, ALENEX and ANALCO (online)
- Program booklet (printed), program and abstracts (online)
- Room set-ups and audio/visual equipment
- Access to the event mobile app (*Whova*)
- SODA Business Meeting (Tuesday, January 17)
- Welcome Reception (Sunday, January 15)

Job Postings

Please visit http://jobs.siam.org/ regarding the availability of job postings.

Name Badges

A space for emergency contact information is provided on the back of your name badge. Help us help you in the event of an emergency!

Get-togethers

Sunday, January 15th, 2017

17:00–20:00 **Welcome Reception**
Las Arenas Foyer

Monday, January 16th, 2017

19:00–19:30 **ANALCO / ALENEX Business Meeting**
Las Arenas II-IV

Tuesday, January 17th, 2017

18:00–19:00 **SODA Business Meeting**
Las Arenas II-IV

20:15–22:15 **Banquet***
Restaurant Marítim

* Ticketed event not included in the registration fee

Restaurant Marítim

Restaurant Marítim al Port
Moll d'Espanya s/n
08039 Barcelona
Spain
http://www.maritimrestaurant.com/
Free parking for customers

Statement on Inclusiveness

As a professional society, SIAM is committed to providing an inclusive climate that encourages the open expression and exchange of ideas, that is free from all forms of discrimination, harassment, and retaliation, and that is welcoming and comfortable to all members and to those who participate in its activities. In pursuit of that commitment, SIAM is dedicated to the philosophy of equality of opportunity and treatment for all participants regardless of gender, gender identity or expression, sexual orientation, race, color, national or ethnic origin, religion or religious belief, age, marital status, disabilities, veteran status, field of expertise, or any other reason not related to scientific merit. This philosophy extends from SIAM conferences, to its publications, and to its governing structures and bodies. We expect all members of SIAM and participants in SIAM activities to work towards this commitment.

Important Notice

Neither SIAM nor UPC are responsible for the safety and security of attendees' computers and other valuables. Do not leave your laptop computers and other personal belongings unattended. Please remember to switch off the sound in your cell phone, tablet, laptop, etc. during sessions.

Recording of Presentations

Audio and video recording of presentations at SIAM meetings is prohibited without the written permission of the presenter and SIAM.

Comments?

Comments about SIAM meetings are encouraged! The SODA 2017 mobile app provides plenty of opportunities for making comments which the local organization will collect and, where applicable, forward to SIAM officers. You might also want to send your comments directly to Cynthia Phillips, SIAM Vice President for Programs (`vpp@siam.org`).

Social Media

SIAM is promoting the use of social media, such as Facebook and Twitter, in order to enhance scientific discussion at its meetings and enable attendees to connect with each other prior to, during and after conferences. If you are tweeting about a conference, please use the designated hashtag to enable other attendees to keep up with the Twitter conversation and to allow better archiving of our conference discussions. The hashtag for SODA is #SIAMDA17. The hashtags for ALENEX and ANALCO are #ALENEX17 and #ANALCO17. Whova makes it very easy to share tweets and post comments in Facebook as you might link your social network accounts to the Whova app.

Sponsors

SIAM and the conference organizing committee wish to extend their thanks and appreciation to the following sponsors for their support of this conference.

Google Inc.

IBM Research

Microsoft

Ministerio de Economía y Competitividad (MINECO)

National Science Foundation (NSF)

ACM Special Interest Group on Algorithms and Computation Theory (SIGACT)

Universitat Politècnica de Catalunya (UPC)

Best Paper and Best Student Paper Award

Best Student Paper
A $(2+\epsilon)$-Approximation for Maximum Weight Matching in the Semi-Streaming Model
Ami Paz and Gregory Schwartzman

Best Paper
Awarded jointly to two papers

A $(2+\epsilon)$-Approximation for Maximum Weight Matching in the Semi-Streaming Model
Ami Paz and Gregory Schwartzman

Subquadratic Algorithms for the Diameter and the Sum of Pairwise Distances in Planar Graphs
Sergio Cabello

The two awarded papers will be presented during the special **SODA Awards** session on Wednesday, January 18th, from 6:00PM to 6:45PM in room Las Arenas II–IV. The certificates and prizes will be given to the winners during the **SODA Business Meeting** on Tuesday, January 17th, also in room Las Arenas II–IV. Cash prizes have been donated by SIAM (Best Student Award) and ACM (Best Paper Award).

Invited Plenary Speakers

All Invited Plenary Presentations will take place in room Las Arenas II–IV

Monday, January 16

9:00 AM – 10:00 AM

A Dynamical View of Convex Optimization Algorithms

Benjamin Recht, *University of California, Berkeley, USA*

2:00 PM – 3:00 PM

Clique colouring random graphs

Colin McDiarmid, *University of Oxford, United Kingdom*

Tuesday, January 17

9:00 AM – 10:00 AM

Systems and Algorithms for the Internet of Things

Dina Katabi, *Massachusetts Institute of Technology, USA*

Thursday, January 19

9:00 AM – 10:00 AM

Uncertainty and Optimization

Nikhil Bansal, *Eindhoven University of Technology, The Netherlands*

SODA, ALENEX & ANALCO At-a-Glance

	Monday 16	Tuesday 17	Wednesday 18	Thursday 19
9-10	Invited Talk	Invited Talk	Tribute D.S. Johnson; SODA 7A/7B/7C ALENEX 3	Invited Talk
10-11	Coffee Break; SODA 1A/1B/1C ANALCO 1	Coffee Break; SODA 4A/4B/4C ANALCO 4	SODA 7A/7B/7C ALENEX 3; Coffee Break	Coffee Break; SODA 11A/11B/11C
11-12	SODA 1A/1B/1C ANALCO 1	SODA 4A/4B/4C ANALCO 4	SODA 8A/8B/8C ALENEX 4	SODA 11A/11B/11C
12-13	SODA 1A/1B/1C ANALCO 1; Lunch	SODA 4A/4B/4C ANALCO 4; Lunch	SODA 8A/8B/8C ALENEX 4; Lunch	SODA 11A/11B/11C; Lunch
13-14	Lunch	Lunch	Lunch	Lunch
14-15	Invited Talk	SODA 5A/5B/5C ALENEX 1	SODA 9A/9B/9C ALENEX 5	SODA 12A/12B/12C
15-16	SODA 2A/2B/2C ANALCO 2	SODA 5A/5B/5C ALENEX 1	SODA 9A/9B/9C ALENEX 5	SODA 12A/12B/12C
16-17	SODA 2A/2B/2C ANALCO 2; Coffee Break	Coffee Break; SODA 6A/6B/6C ALENEX 2	Coffee Break; SODA 10A/10B/10C	Coffee Break; SODA 13A/13B/13C
17-18	SODA 3A/3B/3C | ANALCO 3	SODA 6A/6B/6C ALENEX 2	SODA 10A/10B/10C	SODA 13A/13B/13C
18-19	SODA 3A/3B/3C | ANALCO 3	SODA Business Meeting	SODA Awards	SODA 13A/13B/13C
19-20	ALENEX/ANALCO Business Meeting			
20-22		Conference Banquet		

The drawing is only an approximation to the precise schedule, e.g., sessions 2A/2B/2C run from 15:00 to 16:35. Check the exact timing in the detailed program.

SODA / ALENEX / ANALCO 2017

PROGRAM

siam

acm Association for Computing Machinery

ACM-SIAM Symposium on Discrete Algorithms

January 16-19, 2017
Hotel Porta Fira
Barcelona, Spain

ANALCO17 Meeting on Analytic Algorithmics and Combinatorics

January 16-17, 2017
Hotel Porta Fira
Barcelona, Spain

Organized by:
Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

UPC

ALENEX17

Meeting on
Algorithm Engineering & Experiments

January 17-18, 2017
Hotel Porta Fira
Barcelona, Spain

Program

January 15th, 2017

17:00–20:00 **Registration & Welcome Reception**
Las Arenas Foyer

January 16th, 2017

8:00–17:00 **Registration**
Las Arenas Foyer

Invited Talk 1

9:00–10:00
Chair: Yair Bartal
Room: *Las Arenas II-IV*

9:00 *A Dynamical View of Convex Optimization Algorithms* [1]
Benjamin Recht, University of California, Berkeley, USA

Convex optimization provides a powerful toolkit for robust and efficient solutions of difficult engineering problems. However, it is laborious to extend and generalize the mathematical guarantees of simple convex primitives. In this talk, I will show that much of the analysis and design of optimization algorithms can be automated, and will demonstrate that multiple objectives—such as robustness, accuracy, and speed—can be balanced using tools from dynamical systems.

Noting that all iterative algorithms are dynamical systems, I will illustrate how most of the popular methods in optimization can be cast as a family of feedback systems studied in control theory. Leaning on this abstraction enables us to apply powerful, control-theoretic methods to algorithm analysis. I will show how the convergence rates of most common algorithms—including gradient descent, mirror descent, Nesterov's method, etc.—can be verified using a unified set of potential functions. I will then describe how such potential functions can themselves be found by solving constant-sized semidefinite programming problems. I will close with a discussion of how these techniques can be used to search for optimization algorithms with desired performance characteristics, proposing a new methodology for algorithm design.

10:00–10:30 **Coffee Break**
Las Arenas Foyer

ANALCO 1

10:30–12:30
Chair: Martin Dietzfelbinger
Room: *Saray*

10:30 *The Ordered and Colored Products in Analytic Combinatorics: Application to the Quantitative Study of Synchronizations in Concurrent Processes* [2]
Olivier Bodini, Matthieu Dien, Antoine Genitrini and Frédéric Peschanski

10:55 *On the cycle structure of the product of random maximal cycles* [3]
Miklós Bóna and Boris Pittel

11:20 *An Exact Enumeration of Distance-Hereditary Graphs* [4]
Cédric Chauve, Éric Fusy and Jérémie Lumbroso

11:45 *Isolated cycles of critical random graphs* [5]
Marc Noy, Vonjy Rasendrahasina, Vlady Ravelomanana and Juanjo Rué

12:10 *Cycle Basis Markov Chains for the Ising Model* [6]
Amanda Pascoe Streib and Noah Streib

SODA 1A

10:30–12:30
Chair: Vahab Mirrokni
Room: *Las Arenas II-IV*

10:30 *Proximity in the Age of Distraction: Robust Approximate Nearest Neighbor Search* [7]
Sariel Har-Peled and Sepideh Mahabadi

10:55 *High-dimensional approximate r-nets* [8]
Georgia Avarikioti, Ioannis Z. Emiris, Loukas Kavouras and Ioannis Psarros

11:20 *A Framework for Similarity Search with Space-Time Tradeoffs using Locality-Sensitive Filtering* [9]
Tobias Christiani

11:45 *Optimal Hashing–based Time–Space Trade-offs for Approximate Near Neighbors* [10]
Alexandr Andoni, Thijs Laarhoven, Ilya Razenshteyn and Erik Waingarten

12:10 *LSH Forest: Practical Algorithms Made Theoretical* [11]
Alexandr Andoni, Ilya Razenshteyn and Negev Shekel Nosatzki

SODA 1B

10:30–12:30
Chair: Anne Driemel
Room: *Las Arenas I*

10:30 *Faster approximation schemes for the two-dimensional knapsack problem* [12]
Sandy Heydrich and Andreas Wiese

10:55 *Split Packing: An Algorithm for Packing Circles with Optimal Worst-Case Density* [13]
Sebastian Morr

11:20 *Stochastic k-Center and j-Flat-Center Problems* [14]
Lingxiao Huang and Jian Li

11:45 *Local search for max-sum diversification* [15]
Alfonso Cevallos, Friedrich Eisenbrand and Rico Zenklusen

12:10 *Maximum Scatter TSP in Doubling Metrics* [16]
László Kozma and Tobias Mömke

SODA 1C

10:30–12:30
Chair: Adi Rosén
Room: *Nelva*

10:30 *Matrix Balancing in L_p Norms: Bounding the Convergence rate of Osborne's Iteration* [17]
Rafail Ostrovsky, Yuval Rabani and Arman Yousefi

10:55 *Decidability of the Membership Problem for 2×2 integer matrices* [18]
Igor Potapov and Pavel Semukhin

11:20 *The Identity Problem for Matrix Semigroups in $SL_2(\mathbb{Z})$ is NP-complete* [19]
Paul C. Bell, Mika Hirvensalo and Igor Potapov

11:45 *Exploring an Infinite Space with Finite Memory Scouts* [20]
Lihi Cohen, Yuval Emek, Oren Louidor and Jara Uitto

12:10 *Universal Shape Replicators via Self-Assembly with Attractive and Repulsive Forces* [21]
Cameron Chalk, Erik D. Demaine, Martin L. Demaine, Eric Martinez, Robert Schweller, Luis Vega and Tim Wylie

12:30–14:00 **Lunch**
Nixe I-II / Spiral Restaurant

Invited Talk 2

14:00–15:00
Chair: Mikkel Thorup
Room: *Las Arenas II-IV*

14:00 *Clique colouring random graphs* [22]
Colin McDiarmid, University of Oxford, United Kingdom

Graph colouring occupies a central place in theory and applications, in combinatorics, computer science and OR. We shall discuss clique colouring a graph G; that is, colouring the vertices so that no maximal clique is monochromatic. The least possible number of colours is the clique colouring number $\chi_c(G)$.

Typically, upper bounds on χ_c for random graphs are algorithmic. A key idea for clique colouring certain sparse random graphs is to construct triangle-free colour sets greedily. We shall discuss random perfect graphs, binomial random graphs and random geometric graphs. This is joint work with Nikola Yolov, and with Dieter Mitsche and Pawel Pralat.

ANALCO 2

15:00–16:35
Chair: Hsien-Kuei Hwang
Room: *Saray*

15:00 *Iterative Cutting and Pruning of Planar Trees* [23]
Benjamin Hackl, Sara Kropf and Helmut Prodinger

15:25 *On the distribution of random walk hitting times in random trees* [24]
Joubert Oosthuizen and Stephan Wagner

15:50 *A note on the scaling limits of random Pólya trees* [25]
Bernhard Gittenberger, Emma Yu Jin and Michael Wallner

16:15 *Rates of convergence for balanced irreducible two-color Pólya urns* [26]
Andrea Kuntschik and Ralph Neininger

SODA 2A

15:00–16:35
Chair: Mikkel Thorup
Room: *Las Arenas II-IV*

15:00 *Parameter-free Locality Sensitive Hashing for Spherical Range Reporting* [27]
Thomas D. Ahle, Martin Aumüller and Rasmus Pagh

15:25 *Distance Sensitive Bloom Filters Without False Negatives* [28]
Mayank Goswami, Rasmus Pagh, Francesco Silvestri and Johan Sivertsen

15:50 *Optimal Approximate Polytope Membership* [29]
Sunil Arya, Guilherme D. da Fonseca and David M. Mount

16:15 *Massively-Parallel Similarity Join, Edge-Isoperimetry, and Distance Correlations on the Hypercube* [30]
Paul Beame and Cyrus Rashtchian

SODA 2B

15:00–16:35
Chair: Nicole Megow
Room: *Las Arenas I*

15:00 *Even Delta-Matroids and the Complexity of Planar Boolean CSPs* [31]
Alexandr Kazda, Vladimir Kolmogorov and Michal Rolínek

15:25 *Linear Diophantine Equations, Group CSPs, and Graph Isomorphism* [32]
Christoph Berkholz and Martin Grohe

15:50 *Robust algorithms with polynomial loss for near-unanimity CSPs* [33]
Víctor Dalmau, Marcin Kozik, Adrei Krohkin, Konstantin Makarychev, Yury Makarychev and Jakub Opršal

16:15 *Parameterized Algorithms for Constraint Satisfaction Problems Above Average with Global Cardinality Constraints* [34]
Xue Chen and Yuan Zhou

SODA 2C

15:00–16:35
Chair: Adi Rosén
Room: *Nelva*

15:00 *Hardness of Permutation Pattern Matching* [35]
Vít Jelínek and Jan Kynčl

15:25 *pBWT: Achieving Succinct Data Structures for Parameterized Pattern Matching and Related Problems* [36]
Arnab Ganguly, Rahul Shah and Sharma V. Thankachan

15:50 *Space-Efficient Construction of Compressed Indexes in Deterministic Linear Time* [37]
Ian Munro, Gonzalo Navarro and Yakov Nekrich

16:15 *Sparse Suffix Tree Construction in Optimal Time and Space* [38]
Paweł Gawrychowski and Tomasz Kociumaka

16:35–17:00 **Coffee Break**
Las Arenas Foyer

ANALCO 3

17:00–18:35
Chair: Olivier Bodini
Room: *Saray*

17:00 *The recurrence function of a random Sturmian word* [39]
Pablo Rotondo and Brigitte Vallée

17:25 *An Extended Note on the Comparison-optimal Dual Pivot Quickselect* [40]
Daniel Krenn

17:50 *External Profile of Symmetric Digital Search Trees* [41]
Michael Drmota, Michael Fuchs, Hsien-Kuei Hwang and Ralph Neininger

18:15 *Partial Match Queries in Relaxed K-dt trees* [42]
Amalia Duch and Gustavo Lau

SODA 3A

17:00–19:00
Chair: Vijaya Ramachandran
Room: *Las Arenas II-IV*

17:00 *Fully dynamic all-pairs shortest paths with worst-case update-time revisited* [43]
Ittai Abraham, Shiri Chechik and Sebastian Krinninger

17:25 *Deterministic Partially Dynamic Single Source Shortest Paths for Sparse Graphs* [44]
Aaron Bernstein and Shiri Chechik

17:50 *Fully Dynamic Approximate Maximum Matching and Minimum Vertex Cover in $O(\log^3 n)$ Worst Case Update Time* [45]
Sayan Bhattacharya, Monika Henzinger and Danupon Nanongkai

18:15 *Connectivity Oracles for Graphs Subject to Vertex Failures* [46]
Ran Duan and Seth Pettie

18:40 *Fully Dynamic Connectivity in $O(\log n(\log\log n)^2)$ Amortized Expected Time* [47]
Shang-En Huang, Dawei Huang, Tsvi Kopelowitz and Seth Pettie

SODA 3B

17:00–19:00
Chair: Matt Weinberg
Room: *Las Arenas I*

17:00 *Best-Response Dynamics in Combinatorial Auctions with Item Bidding* [48]
Paul Dütting and Thomas Kesselheim

17:25 *Convergence of Incentive-Driven Dynamics in Fisher Markets* [49]
Krishnamurthy Dvijotham, Yuval Rabani and Leonard Schulman

17:50 *Totally Unimodular Congestion Games* [50]
Alberto Del Pia, Michael Ferris and Carla Michini

18:15 *Playing Anonymous Games using Simple Strategies* [51]
Yu Cheng, Ilias Diakonikolas and Alistair Stewart

18:40 *Computing Walrasian Equilibria: Fast Algorithms and Structural Properties* [52]
Renato Paes Leme and Sam Chiu-wai Wong

SODA 3C

17:00–19:00
Chair: Yair Bartal
Room: *Nelva*

17:00 *Metric embeddings with outliers* [53]
Anastasios Sidiropoulos, Dingkang Wang and Yusu Wang

17:25 *Probabilistic clustering of high dimensional norms* [54]
Assaf Naor

17:50 *Near-Optimal (Euclidean) Metric Compression* [55]
Piotr Indyk and Tal Wagner

18:15 *A Treehouse with Custom Windows: Minimum Distortion Embeddings into Bounded Treewidth Graphs* [56]
Amir Nayyeri and Benjamin Raichel

18:40 *Beyond Metric Embedding: Approximating Group Steiner Trees on Bounded Treewidth Graphs* [57]
Parinya Chalermsook, Syamantak Das, Bundit Laekhanukit and Daniel Vaz

19:00–19:30 **ANALCO/ALENEX Business Meeting**
Las Arenas II-IV

January 17th, 2017

9:00–17:00 **Registration**
Las Arenas Foyer

Invited Talk 3
9:00–10:00
Chair: Philip Klein
Room: *Las Arenas II-IV*

9:00 *Systems and Algorithms for the Internet of Things* [58]
Dina Katabi, Massachusetts Institute of Technology, USA
In this talk, I will describe how we design new IoT devices that analyze the surrounding radio signals to infer our movements, vital signs and even emotions. In particular, I will show how we can accurately track the 3D motion of people from the wireless signals reflected off their bodies, even if they are behind a wall. Such fine-grained tracking can also recognize our gestures, enabling us to control smart home devices simply by pointing at them. I will also describe mechanisms and algorithms to learn humans' breathing, heart rates, and basic emotions from the reflections of wireless signals off their bodies, and without requiring them to hold or wear any device.

10:00–10:30 **Coffee Break**
Las Arenas Foyer

ANALCO 4
10:30–12:30
Chair: Robert Sedgewick
Room: *Saray*

10:30 *Multivariate CLT follows from strong Rayleigh property* [59]
Subhroshekhar Ghosh, Thomas M. Liggett and Robin Pemantle

10:55 *Phase Transitions in Parameter Rich Optimization Problems* [60]
Joachim M. Buhmann, Julien Dumazert, Alexey Gronskiy and Wojciech Szpankowski

11:20 *On Symmetries of Non-Plane Trees in a Non-Uniform Model* [61]
Jacek Cichoń, Abram Magner, Wojciech Szpankowski and Krzysztof Turowski

11:45 *An analogue to Dixon's theorem for automaton groups* [62]
Thibault Godin

12:10 *A Refined Analysis of LSH for Well-dispersed Data Points* [63]
Wenlong Mou and Liwei Wang

SODA 4A
10:30–12:30
Chair: Vijaya Ramachandran
Room: *Las Arenas II-IV*

10:30 *Negative-Weight Shortest Paths and Unit Capacity Minimum Cost Flow in $\tilde{O}(m^{10/7}\log W)$ Time* [64]
Michael B. Cohen, Aleksander Madry, Piotr Sankowski and Adrian Vladu

10:55 *Generalized Preconditioning and Undirected Minimum-Cost Flow* [65]
Jonah Sherman

11:20 *Scaling Algorithms for Weighted Matching in General Graphs* [66]
Ran Duan, Seth Pettie and Hsin-Hao Su

11:45 *Near-Linear Time Approximation Schemes for Some Implicit Fractional Packing Problems* [67]
Chandra Chekuri and Kent Quanrud

12:10 *Fast and Memory-Efficient Algorithms for Evacuation Problems* [68]
Miriam Schlöter and Martin Skutella

SODA 4B
10:30–12:30
Chair: Maria Serna
Room: *Las Arenas I*

10:30 *Approximate Hierarchical Clustering via Sparsest Cut and Spreading Metrics* [69]
Moses Charikar and Vaggos Chatziafratis

10:55 *Approximating Multicut and the demand graph* [70]
Chandra Chekuri and Vivek Madan

11:20 *Minimum Fill-In: Inapproximability and Almost Tight Lower Bounds* [71]
Yixin Cao and R.B. Sandeep

11:45 *Minimizing the Union: Tight Approximations for Small Set Bipartite Vertex Expansion* [72]
Eden Chlamtáč, Michael Dinitz and Yury Makarychev

12:10 *Approximation Algorithms for Label Cover and The Log-Density Threshold* [73]
Eden Chlamtáč, Pasin Manurangsi, Dana Moshkovitz and Aravindan Vijayaraghavan

SODA 4C
10:30–12:30
Chair: Matt Weinberg
Room: *Nelva*

10:30 *An Axiomatic and an Average-Case Analysis of Algorithms and Heuristics for Metric Properties of Graphs* [74]
Michele Borassi, Pierluigi Crescenzi and Luca Trevisan

10:55 *Find Your Place: Simple Distributed Algorithms for Community Detection* [75]
Luca Becchetti, Andrea Clementi, Emanuele Natale, Francesco Pasquale and Luca Trevisan

11:20 *Sorting from Noisier Samples* [76]
Aviad Rubinstein and Shai Vardi

11:45 *Sampling on the Sphere by Mutually Orthogonal Subspaces* [77]
Uri Grupel

12:10 *Exponential Segregation in a Two-Dimensional Schelling Model with Tolerant Individuals* [78]
Nicole Immorlica, Robert Kleinberg, Brendan Lucier and Morteza Zadomighaddam

12:30–14:00 **Lunch**
Nixe I-II / Spiral Restaurant

ALENEX 1 Graphs and hypergraphs

14:00–16:00
Chair: Vijaya Ramachandran
Room: *Saray*

14:00 *Distributed Graph-based Topology Adaptation using Motif Signatures* [79]
Michael Stein, Karsten Weihe, Augustin Wilberg, Roland Kluge, Julian M. Klomp, Mathias Schnee, Lin Wang and Max Mühlhäuser

14:25 *Indexing Variation Graphs* [80]
Jouni Sirén

14:50 *Engineering a direct k-way Hypergraph Partitioning Algorithm* [81]
Yaroslav Akhremtsev, Tobias Heuer, Peter Sanders and Sebastian Schlag

15:15 *Computing Critical Nodes in Directed Graphs* [82]
Nilakantha Paudel, Loukas Georgiadis and Giuseppe F. Italiano

15:40 *I/O-efficient Generation of Massive Graphs Following the LFR Benchmark* [83]
Michael Hamann, Ulrich Meyer, Manuel Penschuck and Dorothea Wagner

SODA 5A

14:00–16:00
Chair: Adi Rosén
Room: *Las Arenas II-IV*

14:00 *Tight Bounds for Online TSP on the Line* [84]
Antje Bjelde, Yann Disser, Jan Hackfeld, Christoph Hansknecht, Maarten Lipmann, Julie Meißner, Kevin Schewior, Miriam Schlöter and Leen Stougie

14:25 *An Improved Upper Bound for the Universal TSP on the Grid* [85]
George Christodoulou and Alkmini Sgouritsa

14:50 *The (h,k)-Server Problem on Bounded-Depth Trees* [86]
Nikhil Bansal, Marek Eliáš, Grigorios Koumoutsos and Łukasz Jeż

15:15 *Online Lower Bounds via Duality* [87]
Yossi Azar, Ilan Reuven Cohen and Alan Roytman

15:40 *Polylogarithmic Bounds on the Competitiveness of Min-cost Perfect Matching with Delays* [88]
Yossi Azar, Ashish Chiplunkar and Haim Kaplan

SODA 5B

14:00–16:00
Chair: Shay Mozes
Room: *Las Arenas I*

14:00 *A Faster Pseudopolynomial Time Algorithm for Subset Sum* [89]
Konstantinos Koiliaris and Chao Xu

14:25 *A Near-Linear Pseudopolynomial Time Algorithm for Subset Sum* [90]
Karl Bringmann

14:50 *Computing minimum cuts in hypergraphs* [91]
Chandra Chekuri and Chao Xu

15:15 *Random Contractions and Sampling for Hypergraph and Hedge Connectivity* [92]
Mohsen Ghaffari, David R. Karger and Debmalya Panigrahi

15:40 *Sandpile prediction on a tree in near linear time* [93]
Akshay Ramachandran and Aaron Schild

SODA 5C

14:00–16:00
Chair: Andrea Pietracaprina
Room: *Nelva*

14:00 *Explicit Resilient Functions Matching Ajtai-Linial* [94]
Raghu Meka (talk given by Parikshit Gopalan)

14:25 *Optimal induced universal graphs for bounded-degree graphs* [95]
Noga Alon and Rajko Nenadov

14:50 *Approximation Algorithms for Finding Maximum Induced Expanders* [96]
Shayan Oveis Gharan and Alireza Rezaei

15:15 *Parallel Algorithms and Concentration Bounds for the Lovasz Local Lemma via Witness-DAGs* [97]
Bernhard Haeupler and David G. Harris

15:40 *Deterministic parallel algorithms for fooling polylogarithmic juntas and the Lovász Local Lemma* [98]
David G. Harris

16:00–16:25 **Coffee Break**
Las Arenas Foyer

ALENEX 2 Sorting and Searching

16:25–18:00
Chair: Subhash Suri
Room: *Saray*

16:25 *CSA++: Fast Pattern Search for Large Alphabets* [99]
Simon Gog, Alistair Moffat and Matthias Petri

16:50 *Robust Massively Parallel Sorting* [100]
Michael Axtmann and Peter Sanders

17:15 *Engineering External Memory Induced Suffix Sorting* [101]
Juha Kärkkäinen, Dominik Kempa, Simon J. Puglisi and Bella Zhukova

17:40 *Compact Dynamic Rewritable (CDRW) Arrays* [102]
Andreas Poyias, Simon J. Puglisi and Rajeev Raman

SODA 6A
16:25–18:00
Chair: Nicole Megow
Room: *Las Arenas II-IV*

16:25 *Online Submodular Maximization with Free Disposal: Randomization Beats 1/4 for Partition Matroids* [103]
T.-H. Hubert Chan, Zhiyi Huang, Shaofeng H.-C. Jiang, Ning Kang and Zhihao Gavin Tang

16:50 *Reordering Buffers with Logarithmic Diameter Dependency for Trees* [104]
Matthias Englert and Harald Räcke

17:15 *$O(depth)$-Competitive Algorithm for Online Multi-level Aggregation* [105]
Niv Buchbinder, Moran Feldman, Joseph Naor and Ohad Talmon

17:40 *Competitive analysis of the top-K ranking problem* [106]
Xi Chen, Sivakanth Gopi, Jieming Mao and Jon Schneider

SODA 6B
16:25–18:00
Chair: Grigory Yaroslavtsev
Room: *Las Arenas I*

16:25 *Statistical Query Algorithms for Mean Vector Estimation and Stochastic Convex Optimization* [107]
Vitaly Feldman, Cristóbal Guzmán and Santosh Vempala

16:50 *Sample Optimal Density Estimation in Nearly-Linear Time* [108]
Jayadev Acharya, Ilias Diakonikolas, Jerry Li and Ludwig Schmidt

17:15 *On Rationality of Nonnegative Matrix Factorization* [109]
Dmitry Chistikov, Stefan Kiefer, Ines Marušić, Mahsa Shirmohammadi and James Worrell

17:40 *Make Up Your Mind: The Price of Online Queries in Differential Privacy* [110]
Mark Bun, Thomas Steinke and Jonathan Ullman

SODA 6C
16:25–18:00
Chair: Matt Weinberg
Room: *Nelva*

16:25 *ETH Hardness for Densest-k-Subgraph with Perfect Completeness* [111]
Mark Braverman, Young Kun Ko, Aviad Rubinstein and Omri Weinstein

16:50 *The Rainbow at the End of the Line - A PPAD Formulation of the Colorful Carathéodory Theorem with Applications* [112]
Frédéric Meunier, Wolfgang Mulzer, Pauline Sarrabezolles and Yannik Stein

17:15 *Hardness of Continuous Local Search: Query Complexity and Cryptographic Lower Bounds* [113]
Pavel Hubáček and Eylon Yogev

17:40 *LP Relaxations of Some NP-Hard Problems Are as Hard as Any LP* [114]
Daniel Průša and Tomáš Werner

18:00–19:00 **SODA Business Meeting**
Las Arenas II-IV

20:15–22:15 **Banquet**
Restaurant Marítim

January 18th, 2017

9:00–12:30 **Registration**
Las Arenas Foyer

9:00–9:20 **Tribute to David Johnson** (speech by Mihalis Yannakakis)
Las Arenas II-IV

ALENEX 3 Text and documents
9:20–10:30
Chair: Rajeev Raman
Room: *Saray*

9:20 *Engineering a Distributed Full-Text Index* [115]
Johannes Fischer, Florian Kurpicz and Peter Sanders

9:45 *Elias-Fano meets Single-Term Top-k Document Retrieval* [116]
Julian Labeit and Simon Gog

10:10 *Efficient Set Intersection Counting Algorithm for Text Similarity Measures* [117]
Preethi Lahoti, Patrick K. Nicholson and Bilyana Taneva

SODA 7A
9:20–10:30
Chair: Philip Klein
Room: *Las Arenas II-IV*

9:20 *Approximation and Kernelization for Chordal Vertex Deletion / Feedback Vertex Set Inspired Kernel for Chordal Vertex Deletion* [118, 119]
Bart M.P. Jansen and Marcin Pilipczuk / Akanksha Agrawal, Daniel Lokshtanov, Pranabendu Misra, Saket Saurabh and Meirav Zehavi

9:45 *Fully polynomial-time parameterized computations for graphs and matrices of low treewidth* [120]
Fedor V. Fomin, Daniel Lokshtanov, Michał Pilipczuk, Saket Saurabh and Marcin Wrochna

10:10 *Spanning Circuits in Regular Matroids* [121]
Fedor V. Fomin, Petr A. Golovach, Daniel Lokshtanov and Saket Saurabh

SODA 7B

9:20–10:30
Chair: Shay Mozes
Room: *Las Arenas I*

9:20 *Algorithmic and Hardness Results for the Hub Labeling Problem* [122]
Haris Angelidakis, Yury Makarychev and Vsevolod Oparin

9:45 *Beyond Highway Dimension: Small Distance Labels Using Tree Skeletons* [123]
Adrian Kosowski and Laurent Viennot

10:10 $(1+\epsilon)$*-Approximate f-Sensitive Distance Oracles* [124]
Shiri Chechik, Sarel Cohen, Amos Fiat and Haim Kaplan

SODA 7C

9:20–10:30
Chair: Mikkel Thorup
Room: *Nelva*

9:20 *On the insertion time of random walk cuckoo hashing* [125]
Alan Frieze and Tony Johansson

9:45 *File Maintenance: When in Doubt, Change the Layout!* [126]
Michael A. Bender, Jeremy T. Fineman, Seth Gilbert, Tsvi Kopelowitz and Pablo Montes

10:10 *Cross-Referenced Dictionaries and the Limits of Write Optimization* [127]
Peyman Afshani, Michael A. Bender, Martín Farach-Colton, Jeremy T. Fineman, Mayank Goswami and Meng-Tsung Tsai

10:30–10:55 **Coffee Break**
Las Arenas Foyer

ALENEX 4 Optimization and randomness

10:55–12:30
Chair: Baruch Schieber
Room: *Saray*

10:55 *The Moser-Tardos Resample algorithm: Where is the limit? (an experimental inquiry)* [128]
Jan Dean Catarata, Scott Corbett, Harry Stern, Mario Szegedy, Tomas Vyskocil and Zheng Zhang

11:20 *Determining Tournament Payout Structures for Daily Fantasy Sports* [129]
Christopher Musco, Maxim Sviridenko and Justin Thaler

11:45 *Map Simplification with Topology Constraints: Exactly and in Practice* [130]
Stefan Funke, Thomas Mendel, Alexander Miller, Sabine Storandt and Maria Wiebe

12:10 *Asynchronous Column Generation* [131]
Saverio Basso and Alberto Ceselli

SODA 8A

10:55–12:30
Chair: Grigory Yaroslavtsev
Room: *Las Arenas II-IV*

10:55 *Testing for Forbidden Order Patterns in an Array* [132]
Ilan Newman, Yuri Rabinovich, Deepak Rajendraprasad and Christian Sohler

11:20 *Sequential measurements, disturbance and property testing* [133]
Aram W. Harrow, Cedric Yen-Yu Lin and Ashley Montanaro

11:45 *A tight bound for Green's arithmetic triangle removal lemma in vector spaces* [134]
László Miklós Lovász and Jacob Fox

12:10 *Fast Permutation Property Testing* [135]
Jacob Fox and Fan Wei

SODA 8B

10:55–12:30
Chair: Mikkel Thorup
Room: *Las Arenas I*

10:55 *Polynomial Kernels and Wideness Properties of Nowhere Dense Graph Classes* [136]
Stephan Kreutzer, Roman Rabinovich and Sebastian Siebertz

11:20 *Partitioning a Graph into Small Pieces with Applications to Path Transversal* [137]
Euiwoong Lee

11:45 *LP-branching algorithms based on biased graphs* [138]
Magnus Wahlström

12:10 *About the Structure of the Integer Cone and its Application to Bin Packing* [139]
Klaus Jansen and Kim-Manuel Klein

SODA 8C

10:55–12:30
Chair: Andrea Pietracaprina
Room: *Nelva*

10:55 *Online and Random-order Load Balancing Simultaneously* [140]
Marco Molinaro

11:20 *Building a Good Team: Secretary Problems and the Supermodular Degree* [141]
Moran Feldman and Rani Izsak

11:45 *Combinatorial Prophet Inequalities* [142]
Aviad Rubinstein and Sahil Singla

12:10 *Adaptivity Gaps for Stochastic Probing: Submodular and XOS Functions* [143]
Anupam Gupta, Viswanath Nagarajan and Sahil Singla

12:30–14:00 **Lunch**
Nixe I-II / Spiral Restaurant

ALENEX 5 Computational Geometry

14:00–16:00
Chair: Sándor Fekete
Room: *Saray*

14:00 *Parallel d-D Delaunay Triangulations in Shared and Distributed Memory* [144]
Daniel Funke and Peter Sanders

14:25 *An Efficient Algorithm for the 1D Total Visibility-Index Problem* [145]
Peyman Afshani, Mark de Berg, Henri Casanova, Benjamin Karsin, Colin Lambrechts, Nodari Sitchinava and Constantinos Tsirogiannis

14:50 *Computing the Expected Value and Variance of Geometric Measures* [146]
Frank Staals and Constantinos Tsirogiannis

15:15 *Growing Balls in $\mathbb{R}^d$* [147]
Daniel Bahrdt, Michael Becher, Stefan Funke, Filip Krumpe, Andre Nusser, Martin Seybold and Sabine Storandt

15:40 *I/O-Efficient Event Based Depression Flood Risk* [148]
Lars Arge, Mathias Rav, Sarfraz Raza and Morten Revsbæk

SODA 9A

14:00–16:00
Chair: Vahab Mirrokni
Room: *Las Arenas II-IV*

14:00 *$(1 + \Omega(1))$-Approximation to MAX-CUT Requires Linear Space* [149]
Michael Kapralov, Sanjeev Khanna, Madhu Sudan and Ameya Velingker

14:25 *On Estimating Maximum Matching Size in Graph Streams* [150]
Sepehr Assadi, Sanjeev Khanna and Yang Li

14:50 *Faster Sublinear Algorithms via Conditional Sampling* [151]
Themistoklis Gouleakis, Christos Tzamos and Manolis Zampetakis

15:15 *Input Sparsity Time Low-Rank Approximation via Ridge Leverage Score Sampling* [152]
Michael B. Cohen, Cameron Musco and Christopher Musco

15:40 *A Hybrid Sampling Scheme for Triangle Counting* [153]
John Kallaugher and Eric Price

SODA 9B

14:00–16:00
Chair: Maria Serna
Room: *Las Arenas I*

14:00 *An FPTAS for Counting Proper Four-Colorings on Cubic Graphs* [154]
Pinyan Lu, Kuan Yang, Chihao Zhang and Minshen Zhu

14:25 *Random cluster dynamics for the Ising model is rapidly mixing* [155]
Heng Guo and Mark Jerrum

14:50 *Approximately Sampling Elements with Fixed Rank in Graded Posets* [156]
Prateek Bhakta, Ben Cousins, Matthew Fahrbach and Dana Randall

15:15 *Random walks with the minimum degree local rule have $O(n^2)$ cover time* [157]
Roee David and Uriel Feige

15:40 *Random Walks and Evolving Sets: Faster Convergences and Limitations* [158]
Siu On Chan, Tsz Chiu Kwok and Lap Chi Lau

SODA 9C

14:00–16:00
Chair: Andrea Pietracaprina
Room: *Nelva*

14:00 *An $O(nm)$ time algorithm for finding the min length directed cycle in a graph* [159]
James B. Orlin and Antonio Sedeño-Noda

14:25 *Strong Connectivity in Directed Graphs under Failures, with Applications* [160]
Loukas Georgiadis, Giuseppe F. Italiano and Nikos Parotsidis

14:50 *Faster Algorithms for Computing Maximal 2-Connected Subgraphs in Sparse Directed Graphs* [161]
Shiri Chechik, Thomas Dueholm Hansen, Giuseppe F. Italiano, Veronika Loitzenbauer and Nikos Parotsidis

15:15 *Local Flow Partitioning for Faster Edge Connectivity* [162]
Monika Henzinger, Satish Rao and Di Wang

15:40 *Doubly Balanced Connected Graph Partitioning* [163]
Saleh Soltan, Mihalis Yannakakis and Gil Zussman

16:00–16:25 **Coffee Break**
Las Arenas Foyer

SODA 10A

16:25–18:00
Chair: Philip Klein
Room: *Las Arenas II-IV*

16:25 *Three Colors Suffice: Conflict-Free Coloring of Planar Graphs* [164]
Zachary Abel, Victor Alvarez, Erik D. Demaine, Sándor P. Fekete, Aman Gour, Adam Hesterberg, Phillip Keldenich and Christian Scheffer

16:50 *LP-Based Robust Algorithms for Noisy Minor-Free and Bounded Treewidth Graphs* [165]
Nikhil Bansal, Daniel Reichman and Seeun William Umboh

17:15 *LR-Drawings of Ordered Rooted Binary Trees and Near-Linear Area Drawings of Outerplanar Graphs* [166]
Fabrizio Frati, Maurizio Patrignani and Vincenzo Roselli

17:40 *Partial and Constrained Level Planarity* [167]
Guido Brückner and Ignaz Rutter

SODA 10B

16:25–18:00
Chair: Grigory Yaroslavtsev
Room: *Las Arenas I*

16:25 *Accurate and Nearly Optimal Sublinear Approximations to Ulam Distance* [168]
Timothy Naumovitz, Michael Saks and C. Ceshadhri

16:50 *A framework for analyzing resparsification algorithms* [169]
Rasmus Kyng, Jakub Pachocki, Richard Peng and Sushant Sachdeva

17:15 *Adaptive Matrix Vector Product* [170]
Santosh Vempala and David P. Woodruff

17:40 *Low-Rank PSD Approximation in Input-Sparsity Time* [171]
Kenneth L. Clarkson and David P. Woodruff

SODA 10C

16:25–18:00
Chair: Gerth Stølting Brodal
Room: *Nelva*

16:25 *Locally testable and locally correctable codes approaching the Gilbert-Varshamov bound* [172]
Sivakanth Gopi, Swastik Kopparty, Rafael Oliveira, Noga Ron-Zewi and Shubhangi Saraf

16:50 *Maximally Recoverable Codes for Grid-like Topologies* [173]
Parikshit Gopalan, Guangda Hu, Swastik Kopparty, Shubhangi Saraf, Carol Wang and Sergey Yekhanin

17:15 *MDS Code Constructions with Small Sub-packetization and Near-optimal Repair Bandwidth* [174]
Venkatesan Guruswami and Ankit Singh Rawat

17:40 *Bridging the Capacity Gap Between Interactive and One-Way Communication* [175]
Bernhard Haeupler and Ameya Velingker

SODA Awards 🏆

18:00–18:45
Chair: Philip Klein
Room: *Las Arenas II-IV*

18:00 *Subquadratic Algorithms for the Diameter and the Sum of Pairwise Distances in Planar Graphs* [176]
Sergio Cabello

18:25 *A $(2+\epsilon)$-Approximation for Maximum Weight Matching in the Semi-Streaming Model* [177]
Ami Paz and Gregory Schwartzman

January 19th, 2017

Invited Talk 4

9:00–10:00
Chair: Grigory Yaroslavtsev
Room: *Las Arenas II-IV*

9:00 *Uncertainty and Optimization* [178]
Nikhil Bansal, Eindhoven University of Technology, The Netherlands

In many settings the input arrives over time and an algorithm must make its current decisions without a precise knowledge of the future. Traditionally, these problems have been studied using mostly disjoint approaches such as competitive anlaysis, regret minimization and under stochastic assumptions. But in recent years, several interesting connections have emerged, often based on methods from optimization and duality. In this talk, we will outline some of these developments as well as some future research directions.

10:00–10:30 **Coffee Break**
Las Arenas Foyer

SODA 11A

10:30–12:30
Chair: Grigory Yaroslavtsev
Room: *Las Arenas II-IV*

10:30 *Completeness for First-Order Properties on Sparse Structures with Algorithmic Applications* [179]
Jiawei Gao, Russell Impagliazzo, Antonina Kolokolova and Ryan Williams

10:55 *Faster Online Matrix-Vector Multiplication* [180]
Kasper Green Larsen and Ryan Williams

11:20 *Beating Brute Force for Systems of Polynomial Equations over Finite Fields* [181]
Daniel Lokshtanov, Ramamohan Paturi, Suguru Tamaki, Ryan Williams and Huacheng Yu

11:45 *The Complexity of Simulation and Matrix Multiplication* [182]
Massimo Cairo and Romeo Rizzi

12:10 *Better Approximations for Tree Sparsity in Nearly-Linear Time* [183]
Arturs Backurs, Piotr Indyk and Ludwig Schmidt

SODA 11B

10:30–12:30
Chair: Neal Young
Room: *Las Arenas I*

10:30 *Counting matchings in irregular bipartite graphs and random lifts* [184]
Marc Lelarge

10:55 *A constant-time algorithm for middle levels Gray codes* [185]
Torsten Mütze and Jerri Nummenpalo

11:20 *On Max-Clique for intersection graphs of sets and the Hadwiger-Debrunner numbers* [186]
Chaya Keller, Shakhar Smorodinky and Gábor Tardos

11:45 *Core congestion is inherent in hyperbolic networks* [187]
Victor Chepoi, Feodor Dragan and Yann Vaxès

12:10 *Better upper bounds on the Füredi–Hajnal limits of permutations* [188]
Josef Cibulka and Jan Kynčl

SODA 11C

10:30–12:30
Chair: Shay Mozes
Room: *Nelva*

10:30 *Popularity, Mixed Matchings, and Self-Duality* [189]
Chien-Chung Huang and Telikepalli Kavitha

10:55 *Constant Approximation Algorithm for Non-Uniform Capacitated Multi-Item Lot-Sizing via Strong Covering Inequalities* [190]
Shi Li (talk given by Mateusz Lewandowski)

11:20 *Small Extended Formulation for Knapsack Cover Inequalities from Monotone Circuits* [191]
Abbas Bazzi, Samuel Fiorini, Sangxia Huang and Ola Svensson

11:45 *Extension Complexity Lower Bounds for Mixed-Integer Extended Formulations* [192]
Robert Hildebrand, Robert Weismantel and Rico Zenklusen

12:10 *Opting Into Optimal Matchings* [193]
Avrim Blum, Ioannis Caragiannis, Nika Haghtalab, Ariel D. Procaccia, Eviatar B. Procaccia and Rohit Vaish

12:30–14:00 **Lunch**
Las Arenas Foyer

SODA 12A

14:00–16:00
Chair: Neal Young
Room: *Las Arenas II-IV*

14:00 *Firefighting on Trees Beyond Integrality Gaps* [194]
David Adjiashvili, Andrea Baggio and Rico Zenklusen

14:25 *Beating Approximation Factor Two for Weighted Tree Augmentation with Bounded Costs* [195]
David Adjiashvili

14:50 *Simplex Transformations and the Multiway Cut Problem* [196]
Niv Buchbinder, Roy Schwartz and Baruch Weizman

15:15 *To Augment or Not to Augment: Solving Unsplittable Flow on a Path by Creating Slack* [197]
Fabrizio Grandoni, Tobias Mömke, Andreas Wiese and Hang Zhou

15:40 *Optimization of Bootstrapping in Circuits* [198]
Fabrice Benhamouda, Tancrède Lepoint, Claire Mathieu and Hang Zhou

SODA 12B

14:00–16:00
Chair: Anne Driemel
Room: *Las Arenas I*

14:00 *Geodesic Spanners for Points on a Polyhedral Terrain* [199]
Mohammad Ali Abam, Mark de Berg and Mohammad Javad Rezaei Seraji

14:25 *Computing the Fréchet Distance between Real-Valued Surfaces* [200]
Bettina Speckmann, Kevin Buchin and Tim Ophelders

14:50 *Incidences with curves and surfaces in three dimensions, with applications to distinct and repeated distances* [201]
Micha Sharir and Noam Solomon

15:15 *Eliminating Depth Cycles among Triangles in Three Dimensions* [202]
Boris Aronov, Edward Y. Miller and Micha Sharir

15:40 *Dynamic Planar Voronoi Diagrams for General Distance Functions and their Algorithmic Applications* [203]
Haim Kaplan, Wolfgang Mulzer, Liam Roditty, Paul Seiferth and Micha Sharir

SODA 12C

14:00–16:00
Chair: Maria Serna
Room: *Nelva*

14:00 *Distributed Degree Splitting, Edge Coloring, and Orientations* [204]
Mohsen Ghaffari and Hsin-Hao Su

14:25 *Tight Network Topology Dependent Bounds on Rounds of Communication* [205]
Arkadev Chattopadhyay, Michael Langberg, Shi Li and Atri Rudra

14:50 *Minimizing Message Size in Stochastic Communication Patterns: Fast Self-Stabilizing Protocols with 3 bits* [206]
Lucas Boczkowski, Amos Korman and Emanuele Natale

15:15 *Time-space Trade-offs in Population Protocols* [207]
Dan Alistarh, James Aspnes, David Eisenstat, Rati Gelashvili and Ronald L. Rivest

15:40 *Fair Coin Flipping: Tighter Analysis and the Many-Party Case* [208]
Niv Buchbinder, Iftach Haitner, Nissan Levi and Eliad Tsfadia

16:00–16:30 **Coffee Break**
Las Arenas Foyer

SODA 13A

16:30–18:30
Chair: Vahab Mirrokni
Room: *Las Arenas II-IV*

16:30 *Fair Scheduling via Iterative Quasi-Uniform Sampling* [209]
Sungjin Im and Benjamin Moseley

16:55 *A Logarithmic Additive Integrality Gap for Bin Packing* [210]
Rebecca Hoberg and Thomas Rothvoss

17:20 *Iterative Partial Rounding for Vertex Cover with Hard Capacities / Tight Algorithms for Vertex Cover with Hard Capacities on Multigraphs and Hypergraphs* [211, 212]
Mong-Jen Kao / Sam Chiu-wai Wong

17:45 *Unrelated Machine Scheduling of Jobs with Uniform Smith Ratios* [213]
Christos Kalaitzis, Ola Svensson and Jakub Tarnawski

18:10 *On the Configuration-LP of the Restricted Assignment Problem* [214]
Klaus Jansen and Lars Rohwedder

SODA 13B

16:30–18:30
Chair: Neal Young
Room: *Las Arenas I*

16:30 *Approximating Spanners and Directed Steiner Forest: Upper and Lower Bounds* [215]
Eden Chlamtáč, Michael Dinitz, Guy Kortsarz and Bundit Laekhanukit

16:55 *A Hierarchy of Lower Bounds for Sublinear Additive Spanners* [216]
Amir Abboud, Greg Bodwin and Seth Pettie

17:20 *LAST but not Least: Buy-at-Bulk via Online Spanners* [217]
Anupam Gupta, R. Ravi, Kunal Talwar and Seeun William Umboh

17:45 *Linear Size Distance Preservers* [218]
Greg Bodwin

18:10 *Efficient Algorithms for Constructing Very Sparse Spanners and Emulators* [219]
Michael Elkin and Ofer Neiman

SODA 13C

16:30–18:30
Chair: Anne Driemel
Room: *Nelva*

16:30 *When and Why the Topological Coverage Criterion Works* [220]
Nicholas J. Cavanna, Kirk P. Gardner and Donald R. Sheehy

16:55 *Deciding Contractibility of a Non-Simple Curve on the Boundary of a 3-Manifold* [221]
Éric Colin de Verdière and Salman Parsa

17:20 *An Efficient Representation for Filtrations of Simplicial Complexes* [222]
Jean-Daniel Boissonnat and Karthik C. S.

17:45 *A polynomial time algorithm to compute quantum invariants of 3-manifolds with bounded first Betti number* [223]
Clément Maria and Jonathan Spreer

18:10 *Parameter-free Topology Inference and Sparsification for Data on Manifolds* [224]
Tamal Dey, Zhe Dong and Yusu Wang

Speaker Index

SODA / ALENEX / ANALCO 2017

ABSTRACTS

ACM-SIAM Symposium on Discrete Algorithms
January 16-19, 2017
Hotel Porta Fira
Barcelona, Spain

ANALCO17 Meeting on Analytic Algorithmics and Combinatorics
January 16-17, 2017
Hotel Porta Fira
Barcelona, Spain

Organized by:
Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

ALENEX17
Meeting on
Algorithm Engineering & Experiments
January 17-18, 2017
Hotel Porta Fira
Barcelona, Spain

These abstracts appear as sent by the authors to the local organization. They should correspond to the final version of the abstracts as they appear in the proceedings of the three conferences. Here, the abstracts are given in chronological order of sessions, then in chronological order within the corresponding session. Numbers are cross-references to the detailed program. The name of speakers (whenever that information was available) appear underlined.

Abstracts

[1] Benjamin Recht. A Dynamical View of Convex Optimization Algorithms. **Invited Talk 1** (Las Arenas Foyer). January 16th, 2017, 09:00-10:00

Convex optimization provides a powerful toolkit for robust and efficient solutions of difficult engineering problems. However, it is laborious to extend and generalize the mathematical guarantees of simple convex primitives. In this talk, I will show that much of the analysis and design of optimization algorithms can be automated, and will demonstrate that multiple objectives—such as robustness, accuracy, and speed—can be balanced using tools from dynamical systems.

Noting that all iterative algorithms are dynamical systems, I will illustrate how most of the popular methods in optimization can be cast as a family of feedback systems studied in control theory. Leaning on this abstraction enables us to apply powerful, control-theoretic methods to algorithm analysis. I will show how the convergence rates of most common algorithms—including gradient descent, mirror descent, Nesterov's method, etc.—can be verified using a unified set of potential functions. I will then describe how such potential functions can themselves be found by solving constant-sized semidefinite programming problems. I will close with a discussion of how these techniques can be used to search for optimization algorithms with desired performance characteristics, proposing a new methodology for algorithm design.

.

[2] Olivier Bodini, Matthieu Dien, Antoine Genitrini and Frédéric Peschanski. The Ordered and Colored Products in Analytic Combinatorics: Application to the Quantitative Study of Synchronizations in Concurrent Processes. **ANALCO 1** (Saray). January 16th, 2017, 10:30-10:50

In this paper, we study two operators for composing combinatorial classes: the *ordered product* and its dual, the *colored product*. These operators have a natural interpretation in terms of *Analytic Combinatorics*, in relation with combinations of Borel and Laplace transforms. Based on these new constructions, we exhibit a set of *transfer theorems* and closure properties. We also illustrate the use of these operators to specify increasingly labeled structures tightly related to Series-Parallel constructions and concurrent processes. In particular, we provide a quantitative analysis of Fork/Join (FJ) parallel processes, a particularly expressive example of such a class.

.

[3] Miklós Bóna and Boris Pittel. On the cycle structure of the product of random maximal cycles. **ANALCO 1** (Saray). January 16th, 2017, 10:55-11:15

The subject of this paper is the cycle structure of the random permutation σ of $[N]$, which is the product of k independent random cycles of maximal length N. We use the character-based Fourier transform to study the number of cycles of σ and also the distribution of the elements of the subset $[\ell]$ among the cycles of σ.

.

[4] Cédric Chauve, Éric Fusy and Jérémie Lumbroso. An Exact Enumeration of Distance-Hereditary Graphs. **ANALCO 1** (Saray). January 16th, 2017, 11:20-11:40

Distance-hereditary graphs form an important class of graphs, from the theoretical point of view, due to the fact that they are the totally decomposable graphs for the split-decomposition. The previous best enumerative result for these graphs is from Nakano *et al.* (J. Comp. Sci. Tech., 2007), who have proven that the number of distance-hereditary graphs on n vertices is bounded by $2^{\lceil 3.59n \rceil}$.

In this paper, using classical tools of enumerative combinatorics, we improve on this result by providing an *exact* enumeration and *full asymptotic* of distance-hereditary graphs, which allows to show that the number of distance-hereditary graphs on n vertices is tightly bounded by $(7.24975\ldots)^n$—opening the perspective such graphs could be encoded on $3n$ bits. We also provide the exact enumeration and full asymptoticss of an important subclass, the 3-leaf power graphs. Our work illustrates the power of revisiting graph decomposition results through the framework of analytic combinatorics.

.

[5] Marc Noy, Vonjy Rasendrahasina, Vlady Ravelomanana and Juanjo Rué. Isolated cycles of critical random graphs. **ANALCO 1** (Saray). January 16th, 2017, 11:45-12:05

Consider the Erdős-Rényi random graph $G(n, M)$ built with n vertices and M edges uniformly randomly chosen from the set of $\binom{n}{2}$ edges. Let L be a set of positive integers. For any number of edges $M \leq \frac{n}{2} + O(n^{2/3})$, we compute – via analytic combinatorics – the number of isolated cycles of $G(n, M)$ whose length is in L.

.

[6] Amanda Pascoe Streib and Noah Streib. Cycle Basis Markov Chains for the Ising Model. **ANALCO 1** (Saray). January 16th, 2017, 12:10-12:30

A very challenging problem from statistical physics is computing the partition function of the ferromagnetic Ising model, even in the relatively simple case of no applied field. In this case, the partition function can be written as a function of the subgraphs of the underlying graph in which all vertices have even degree. In their seminal work, Jerrum and Sinclair showed that this quantity can be approximated by a rapidly converging Markov chain on all subgraphs. However, their chain frequently leaves the space of even subgraphs. Our aim is to devise and analyze a new class of Markov chains that do not leave this space, in the hopes of finding a faster sampling algorithm. We define Markov chains by viewing the even subgraphs as a vector space (often called the cycle space) whose transitions are defined by the addition of basis elements. The rate of convergence depends on the basis chosen, and our analysis proceeds by dividing bases into two types, short and long. The classical single-site update Markov chain known as Glauber dynamics is a special case of our short cycle basis Markov chains. We show that for any graph and any long basis, there is a temperature for which the corresponding Markov chain requires an exponential time to mix. Moreover, we show that for d-dimensional grids with $d \geq 2$—those of the most physical importance—all fundamental bases (a natural class of bases derived from spanning trees) are long. For the 2-dimensional grid on the torus, we show that there is a temperature for which the Markov chain requires exponential time for any chosen basis.

.

[7] Sariel Har-Peled and Sepideh Mahabadi. Proximity in the Age of Distraction: Robust Approximate Nearest Neighbor Search. **SODA 1A** (Las Arenas II-IV). January 16th, 2017, 10:30-10:50

We introduce a new variant of the nearest neighbor search problem, which allows for some coordinates of the dataset to be arbitrarily corrupted or unknown. Formally, given a dataset of n points $P = P_1, ..., P_n$ in high-dimensions, and a parameter k, the goal is to preprocess the dataset, such that given a query point q, one can compute quickly a point $p \in P$, such that the distance of the query to the point p is minimized, when ignoring the "optimal" k coordinates. Note, that the coordinates being ignored are a function of both the query point and the point returned. We present a general reduction from this problem to answering ANN queries, which is similar in spirit to LSH (locality sensitive hashing) [IM98]. Specifically, we give a sampling technique which achieves a bi-criterion approximation for this problem. If the distance to the nearest neighbor after ignoring k coordinates is r, the data-structure returns a point that is within a distance of $O(r)$ after ignoring $O(k)$ coordinates. We also present other applications and further extensions and refinements of the above result. The new data-structures are simple and (arguably) elegant, and should be practical – specifically, all bounds are polynomial in all relevant parameters (including the dimension of the space, and the robustness parameter k).

.

[8] Georgia Avarikioti, Ioannis Z. Emiris, Loukas Kavouras and Ioannis Psarros. High-dimensional approximate r-nets. **SODA 1A** (Las Arenas II-IV). January 16th, 2017, 10:55-11:15

The construction of r-nets offers a powerful tool in computational and metric geometry. We focus on high-dimensional spaces and present a new randomized algorithm which efficiently computes approximate r-nets with respect to Euclidean distance. For any fixed $\epsilon > 0$, the approximation factor is $1 + \epsilon$ and the complexity is polynomial in the dimension and subquadratic in the number of points. The algorithm succeeds with high probability. Specifically, we improve upon the best previously known (LSH-based) construction of Eppstein et al. (D. Eppstein, S. Har-Peled, and A. Sidiropoulos. Approximate greedy clustering and distance selection for graph metrics. CoRR, abs/1507.01555, 2015) in terms of complexity, by reducing the dependence on ϵ, provided that ϵ is sufficiently small. Our method does not require LSH but, instead, follows Valiant's (G. Valiant. Finding correlations in subquadratic time, with applications to learning parities and the closest pair problem. J. ACM, 62(2):13, 2015) approach in designing a sequence of reductions of our problem to other problems in different spaces, under Euclidean distance or inner product, for which r-nets are computed efficiently and the error can be controlled. Our result immediately implies efficient solutions to a number of geometric problems in high dimension, such as finding the $(1 + \epsilon)$-approximate kth nearest neighbor distance in time subquadratic in the size of the input.

.

[9] Tobias Christiani. A Framework for Similarity Search with Space-Time Tradeoffs using Locality-Sensitive Filtering. **SODA 1A** (Las Arenas II-IV). January 16th, 2017, 11:20-11:40

We present a framework for similarity search based on Locality-Sensitive Filtering (LSF), generalizing the Indyk-Motwani (STOC 1998) Locality-Sensitive Hashing (LSH) framework to support space-time tradeoffs. Given a family of filters, defined as a distribution over pairs of subsets of space that satisfies certain locality-sensitivity properties, we can construct a dynamic data structure that solves the approximate near neighbor problem in d-dimensional space with query time $dn^{\rho_q + o(1)}$, update time $dn^{\rho_u + o(1)}$, and space usage $dn + n^{1+\rho_u+o(1)}$ where n denotes the number of points in the data structure. The space-time tradeoff is tied to the tradeoff between query time and update time (insertions/deletions), controlled by the exponents ρ_q, ρ_u that are determined by the filter family.
Locality-sensitive filtering was introduced by Becker et al. (SODA 2016) together with a framework yielding a single, balanced, tradeoff between query time and space, further relying on the assumption of an efficient oracle for the filter evaluation algorithm. We extend the LSF framework to support space-time tradeoffs and through a combination of existing

techniques we remove the oracle assumption.
Laarhoven (arXiv 2015), building on Becker et al., introduced a family of filters with space-time tradeoffs for the high-dimensional unit sphere under inner product similarity and analyzed it for the important special case of random data. We show that a small modification to the family of filters gives a simpler analysis that we use, together with our framework, to provide guarantees for worst-case data. Through an application of Bochner's Theorem from harmonic analysis by Rahimi & Recht (NIPS 2007), we are able to extend our solution on the unit sphere to $\mathbb{R}^d$ under the class of similarity measures corresponding to real-valued characteristic functions. For the characteristic functions of s-stable distributions we obtain a solution to the (r, cr)-near neighbor problem in ℓ_s^d-spaces with query and update exponents $\rho_q = \frac{c^s(1+\lambda)^2}{(c^s+\lambda)^2}$ and $\rho_u = \frac{c^s(1-\lambda)^2}{(c^s+\lambda)^2}$ where $\lambda \in [-1, 1]$ is a tradeoff parameter. This result improves upon the space-time tradeoff of Kapralov (PODS 2015) and is shown to be optimal in the case of a balanced tradeoff, matching the LSH lower bound by O'Donnell et al. (ITCS 2011) and a similar LSF lower bound proposed in this paper. Finally, we show a lower bound for the space-time tradeoff on the unit sphere that matches Laarhoven's and our own upper bound in the case of random data.

.

[10] Alexandr Andoni, Thijs Laarhoven, Ilya Razenshteyn and Erik Waingarten. Optimal Hashing–based Time–Space Trade-offs for Approximate Near Neighbors. **SODA 1A** (Las Arenas II-IV). January 16th, 2017, 11:45-12:05

We show tight upper and lower bounds for time–space trade-offs for the c-approximate Near Neighbor Search problem. For the d-dimensional Euclidean space and n-point datasets, we develop a data structure with space $n^{1+\rho_u+o(1)} + O(dn)$ and query time $n^{\rho_q+o(1)} + dn^{o(1)}$ for every $\rho_u, \rho_q \geq 0$ with:

$$c^2\sqrt{\rho_q} + (c^2 - 1)\sqrt{\rho_u} = \sqrt{2c^2 - 1}. \tag{1}$$

In particular, for the approximation $c = 2$ we get:

- Space $n^{1.77\ldots}$ and query time $n^{o(1)}$, significantly improving upon known data structures that support very fast queries [Indyk, Motwani 1998] [Kushilevitz, Ostrovsky, Rabani 2000];
- Space $n^{1.14\ldots}$ and query time $n^{0.14\ldots}$, matching the optimal data-dependent Locality-Sensitive Hashing (LSH) from [Andoni, Razenshteyn 2015];
- Space $n^{1+o(1)}$ and query time $n^{0.43\ldots}$, making significant progress in the regime of near-linear space, which is arguably of the most interest for practice [Lv, Josephson, Wang, Charikar, Li 2007].

This is the first data structure that achieves sublinear query time and near-linear space for *every* approximation factor $c > 1$, improving upon [Kapralov 2015]. The data structure is a culmination of a long line of work on the problem for all space regimes; it builds on Spherical Locality-Sensitive Filtering [Becker, Ducas, Gama, Laarhoven 2016] and data-dependent hashing [Andoni, Indyk, Nguyen, Razenshteyn 2014][Andoni, Razenshteyn 2015]. Our matching lower bounds are of two types: conditional and unconditional. First, we prove tightness of the *whole* trade-off in a restricted model of computation, which captures all known hashing-based approaches. We then show *unconditional* cell-probe lower bounds for one and two probes that match the trade-off for $\rho_q = 0$, improving upon the best known lower bounds from [Panigrahy, Talwar, Wieder 2010]. In particular, this is the first space lower bound (for *any* static data structure) for two probes which is not polynomially smaller than the one-probe bound. To show the result for two probes, we establish and exploit a connection to *locally-decodable codes*.

.

[11] Alexandr Andoni, Ilya Razenshteyn and Negev Shekel Nosatzki. LSH Forest: Practical Algorithms Made Theoretical. **SODA 1A** (Las Arenas II-IV). January 16th, 2017, 12:10-12:30

We analyze *LSH Forest* [Bawa, Condie, Ganesan 2005]—a popular heuristic for the nearest neighbor search—and show that a careful yet simple modification of it outperforms "vanilla" LSH algorithms. The end result is the first instance of a simple, practical algorithm that provably leverages data-dependent hashing to improve upon data-oblivious LSH. Here is the entire algorithm for the d-dimensional Hamming space. The LSH Forest, for a given dataset, applies a random permutation to all the d coordinates, and builds a *trie* on the resulting strings. In our modification, we further augment this trie: for each node, we store a *constant* number of points close to the mean of the corresponding subset of the dataset, which are compared to any query point reaching that node. The overall data structure is simply several such tries sampled independently. While the new algorithm does not *quantitatively* improve upon the best data-dependent hashing algorithms from [Andoni, Razenshteyn 2015] (which are known to be optimal), it is significantly simpler, being based on a practical heuristic, and is provably better than the best LSH algorithm for the Hamming space [Indyk, Motwani 1998] [Har-Peled, Indyk, Motwani 2012].

.

[12] Sandy Heydrich and Andreas Wiese. Faster approximation schemes for the two-dimensional knapsack problem. **SODA 1B** (Las Arenas I). January 16th, 2017, 10:30-10:50

An important question in theoretical computer science is to determine the best possible running time for solving a problem at hand. For geometric optimization problems, we often understand their complexity on a rough scale, but not very well on a finer scale. One such example is the two-dimensional knapsack problem for squares. There is a polynomial time $(1+\epsilon)$-approximation algorithm for it (i.e., a PTAS) but the running time of this algorithm is triple exponential in $1/\epsilon$, i.e., $\Omega(n^{2^{2^{1/\epsilon}}})$. A double or triple exponential dependence on $1/\epsilon$ is inherent in how this and several other algorithms for other geometric problems work. In this paper, we present an EPTAS for knapsack for squares, i.e., a $(1+\epsilon)$-approximation algorithm with a running time of $O_\epsilon(1) \cdot n^{O(1)}$. In particular, the exponent of n in the running time does not depend on ϵ at all! Since there can be no FPTAS for the problem (unless P = NP) this is the best kind of approximation scheme we can hope for. To achieve this improvement, we introduce two new key ideas: We present a fast method to guess the $\Omega(2^{2^{1/\epsilon}})$ relatively large squares of a suitable near-optimal packing instead of using brute-force enumeration. Secondly, we introduce an *indirect guessing* framework to define sizes of cells for the remaining squares. In the previous PTAS each of these steps needs a running time of $\Omega(n^{2^{2^{1/\epsilon}}})$ and we improve both to $O_\epsilon(1) \cdot n^{O(1)}$. We complete our result by giving an algorithm for two-dimensional knapsack for rectangles under $(1+\epsilon)$-resource augmentation. In this setting, we also improve the best known running time of $\Omega(n^{1/\epsilon^{1/\epsilon}})$ to $O_\epsilon(1) \cdot n^{O(1)}$ and compute even a solution with optimal profit, in contrast to the best previously known polynomial time algorithm for this setting that computes only an approximation. We believe that our new techniques have the potential to be useful for other settings as well.

.

[13] Sebastian Morr. Split Packing: An Algorithm for Packing Circles with Optimal Worst-Case Density. **SODA 1B** (Las Arenas I). January 16th, 2017, 10:55-11:15

In the classic *circle packing problem*, one asks whether a given set of circles can be packed into the unit square. This problem is known to be NP-hard. In this paper, we present a new sufficient condition using only the circles' combined area: It is possible to pack any circle instance with a combined area of up to ≈ 0.5390. This bound is tight, in the sense that for any larger combined area, there are instances which cannot be packed, which is why we call this number the problem's *critical density*. Similar results have long been known for squares, but to the best of our knowledge, this paper gives the first results of this type for circular objects. Our proof is constructive: We describe a subdivision scheme which recursively splits the circles into groups and then packs these into subcontainers. We call this algorithm *Split Packing*. Beside realizing all packings up to the critical density bound, Split Packing also serves as a constant-factor approximation algorithm when looking for the smallest square in which a given set of circles can be packed. We believe that the ideas behind Split Packing are interesting and elegant on their own, and we see many opportunities to apply this technique in the context of other packing and covering problems. A browser-based, interactive visualization of the Split Packing approach and other related material can be found at https://morr.cc/split-packing/.

.

[14] Lingxiao Huang and Jian Li. Stochastic k-Center and j-Flat-Center Problems. **SODA 1B** (Las Arenas I). January 16th, 2017, 11:20-11:40

Solving geometric optimization problems over uncertain data has become increasingly important in many applications and has attracted a lot of attentions in recent years. In this paper, we study two important geometric optimization problems, the k-center problem and the j-flat-center problem, over stochastic/uncertain data points in Euclidean spaces. For the stochastic k-center problem, we would like to find k points in a fixed dimensional Euclidean space, such that the expected value of the k-center objective is minimized. For the stochastic j-flat-center problem, we seek a j-flat (i.e., a j-dimensional affine subspace) such that the expected value of the maximum distance from any point to the j-flat is minimized. We consider both problems under two popular stochastic geometric models, the existential uncertainty model, where the existence of each point may be uncertain, and the locational uncertainty model, where the location of each point may be uncertain. We provide the first PTAS (Polynomial Time Approximation Scheme) for both problems under the two models. Our results generalize the previous results for stochastic minimum enclosing ball and stochastic enclosing cylinder.

.

[15] Alfonso Cevallos, Friedrich Eisenbrand and Rico Zenklusen. Local search for max-sum diversification. **SODA 1B** (Las Arenas I). January 16th, 2017, 11:45-12:05

We provide simple and fast polynomial-time approximation schemes (PTASs) for several variants of the max-sum diversification problem which, in its most basic form, is as follows: given n points in a Euclidean space and an integer k, select k points such that the average distance among the chosen points is maximized. This problem is commonly applied in web search and information retrieval in order to select a diverse set of representative points from the input. In this context, it has recently received a lot of attention. We present new techniques to analyze natural local search algorithms. This leads to a $(1-O(1/k))$-approximation for distances of negative type, even subject to a general matroid constraint of rank k, in time $O(nk^2 \log k)$, when assuming that distance evaluations and calls to the independence oracle are constant time. Negative-type distances include as special cases Euclidean and Manhattan distances, among other natural distances. Our result easily transforms into a PTAS. It improves on the only previously known PTAS for this setting, which relies on convex optimization techniques in an n-dimensional space and is impractical for large data sets. In contrast, our procedure has an (optimal) linear dependence on n. Using generalized exchange properties of matroid intersection, we show that a PTAS can be obtained for matroid intersection constraints as well. Moreover, our techniques, being based on local search, are conceptually simple and allow for various extensions. In particular, we get asymptotically optimal $O(1)$-approximations when combining the classic dispersion function with a monotone submodular objective, which is a very common class of functions to measure diversity and relevance. This result leverages recent advances on local-search techniques based on proxy functions to obtain optimal approximations for monotone submodular function maximization subject to a matroid constraint.

.

[16] László Kozma and Tobias Mömke. Maximum Scatter TSP in Doubling Metrics. **SODA 1B** (Las Arenas I). January 16th, 2017, 12:10-12:30

We study the problem of finding a tour of n points in which *every edge* is *long*. More precisely, we wish to find a tour that visits every point exactly once, maximizing the length of the shortest edge in the tour. The problem is known as Maximum Scatter TSP, and was introduced by Arkin et al. (SODA 1997), motivated by applications in manufacturing and medical imaging. Arkin et al. gave a 0.5-approximation for the metric version of the problem and showed that this is the best possible ratio achievable in polynomial time (assuming P $\neq$ NP). Arkin et al. raised the question of whether a better approximation ratio can be obtained in the Euclidean plane. We answer this question in the affirmative in a more general setting, by giving a $(1-\epsilon)$-approximation algorithm for d-dimensional doubling metrics, with running time $\tilde{O}\left(n^3 + 2^{O(K \log K)}\right)$, where $K \leq \left(\frac{13}{\epsilon}\right)^d$. As a corollary we obtain (i) an efficient polynomial-time approximation scheme (EPTAS) for all constant dimensions d, (ii) a polynomial-time approximation scheme (PTAS) for dimension $d = \log\log n/c$, for a sufficiently large constant c, and (iii) a PTAS for constant d and $\epsilon = \Omega(1/\log\log n)$. Furthermore, we show the dependence on d in our approximation scheme to be essentially optimal, unless Satisfiability can be solved in subexponential time.

.

[17] Rafail Ostrovsky, Yuval Rabani and Arman Yousefi. Matrix Balancing in L_p Norms: Bounding the Convergence rate of Osborne's Iteration. **SODA 1C** (Nelva). January 16th, 2017, 10:30-10:50

We study an iterative matrix conditioning algorithm due to Osborne (1960). The goal of the algorithm is to convert a square matrix into a *balanced* matrix where every row and corresponding column have the same norm. The original algorithm was proposed for balancing rows and columns in the L_2 norm, and it works by iterating over balancing a row-column pair in fixed round-robin order. Variants of the algorithm for other norms have been heavily studied and are implemented as standard preconditioners in many numerical linear algebra packages. Recently, Schulman and Sinclair (2015), in a first result of its kind for any norm, analyzed the rate of convergence of a variant of Osborne's algorithm that uses the L_∞ norm and a different order of choosing row-column pairs. In this paper we study matrix balancing in the L_1 norm and other L_p norms. We show the following results for any matrix $A = (a_{ij})_{i,j=1}^n$, resolving in particular a main open problem mentioned by Schulman and Sinclair.

1. We analyze the iteration for the L_1 norm under a greedy order of balancing. We show that it converges to an ϵ-balanced matrix in $K = O(\min\{\epsilon^{-2}\log w, \epsilon^{-1}n^{3/2}\log(w/\epsilon)\})$ iterations that cost a total of $O(m + Kn\log n)$ arithmetic operations over $O(n\log(w/\epsilon))$-bit numbers. Here m is the number of non-zero entries of A, and $w = \sum_{i,j}|a_{ij}|/a_{\min}$ with $a_{\min} = \min\{|a_{ij}| : a_{ij} \neq 0\}$.
2. We show that the original round-robin implementation converges to an ϵ-balanced matrix in $O(\epsilon^{-2}n^2\log w)$ iterations totaling $O(\epsilon^{-2}mn\log w)$ arithmetic operations over $O(n\log(w/\epsilon))$-bit numbers.
3. We show that a random implementation of the iteration converges to an ϵ-balanced matrix in $O(\epsilon^{-2}\log w)$ iterations using $O(m + \epsilon^{-2}n\log w)$ arithmetic operations over $O(\log(wn/\epsilon))$-bit numbers.
4. We demonstrate a lower bound of $\Omega(1/\sqrt{\epsilon})$ on the convergence rate of any implementation of the iteration.
5. We observe, through a known trivial reduction, that our results for L_1 balancing apply to any L_p norm for all finite p, at the cost of increasing the number of iterations by only a factor of p.

We note that our techniques are very different from those used by Schulman and Sinclair.

.

[18] Igor Potapov and Pavel Semukhin. Decidability of the Membership Problem for 2×2 integer matrices. **SODA 1C** (Nelva). January 16th, 2017, 10:55-11:15

The main result of this paper is the decidability of the membership problem for 2×2 nonsingular integer matrices. Namely, we will construct the first algorithm that for any nonsingular 2×2 integer matrices $M_1, \ldots, M_n$ and M decides whether M belongs to the semigroup generated by $\{M_1, \ldots, M_n\}$. Our algorithm relies on a translation of numerical problems on matrices into combinatorial problems on words. It also makes use of some algebraic properties of well-known subgroups of $\mathrm{GL}(2, \mathbb{Z})$ and various new techniques and constructions that help to convert matrix equations into the emptiness problem for intersection of regular languages.

.

[19] Paul C. Bell, Mika Hirvensalo and Igor Potapov. The Identity Problem for Matrix Semigroups in $\mathsf{SL}_2(\mathbb{Z})$ is NP-complete. **SODA 1C** (Nelva). January 16th, 2017, 11:20-11:40

In this paper, we show that the problem of determining if the identity matrix belongs to a finitely generated semigroup of 2×2 matrices from the modular group $\mathsf{PSL}_2(\mathbb{Z})$ and thus the Special Linear group $\mathsf{SL}_2(\mathbb{Z})$ is solvable in $\mathbf{NP}$. From this fact, we can immediately derive that the fundamental problem of whether a given finite set of matrices from $\mathsf{SL}_2(\mathbb{Z})$ or $\mathsf{PSL}_2(\mathbb{Z})$ generates a group or free semigroup is also decidable in $\mathbf{NP}$. The previous algorithm for these problems, shown in 2005 by Choffrut and Karhumäki, was in $\mathbf{EXPSPACE}$ mainly due to the translation of matrices into exponentially long words over a binary alphabet $\{s, r\}$ and further constructions with a large nondeterministic finite state automaton that is built on these words. Our algorithm is based on various new techniques that allow us to operate with compressed word representations of matrices without explicit expansions. When combined with the known $\mathbf{NP}$-hard lower bound, this proves that the membership problem for the identity problem, the group problem and the freeness problem in $\mathsf{SL}_2(\mathbb{Z})$ are $\mathbf{NP}$-complete.

.

[20] Lihi Cohen, Yuval Emek, Oren Louidor and Jara Uitto. Exploring an Infinite Space with Finite Memory Scouts. **SODA 1C** (Nelva). January 16th, 2017, 11:45-12:05

Consider a small number of *scouts* exploring the infinite d-dimensional grid with the aim of hitting a hidden target point. Each scout is controlled by a probabilistic finite automaton that determines its movement (to a neighboring grid point) based on its current state. The scouts, that operate under a fully synchronous schedule, communicate with each other (in a way that affects their respective states) when they share the same grid point and operate independently otherwise. Our main research question is: How many scouts are required to guarantee that the target admits a *finite mean hitting time*? Recently, it was shown that $d+1$ is an upper bound on the answer to this question for any dimension $d \geq 1$ and the main contribution of this paper comes in the form of proving that this bound is tight for $d \in \{1, 2\}$.

.

[21] Cameron Chalk, Erik D. Demaine, Martin L. Demaine, Eric Martinez, Robert Schweller, Luis Vega and Tim Wylie. Universal Shape Replicators via Self-Assembly with Attractive and Repulsive Forces. **SODA 1C** (Nelva). January 16th, 2017, 12:10-12:30

We show how to design a universal *shape replicator* in a self-assembly system with both attractive and repulsive forces. More precisely, we show that there is a universal set of constant-size objects that, when added to *any unknown* hole-free polyomino shape, produces an unbounded number of copies of that shape (plus constant-size garbage objects). The constant-size objects can be easily constructed from a constant number of individual tile types using a constant number of preprocessing self-assembly steps. Our construction uses the well-studied 2-Handed Assembly Model (2HAM) of tile self-assembly, in the simple model where glues interact only with identical glues, allowing glue strengths that are either positive (attractive) or negative (repulsive), and constant temperature (required glue strength for parts to hold together). We also require that the given shape has specified glue types on its surface, and that the feature size (smallest distance between nonincident edges) is bounded below by a constant. Shape replication necessarily requires a self-assembly model where parts can both attach and detach, and this construction is the first to do so using the natural model of negative/repulsive glues (also studied before for other problems such as fuel-efficient computation); previous replication constructions require more powerful global operations such as an "enzyme" that destroys a subset of the tile types.

.

[22] Colin McDiarmid. Clique colouring random graphs. **Invited Talk 2** (Las Arenas II-IV). January 16th, 2017, 14:00-15:00

Graph colouring occupies a central place in theory and applications, in combinatorics, computer science and OR. We shall discuss clique colouring a graph G; that is, colouring the vertices so that no maximal clique is monochromatic. The least possible number of colours is the clique colouring number $\chi_c(G)$.
Typically, upper bounds on χ_c for random graphs are algorithmic. A key idea for clique colouring certain sparse random graphs is to construct triangle-free colour sets greedily. We shall discuss random perfect graphs, binomial random graphs and random geometric graphs. This is joint work with Nikola Yolov, and with Dieter Mitsche and Pawel Pralat.

.

[23] Benjamin Hackl, Sara Kropf and Helmut Prodinger. Iterative Cutting and Pruning of Planar Trees. **ANALCO 2** (Saray). January 16th, 2017, 15:00-15:20

Rooted plane trees are reduced by four different operations on the fringe. The number of surviving nodes after reducing the tree repeatedly for a fixed number of times is asymptotically analyzed. The four different operations include cutting all or only the leftmost leaves or maximal paths. This generalizes the concept of pruning a tree. The results include exact expressions and asymptotic expansions for the expected value and the variance as well as central limit theorems.

.

[24] Joubert Oosthuizen and Stephan Wagner. On the distribution of random walk hitting times in random trees. **ANALCO 2** (Saray). January 16th, 2017, 15:25-15:45

The hitting time H_{xy} between two vertices x and y of a graph is the average time that the standard simple random walk takes to get from x to y. In this paper, we study the distribution of the hitting time between two randomly chosen vertices of a random tree. We consider both uniformly random labelled trees and a more general model with vertex weights akin to simply generated trees. We show that the r-th moment of the hitting time is of asymptotic order $n^{3r/2}$ in trees of order n, and we describe the limiting distribution upon normalisation by means of its moments. Moreover, we also obtain joint moments with the distance between the two selected vertices. Finally, we discuss a somewhat different model of randomness, namely random recursive trees. In this setup, the root is of special importance, and so we study the hitting time from the root to a random vertex or from a random vertex to the root. Interestingly, the hitting time from the root is of order $n \log n$, with a normal limit law, while the hitting time to the root is only of linear order and has a non-Gaussian limit law.

.

[25] Bernhard Gittenberger, Emma Yu Jin and Michael Wallner. A note on the scaling limits of random Pólya trees. **ANALCO 2** (Saray). January 16th, 2017, 15:50-16:10

Panagiotou and Stufler (arXiv:1502.07180v2) recently proved one important fact on their way to establish the scaling limits of random Pólya trees: a uniform random Pólya tree of size n consists of a conditioned critical Galton-Watson tree C_n and many small forests, where with probability tending to one as n tends to infinity, any forest $F_n(v)$, that is attached to a node v in C_n, is maximally of size $|F_n(v)| = O(\log n)$. Their proof used the framework of a Boltzmann sampler and deviation inequalities. In this paper, first, we employ a unified framework in analytic combinatorics to prove this fact with additional improvements on the bound of $|F_n(v)|$, namely $|F_n(v)| = \Theta(\log n)$. Second, we give a combinatorial interpretation of the rational weights of these forests and the defining substitution process in terms of automorphisms associated to a given Pólya tree. Finally, we derive the limit probability that for a random node v the attached forest $F_n(v)$ is of a given size.

.

[26] Andrea Kuntschik and Ralph Neininger. Rates of convergence for balanced irreducible two-color Pólya urns. **ANALCO 2** (Saray). January 16th, 2017, 16:15-16:35

For balanced Pólya urns with two colors the (normalized) number of balls of each color satisfies a limit law with two possible regimes: with weak convergence towards the normal distribution and almost sure convergence towards distributions that can be characterized by moments or by recursive distributional equations. We bound the rate of convergence in these limit theorems for such irreducible urn schemes. The bounds are sufficiently tight to confirm a conjecture of S. Janson for a subclass of these urns.

.

[27] Thomas D. Ahle, Martin Aumüller and Rasmus Pagh. Parameter-free Locality Sensitive Hashing for Spherical Range Reporting. **SODA 2A** (Las Arenas II-IV). January 16th, 2017, 15:00-15:20

We present a data structure for *spherical range reporting* on a point set S, i.e., reporting all points in S that lie within radius r of a given query point q (with a small probability of error). Our solution builds upon the Locality-Sensitive Hashing (LSH) framework of Indyk and Motwani, which represents the asymptotically best solutions to near neighbor problems in high dimensions. While traditional LSH data structures have several parameters whose optimal values depend on the distance distribution from q to the points of S (and in particular on the number of points to report), our data structure is essentially parameter-free and only takes as parameter the space the user is willing to allocate. Nevertheless, its expected query time basically matches that of an LSH data structure whose parameters have been *optimally chosen for the data and query* in question under the given space constraints. In particular, our data structure provides a smooth trade-off between hard queries (typically addressed by standard LSH parameter settings) and easy queries such as those where the number of points to report is a constant fraction of S, or where almost all points in S are far away from the query point. In contrast, known data structures fix LSH parameters based on certain parameters of the input alone. The algorithm has expected query time bounded by $O(t(n/t)^\rho)$, where t is the number of points to report and $\rho \in (0, 1)$ depends on the data distribution and the strength of the LSH family used. The previously best running time in high dimensions was $\Omega(tn^\rho)$, achieved by traditional LSH-based data structures where parameters are tuned for outputting a single point within distance r. Further, for many data distributions where the intrinsic dimensionality of the point set close to q is low, we can give improved upper bounds on the expected query time. We finally present a parameter-free way of using multi-probing, for LSH families that support it, and show that for many such families this approach allows us to get expected query time close to $O(n^\rho + t)$, which is the best we can hope to achieve using LSH.

.

[28] Mayank Goswami, Rasmus Pagh, Francesco Silvestri and Johan Sivertsen. Distance Sensitive Bloom Filters Without False Negatives. **SODA 2A** (Las Arenas II-IV). January 16th, 2017, 15:25-15:45

A Bloom filter is a widely used data-structure for representing a set S and answering queries of the form "Is x in S?". By allowing some false positive answers (saying 'yes' when the answer is in fact 'no') Bloom filters use space significantly below what is required for storing S. In the *distance sensitive* setting we work with a set S of (Hamming) vectors and seek a data structure that offers a similar trade-off, but answers queries of the form "Is x *close* to an element of S?" (in Hamming distance). Previous work on distance sensitive Bloom filters have accepted false positive *and* false negative answers. Absence of false negatives is of critical importance in many applications of Bloom filters, so it is natural to ask if this can be also achieved in the distance sensitive setting. Our main contributions are upper and lower bounds (that are tight in several cases) for space usage in the distance sensitive setting where false negatives are not allowed.

.

[29] Sunil Arya, Guilherme D. da Fonseca and David M. Mount. Optimal Approximate Polytope Membership. **SODA 2A** (Las Arenas II-IV). January 16th, 2017, 15:50-16:10

In the polytope membership problem, a convex polytope K in $\mathbb{R}^d$ is given, and the objective is to preprocess K into a data structure so that, given a query point $q \in \mathbb{R}^d$, it is possible to determine efficiently whether $q \in K$. We consider this problem in an approximate setting and assume that d is a constant. Given an approximation parameter $\epsilon > 0$, the query can be answered either way if the distance from q to K's boundary is at most ϵ times K's diameter. Previous solutions to the problem were on the form of a space-time trade-off, where logarithmic query time demands $O(1/\epsilon^{d-1})$ storage, whereas storage $O(1/\epsilon^{(d-1)/2})$ admits roughly $O(1/\epsilon^{(d-1)/8})$ query time. In this paper, we present a data structure that achieves logarithmic query time with storage of only $O(1/\epsilon^{(d-1)/2})$, which matches the worst-case lower bound on the complexity of any ϵ-approximating polytope. Our data structure is based on a new technique, a hierarchy of ellipsoids defined as approximations to Macbeath regions. As an application, we obtain major improvements to approximate Euclidean nearest neighbor searching. Notably, the storage needed to answer ϵ-approximate nearest neighbor queries for a set of n points in $O(\log(n/\epsilon))$ time is reduced to $O(n/\epsilon^{d/2})$. This halves the exponent in the ϵ-dependency of the existing space bound of roughly $O(n/\epsilon^d)$, which has stood for 15 years (Har-Peled, 2001).

.

[30] Paul Beame and Cyrus Rashtchian. Massively-Parallel Similarity Join, Edge-Isoperimetry, and Distance Correlations on the Hypercube. **SODA 2A** (Las Arenas II-IV). January 16th, 2017, 16:15-16:35

We study distributed protocols for finding all pairs of similar vectors in a large dataset. Our results pertain to a variety of discrete metrics, and we give concrete instantiations for Hamming distance. In particular, we give improved upper bounds on the overhead required for similarity defined by Hamming distance $r > 1$ and prove a lower bound showing qualitative optimality of the overhead required for similarity over any Hamming distance r. Our main conceptual contribution is a connection between similarity search algorithms and certain graph-theoretic quantities. For our upper bounds, we exhibit a general method for designing one-round protocols using edge-isoperimetric shapes in similarity graphs. For our lower bounds, we define a new combinatorial optimization problem, which can be stated in purely graph-theoretic terms yet also captures the core of the analysis in previous theoretical work on distributed similarity joins. As one of our main technical results, we prove new bounds on distance correlations in subsets of the Hamming cube.

.

[31] Alexandr Kazda, Vladimir Kolmogorov and Michal Rolínek. Even Delta-Matroids and the Complexity of Planar Boolean CSPs. **SODA 2B** (Las Arenas I). January 16th, 2017, 15:00-15:20

The main result of this paper is a generalization of the classical blossom algorithm for finding perfect matchings. Our algorithm can efficiently solve Boolean CSPs where each variable appears in exactly two constraints (we call it edge CSP) and all constraints are *even Δ-matroid* relations (represented by lists of tuples). As a consequence of this, we settle the complexity classification of planar Boolean CSPs started by Dvořák and Kupec. Knowing that edge CSP is tractable for even Δ-matroid constraints allows us to extend the tractability result to a larger class of Δ-matroids that includes many classes that were known to be tractable before, namely *co-independent*, *compact*, *local* and *binary*.

.

[32] Christoph Berkholz and Martin Grohe. Linear Diophantine Equations, Group CSPs, and Graph Isomorphism. **SODA 2B** (Las Arenas I). January 16th, 2017, 15:25-15:45

In recent years, we have seen several approaches to the graph isomorphism problem based on "generic" mathematical programming or algebraic (Gröbner basis) techniques. For most of these, lower bounds have been established. In fact, it has been shown that the pairs of non-isomorphic CFI-graphs (introduced by Cai, Fürer, and Immerman in 1992 as hard examples for the combinatorial Weisfeiler-Leman algorithm) cannot be distinguished by these mathematical algorithms. A notable exception were the algebraic algorithms over the field $\mathbb{F}_2$, for which no lower bound was known. Another, in some way even stronger, approach to graph isomorphism testing is based on solving systems of linear Diophantine equations (that is, linear equations over the integers), which is known to be possible in polynomial time. So far, no lower bounds for this approach were known. Lower bounds for the algebraic algorithms can best be proved in the framework of proof complexity, where they can be phrased as lower bounds for algebraic proof systems such as Nullstellensatz or the (more powerful) polynomial calculus. We give new hard examples for these systems: families of pairs of non-isomorphic graphs that are hard to distinguish by polynomial calculus proofs simultaneously over all prime fields, including $\mathbb{F}_2$, as well as examples that are hard to distinguish by the systems-of-linear-Diophantine-equations approach. In a previous paper, we observed that the CFI-graphs are closely related to what we call "group CSPs": constraint satisfaction problems where the constraints are membership tests in some coset of a subgroup of a cartesian power of a base group ($\mathbb{Z}_2$ in the case of the classical CFI-graphs). Our new examples are also based on group CSPs (for Abelian groups), but here we extend the CSPs by a few non-group constraints to obtain even harder instances for graph isomorphism.

.

[33] Víctor Dalmau, Marcin Kozik, Adrei Krohkin, Konstantin Makarychev, Yury Makarychev and Jakub Opršal. Robust algorithms with polynomial loss for near-unanimity CSPs. **SODA 2B** (Las Arenas I). January 16th, 2017, 15:50-16:10

An instance of the Constraint Satisfaction Problem (CSP) is given by a family of constraints on overlapping sets of variables, and the goal is to assign values from a fixed domain to the variables so that all constraints are satisfied. In the optimization version, the goal is to maximize the number of satisfied constraints. An approximation algorithm for CSP is called robust if it outputs an assignment satisfying a $(1 - g(\varepsilon))$-fraction of constraints on any $(1 - \varepsilon)$-satisfiable instance, where the loss function g is such that $g(\varepsilon) \to 0$ as $\varepsilon \to 0$. We study how the robust approximability of CSPs depends on the set of constraint relations allowed in instances, the so-called constraint language. All constraint languages admitting a robust polynomial-time algorithm (with some g) have been characterised by Barto and Kozik, with the general bound on the loss g being doubly exponential, specifically $g(\varepsilon) = O((\log\log(1/\varepsilon))/\log(1/\varepsilon))$. It is natural to ask when a better loss can be achieved: in particular, polynomial loss $g(\varepsilon) = O(\varepsilon^{1/k})$ for some constant k. In this paper, we consider CSPs with a constraint language having a near-unanimity polymorphism. We give two randomized robust algorithms with polynomial loss for such CSPs: one works for any near-unanimity polymorphism and the parameter k in the loss depends on the size of the domain and the arity of the relations in Γ, while the other works for a special ternary near-unanimity operation called dual discriminator with $k = 2$ for any domain size. In the latter case, the CSP is a common generalisation of UNIQUE GAMES with a fixed domain and 2-SAT. In the former case, we use the algebraic approach to the CSP. Both cases use the standard semidefinite programming relaxation for CSP.

.

[34] Xue Chen and Yuan Zhou. Parameterized Algorithms for Constraint Satisfaction Problems Above Average with Global Cardinality Constraints. **SODA 2B** (Las Arenas I). January 16th, 2017, 16:15-16:35

Given a constraint satisfaction problem (CSP) on n variables, $x_1, x_2, \ldots, x_n \in \{\pm 1\}$, and m constraints, a global cardinality constraint has the form of $\sum_{i=1}^{n} x_i = (1-2p)n$, where $p \in (\Omega(1), 1-\Omega(1))$ and pn is an integer. Let AVG be the expected number of constraints satisfied by randomly choosing an assignment to $x_1, x_2, \ldots, x_n$, complying with the global cardinality constraint. The CSP above average with the global cardinality constraint problem asks whether there is an assignment (complying with the cardinality constraint) that satisfies more than $(AVG + t)$ constraints, where t is an input parameter. In this paper, we present an algorithm that finds a valid assignment satisfying more than $(AVG + t)$ constraints (if there exists one) in time $(2^{O(t^2)} + n^{O(d)})$. Therefore, the CSP above average with the global cardinality constraint problem is fixed-parameter tractable.

.

[35] Vít Jelínek and Jan Kynčl. Hardness of Permutation Pattern Matching. **SODA 2C** (Nelva). January 16th, 2017, 15:00-15:20

Permutation Pattern Matching (or PPM) is a decision problem whose input is a pair of permutations π and τ, represented as sequences of integers, and the task is to determine whether τ contains a subsequence order-isomorphic to π. Bose, Buss and Lubiw proved that PPM is NP-complete on general inputs. We show that PPM is NP-complete even when π has no decreasing subsequence of length 3 and τ has no decreasing subsequence of length 4. This provides the first known example of PPM being hard when one or both of π and τ are restricted to a proper hereditary class of permutations. This hardness result is tight in the sense that PPM is known to be polynomial when both π and τ avoid a decreasing subsequence of length 3, as well as when π avoids a decreasing subsequence of length 2. The result is also tight in another sense: we will show that for any hereditary proper subclass $\mathcal{C}$ of the class of permutations avoiding a decreasing sequence of length 3, there is a polynomial algorithm solving PPM instances where π is from $\mathcal{C}$ and τ is arbitrary. We also obtain analogous hardness and tractability results for the class of so-called skew-merged patterns. From these results, we deduce a complexity dichotomy for the PPM problem restricted to π belonging to $\mathrm{Av}(\alpha)$, where $\mathrm{Av}(\alpha)$ denotes the class of permutations avoiding a permutation α. Specifically, we show that the problem is polynomial when α is in the set $\{1, 12, 21, 132, 213, 231, 312\}$, and it is NP-complete for any other α.

.

[36] Arnab Ganguly, Rahul Shah and Sharma V. Thankachan. pBWT: Achieving Succinct Data Structures for Parameterized Pattern Matching and Related Problems. **SODA 2C** (Nelva). January 16th, 2017, 15:25-15:45

The fields of succinct data structures and compressed text indexing have seen quite a bit of progress over the last two decades. An important achievement, primarily using techniques based on the Burrows-Wheeler Transform (BWT), was obtaining the full functionality of the suffix tree in the optimal number of bits. A crucial property that allows the use of BWT for designing compressed indexes is *order-preserving suffix links*. Specifically, the relative order between two suffixes in the subtree of an internal node is same as that of the suffixes obtained by truncating the first character of the two suffixes. Unfortunately, in many variants of the text-indexing problem, for e.g., parameterized pattern matching, 2D pattern matching, and order-isomorphic pattern matching, this property does not hold. Consequently, the compressed indexes based on BWT do not directly apply. Furthermore, a compressed index for any of these variants has been elusive throughout the advancement of the field of succinct data structures. We achieve a positive breakthrough on one such problem, namely the *Parameterized Pattern Matching* problem. Let T be a text that contains n characters from an alphabet Σ, which is the union of two disjoint sets: Σ_s containing static characters (s-characters) and Σ_p containing parameterized characters (p-characters). A pattern P (also over Σ) matches an equal-length substring S of T iff the s-characters match exactly, and there exists a one-to-one function that renames the p-characters in S to that in P. The task is to find the starting positions (occurrences) of all such substrings S. Previous index [Baker, STOC 1993], known as *Parameterized Suffix Tree*, requires $\Theta(n \log n)$ bits of space, and can find all occ occurrences in time $O(|P| \log \sigma + occ)$, where $\sigma = |\Sigma|$. We introduce an $n \log \sigma + O(n)$-bit index with $O(|P| \log \sigma + occ \cdot \log n \log \sigma)$ query time. At the core, lies a new BWT-like transform, which we call the *Parameterized Burrows-Wheeler Transform* (pBWT). The techniques are extended to obtain a succinct index for the *Parameterized Dictionary Matching* problem of Idury and Schäffer [CPM, 1994].

.

[37] Ian Munro, Gonzalo Navarro and Yakov Nekrich. Space-Efficient Construction of Compressed Indexes in Deterministic Linear Time. **SODA 2C** (Nelva). January 16th, 2017, 15:50-16:10

We show that the compressed suffix array and the compressed suffix tree of a string T can be built in $O(n)$ deterministic time using $O(n \log \sigma)$ bits of space, where n is the string length and σ is the alphabet size. Previously described deterministic algorithms either run in time that depends on the alphabet size or need $\omega(n \log \sigma)$ bits of working space. Our result has immediate applications to other problems, such as yielding the first deterministic linear-time LZ77 and LZ78 parsing algorithms that use $O(n \log \sigma)$ bits.

.

[38] Paweł Gawrychowski and Tomasz Kociumaka. Sparse Suffix Tree Construction in Optimal Time and Space. **SODA 2C** (Nelva). January 16th, 2017, 16:15-16:35

Suffix tree (and the closely related suffix array) are fundamental structures capturing all substrings of a given text essentially by storing all its suffixes in the lexicographical order. In some applications, such as sparse text indexing, we work with a subset of b interesting suffixes, which are stored in the so-called sparse suffix tree. Because the size of this structure is $\Theta(b)$, it is natural to seek a construction algorithm using only $O(b)$ words of space assuming read-only random access to the text. We design a linear-time Monte Carlo algorithm for this problem, hence resolving an open question explicitly stated by Bille et al. [TALG 2016]. The best previously known algorithm by I et al. [STACS 2014] works in $O(n \log b)$ time. As opposed to previous solutions, which were based on the divide-and-conquer paradigm, our solution proceeds in n/b rounds. In the r-th round, we consider all suffixes starting at positions congruent to r modulo n/b. By maintaining rolling hashes, we can lexicographically sort all interesting suffixes starting at such positions, and then we can merge them with the already considered suffixes. For efficient merging, we also need to answer LCE queries efficiently (and in small space). By plugging in the structure of Bille et al. [CPM 2015] we obtain $O(n + b \log b)$ time complexity. We improve this structure by a recursive application of the so-called difference covers, which then implies a linear-time sparse suffix tree construction algorithm. We complement our Monte Carlo algorithm with a deterministic verification procedure. The verification takes $O(n\sqrt{\log b})$ time, which improves upon the bound of $O(n \log b)$ obtained by I et al. [STACS 2014]. This is obtained by first observing that the pruning done inside the previous solution has a rather clean description using the notion of graph spanners with small multiplicative stretch. Then, we are able to decrease the verification time by applying difference covers twice. Combined with the Monte Carlo algorithm, this gives us an $O(n\sqrt{\log b})$-time and $O(b)$-space Las Vegas algorithm.

.

[39] Pablo Rotondo and Brigitte Vallée. The recurrence function of a random Sturmian word. **ANALCO 3** (Saray). January 16th, 2017, 17:00-17:20

This paper describes the probabilistic behaviour of a random Sturmian word. It performs the probabilistic analysis of the recurrence function which can be viewed as a waiting time to discover all the factors of length n of the Sturmian word. This parameter is central to combinatorics of words. Having fixed a possible length n for the factors, we let α to be drawn uniformly from the unit interval $[0, 1]$, hence defining a random Sturmian word of slope α. Thus the waiting time for these factors becomes a random variable, for which we study the limit distribution and the limit density.

.

[40] Daniel Krenn. An Extended Note on the Comparison-optimal Dual Pivot Quickselect. **ANALCO 3** (Saray). January 16th, 2017, 17:25-17:45

In this note the precise minimum number of key comparisons any dual-pivot quickselect algorithm (without sampling) needs on average is determined. The result is in the form of exact as well as asymptotic formulæ of this number of a comparison-optimal algorithm. It turns out that the main terms of these asymptotic expansions coincide with the main terms of the corresponding analysis of the classical quickselect, but still—as this was shown for Yaroslavskiy quickselect—more comparisons are needed in the dual-pivot variant. The results are obtained by solving a second order differential equation for the generating function obtained from a recursive approach.

.

[41] Michael Drmota, Michael Fuchs, Hsien-Kuei Hwang and Ralph Neininger. External Profile of Symmetric Digital Search Trees. **ANALCO 3** (Saray). January 16th, 2017, 17:50-18:10

The external profile is among the first examined shape parameters of digital search trees in connection with the performance of unsuccessful search of a random query in the early 1970s. However, finer and important properties beyond the mean such as the variance and the limit law have remained unknown. In this extended abstract, we describe the first results for the asymptotic variance and the limit law of the external profile. In particular, the analysis of the variance turns out to be highly demanding and nontrivial, and we need diverse techniques from analytic combinatorics to unveil its asymptotic behaviors.

.

[42] Amalia Duch and Gustavo Lau. Partial Match Queries in Relaxed K-dt trees. **ANALCO 3** (Saray). January 16th, 2017, 18:15-18:35

The study of partial match queries on random hierarchical multidimensional data structures dates back to Ph. Flajolet and C. Puech's 1986 seminal paper on partial match retrieval. It was not until recently that fixed (as opposed to random) partial match queries were studied for random relaxed K-d trees, random standard K-d trees, and random 2-dimensional quad trees. Based on those results it seemed natural to classify the general form of the cost of fixed partial match queries into two families: that of either random hierarchical structures or perfectly balanced structures, as conjectured by Duch, Lau and Martínez (On the Cost of Fixed Partial Queries in K-d trees *Algorithmica*, 75(4):684–723, 2016). Here we show that the conjecture just mentioned does not hold by introducing relaxed K-dt trees and providing the average-case analysis for random partial match queries as well as some advances on the average-case analysis for fixed partial match queries on them. In fact this cost –for fixed partial match queries– does not follow the conjectured forms.

.

[43] Ittai Abraham, Shiri Chechik and Sebastian Krinninger. Fully dynamic all-pairs shortest paths with worst-case update-time revisited. **SODA 3A** (Las Arenas II-IV). January 16th, 2017, 17:00-17:20

We revisit the classic problem of dynamically maintaining shortest paths between all pairs of nodes of a directed weighted graph. The allowed updates are insertions and deletions of nodes and their incident edges. We give worst-case guarantees on the time needed to process a single update (in contrast to related results, the update time is *not* amortized over a sequence of updates). Our main result is a simple randomized algorithm that for any parameter $c > 1$ has a worst-case update time of $O(cn^{2+2/3}\log^{4/3} n)$ and answers distance queries correctly with probability $1 - 1/n^c$, against an adaptive online adversary if the graph contains no negative cycle. The best deterministic algorithm is by Thorup [STOC 2005] with a worst-case update time of $\tilde{O}(n^{2+3/4})$ and assumes non-negative weights. This is the first improvement for this problem for more than a decade. Conceptually, our algorithm shows that randomization along with a more direct approach can provide better bounds.

.

[44] Aaron Bernstein and Shiri Chechik. Deterministic Partially Dynamic Single Source Shortest Paths for Sparse Graphs. **SODA 3A** (Las Arenas II-IV). January 16th, 2017, 17:25-17:45

In this paper we consider the decremental single-source shortest paths (SSSP) problem, where given a graph G and a source node s the goal is to maintain shortest paths between s and all other nodes in G under a sequence of online adversarial edge deletions. (Our algorithm can also be modified to work in the incremental setting, where the graph is initially empty and subject to a sequence of online adversarial edge insertions.) In their seminal work, Even and Shiloach [JACM 1981] presented an exact solution to the problem with only $O(mn)$ total update time over all edge deletions. Later papers presented conditional lower bounds showing that $O(mn)$ is optimal up to log factors. In SODA 2011, Bernstein and Roditty showed how to bypass these lower bounds and improve upon the Even and Shiloach $O(mn)$ total update time bound by allowing a $(1+\epsilon)$ approximation. This triggered a series of new results, culminating in a recent breakthrough of Henzinger, Krinninger and Nanongkai [FOCS 14], who presented a $(1+\epsilon)$-approximate algorithm whose total update time is near linear: $O(m^{1+o(1)})$. However, every single one of these improvements over the Even-Shiloach algorithm was randomized and assumed a non-adaptive adversary. This additional assumption meant that the algorithms were not suitable for certain settings and could not be used as a black box data structure. Very recently Bernstein and Chechik presented in STOC 2016 the first *deterministic* improvement over Even and Shiloach, that did not rely on randomization or assumptions about the adversary: in an undirected unweighted graph the algorithm maintains $(1+\epsilon)$-approximate distances and has total update time $\tilde{O}(n^2)$. In this paper, we present a new deterministic algorithm for the problem with total update time $\tilde{O}(n^{1.25}\sqrt{m}) = \tilde{O}(mn^{3/4})$: it returns a $(1+\epsilon)$ approximation, and is limited to undirected unweighted graphs. Although this result is still far from matching the randomized near-linear total update time, it presents important progress towards that direction, because unlike the STOC 2016 $\tilde{O}(n^2)$ algorithm it beats the Even and Shiloach $O(mn)$ bound for all graphs, not just sufficiently dense ones. In particular, the $\tilde{O}(n^2)$ algorithm relied entirely on a new sparsification technique, and so could not hope to yield an improvement for sparse graphs. We present the first deterministic improvement for sparse graphs by significantly extending some of the ideas from the $\tilde{O}(n^2)$ algorithm and combining them with the hop-set technique used in several earlier dynamic shortest path papers.

.

[45] Sayan Bhattacharya, Monika Henzinger and Danupon Nanongkai. Fully Dynamic Approximate Maximum Matching and Minimum Vertex Cover in $O(\log^3 n)$ Worst Case Update Time. **SODA 3A** (Las Arenas II-IV). January 16th, 2017, 17:50-18:10

We consider the problem of maintaining an approximately maximum (fractional) matching and an approximately minimum vertex cover in a dynamic graph. Starting with the seminal paper by Onak and Rubinfeld [STOC 2010], this problem has received significant attention in recent years. There remains, however, a polynomial gap between the best known worst case update time and the best known amortised update time for this problem, even after allowing for randomisation. Specifically, Bernstein and Stein [ICALP 2015, SODA 2016] have the best known worst case update time. They present a deterministic data structure with approximation ratio $(3/2+\epsilon)$ and worst case update time $O(m^{1/4}/\epsilon^2)$, where m is the number of edges in the graph. In recent past, Gupta and Peng [FOCS 2013] gave a deterministic data structure with approximation ratio $(1+\epsilon)$ and worst case update time $O(\sqrt{m}/\epsilon^2)$. No known randomised data structure beats the worst case update times of these two results. In contrast, the paper by Onak and Rubinfeld [STOC 2010] gave a randomised data structure with approximation ratio $O(1)$ and amortised update time $O(\log^2 n)$, where n is the number of nodes in the graph. This was later improved by Baswana, Gupta and Sen [FOCS 2011] and Solomon [FOCS 2016], leading to a randomised date structure with approximation ratio 2 and amortised update time $O(1)$. We bridge the polynomial gap between the worst case and amortised update times for this problem, without using any randomisation. We present a deterministic data structure with approximation ratio $(2+\epsilon)$ and worst case update time $O(\log^3 n)$, for all sufficiently small constants ϵ.

.

[46] Ran Duan and Seth Pettie. Connectivity Oracles for Graphs Subject to Vertex Failures. **SODA 3A** (Las Arenas II-IV). January 16th, 2017, 18:15-18:35

We introduce new data structures for answering connectivity queries in graphs subject to batched *vertex failures*. Our deterministic structure processes a batch of $d \leq d_*$ failed vertices in $\tilde{O}(d^3)$ time and thereafter answers connectivity queries in $O(d)$ time. It occupies space $O(d_* m \log n)$. We develop a randomized Monte Carlo version of our data structure with update time $\tilde{O}(d^2)$, query time $O(d)$, and space $\tilde{O}(m)$ for any d_*. This is the first connectivity oracle for general graphs that can efficiently deal with an unbounded number of vertex failures. Our data structures are based on a new decomposition theorem for an undirected graph $G=(V,E)$, which is of independent interest. It states that for any terminal set $U \subseteq V$ we can remove a set B of $|U|/(s-2)$ vertices such that the remaining graph contains a Steiner forest for $U-B$ with maximum degree s.

.

[47] Shang-En Huang, Dawei Huang, Tsvi Kopelowitz and Seth Pettie. Fully Dynamic Connectivity in $O(\log n(\log\log n)^2)$ Amortized Expected Time. **SODA 3A** (Las Arenas II-IV). January 16th, 2017, 18:40-19:00

Dynamic connectivity is one of the most fundamental problems in dynamic graph algorithms. We present a new randomized dynamic connectivity structure with $O(\log n(\log\log n)^2)$ amortized expected update time and $O(\log n/\log\log\log n)$ query time, which comes within an $O((\log\log n)^2)$ factor of a lower bound due to Pătrașcu and Demaine. The new structure is based on a dynamic connectivity algorithm proposed by Thorup in an extended abstract at STOC 2000, which left out some important details.

.

[48] Paul Dütting and Thomas Kesselheim. Best-Response Dynamics in Combinatorial Auctions with Item Bidding. **SODA 3B** (Las Arenas I). January 16th, 2017, 17:00-17:20

In a combinatorial auction with item bidding, agents participate in multiple single-item second-price auctions at once. As some items might be substitutes, agents need to strategize in order to maximize their utilities. A number of results indicate that high welfare can be achieved this way, giving bounds on the welfare at equilibrium. Recently, however, criticism has been raised that equilibria are hard to compute and therefore unlikely to be attained. In this paper, we take a different perspective. We study simple best-response dynamics. That is, agents are activated one after the other and each activated agent updates his strategy myopically to a best response against the other agents' current strategies. Often these dynamics may take exponentially long before they converge or they may not converge at all. However, as we show, convergence is not even necessary for good welfare guarantees. Given that agents' bid updates are aggressive enough but not too aggressive, the game will remain in states of good welfare after each agent has updated his bid at least once. In more detail, we show that if agents have fractionally subadditive valuations, natural dynamics reach and remain in a state that provides a $1/3$ approximation to the optimal welfare after each agent has updated his bid at least once. For subadditive valuations, we can guarantee an $\Omega(1/\log m)$ approximation in case of m items that applies after each agent has updated his bid at least once and at any point after that. The latter bound is complemented by a negative result, showing that no kind of best-response dynamics can guarantee more than a an $o(\log\log m/\log m)$ fraction of the optimal social welfare.

.

[49] Krishnamurthy Dvijotham, Yuval Rabani and Leonard Schulman. Convergence of Incentive-Driven Dynamics in Fisher Markets. **SODA 3B** (Las Arenas I). January 16th, 2017, 17:25-17:45

In both general equilibrium theory and game theory, the dominant mathematical models rest on a fully rational solution concept in which every player's action is a best-response to the actions of the other players. In both theories there is less agreement on suitable out-of-equilibrium modeling, but one attractive approach is the level k model in which a level 0 player adopts a very simple response to current conditions, a level 1 player best-responds to a model in which others take level 0 actions, and so forth. (This is analogous to k-ply exploration of game trees in AI, and to receding-horizon control in control theory.) If players have deterministic mental models with this kind of nite-level response,there is obviously no way their mental models can all be consistent. Nevertheless, there is experimental evidence that people act this way in many situations, motivating the question of what the dynamics of such interactions lead to. We address the problem of out-of-equilibrium price dynamics in the setting of Fisher markets. We develop a general framework in which sellers have (a) a set of atomic price update rules which are simple responses to a price vector; (b) a belief-formation procedure that simulates actions of other sellers (themselves using the atomic price updates) to some nite horizon in the future. In this framework, sellers use an atomic price up-date rule to respond to a price vector they generate with the belief formation procedure. The framework is general and allows sellers to have inconsistent and time-varying beliefs about each other. Under certain assumptions on the atomic update rules, we show that despite the inconsistent and time-varying nature of beliefs, the market converges to a unique equilibrium. (If the price updates are driven by weak-gross substitutes demands,this is the same equilibrium point predicted by those demands.) This result holds for both synchronous and asynchronous discrete-time updates. Moreover, the result is computationally feasible in the sense that the convergence rate is linear, i.e., the distance to equilibrium decays exponentially fast. To the best of our knowledge,this is the first result that demonstrates, in Fisher markets, convergence at any rate for dynamics driven by a plausible model of seller incentives. We then specialize our results to Fisher markets with elastic demands (a further special case corresponds to demand generated by buyers with constant elasticity of substitution (CES) utilities, in the weak gross substitutes (WGS) regime) and show that the atomic update rule in which a seller uses the best-response (=prot-maximizing) update given the prices of all other sellers, satifises the assumptions required on atomic price update rules in our framework. We can even characterizethe convergence rate (as a function of elasticity parameters of the demand function). Our results apply also to settings where, to the best of our knowledge, there exists no previous demonstration of efficient convergence of any discrete dynamic of price updates. Even for the simple case of (level 0) best-response dynamics, our result is the first to demonstratea linear rate of convergence.

.

[50] Alberto Del Pia, Michael Ferris and Carla Michini. Totally Unimodular Congestion Games. **SODA 3B** (Las Arenas I). January 16th, 2017, 17:50-18:10

We investigate a new class of congestion games, called *Totally Unimodular (TU) Congestion Games*, where the players' strategies are binary vectors inside polyhedra defined by totally unimodular constraint matrices. Network congestion games belong to this class. In the symmetric case, when all players have the same strategy set, we design an algorithm that finds an optimal aggregated strategy and then decomposes it into the single players' strategies. This approach yields strongly polynomial-time algorithms to (i) find a pure Nash equilibrium, and (ii) compute a socially optimal state, if the delay functions are weakly convex. We also show how this technique can be extended to matroid congestion games. We then introduce some combinatorial TU congestion games, where the players' strategies are matchings, vertex covers, edge covers, and stable sets of a given bipartite graph. In the asymmetric case, we show that for these games (i) it is PLS-complete to find a pure Nash equilibrium even in case of linear delay functions, and (ii) it is NP-hard to compute a socially optimal state, even in case of weakly convex delay functions.

.

[51] Yu Cheng, Ilias Diakonikolas and Alistair Stewart. Playing Anonymous Games using Simple Strategies. **SODA 3B** (Las Arenas I). January 16th, 2017, 18:15-18:35

We investigate the complexity of computing approximate Nash equilibria in anonymous games. Our main algorithmic result is the following: For any n-player anonymous game with a bounded number of strategies and any constant $\delta > 0$, an $O(1/n^{1-\delta})$-approximate Nash equilibrium can be computed in polynomial time. Complementing this positive result, we show that if there exists any constant $\delta > 0$ such that an $O(1/n^{1+\delta})$-approximate equilibrium can be computed in polynomial time, then there is a fully polynomial-time approximation scheme (FPTAS) for this problem. We also present a faster algorithm that, for any n-player k-strategy anonymous game, runs in time $\tilde{O}((n+k)kn^k)$ and computes an $\tilde{O}(n^{-1/3}k^{11/3})$-approximate equilibrium. This algorithm follows from the existence of simple approximate equilibria of anonymous games, where each player plays one strategy with probability $1-\delta$, for some small δ, and plays uniformly at random with probability δ. Our approach exploits the connection between Nash equilibria in anonymous games and Poisson multinomial distributions (PMDs). Specifically, we prove a new probabilistic lemma establishing the following: Two PMDs, with large variance in each direction, whose first few moments are approximately matching are close in total variation distance. Our structural result strengthens previous work by providing a smooth tradeoff between the variance bound and the number of matching moments.

.

[52] Renato Paes Leme and Sam Chiu-wai Wong. Computing Walrasian Equilibria: Fast Algorithms and Structural Properties. **SODA 3B** (Las Arenas I). January 16th, 2017, 18:40-19:00

We present the first polynomial time algorithm for computing Walrasian equilibrium in an economy with indivisible goods and *general* buyer valuations having only access to an *aggregate demand oracle*, i.e., an oracle that given prices on all goods, returns the aggregated demand over the entire population of buyers. For the important special case of gross substitute valuations, our algorithm queries the aggregate demand oracle $\widetilde{O}(n)$ times and takes $\widetilde{O}(n^3)$ time, where n is the number of goods. At the heart of our solution is a method for exactly minimizing certain convex functions which cannot be evaluated but for which the subgradients can be computed. We also give the fastest known algorithm for computing Walrasian equilibrium for gross substitute valuations in the *value oracle model*. Our algorithm has running time $\widetilde{O}((mn + n^3)T_V)$ where T_V is the cost of querying the value oracle. A key technical ingredient is to regularize a convex programming formulation of the problem in a way that subgradients are cheap to compute. En route, we give necessary and sufficient conditions for the existence of *robust Walrasian prices*, i.e., prices for which each agent has a unique demanded bundle and the demanded bundles clear the market. When such prices exist, the market can be perfectly coordinated by solely using prices.

.

[53] Anastasios Sidiropoulos, Dingkang Wang and Yusu Wang. Metric embeddings with outliers. **SODA 3C** (Nelva). January 16th, 2017, 17:00-17:20

We initiate the study of metric embeddings with *outliers*. Given some finite metric space we wish to remove a small set of points and to find either an isometric or a low-distortion embedding of the remaining points into some host metric space. This is a natural problem that captures scenarios where a small fraction of points in the input corresponds to noise. We present polynomial-time approximation algorithms for computing outlier embeddings into Euclidean space, trees, and ultrametrics. In the case of isometric embeddings the objective is to minimize the number of outliers, while in the case of non-isometries we have a bi-criteria optimization problem where the goal is to minimize both the number of outliers and the distortion. We complement our approximation algorithms with NP-hardness results for these problems. We conclude with a brief experimental evaluation of our non-isometric outlier embedding on synthetic and real-world data sets.

.

[54] Assaf Naor. Probabilistic clustering of high dimensional norms. **SODA 3C** (Nelva). January 16th, 2017, 17:25-17:45

Separating decompositions of metric spaces are an important randomized clustering paradigm that was formulated by Bartal in (*Probabilistic approximation of metric spaces and its algorithmic applications*, FOCS '96 184–193, 1996) and is defined as follows. Given a metric space (X, d_X), its *modulus of separated decomposability*, denoted $\mathsf{SEP}(X, d_X)$, is the infimum over those $\sigma \in (0, \infty]$ such that for every finite subset $S \subseteq X$ and every $\Delta > 0$ there exists a distribution over random partitions $\mathcal{P}$ of S into sets of diameter at most Δ such that for every $x, y \in S$ the probability that both x and y do not fall into the same cluster of the random partition $\mathcal{P}$ is at most $\sigma d_X(x, y)/\Delta$. Here we obtain new bounds on $\mathsf{SEP}(X, \|\cdot\|_X)$ when $(X, \|\cdot\|_X)$ is a finite dimensional normed space, yielding, as a special case, that $\sqrt{n} \lesssim \mathsf{SEP}(\ell_\infty^n) \lesssim \sqrt{n \log n}$ for every $n \in \mathbb{N}$. More generally, $\sqrt{n} \lesssim \mathsf{SEP}(\ell_p^n) \lesssim \sqrt{n \min\{p, \log n\}}$ for every $p \in [2, \infty]$. This improves over the work of Charikar, Chekuri, Goel, Guha, and Plotkin (*Approximating a finite metric by a small number of tree metrics*, FOCS '98, 379–388, 1998), who obtained this bound when $p = 2$, yet for $p \in (2, \infty]$ they obtained the asymptotically weaker estimate $\mathsf{SEP}(\ell_p^n) \lesssim n^{1-1/p}$. One should note that it was claimed in (M. Charikar *et al.*, FOCS '98) that the bound $\mathsf{SEP}(\ell_p^n) \lesssim n^{1-1/p}$ is sharp for every $p \in [2, \infty]$, and in particular it was claimed in (M. Charikar *et al.*, FOCS '98) that $\mathsf{SEP}(\ell_\infty^n) \asymp n$. However, the above results show that this claim of (M. Charikar *et al.*, FOCS '98) is incorrect for every $p \in (2, \infty]$. Our new bounds on the modulus of separated decomposability rely on extremal results for orthogonal hyperplane projections of convex bodies, specifically using the work of Barthe and the author in (F. Barthe and A. Naor, *Hyperplane projections of the unit ball of* ℓ_p^n, Discrete Comput. Geom., 27(2):215–226, 2002). This yields additional refined estimates, an example of which is that for every $n \in \mathbb{N}$ and $k \in \{1, \ldots, n\}$ we have $\mathsf{SEP}\big((\ell_2^n)_{\leqslant k}\big) \lesssim \sqrt{k \log(en/k)}$, where $(\ell_2^n)_{\leqslant k}$ denotes the subset of $\mathbb{R}^n$ consisting of all those vectors that have at most k nonzero entries, equipped with the Euclidean metric. The above statements have implications to the Lipschitz extension problem through its connection to random partitions that was developed by Lee and the author in (J. R. Lee and A. Naor, *Extending Lipschitz functions via random metric partitions*, Invent. Math., 160(1):59–95, 2005). Given a metric space (X, d_X), let $\mathsf{e}(X)$ denote the infimum over those $K \in (0, \infty]$ such that for every Banach space Y and every subset $S \subset X$, every 1-Lipschitz function $f : S \to Y$ has a K-Lipschitz extension to all of X. Johnson, Lindenstrauss and Schechtman proved in (*Extensions of Lipschitz maps into Banach spaces*, Israel J. Math., 54(2):129–138, 1986) that $\mathsf{e}(X) \lesssim \dim(X)$ for every finite dimensional normed space $(X, \|\cdot\|_X)$. It is a longstanding open problem to determine the correct asymptotic dependence on $\dim(X)$ in this context, with the best known lower bound, due to Johnson and Lindenstrauss (*Extensions of Lipschitz mappings into a Hilbert space*, in Conference in Modern Analysis and Probability (New Haven, Conn., 1982), volume 26 of Contemp. Math., pages 189–206. Amer. Math. Soc., 1984), being that the quantity $\mathsf{e}(X)$ must sometimes be at least a constant multiple of $\sqrt{\dim(X)}$. In particular, the previously best known upper bound on $\mathsf{e}(\ell_\infty^n)$ was the $O(n)$ estimate of (W. B. Johnson *et al.*, 1986). It is shown here that for every $n \in \mathbb{N}$ we have $\sqrt{n} \lesssim \mathsf{e}(\ell_\infty^n) \lesssim \sqrt{n \log n}$, thus answering (up to logarithmic factors) a question that was posed by Brudnyi and Brudnyi as Problem 2 in (A. Brudnyĭ and Y. Brudnyĭ, *Simultaneous extensions of Lipschitz functions*, Uspekhi Mat. Nauk, 60(6(366)):53–72, 2005). More generally, $\mathsf{e}(\ell_p^n) \lesssim \sqrt{n \min\{p, \log n\}}$ for every $p \in [2, \infty]$, thus resolving (negatively) a conjecture that Brudnyi and Brudnyi posed as Conjecture 5 in (A. Brudnyĭ and Y. Brudnyĭ, 2005).

.

[55] Piotr Indyk and Tal Wagner. Near-Optimal (Euclidean) Metric Compression. **SODA 3C** (Nelva). January 16th, 2017, 17:50-18:10

The metric sketching problem is defined as follows. Given a metric on n points, and $\epsilon > 0$, we wish to produce a small size data structure (sketch) that, given any pair of point indices, recovers the distance between the points up to a $1+\epsilon$ distortion. In this paper we consider metrics induced by ℓ_2 and ℓ_1 norms whose spread (the ratio of the diameter to the closest pair distance) is bounded by $\Phi > 0$. A well-known dimensionality reduction theorem due to Johnson and Lindenstrauss yields a sketch of size $O(\epsilon^{-2}\log(\Phi n)n\log n)$, i.e., $O(\epsilon^{-2}\log(\Phi n)\log n)$ bits per point. We show that this bound is not optimal, and can be substantially improved to $O(\epsilon^{-2}\log(1/\epsilon)\cdot\log n + \log\log\Phi)$ bits per point. Furthermore, we show that our bound is tight up to a factor of $\log(1/\epsilon)$. We also consider sketching of general metrics and provide a sketch of size $O(n\log(1/\epsilon) + \log\log\Phi)$ bits per point, which we show is optimal.

.

[56] Amir Nayyeri and Benjamin Raichel. A Treehouse with Custom Windows: Minimum Distortion Embeddings into Bounded Treewidth Graphs. **SODA 3C** (Nelva). January 16th, 2017, 18:15-18:35

We describe a $(1+\varepsilon)$-approximation algorithm for finding the minimum distortion embedding of an n-point metric space X into the shortest path metric space of a weighted graph G with m vertices. The running time of our algorithm is

$$m^{O(1)} \cdot n^{O(\omega)} \cdot (\delta_{opt}\Delta)^{\omega\cdot(1/\varepsilon)^{\lambda+2}\cdot\lambda\cdot\left(O(\delta_{opt})\right)^{2\lambda}}$$

parametrized by the values of the minimum distortion, δ_{opt}, the spread, Δ, of the points of X, the treewidth, ω, of G, and the doubling dimension, λ, of G. In particular, our result implies a PTAS provided an X with polynomial spread, and the doubling dimension of G, the treewidth of G, and δ_{opt}, are all constant. For example, if X has a polynomial spread and δ_{opt} is a constant, we obtain PTAS's for embedding X into the following spaces: the line, a cycle, a tree of bounded doubling dimension, and a k-outer planar graph of bounded doubling dimension (for a constant k).

.

[57] Parinya Chalermsook, Syamantak Das, Bundit Laekhanukit and Daniel Vaz. Beyond Metric Embedding: Approximating Group Steiner Trees on Bounded Treewidth Graphs. **SODA 3C** (Nelva). January 16th, 2017, 18:40-19:00

The *Group Steiner Tree* (GST) problem is a classical problem in combinatorial optimization and theoretical computer science. In the Edge-Weighted Group Steiner Tree (EW-GST) problem, we are given an undirected graph $G=(V,E)$ on n vertices with edge costs $c: E \rightarrow \mathbb{R}_{\geq 0}$, a source vertex s and a collection of subsets of vertices, called *groups*, $S_1,\ldots,S_k \subseteq V$. The goal is to find a minimum-cost tree $H \subseteq G$ that connects s to some vertex from each group S_i, for all $i=1,2,\ldots,k$. The Node-Weighted Group Steiner Tree (NW-GST) problem has the same setting, but the costs are associated with nodes. The goal is to find a minimum-cost node set $X \subseteq V$ such that $G[X]$ connects every group to the source. When G is a tree, both EW-GST and NW-GST admit a polynomial-time $O(\log n \log k)$ approximation algorithm due to the seminal result of (N. Garg, G. Konjevod, and R. Ravi, *A polylogarithmic approximation algorithm for the group steiner tree problem*, J. Algorithms 37(1):66–84, 2000, Preliminary version in SODA'98). The matching hardness of $\log^{2-\epsilon} n$ is known even for tree instances of EW-GST and NW-GST (E. Halperin and R. Krauthgamer, *Polylogarithmic inapproximability*, In Proceedings of the 35th Annual ACM Symposium on Theory of Computing, June 9-11, 2003, San Diego, CA, USA, pages 585–594, 2003). In general graphs, most of polynomial-time approximation algorithms for EW-GST reduce the problem to a tree instance using the metric-tree embedding, incurring a loss of $O(\log n)$ on the approximation factor (Y. Bartal, *Probabilistic approximations of metric spaces and its algorithmic applications*, In 37th Annual Symposium on Foundations of Computer Science, FOCS '96, Burlington, Vermont, USA, 14-16 October, 1996, pages 184–193, 1996), (J. Fakcharoenphol, S. Rao, and K. Talwar, *A tight bound on approximating arbitrary metrics by tree metrics*, J. Comput. Syst. Sci. 69(3):485–497, 2004, Preliminary version in STOC'03). This yields an approximation ratio of $O(\log^2 n \log k)$ for EW-GST. Using metric-tree embedding, this factor cannot be improved: The loss of $\Omega(\log n)$ is necessary on some input graphs (e.g., grids and expanders). There are alternative approaches that avoid metric-tree embedding, e.g., the algorithm of (C. Chekuri and M. Pál, *A recursive greedy algorithm for walks in directed graphs*, In 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23-25 October 2005, Pittsburgh, PA, USA, Proceedings, pages 245–253, 2005), which gives a tight approximation ratio, but none of which achieves polylogarithmic approximation in polynomial-time. This state of the art shows a clear lack of understanding of GST in general graphs beyond the metric-tree embedding technique. For NW-GST (for which the metric-tree embedding does not apply), not even a polynomial-time polylogarithmic approximation algorithm is known. In this paper, we present $O(\log n \log k)$ approximation algorithms that run in time $n^{\tilde{O}(tw(G)^2)}$ for both NW-GST and EW-GST, where $tw(G)$ denotes the treewidth of graph G. The key to both results is a different type of "tree-embedding" that produces a tree of much bigger size, but *does not cause any loss on the approximation factor.* Our embedding is inspired by dynamic programming, a technique which is typically not applicable to Group Steiner problems.

.

[58] Dina Katabi. Systems and Algorithms for the Internet of Things. **Invited Talk 3** (Las Arenas Foyer). January 17th, 2017, 09:00-10:00

In this talk, I will describe how we design new IoT devices that analyze the surrounding radio signals to infer our movements, vital signs and even emotions. In particular, I will show how we can accurately track the 3D motion of people from the wireless signals reflected off their bodies, even if they are behind a wall. Such fine-grained tracking can also recognize our gestures, enabling us to control smart home devices simply by pointing at them. I will also describe mechanisms and algorithms to learn humans' breathing, heart rates, and basic emotions from the reflections of wireless signals off their bodies, and without requiring them to hold or wear any device.

.

[59] Subhroshekhar Ghosh, Thomas M. Liggett and Robin Pemantle. Multivariate CLT follows from strong Rayleigh property. **ANALCO 4** (Saray). January 17th, 2017, 10:30-10:50

Let $(X_1, \ldots, X_d)$ be a random nonnegative integer vector. Many conditions are known to imply a central limit theorem for a sequence of such random vectors, for example, independence and convergence of the normalized covariances, or various combinatorial conditions allowing the application of Stein's method, couplings, etc. Here, we prove a central limit theorem directly from hypotheses on the probability generating function $f(z_1, \ldots, z_d)$. In particular, we show that the f being *real stable* (meaning no zeros with all coordinates in the open upper half plane) is enough to imply a CLT under a nondegeneracy condition on the variance. Known classes of distributions with real stable generating polynomials include spanning tree measures, conditioned Bernoullis and counts for determinantal point processes. Soshnikov (2002) showed that occupation counts of disjoint sets by a determinantal point process satisfy a multivariate CLT. Our results extend Soshnikov's to the class of real stable laws. The class of real stable laws is much larger than the class of determinantal laws, being defined by inequalities rather than identities. Along the way we investigate the related problem of *stable multiplication*.

.

[60] Joachim M. Buhmann, Julien Dumazert, Alexey Gronskiy and Wojciech Szpankowski. Phase Transitions in Parameter Rich Optimization Problems. **ANALCO 4** (Saray). January 17th, 2017, 10:55-11:15

Most real world combinatorial optimization problems are affected by noise in the input data. Search algorithms to identify "good" solutions with low costs behave like the dynamics of large disordered particle systems, e.g., random networks or spin glasses. Such solutions to noise perturbed optimization problems are characterized by *Gibbs distributions* when the optimization algorithm searches for typical solutions by stochastically minimizing costs. The *free energy* that determines the normalization of the Gibbs distribution balances cost minimization relative to entropy maximization. The problem to analytically compute the free energy of disordered systems has been known as a notoriously difficult mathematical challenge for at least half a century (see M. Talagrand. *Spin Glasses: A Challenge for Mathematicians: Cavity and Mean Field Models.* Springer Verlag, 2003.). We provide rigorous, matching upper and lower bounds on the free energy for two disordered combinatorial optimization problems of random graph instances, the sparse Minimum Bisection Problem (sMBP) and Lawler's Quadratic Assignment Problem (LQAP). These two problems exhibit phase transitions that are equivalent to the statistical behavior of Derrida's Random Energy Model (REM). Both optimization problems can be characterized as *parameter rich* since individual solutions depend on more parameters than the logarithm of the solution space cardinality would suggest for e.g. a coordinate representation.

.

[61] Jacek Cichoń, Abram Magner, Wojciech Szpankowski and Krzysztof Turowski. On Symmetries of Non-Plane Trees in a Non-Uniform Model. **ANALCO 4** (Saray). January 17th, 2017, 11:20-11:40

Binary trees come in two varieties: plane trees, often simply called binary trees, and non-plane trees, in which the order of subtrees does not matter. Non-plane trees find many applications; for example in modeling epidemics, in studying phylogenetic trees, and as models in data compression. While binary trees have been studied very extensively, non-plane trees still pose some challenges. Moreover, in most analyses a uniform probabilistic model is assumed; that is, a tree is selected uniformly from among all trees. Such a model limits significantly applications of the analysis. In this paper we study by analytic techniques non-plane trees in a non-uniform model. In our model, we grow the tree on n leaves by selecting randomly a leaf and appending two children to it. We can show that this is equivalent to an alternative model, also used to analyze the average-case performance of binary search trees, that is more easily amenable to study by recurrences and generating functions. Here, one of the most important questions is the number of symmetries in such trees (i.e., the number of internal nodes with two isomorphic subtrees), or the sizes of such symmetric subtrees. We first present a functional-differential equation characterizing tree symmetries, and then analyze it. In this conference paper we focus on the expected number of symmetries, the size of symmetric subtrees, and the tree entropy.

.

[62] Thibault Godin. An analogue to Dixon's theorem for automaton groups. **ANALCO 4** (Saray). January 17th, 2017, 11:45-12:05

Dixon's theorem states that the group generated by two random permutations of a finite set is generically either the whole symmetric group or the alternating group. In the context of random generation of finite groups, this means that it is hopeless to wish for a uniform distribution—or even a non-trivial one—by drawing random permutations and looking at the generated group. Mealy automata are a powerful tool to generate groups, including all finite groups and many interesting infinite ones, whence the idea of generating random finite groups by drawing random Mealy automata. In this paper we show that, for a special class of Mealy automata that generate only finite groups, the distribution is far from being uniform since the obtained groups are generically a semi-direct product between a direct product of alternating groups and a group generated by a tuple of transpositions.

.

[63] Wenlong Mou and Liwei Wang. A Refined Analysis of LSH for Well-dispersed Data Points. **ANALCO 4** (Saray). January 17th, 2017, 12:10-12:30

Near neighbor problems are fundamental in algorithms for high-dimensional Euclidean spaces. While classical approaches suffer from the curse of dimensionality, locality sensitive hashing (LSH) can effectively solve α-approximate r-near neighbor problem, and has been proven to be optimal in the worst case. However, for real-world data sets, LSH can naturally benefit from well-dispersed data and low doubling dimension, leading to significantly improved performance.

In this paper, we address this issue and propose a refined analyses for running time of approximating near neighbors queries via LSH. We characterize dispersion of data using N_β, the number of βr-near pairs among the data points. Combined with optimal data-oblivious LSH scheme, we get a $O\left(\left(1+\frac{4\sqrt{2}\alpha}{\beta}\right)^{\frac{d}{2\alpha^2}}(n+N_\beta)^{\frac{1}{2\alpha^2}}\right)$ bound for expected query time. For many natural scenarios where points are well-dispersed or lying in a low-doubling-dimension space, our result leads to sharper performance than existing worst-case analysis. This paper not only presents the *first* rigorous proof on how LSHs make use of the structure of data points, but also provides important insights into parameter setting in the practice of LSH beyond worst case. Besides, the techniques in our analysis involve a generalized version of sphere packing problem, which might be of some independent interest.

.

[64] Michael B. Cohen, Aleksander Madry, Piotr Sankowski and Adrian Vladu. Negative-Weight Shortest Paths and Unit Capacity Minimum Cost Flow in $\tilde{O}(m^{10/7}\log W)$ Time. **SODA 4A** (Las Arenas II-IV). January 17th, 2017, 10:30-10:50

In this paper, we study a set of combinatorial optimization problems on weighted graphs: the shortest path problem with negative weights, the weighted perfect bipartite matching problem, the unit-capacity minimum-cost maximum flow problem, and the weighted perfect bipartite b-matching problem under the assumption that $\|b\|_1 = O(m)$. We show that each of these four problems can be solved in $\tilde{O}(m^{10/7}\log W)$ time, where W is the absolute maximum weight of an edge in the graph, providing the first polynomial improvement in their sparse-graph time complexity in over 25 years.

.

[65] Jonah Sherman. Generalized Preconditioning and Undirected Minimum-Cost Flow. **SODA 4A** (Las Arenas II-IV). January 17th, 2017, 10:55-11:15

We present a nearly-linear time approximation algorithm for uncapacitated minimum-cost flow in undirected graphs, along with a more general framework for approximately solving problems of the form: find x satisfying $Ax = b$ with minimal norm $\|x\|$, where the norm is generally non-Euclidean. For most of the extensive applications of the latter problem, the exact constraints are essential, so an x satisfying $Ax = b$ with almost-minimal norm is acceptable, while relaxing $Ax = b$ to $Ax \approx b$ significantly beyond numerical precision is not. On the other hand, existing nearly-linear time solvers for non-Euclidean norms use dual or penalty methods, yielding the opposite notion where $\|x\|$ is minimal while $\|b - Ax\| \leq t^{-\Omega(1)}$ after t iterations. We show that by composing solvers of the latter type, we may obtain solvers of the more-useful former type. Convergence of the composed solvers depends strongly on a generalization of the classical condition number to general norms. Following our framework, the task of the algorithm designer for such problems is reduced to that of designing a *generalized preconditioner* for A. Applying the framework to uncapacitated minimum-cost flow, we present an algorithm that, given an undirected graph with m edges labelled with costs, and n vertices labelled with demands, takes $\epsilon^{-2}m^{1+o(1)}$-time and outputs a flow routing the demands with total cost at most $(1+\epsilon)$ times larger than minimal, along with a dual solution proving near-optimality. The generalized preconditioner is obtained by embedding the cost metric into ℓ_1, and then considering a simple hierarchical routing scheme in ℓ_1 where demands initially supported on a dense lattice are pulled from a sparser lattice by randomly rounding unaligned coordinates to their aligned neighbors. Analysis of the generalized condition number for the corresponding preconditioner follows that of the classical multigrid algorithm for lattice Laplacian systems.

.

[66] Ran Duan, Seth Pettie and Hsin-Hao Su. Scaling Algorithms for Weighted Matching in General Graphs. **SODA 4A** (Las Arenas II-IV). January 17th, 2017, 11:20-11:40

We present a new scaling algorithm for maximum (or minimum) weight perfect matching on general, edge weighted graphs. Our algorithm runs in $O(m\sqrt{n}\log(nN))$ time, $O(m\sqrt{n})$ per scale, which matches the running time of the best cardinality matching algorithms on sparse graphs. Here m, n, and N bound the number of edges, vertices, and magnitude of any integer edge weight. Our result improves on a 25-year old algorithm of Gabow and Tarjan, which runs in $O(m\sqrt{n \log n \alpha(m,n)} \log(nN))$ time.

.

[67] Chandra Chekuri and Kent Quanrud. Near-Linear Time Approximation Schemes for Some Implicit Fractional Packing Problems. **SODA 4A** (Las Arenas II-IV). January 17th, 2017, 11:45-12:05

We consider several *implicit* fractional packing problems and obtain faster implementations of approximation schemes based on multiplicative-weight updates. This leads to new algorithms with near-linear running times for some fundamental problems in combinatorial optimization. We highlight two concrete applications. The first is to find the maximum fractional packing of spanning trees in a capacitated graph; we obtain a $(1-\epsilon)$-approximation in $\tilde{O}(m/\epsilon^2)$ time, where m is the number of edges in the graph. Second, we consider the LP relaxation of the weighted unsplittable flow problem on a path and obtain a $(1-\epsilon)$-approximation in $\tilde{O}(n/\epsilon^2)$ time, where n is the number of demands.

.

[68] Miriam Schlöter and Martin Skutella. Fast and Memory-Efficient Algorithms for Evacuation Problems. **SODA 4A** (Las Arenas II-IV). January 17th, 2017, 12:10-12:30

We study two classical flow over time problems that capture the essence of evacuation planning. Given a network with capacities and transit times on the arcs and sources/sinks with supplies/demands, a *quickest transshipment* sends the supplies from the sources to meet the demands at the sinks as quickly as possible. In a 1995 landmark paper, Hoppe and Tardos describe the first strongly polynomial time algorithm solving the quickest transshipment problem. Their algorithm relies on repeatedly calling an oracle for parametric submodular function minimization. We present a somewhat simpler and more efficient algorithm for the quickest transshipment problem. Our algorithm (i) relies on only one parametric submodular function minimization and, as a consequence, has considerably improved running time, (ii) uses not only the solution of a submodular function minimization but actually exploits the underlying algorithmic approach to determine a quickest transshipment as a convex combination of simple lex-max flows over time, and (iii) in this way determines a structurally easier solution in the form of a generalized temporally repeated flow. Our second main result is an entirely novel algorithm for computing *earliest arrival transshipments*, which feature a particularly desirable property in the context of evacuation planning. An earliest arrival transshipment – which in general only exists in networks with a single sink – is a quickest transshipment maximizing the amount of flow which has reached the sink for every point in time simultaneously. In contrast to previous approaches, our algorithm solely works on the given network and, as a consequence, requires only polynomial space.

.

[69] Moses Charikar and Vaggos Chatziafratis. Approximate Hierarchical Clustering via Sparsest Cut and Spreading Metrics. **SODA 4B** (Las Arenas I). January 17th, 2017, 10:30-10:50

Dasgupta recently introduced a cost function for the hierarchical clustering of a set of points given pairwise similarities between them. He showed that this function is NP-hard to optimize, but a top-down recursive partitioning heuristic based on an α_n-approximation algorithm for uniform sparsest cut gives an approximation of $O(\alpha_n \log n)$ (the current best algorithm has $\alpha_n = O(\sqrt{\log n})$). We show that the aforementioned sparsest cut heuristic in fact obtains an $O(\alpha_n)$-approximation. The algorithm also applies to a generalized cost function studied by Dasgupta. Moreover, we obtain a strong inapproximability result, showing that the Hierarchical Clustering objective is hard to approximate to within any constant factor assuming the *Small-Set Expansion (SSE) Hypothesis*. Finally, we discuss approximation algorithms based on convex relaxations. We present a spreading metric SDP relaxation for the problem and show that it has integrality gap at most $O(\sqrt{\log n})$. The advantage of the SDP relative to the sparsest cut heuristic is that it provides an explicit lower bound on the optimal solution and could potentially yield an even better approximation for hierarchical clustering. In fact our analysis of this SDP served as the inspiration for our improved analysis of the sparsest cut heuristic. We also show that a spreading metric LP relaxation gives an $O(\log n)$-approximation.

.

[70] Chandra Chekuri and Vivek Madan. Approximating Multicut and the demand graph. **SODA 4B** (Las Arenas I). January 17th, 2017, 10:55-11:15

In the minimum Multicut problem, the input is an edge-weighted supply graph $G = (V, E)$ and a demand graph $H = (V, F)$. Either G and H are directed (DMulC) or both are undirected (UMulC). The goal is to remove a minimum weight set of supply edges $E' \subseteq E$ such that in $G - E'$ there is no path from s to t for any demand edge $(s,t) \in F$. UMulC admits $O(\log k)$-approximation where k is the number of edges in H while the best known approximation for DMulC is $\min\{k, \tilde{O}(|V|^{11/23})\}$. These approximations are obtained by proving corresponding results on the multicommodity flow-cut gap. In this paper we consider the role that the structure of the demand graph plays in determining the approximability of Multicut. We obtain several new positive and negative results. In undirected graphs our main result is a 2-approximation in $n^{O(t)}$ time when the demand graph excludes an induced matching of size t. This gives a constant factor approximation for a specific demand graph that motivated this work, and is based on a reduction to uniform metric labeling and not via the flow-cut gap. In contrast to the positive result for undirected graphs, we prove that in directed graphs such approximation algorithms can not exist. We prove that, assuming the Unique Games Conjecture (UGC), that for a large class of fixed demand graphs DMulC cannot be approximated to a factor better than the worst-case flow-cut gap. As a consequence we prove that for any fixed k, assuming UGC, DMulC with k demand pairs is hard to approximate to within a factor better than k. On the positive side, we obtain a k approximation when the demand graph excludes certain graphs as an induced subgraph. This generalizes the known 2 approximation for directed Multiway Cut to a larger class of demand graphs.

.

[71] Yixin Cao and R.B. Sandeep. Minimum Fill-In: Inapproximability and Almost Tight Lower Bounds. **SODA 4B** (Las Arenas I). January 17th, 2017, 11:20-11:40

Performing Gaussian elimination to a sparse matrix may turn some zeroes into nonzero values, so called fill-ins, which we want to minimize to keep the matrix sparse. Let n denote the rows of the matrix and k the number of fill-ins. For the minimum fill-in problem, we exclude the existence of polynomial time approximation schemes, assuming P≠NP, and the existence of $2^{O(n^{1-\delta})}$-time approximation schemes for any positive δ, assuming the Exponential Time Hypothesis. Also implied is a $2^{O(k^{1/2-\delta})} \cdot n^{O(1)}$ parameterized lower bound. Behind these results is a new reduction from vertex cover, which might be of its own interest: All previous reductions for similar problems are from some kind of graph layout problems.

.

[72] Eden Chlamtáč, Michael Dinitz and Yury Makarychev. Minimizing the Union: Tight Approximations for Small Set Bipartite Vertex Expansion. **SODA 4B** (Las Arenas I). January 17th, 2017, 11:45-12:05

In the Minimum k-Union problem (MkU) we are given a set system with n sets and are asked to select k sets in order to minimize the size of their union. Despite being a very natural problem, it has received surprisingly little attention: the only known approximation algorithm is an $O(\sqrt{n})$-approximation due to [Chlamtáč et al APPROX '16]. This problem can also be viewed as the bipartite version of the Small Set Vertex Expansion problem (SSVE), which we call the Small Set Bipartite Vertex Expansion problem (SSBVE). SSVE, in which we are asked to find a set of k nodes to minimize their vertex expansion, has not been as well studied as its edge-based counterpart Small Set Expansion (SSE), but has recently received significant attention, e.g. [Louis-Makarychev APPROX '15]. However, due to the connection to Unique Games and hardness of approximation the focus has mostly been on sets of size $k = \Omega(n)$, while we focus on the case of general k, for which no polylogarithmic approximation is known. We improve the upper bound for this problem by giving an $n^{1/4+\varepsilon}$ approximation for SSBVE for any constant $\varepsilon > 0$. Our algorithm follows in the footsteps of Densest k-Subgraph (DkS) and related problems, by designing a tight algorithm for random models, and then extending it to give the same guarantee for arbitrary instances. Moreover, we show that this is tight under plausible complexity conjectures: it cannot be approximated better than $O(n^{1/4})$ assuming an extension of the so-called "Dense versus Random"" conjecture for DkS to hypergraphs. In addition to conjectured hardness via our reduction, we show that the same lower bound is also matched by an integrality gap for a super-constant number of rounds of the Sherali-Adams LP hierarchy, and an even worse integrality gap for the natural SDP relaxation. Finally, we note that there exists a simple bicriteria $\tilde{O}(\sqrt{n})$ approximation for the more general SSVE problem (where no non-trivial approximations were known for general k).

.

[73] Eden Chlamtáč, Pasin Manurangsi, Dana Moshkovitz and Aravindan Vijayaraghavan. Approximation Algorithms for Label Cover and The Log-Density Threshold. **SODA 4B** (Las Arenas I). January 17th, 2017, 12:10-12:30

Many known optimal NP-hardness of approximation results are reductions from a problem called LABEL-COVER. The input is a bipartite graph $G = (L, R, E)$ and each edge $e = (x, y) \in E$ carries a projection π_e that maps labels to x to labels to y. The objective is to find a labeling of the vertices that satisfies as many of the projections as possible. It is believed that the best approximation ratio efficiently achievable for LABEL-COVER is of the form N^{-c} where $N = nk$, n is the number of vertices, k is the number of labels, and $0 < c < 1$ is *some* constant. Inspired by a framework originally developed for DENSEST k-SUBGRAPH, we propose a "log density threshold" for the approximability of Label-Cover. Specifically, we suggest the possibility that the Label-Cover approximation problem undergoes a computational phase transition at the same threshold at which local algorithms for its random counterpart fail. This threshold is $N^{3-2\sqrt{2}} \approx N^{-0.17}$. We then design, for any $\varepsilon > 0$, a polynomial-time approximation algorithm for *semi-random* LABEL-COVER whose approximation ratio is $N^{3-2\sqrt{2}+\varepsilon}$. In our semi-random model, the input graph is random (or even just expanding), and the projections on the edges are arbitrary. For *worst-case* LABEL-COVER we show a polynomial-time algorithm whose approximation ratio is roughly $N^{-0.233}$. The previous best efficient approximation ratio was $N^{-0.25}$. We present some evidence towards an N^{-c} threshold by constructing integrality gaps for $N^{\Omega(1)}$ rounds of the Sum-of-squares/Lasserre hierarchy of the natural relaxation of Label Cover. For general 2CSP the "log density threshold" is $N^{-0.25}$, and we give a polynomial-time algorithm in the semi-random model whose approximation ratio is $N^{-0.25+\varepsilon}$ for any $\varepsilon > 0$.

.

[74] Michele Borassi, Pierluigi Crescenzi and Luca Trevisan. An Axiomatic and an Average-Case Analysis of Algorithms and Heuristics for Metric Properties of Graphs. **SODA 4C** (Nelva). January 17th, 2017, 10:30-10:50

In recent years, researchers proposed several algorithms that compute metric quantities of real-world complex networks, and that are very efficient in practice, although there is no worst-case guarantee. In this work, we propose an axiomatic framework to analyze the performances of these algorithms, by proving that they are efficient on the class of graphs satisfying certain properties. Furthermore, we prove that these properties are verified asymptotically almost surely by several probabilistic models that generate power law random graphs, such as the Configuration Model, the Chung-Lu model, and the Norros-Reittu model. Thus, our results imply average-case analyses in these models. For example, in our framework, existing algorithms can compute the diameter and the radius of a graph in subquadratic time, and sometimes even in time $n^{1+o(1)}$. Moreover, in some regimes, it is possible to compute the k most central vertices according to closeness centrality in subquadratic time, and to design a distance oracle with sublinear query time and subquadratic space occupancy. In the worst case, it is impossible to obtain comparable results for any of these problems, unless widely-believed conjectures are false.

.

[75] Luca Becchetti, Andrea Clementi, Emanuele Natale, Francesco Pasquale and Luca Trevisan. Find Your Place: Simple Distributed Algorithms for Community Detection. **SODA 4C** (Nelva). January 17th, 2017, 10:55-11:15

Given an underlying graph, we consider the following *dynamics*: Initially, each node locally chooses a value in $\{-1, 1\}$, uniformly at random and independently of other nodes. Then, in each consecutive round, every node updates its local value to the average of the values held by its neighbors, at the same time applying an elementary, local clustering rule that only depends on the current and the previous values held by the node. We prove that the process resulting from this dynamics produces a clustering that exactly or approximately (depending on the graph) reflects the underlying cut in logarithmic time, under various graph models that exhibit a sparse balanced cut, including the stochastic block model. We also prove that a natural extension of this dynamics performs community detection on a regularized version of the stochastic block model with multiple communities. Rather surprisingly, our results provide rigorous evidence for the ability of an extremely simple and natural dynamics to address a computational problem that is non-trivial even in a centralized setting.

.

[76] Aviad Rubinstein and Shai Vardi. Sorting from Noisier Samples. **SODA 4C** (Nelva). January 17th, 2017, 11:20-11:40

We study the problem of constructing an order over a set of elements given noisy samples. We consider two models for generating the noisy samples; in both, the distribution of samples is induced by an unknown state of nature: a permutation ρ. In *Mallow's model*, r permutations π_i are generated independently from ρ, each with probability proportional to $e^{-\beta d_K(\rho,\pi_i)}$, where $d_K(\rho, \pi_i)$ is the Kemeny distance between ρ and π_i - the number of pairs they order differently. In the *noisy comparisons* model, we are given a tournament, generated from ρ as follows: if i is before j in ρ, then with probability $1/2 + \gamma$, the edge between them is oriented from i to j. Both of these problems were studied by Braverman and Mossel [BM09]; they showed how to construct a maximum-likelihood permutation when the noise parameter (β or γ, respectively) is constant. In this work, we obtain algorithms that work in the presence of stronger noise ($\beta^2 r = \tilde{\Omega}\left(\frac{1}{\log^2 n}\right)$ or $\gamma = \tilde{\Omega}\left(\frac{1}{\log^{1/6} n}\right)$, respectively). In Mallow's model, our algorithm works for a relaxed solution concept: *likelier than nature*. That is, rather than requiring that our output maximizes the likelihood over the entire domain, we guarantee that the likelihood of our output is, w.h.p., greater than or equal to that of the true state of nature (ρ). An interesting feature of our algorithm is that it handles noise by adding more noise.

.

[77] Uri Grupel. Sampling on the Sphere by Mutually Orthogonal Subspaces. **SODA 4C** (Nelva). January 17th, 2017, 11:45-12:05

The purpose of this paper is twofold. First, we provide an optimal $\Omega(\sqrt{n})$ bits lower bound for any two-way protocol for the Vector in Subspace Communication Problem which is of bounded total rank. This result complements Raz's $O(\sqrt{n})$ protocol, which has a simple variant of bounded total rank. Second, we present a plausible mathematical conjecture on a measure concentration phenomenon that implies an $\Omega(\sqrt{n})$ lower bound for a general protocol. We prove the conjecture for the subclass of sets that depend only on $O(\sqrt{n})$ directions.

.

[78] Nicole Immorlica, Robert Kleinberg, Brendan Lucier and Morteza Zadomighaddam. Exponential Segregation in a Two-Dimensional Schelling Model with Tolerant Individuals. **SODA 4C** (Nelva). January 17th, 2017, 12:10-12:30

We prove that the two-dimensional Schelling segregation model yields monochromatic regions of size exponential in the area of individuals' neighborhoods, provided that the tolerance parameter is a constant strictly less than 1/2 but sufficiently close to it. Our analysis makes use of a connection with the first-passage percolation model from the theory of stochastic processes.

.

[79] Michael Stein, Karsten Weihe, Augustin Wilberg, Roland Kluge, Julian M. Klomp, Mathias Schnee, Lin Wang and Max Mühlhäuser. Distributed Graph-based Topology Adaptation using Motif Signatures. **ALENEX 1** (Saray). January 17th, 2017, 14:00-14:20

A motif is a small graph pattern, and a motif signature counts the occurrences of selected motifs in a network. The motif signature of a real-world network is an important characteristic because it is closely related to a variety of semantic and functional aspects. In recent years, motif analysis has been successfully applied for adapting topologies of communication networks: The motif signatures of very good networks (e.g., in terms of load balancing) are determined a priori to derive a target motif signature. Then, a given network is adapted in iterative steps, subject to side constraints and in a distributed way, such that its motif signature approximates the target motif signature. In this paper, we formalize this adaptation problem and show that it is $\mathcal{NP}$-hard. We present LoMbA, a generic approach for motif-based graph adaptation: All types of networks, all selections of motifs, and all types of consistency-maintaining constraints can be incorporated. To evaluate LoMbA, we conduct a simulation study based on several scenarios of topology adaptation from the domain of communication networks. We consider topology control in wireless ad-hoc networks, balancing of video streaming trees, and load balancing of peer-to-peer overlays. In each considered application scenario, the simulation results are remarkably good, although the implementation was not tuned toward these scenarios.

.

[80] Jouni Sirén. Indexing Variation Graphs. **ALENEX 1** (Saray). January 17th, 2017, 14:25-14:45

Variation graphs, which represent genetic variation within a population, are replacing sequences as reference genomes. Path indexes are one of the most important tools for working with variation graphs. They generalize text indexes to graphs, allowing one to find the paths matching the query string. We propose using de Bruijn graphs as path indexes, compressing them by merging redundant subgraphs, and encoding them with the Burrows-Wheeler transform. The resulting fast, space-efficient, and versatile index is used in the variation graph toolkit vg.

.

[81] Yaroslav Akhremtsev, Tobias Heuer, Peter Sanders and Sebastian Schlag. Engineering a direct k-way Hypergraph Partitioning Algorithm. **ALENEX 1** (Saray). January 17th, 2017, 14:50-15:10

We develop a fast and high quality multilevel algorithm that directly partitions hypergraphs into k balanced blocks – without the detour over recursive bipartitioning. In particular, our algorithm efficiently implements the powerful FM local search heuristics for the complicated k-way case. This is important for objective functions which depend on the number of blocks connected by a hyperedge. We also remove several further bottlenecks in processing large hyperedges, develop a faster contraction algorithm, and a new adaptive stopping rule for local search. To further reduce the size of hyperedges, we develop a *pin-sparsifier* based on the min-hashing technique that clusters vertices with similar neighborhood. Extensive experiments indicate that our KaHyPar-partitioner compares favorably with the best previous systems. KaHyPar is faster than hMetis *and* computes better solutions. KaHyPar's results are considerably better than the (faster) PaToH partitioner.

.

[82] Nilakantha Paudel, Loukas Georgiadis and Giuseppe F. Italiano. Computing Critical Nodes in Directed Graphs. **ALENEX 1** (Saray). January 17th, 2017, 15:15-15:35

We consider the critical node detection problem (CNDP) in directed graphs. Given a directed graph G and a parameter k, we wish to remove a subset S of at most k vertices of G such that the residual graph $G \setminus S$ has minimum pairwise strong connectivity. This problem is NP-hard, and thus we are interested in practical heuristics. We present a sophisticated linear-time algorithm for the $k = 1$ case, and, based on this algorithm, give an efficient heuristic for the general case. Then, we conduct a thorough experimental evaluation of various heuristics for CNDP. Our experimental results suggest that our heuristic performs very well in practice, both in terms of running time and of solution quality.

.

[83] Michael Hamann, Ulrich Meyer, Manuel Penschuck and Dorothea Wagner. I/O-efficient Generation of Massive Graphs Following the LFR Benchmark. **ALENEX 1** (Saray). January 17th, 2017, 15:40-16:00

LFR is a popular benchmark graph generator used to evaluate community detection algorithms. We present EM-LFR, the first external memory algorithm able to generate massive complex networks following the LFR benchmark. Its most expensive component is the generation of random graphs with prescribed degree sequences which can be divided into two steps: the graphs are first materialized deterministically using the Havel-Hakimi algorithm, and then randomized. Our main contributions are EM-HH and EM-ES, two I/O-efficient external memory algorithms for these two steps. In an experimental evaluation we demonstrate their performance: our implementation is able to handle graphs with more than 37 billion edges on a single machine, is competitive with a massive parallel distributed algorithm, and is faster than a state-of-the-art internal memory implementation even on instances fitting in main memory. EM-LFR's implementation is capable of generating large graph instances orders of magnitude faster than the original implementation. We give evidence that both implementations yield graphs with matching properties by applying clustering algorithms to generated instances.

.

[84] Antje Bjelde, Yann Disser, Jan Hackfeld, Christoph Hansknecht, Maarten Lipmann, Julie Meißner, Kevin Schewior, Miriam Schlöter and Leen Stougie. Tight Bounds for Online TSP on the Line. **SODA 5A** (Las Arenas II-IV). January 17th, 2017, 14:00-14:20

We consider the online traveling salesperson problem (TSP), where requests appear online over time on the real line and need to be visited by a server initially located at the origin. We distinguish between closed and open online TSP, depending on whether the server eventually needs to return to the origin or not. While online TSP on the line is a very natural online problem that was introduced more than two decades ago, no tight competitive analysis was known to date. We settle this problem by providing tight bounds on the competitive ratios for both the closed and the open variant of the problem. In particular, for closed online TSP, we provide a 1.64-competitive algorithm, thus matching a known lower bound. For open online TSP, we give a new upper bound as well as a matching lower bound that establish the remarkable competitive ratio of 2.04.

Additionally, we consider the online DIAL-A-RIDE problem on the line, where each request needs to be transported to a specified destination. We provide an improved non-preemptive lower bound of 1.75 for this setting, as well as an improved preemptive algorithm with competitive ratio 2.41.

Finally, we generalize known and give new complexity results for the underlying offline problems. In particular, we give an algorithm with running time $O(n^2)$ for closed offline TSP on the line with release dates and show that both variants of offline DIAL-A-RIDE on the line are NP-hard for any capacity $c \geq 2$ of the server.

.

[85] George Christodoulou and Alkmini Sgouritsa. An Improved Upper Bound for the Universal TSP on the Grid. **SODA 5A** (Las Arenas II-IV). January 17th, 2017, 14:25-14:45

We study the universal Traveling Salesman Problem in an $n \times n$ grid with the shortest path metric. The goal is to define a (universal) total ordering over the set of grid's vertices, in a way that for any input (subset of vertices), the tour, which visits the points in this ordering, is a good approximation of the optimal tour, i.e. has low competitive ratio. This problem was first studied by Platzman and Bartholdi 1989. They proposed a heuristic, which was based on the Sierpinski space-filling curve, in order to define a universal ordering of the unit square $[0, 1]^2$ under the Euclidean metric. Their heuristic visits the points of the unit square in the order of their appearance along the space-filling curve. They provided a logarithmic upper bound which was shown to be tight up to a constant by Bertsimas and Grigni 1989. Bertsimas and Grigni further showed logarithmic lower bounds for other space-filling curves and they conjectured that any universal ordering has a logarithmic lower bound for the $n \times n$ grid. In this work, we disprove this conjecture by showing that there exists a universal ordering of the $n \times n$ grid with competitive ratio of $O\left(\frac{\log n}{\log \log n}\right)$. The heuristic we propose defines a universal ordering of the grid's vertices based on a generalization of the Lebesgue space-filling curve. In order to analyze the competitive ratio of our heuristic, we employ techniques from the theory of geometric spanners in Euclidean spaces. We finally show that our analysis is tight up to a constant.

.

[86] Nikhil Bansal, Marek Eliáš, Grigorios Koumoutsos and Łukasz Jeż. The (h,k)-Server Problem on Bounded-Depth Trees. **SODA 5A** (Las Arenas II-IV). January 17th, 2017, 14:50-15:10

We study the k-server problem in the resource augmentation setting i.e., when the performance of the online algorithm with k servers is compared to the offline optimal solution with $h \leq k$ servers. The problem is very poorly understood beyond uniform metrics. For this special case, the classic k-server algorithms are roughly $(1 + 1/\epsilon)$-competitive when $k = (1 + \epsilon)h$, for any $\epsilon > 0$. Surprisingly however, no $o(h)$-competitive algorithm is known even for HSTs of depth 2 and even when k/h is arbitrarily large. We obtain several new results for the problem. First we show that the known k-server algorithms do not work even on very simple metrics. In particular, the Double Coverage algorithm has competitive ratio $\Omega(h)$ irrespective of the value of k, even for depth-2 HSTs. Similarly the Work Function Algorithm, that is believed to be optimal for all metric spaces when $k = h$, has competitive ratio $\Omega(h)$ on depth-3 HSTs even if $k = 2h$. Our main result is a new algorithm that is $O(1)$-competitive for constant depth trees, whenever $k = (1 + \epsilon)h$ for any $\epsilon > 0$. Finally, we give a general lower bound that any deterministic online algorithm has competitive ratio at least 2.4 even for depth-2 HSTs and when k/h is arbitrarily large. This gives a surprising qualitative separation between uniform metrics and depth-2 HSTs for the (h, k)-server problem, and gives the strongest known lower bound for the problem on general metrics.

.

[87] Yossi Azar, Ilan Reuven Cohen and Alan Roytman. Online Lower Bounds via Duality. **SODA 5A** (Las Arenas II-IV). January 17th, 2017, 15:15-15:35

In this paper, we exploit linear programming duality in the online setting, where input arrives on the fly, from the unique perspective of designing lower bounds (i.e., hardness results) on the competitive ratio. In particular, we provide a systematic method (as opposed to ad hoc case analysis that is typically done) for obtaining online deterministic and randomized lower bounds on the competitive ratio for a wide variety of problems. We show the usefulness of our approach by providing new, tight hardness results for three diverse online problems: the Vector Bin Packing problem, Ad-auctions (and various online matching problems), and the Capital Investment problem. Our methods are sufficiently general that they can also be used to reconstruct existing lower bounds. Our approach is in stark contrast to previous works, which exploit linear programming duality to obtain positive results, often via the useful primal-dual scheme. We design a general recipe with the opposite aim of obtaining negative results via duality. The general idea behind our approach is to construct a parameterized family of primal linear programs based on a candidate collection of input sequences for proving the lower bound, where the objective function corresponds to optimizing the competitive ratio. Solving the parameterized family of primal linear programs optimally would yield a valid lower bound, but is a challenging task and limits the tools that can be applied, since analysis must be done precisely and exact optimality needs to be proved. To this end, we consider the corresponding parameterized family of dual linear programs and provide feasible solutions, where the objective function yields a lower bound on the competitive ratio. This opens up additional doors for analysis, including some of the techniques we employ (e.g., continuous analysis, differential equations, etc.), as we need not be so careful about exact optimality. We are confident that our methods can be successfully applied to produce many more lower bounds for a wide array of online problems.

.

[88] Yossi Azar, Ashish Chiplunkar and Haim Kaplan. Polylogarithmic Bounds on the Competitiveness of Min-cost Perfect Matching with Delays. **SODA 5A** (Las Arenas II-IV). January 17th, 2017, 15:40-16:00

We consider the problem of online Min-cost Perfect Matching with Delays (MPMD) recently introduced by Emek et al, (STOC 2016). This problem is defined on an underlying n-point metric space. An adversary presents real-time requests online at points of the metric space, and the algorithm is required to match them, possibly after keeping them waiting for some time. The cost incurred is the sum of the distances between matched pairs of requests (the connection cost), and the sum of the waiting times of the requests (the delay cost). We prove the first logarithmic upper bound and the first polylogarithmic lower bound on the randomized competitive ratio of this problem. We present an algorithm with a competitive ratio of $O(\log n)$, which improves the upper bound of $O(\log^2 n + \log \Delta)$ of Emek et al, by removing the dependence on Δ, the aspect ratio of the metric space (which can be unbounded as a function of n). The core of our algorithm is a deterministic algorithm for MPMD on metrics induced by edge-weighted trees of height h, whose cost is guaranteed to be at most $O(1)$ times the connection cost plus $O(h)$ times the delay cost of every feasible solution. The reduction from MPMD on arbitrary metrics to MPMD on trees is achieved using the result on embedding n-point metric spaces into distributions over weighted hierarchically separated trees of height $O(\log n)$, with distortion $O(\log n)$. We also prove a lower bound of $\Omega(\sqrt{\log n})$ on the competitive ratio of any randomized algorithm. This is the first lower bound which increases with n, and is attained on the metric of n equally spaced points on a line.

.

[89] Konstantinos Koiliaris and Chao Xu. A Faster Pseudopolynomial Time Algorithm for Subset Sum. **SODA 5B** (Las Arenas I). January 17th, 2017, 14:00-14:20

Given a multiset S of n positive integers and a target integer t, the subset sum problem is to decide if there is a subset of S that sums up to t. We present a new divide-and-conquer algorithm that computes *all* the realizable subset sums up to an integer u in $\widetilde{O}(\min\{\sqrt{n}u, u^{4/3}, \sigma\})$, where σ is the sum of all elements in S and $\widetilde{O}$ hides polylogarithmic factors. This result improves upon the standard dynamic programming algorithm that runs in $O(nu)$ time. To the best of our knowledge, the new algorithm is the fastest general deterministic algorithm for this problem. We also present a modified algorithm for finite cyclic groups, which computes all the realizable subset sums within the group in $\widetilde{O}(\min\{\sqrt{n}m, m^{5/4}\})$ time, where m is the order of the group.

.

[90] Karl Bringmann. A Near-Linear Pseudopolynomial Time Algorithm for Subset Sum. **SODA 5B** (Las Arenas I). January 17th, 2017, 14:25-14:45

Given a set Z of n positive integers and a target value t, the Subset Sum problem asks whether any subset of Z sums to t. A textbook pseudopolynomial time algorithm by Bellman from 1957 solves Subset Sum in time $O(nt)$. This has been improved to $O(n \max Z)$ by Pisinger [J. Algorithms'99] and recently to $\tilde{O}(\sqrt{n}t)$ by Koiliaris and Xu [SODA'17]. Here we present a simple and elegant randomized algorithm running in time $\tilde{O}(n+t)$. This improves upon a classic algorithm and is likely to be near-optimal, since it matches conditional lower bounds from Set Cover and k-Clique. We then use our new algorithm and additional tricks to improve the best known polynomial space solution from time $\tilde{O}(n^3 t)$ and space $\tilde{O}(n^2)$ to time $\tilde{O}(nt)$ and space $\tilde{O}(n \log t)$, assuming the Extended Riemann Hypothesis. Unconditionally, we obtain time $\tilde{O}(nt^{1+\varepsilon})$ and space $\tilde{O}(nt^{\varepsilon})$ for any constant $\varepsilon > 0$.

.

[91] Chandra Chekuri and Chao Xu. Computing minimum cuts in hypergraphs. **SODA 5B** (Las Arenas I). January 17th, 2017, 14:50-15:10

We study algorithmic and structural aspects of connectivity in hypergraphs. Given a hypergraph $H = (V, E)$ with $n = |V|$, $m = |E|$ and $p = \sum_{e \in E} |e|$ the fastest known algorithm to compute a global minimum cut in H runs in $O(np)$ time for the uncapacitated case, and in $O(np + n^2 \log n)$ time for the capacitated case. We show the following new results.

- Given an uncapacitated hypergraph H and an integer k we describe an algorithm that runs in $O(p)$ time to find a subhypergraph H' with sum of degrees $O(kn)$ that preserves all edge-connectivities up to k (a k-sparsifier). This generalizes the corresponding result of Nagamochi and Ibaraki from graphs to hypergraphs. Using this sparsification we obtain an $O(p+\lambda n^2)$ time algorithm for computing a global minimum cut of H where λ is the minimum cut value.
- We generalize Matula's argument for graphs to hypergraphs and obtain a $(2+\epsilon)$-approximation to the global minimum cut in a capacitated hypergraph in $O(\frac{1}{\epsilon}(p\log n+n\log^2 n))$ time, and in in $O(p/\epsilon)$ time for uncapacitated hypergraphs.
- We show that a hypercactus representation of *all* the global minimum cuts of a capacitated hypergraph can be computed in $O(np+n^2\log n)$ time and $O(p)$ space.

Our results build upon properties of vertex orderings that were inspired by the maximum adjacency ordering for graphs due to Nagamochi and Ibaraki. Unlike graphs we observe that there are several different orderings for hypergraphs which yield different insights.

.

[92] Mohsen Ghaffari, David R. Karger and Debmalya Panigrahi. Random Contractions and Sampling for Hypergraph and Hedge Connectivity. **SODA 5B** (Las Arenas I). January 17th, 2017, 15:15-15:35

We initiate the study of hedge connectivity of undirected graphs, motivated by dependent edge failures in real-world networks. In this model, edges are partitioned into groups called hedges that fail together. The hedge connectivity of a graph is the minimum number of hedges whose removal disconnects the graph. We give a polynomial-time approximation scheme and a quasi-polynomial exact algorithm for hedge connectivity. This provides strong evidence that the hedge connectivity problem is tractable, which contrasts with prior work that established the intractability of the corresponding $s-t$ min-cut problem. Our techniques also yield new combinatorial and algorithmic results in hypergraph connectivity. Next, we study the behavior of hedge graphs under uniform random sampling of hedges. We show that unlike graphs, all cuts in the sample do not converge to their expected value in hedge graphs. Nevertheless, the min-cut of the sample does indeed concentrate around the expected value of the original min-cut. This leads to a sharp threshold on hedge survival probabilities for graph disconnection. To the best of our knowledge, this is the first network reliability analysis under dependent edge failures.

.

[93] Akshay Ramachandran and Aaron Schild. Sandpile prediction on a tree in near linear time. **SODA 5B** (Las Arenas I). January 17th, 2017, 15:40-16:00

In the *sandpile model*, we are given an undirected graph G and an initial list of chip counts on each vertex of G and we may fire degree(v) chips from any vertex v to its neighbors. Doing chip moves either results in a unique terminal configuration or recurs forever. On many families of graphs – including trees – the problem of computing the final configuration is P-complete and simulation can take as long as $\Theta(n^3)$ time. We give a $O(n\log^5 n)$ time algorithm for trees that computes the terminal configuration or shows that chip firing will not terminate.

.

[94] Raghu Meka. Explicit Resilient Functions Matching Ajtai-Linial. **SODA 5C** (Nelva). January 17th, 2017, 14:00-14:20

A Boolean function on n variables is q-resilient if for any subset of at most q variables, the function is very likely to be determined by a uniformly random assignment to the remaining $n-q$ variables; in other words, no coalition of at most q variables has significant influence on the function. Resilient functions have been extensively studied with a variety of applications in cryptography, distributed computing, and pseudorandomness. The best known balanced resilient function on n variables due to Ajtai and Linial [AL93] is $\Omega(n/(\log^2 n))$-resilient. However, the construction of Ajtai and Linial is by the probabilistic method and does not give an efficiently computable function. In this work we give an explicit monotone depth three almost-balanced Boolean function on n bits that is $\Omega(n/(\log^2 n))$-resilient matching the work of Ajtai and Linial. The best previous explicit construction due to Meka [Meka09] (which only gives a logarithmic depth function) and Chattopadhyay and Zuckermman [CZ15] were only n^{1-c}-resilient for any constant $c<1$. Our construction and analysis are motivated by (and simplifies parts of) the recent breakthrough of [CZ15] giving explicit two-sources extractors for polylogarithmic min-entropy; a key ingredient in their result was the construction of explicit constant-depth resilient functions. An important ingredient in our construction is a new randomness optimal oblivious sampler which preserves moment generating functions of sums of variables and could be useful elsewhere.

.

[95] Noga Alon and Rajko Nenadov. Optimal induced universal graphs for bounded-degree graphs. **SODA 5C** (Nelva). January 17th, 2017, 14:25-14:45

We show that for any constant $\Delta \geq 2$, there exists a graph Γ with $O(n^{\Delta/2})$ vertices which contains every n-vertex graph with maximum degree Δ as an induced subgraph. For odd Δ this significantly improves the best-known earlier bound of Esperet et al. and is optimal up to a constant factor, as it is known that any such graph must have at least $\Omega(n^{\Delta/2})$ vertices. Our proof builds on the approach of Alon and Capalbo (SODA 2008) together with several additional ingredients. The construction of Γ is explicit and is based on an appropriately defined composition of high-girth expander graphs. The proof also provides an efficient deterministic procedure for finding, for any given input graph H on n vertices with maximum degree at most Δ, an induced subgraph of Γ isomorphic to H.

.

[96] Shayan Oveis Gharan and Alireza Rezaei. Approximation Algorithms for Finding Maximum Induced Expanders. **SODA 5C** (Nelva). January 17th, 2017, 14:50-15:10

We initiate the study of approximating the largest induced expander in a given graph G. Given a Δ-regular graph G with n vertices, the goal is to find the set with the largest induced expansion of size at least $\delta{\cdot}n$. We design a bi-criteria approximation algorithm for this problem; if the optimum has induced spectral expansion λ our algorithm returns a $\frac{\lambda}{\log^2 \delta \exp(\Delta/\lambda)}$-(spectral) expander of size at least δn (up to constants). Our proof introduces and employs a novel semidefinite programming relaxation for the largest induced expander problem. We expect to see further applications of our SDP relaxation in graph partitioning problems. In particular, because of the close connection to the small set expansion problem, one may be able to obtain new insights into the unique games problem.

.

[97] Bernhard Haeupler and David G. Harris. Parallel Algorithms and Concentration Bounds for the Lovasz Local Lemma via Witness-DAGs. **SODA 5C** (Nelva). January 17th, 2017, 15:15-15:35

The Lovász Local Lemma (LLL) is a cornerstone principle in the probabilistic method of combinatorics, and a seminal algorithm of Moser & Tardos (2010) provided an efficient randomized algorithm to implement it. This algorithm could be parallelized to give an algorithm that uses polynomially many processors and runs in $O(\log^3 n)$ time, stemming from $O(\log n)$ adaptive computations of a maximal independent set (MIS). Chung et. al. (2014) developed faster local and parallel algorithms, potentially running in time $O(\log^2 n)$, but these algorithms work under significantly more stringent conditions than the LLL. We give a new parallel algorithm, that works under essentially the same conditions as the original algorithm of Moser & Tardos, but uses only a single MIS computation, thus running in $O(\log^2 n)$ time. This conceptually new algorithm also gives a clean combinatorial description of a satisfying assignment which might be of independent interest. Our techniques extend to the deterministic LLL algorithm given by Chandrasekaran et al. (2013) leading to an NC-algorithm running in time $O(\log^2 n)$ as well. We also provide improved bounds on the run-times of the sequential and parallel resampling-based algorithms originally developed by Moser & Tardos. These bounds extend to any problem instance in which the tighter Shearer LLL criterion is satisfied. We also improve on the analysis of Kolipaka & Szegedy (2011) to give tighter concentration results. Interestingly, we are able to give bounds which are independent of the (somewhat mysterious) weighting function used in formulations of the asymmetric LLL.

.

[98] David G. Harris. Deterministic parallel algorithms for fooling polylogarithmic juntas and the Lovász Local Lemma. **SODA 5C** (Nelva). January 17th, 2017, 15:40-16:00

Many randomized algorithms can be derandomized efficiently using either the method of conditional expectations or probability spaces with low (almost-) independence. A series of papers, beginning with work by Luby (1988) and continuing with Berger & Rompel (1991) and Chari et al. (1994), showed that these techniques can be combined to give deterministic parallel algorithms for combinatorial optimization problems involving sums of w-juntas. We improve these algorithms through derandomized variable partitioning. This reduces the processor complexity to essentially independent of w while the running time is reduced from exponential in w to linear in w. For example, we improve the time complexity of an algorithm of Berger & Rompel (1991) for rainbow hypergraph coloring by a factor of approximately $\log^2 n$ and the processor complexity by a factor of approximately $m^{\ln 2}$. As a major application of this, we give an NC algorithm for the Lovász Local Lemma. Previous NC algorithms, including the seminal algorithm of Moser & Tardos (2010) and the work of Chandrasekaran et. al (2013), required that (essentially) the bad-events could span only $O(\log n)$ variables; we relax this to allowing polylog(n) variables. As two applications of our new algorithm, we give algorithms for defective vertex coloring and domatic graph partition. One main sub-problem encountered in these algorithms is to generate a probability space which can "fool" a given list of $GF(2)$ Fourier characters. Schulman (1992) gave an NC algorithm for this; we dramatically improve its efficiency to near-optimal time and processor complexity and code dimension. This leads to a new algorithm to solve the heavy-codeword problem, introduced by Naor & Naor (1993), with a near-linear processor complexity $(mn)^{1+o(1)}$.

.

[99] Simon Gog, Alistair Moffat and Matthias Petri. CSA++: Fast Pattern Search for Large Alphabets. **ALENEX 2** (Saray). January 17th, 2017, 16:25-16:45

Indexed pattern search in text has been studied for many decades. For small alphabets, the FM-Index provides unmatched performance for Count operations, in terms of both space required and search speed. For large alphabets – for example, when the tokens are words – the situation is more complex, and FM-Index representations are compact, but potentially slow. In this paper we apply recent innovations from the field of inverted indexing and document retrieval to compressed pattern search, including for alphabets into the millions. Commencing with the practical compressed suffix array structure developed by Sadakane, we show that the Elias-Fano code-based approach to document indexing can be adapted to provide new trade- off options in indexed pattern search, and offers significantly faster pattern processing compared to previous implementations, as well as reduced space requirements. We report a detailed experimental evaluation that demonstrates the relative advantages of the new approach, using the standard Pizza&Chili methodology and files, as well as applied use-cases derived from large-scale data compression, and from natural language processing. For large alphabets, the new structure gives rise to space requirements that are close to those of the most highly-compressed FM-Index variants, in conjunction with unparalleled Count throughput rates.

.

[100] Michael Axtmann and Peter Sanders. Robust Massively Parallel Sorting. **ALENEX 2** (Saray). January 17th, 2017, 16:50-17:10

We investigate distributed memory parallel sorting algorithms that scale to the largest available machines and are robust with respect to input size, duplicate keys, and distribution of the input elements. The main outcome is that four sorting algorithms cover the entire range of possible input sizes. For three algorithms we devise new low overhead mechanisms to make them robust with respect to duplicate keys and skewed input distributions. One of these, designed for medium sized inputs, is a new variant of quicksort with fast high-quality pivot selection. At the same time asymptotic analysis provides performance guarantees and guides the selection and configuration of the algorithms. We validate these hypotheses using extensive experiments on 7 algorithms, 10 input distributions, up to 262 144 cores, and varying input sizes over 9 orders of magnitude. For "difficult" input distributions, our algorithms are the only ones that work at all. For all but the largest input sizes, we are the first to perform experiments on such large machines at all and our algorithms significantly outperform the ones one would conventionally have considered.

.

[101] Juha Kärkkäinen, Dominik Kempa, Simon J. Puglisi and Bella Zhukova. Engineering External Memory Induced Suffix Sorting. **ALENEX 2** (Saray). January 17th, 2017, 17:15-17:35

Suffix sorting — determining the lexicographical order of all the suffixes of a string — is one of the most important problems in string processing. The resulting data structure is called the suffix array (SA) and underpins dozens of applications in bioinformatics, data compression, and information retrieval. When the size of the input string or the SA exceeds that of internal memory (RAM), an external memory (EM) suffix sorting algorithm must be used. The most scalable of these EM methods is due to Bingmann et al. (Proc. ALENEX 2013), and is essentially a careful disk-based implementation of the so-called induced sorting technique used by the fastest RAM suffix sorting algorithms. In this paper we show how to greatly improve the efficiency of induced suffix sorting in external memory via a non-trivial reorganization of the computation involved. Our experiments show this new approach to be twice as fast as state-of-the-art methods, while, just as significantly, using a third of the disk memory. We also demonstrate the efficacy of our implementation for handling strings on large alphabets (with many millions of distinct symbols), which is important, e.g., for applications in natural language processing and information retrieval, but unaddressed by previous EM suffix sorting implementations. Our implementation uses a (EM) radix heap data structure and, as a side result of independent interest, we introduce a new operation for radix heaps and other monotone priority queues called min-comp, which we believe to be useful for many other applications, including discrete event simulation and sweep line algorithms, even in internal memory.

.

[102] Andreas Poyias, Simon J. Puglisi and Rajeev Raman. Compact Dynamic Rewritable (CDRW) Arrays. **ALENEX 2** (Saray). January 17th, 2017, 17:40-18:00

In this paper we consider the problem of compactly representing a rewritable array of bit-strings. The operations supported are: create (N, k), which creates a new array of size N, where each entry is of size at most k bits and equal to 0; set (i, v), which sets $A[i]$ to v, provided that v is at most k bits long and get (i) which returns the value of $A[i]$. Our aim is to approach the minimum possible space bound of $S = \sum_{i=0}^{N-1} |A[i]|$, where $|A[i]| \geq 1$ is the length in bits of the number in $A[i]$, while simultaneously supporting operations in $O(1)$ time. We call such a data structure a *Compact Dynamic Rewriteable Array (CDRW)* array. On the word RAM model with word size w, for $n < 2^w$ and $k \leq w$, we give practical solutions based on *compact hashing* that achieve $O(1/\epsilon)$ expected time for get and set and use $(1 + \epsilon)S + O(N)$ bits, for any constant $\epsilon > 0$. Experimental evaluation of our (preliminary, only somewhat optimized) implementations shows excellent performance in terms of both space and time, particularly when heuristics are added to our base algorithms.

.

[103] T.-H. Hubert Chan, Zhiyi Huang, Shaofeng H.-C. Jiang, Ning Kang and Zhihao Gavin Tang. Online Submodular Maximization with Free Disposal: Randomization Beats 1/4 for Partition Matroids. **SODA 6A** (Las Arenas II-IV). January 17th, 2017, 16:25-16:45

We study the online submodular maximization problem with free disposal under a matroid constraint. Elements from some ground set arrive one by one in rounds, and the algorithm maintains a feasible set that is independent in the underlying matroid. In each round when a new element arrives, the algorithm may accept the new element into its feasible set and possibly remove elements from it, provided that the resulting set is still independent. The goal is to maximize the value of the final feasible set under some monotone submodular function, to which the algorithm has oracle access. For k-uniform matroids, we give a deterministic algorithm with competitive ratio at least 0.2959, and the ratio approaches $\frac{1}{\alpha_\infty} \approx 0.3178$ as k approaches infinity, improving the previous best ratio of 0.25 by Chakrabarti and Kale (IPCO 2014), Buchbinder et al. (SODA 2015) and Chekuri et al. (ICALP 2015). We also show that our algorithm is optimal among a class of deterministic monotone algorithms that accept a new arriving element only if the objective is strictly increased. Further, we prove that no deterministic monotone algorithm can be strictly better than 0.25-competitive even for partition matroids, the most modest generalization of k-uniform matroids, matching the competitive ratio by Chakrabarti and Kale (IPCO 2014) and Chekuri et al. (ICALP 2015). Interestingly, we show that randomized algorithms are strictly more powerful by giving a (non-monotone) randomized algorithm for partition matroids with ratio $\frac{1}{\alpha_\infty} \approx 0.3178$. Finally, our techniques can be extended to a more general problem that generalizes both the online submodular maximization problem and the online bipartite matching problem with free disposal. Using the techniques developed in this paper, we give constant-competitive algorithms for the submodular online bipartite matching problem.

.

[104] Matthias Englert and Harald Räcke. Reordering Buffers with Logarithmic Diameter Dependency for Trees. **SODA 6A** (Las Arenas II-IV). January 17th, 2017, 16:50-17:10

In the reordering buffer problem a sequence of items located in a metric space arrive online, and have to be processed by a single server moving within the metric space. At any point in time, the first k still unprocessed items from the sequence are available for processing and the server has to select one of these items and process it by visiting its location. The goal is to process all items while minimizing the total distance the server moves. Englert, Räcke, Westermann (STOC'07) gave a deterministic $O(D \cdot \log k)$-competitive online algorithm for weighted tree metrics with hop-diameter D. We improve the analysis of this algorithm and significantly improve the dependency on D. Specifically, we show that the algorithm is in fact $O(\log D + \log k)$-competitive. Our analysis is quite robust. Even when an optimal algorithm, to which we compare the online algorithm, is allowed to choose between the first $h > k$ unprocessed items, the online algorithm is still $O(h \cdot (\log D + \log h)/k)$-competitive. For $h = (1+\varepsilon) \cdot k$, with constant $\varepsilon > 0$, this is optimal. Our results also imply better competitive ratio for general metric spaces, improving the randomized $O(\log n \cdot \log^2 k)$ result for n-point metric spaces from STOC'07 to $O(\log n \cdot \log k)$.

.

[105] Niv Buchbinder, Moran Feldman, Joseph Naor and Ohad Talmon. O(depth)-Competitive Algorithm for Online Multi-level Aggregation. **SODA 6A** (Las Arenas II-IV). January 17th, 2017, 17:15-17:35

We consider a multi-level aggregation problem in a weighted rooted tree, studied recently by Bienkowski et al. in (Marcin Bienkowski, Martin Böhm, Jaroslaw Byrka, Marek Chrobak, Christoph Dürr, Lukáš Folwarczný, Lukasz Jez, Jirí Sgall, Nguyen Kim Thang, and Pavel Veselý,*Online algorithms for multi-level aggregation*, ESA 2016). In this problem requests arrive over time at the nodes of the tree, and each request specifies a deadline. A request is served by sending it to the root before its deadline at a cost equal to the weight of the path from the node in which it resides to the root. However, requests from different nodes can be aggregated, and served together, so as to save on cost. The cost of serving an aggregated set of requests is equal to the weight of the subtree spanning the nodes in which the requests reside. Thus, the problem is to find a competitive online aggregation algorithm that minimizes the total cost of the aggregated requests. This problem arises naturally in many scenarios, including multicasting, supply-chain management and sensor networks. It is also related to the well studied TCP-acknowledgement problem and the online joint replenishment problem. We present an online $O(D)$-competitive algorithm for the problem, where D is the depth, or number of levels, of the aggregation tree. This result improves upon the $D^2 2^D$-competitive algorithm obtained recently by Bienkowski et al. in in (Marcin Bienkowski, Martin Böhm, Jaroslaw Byrka, Marek Chrobak, Christoph Dürr, Lukáš Folwarczný, Lukasz Jez, Jirí Sgall, Nguyen Kim Thang, and Pavel Veselý,*Online algorithms for multi-level aggregation*, ESA 2016).

.

[106] Xi Chen, Sivakanth Gopi, Jieming Mao and Jon Schneider. Competitive analysis of the top-K ranking problem. **SODA 6A** (Las Arenas II-IV). January 17th, 2017, 17:40-18:00

Motivated by applications in recommender systems, web search, social choice and crowdsourcing, we consider the problem of identifying the set of top K items from noisy pairwise comparisons. In our setting, we are non-actively given r pairwise comparisons between each pair of n items, where each comparison has noise constrained by a very general noise model called the strong stochastic transitivity (SST) model. We analyze the competitive ratio of algorithms for the top-K problem. In particular, we present a linear time algorithm for the top-K problem which has a competitive ratio of $\tilde{O}(\sqrt{n})$; i.e. to solve any instance of top-K, our algorithm needs at most $\tilde{O}(\sqrt{n})$ times as many samples needed as the best possible algorithm for that instance (in contrast, all previous known algorithms for the top-K problem have competitive ratios of $\tilde{\Omega}(n)$ or worse). We further show that this is tight: any algorithm for the top-K problem has competitive ratio at least $\tilde{\Omega}(\sqrt{n})$.

.

[107] Vitaly Feldman, Cristóbal Guzmán and Santosh Vempala. Statistical Query Algorithms for Mean Vector Estimation and Stochastic Convex Optimization. **SODA 6B** (Las Arenas I). January 17th, 2017, 16:25-16:45

Stochastic convex optimization, where the objective is the expectation of a random convex function, is an important and widely used method with numerous applications in machine learning, statistics, operations research and other areas. We study the complexity of stochastic convex optimization given only *statistical query* (SQ) access to the objective function. We show that well-known and popular first-order iterative methods can be implemented using only statistical queries. For many cases of interest we derive nearly matching upper and lower bounds on the estimation (sample) complexity including linear optimization in the most general setting. We then present several consequences for machine learning, differential privacy and proving concrete lower bounds on the power of convex optimization based methods. The key ingredient of our work is SQ algorithms and lower bounds for estimating the mean vector of a distribution over vectors supported on a convex body in $\mathbb{R}^d$. This natural problem has not been previously studied and we show that our solutions can be used to get substantially improved SQ versions of Perceptron and other online algorithms for learning halfspaces.

.

[108] Jayadev Acharya, Ilias Diakonikolas, Jerry Li and Ludwig Schmidt. Sample Optimal Density Estimation in Nearly-Linear Time. **SODA 6B** (Las Arenas I). January 17th, 2017, 16:50-17:10

We design a new, fast algorithm for agnostically learning univariate probability distributions whose densities are well-approximated by piecewise polynomial functions. Let f be the density function of an arbitrary univariate distribution, and suppose that f is OPT-close in L_1-distance to an unknown piecewise polynomial function with t interval pieces and degree d. For any $\gamma > 0$, our algorithm draws $n = \widetilde{O}_\gamma(t(d+1)/\epsilon^2)$ samples from f, runs in time $\widetilde{O}(n)$, and with probability at least $9/10$ outputs an $O_\gamma(t)$-piecewise degree-d hypothesis h that is $(3+\gamma)\cdot\mathrm{OPT}+\epsilon$ close to f. Our approximation factor almost matches the best known information–theoretic (but computationally inefficient) upper bound of 3. Our general algorithm yields (nearly) sample-optimal and *nearly-linear time* estimators for a wide range of structured distribution families over both continuous and discrete domains in a unified way. For most of our applications, these are the *first* sample-optimal and nearly-linear time estimators in the literature. As a consequence, our work resolves the sample and computational complexities of a broad class of inference tasks via a single "meta-algorithm". Moreover, we demonstrate that our algorithm performs well in experiments. Our algorithm consists of three levels: (i) At the top level, we employ an iterative greedy algorithm for finding a good partition of the real line into the pieces of a piecewise polynomial. (ii) For each piece, we show that the sub-problem of finding a good polynomial fit on the current interval can be solved efficiently with a separation oracle method. (iii) We reduce the task of finding a separating hyperplane to a combinatorial problem and design a nearly-linear algorithm for this problem. Combining these three procedures gives a density estimation algorithm with the claimed guarantees.

.

[109] Dmitry Chistikov, Stefan Kiefer, Ines Marušić, Mahsa Shirmohammadi and James Worrell. On Rationality of Nonnegative Matrix Factorization. **SODA 6B** (Las Arenas I). January 17th, 2017, 17:15-17:35

Nonnegative matrix factorization (NMF) is the problem of decomposing a given nonnegative $n \times m$ matrix M into a product of a nonnegative $n \times d$ matrix W and a nonnegative $d \times m$ matrix H. NMF has a wide variety of applications, including bioinformatics, chemometrics, communication complexity, machine learning, polyhedral combinatorics, among many others. A longstanding open question, posed by Cohen and Rothblum in 1993, is whether every rational matrix M has an NMF with minimal d whose factors W and H are also rational. We answer this question negatively, by exhibiting a matrix M for which W and H require irrational entries. As an application of this result, we show that state minimization of labeled Markov chains can require the introduction of irrational transition probabilities. We complement these irrationality results with an NP-complete version of NMF for which rational numbers suffice.

.

[110] Mark Bun, Thomas Steinke and Jonathan Ullman. Make Up Your Mind: The Price of Online Queries in Differential Privacy. **SODA 6B** (Las Arenas I). January 17th, 2017, 17:40-18:00

We consider the problem of answering queries about a sensitive dataset subject to differential privacy. The queries may be chosen adversarially from a larger set Q of allowable queries in one of three ways, which we list in order from easiest to hardest to answer:

- *Offline:* The queries are chosen all at once and the differentially private mechanism answers the queries in a single batch.
- *Online:* The queries are chosen all at once, but the mechanism only receives the queries in a streaming fashion and must answer each query before seeing the next query.
- *Adaptive:* The queries are chosen one at a time and the mechanism must answer each query before the next query is chosen. In particular, each query may depend on the answers given to previous queries.

Many differentially private mechanisms are just as efficient in the adaptive model as they are in the offline model. Meanwhile, most lower bounds for differential privacy hold in the offline setting. This suggests that the three models may be equivalent. We prove that these models are all, in fact, distinct. Specifically, we show that there is a family of statistical queries such that exponentially more queries from this family can be answered in the offline model than in the online model. We also exhibit a family of search queries such that exponentially more queries from this family can be answered in the online model than in the adaptive model. We also investigate whether such separations might hold for simple queries like threshold queries over the real line.

.

[111] Mark Braverman, Young Kun Ko, Aviad Rubinstein and Omri Weinstein. ETH Hardness for Densest-k-Subgraph with Perfect Completeness. **SODA 6C** (Nelva). January 17th, 2017, 16:25-16:45

We show that, assuming the (deterministic) Exponential Time Hypothesis, distinguishing between a graph with an induced k-clique and a graph in which all k-subgraphs have density at most $1-\epsilon$, requires $n^{\tilde{\Omega}(\log n)}$ time. Our result essentially matches the quasi-polynomial algorithms of Feige and Seltser [FS97] and Barman [Bar14] for this problem, and is the first one to rule out an additive PTAS for Densest k-Subgraph. We further strengthen this result by showing that our lower bound continues to hold when, in the soundness case, even subgraphs smaller by a near-polynomial factor ($k' = k \cdot 2^{-\tilde{\Omega}(\log n)}$) are assumed to be at most $(1-\epsilon)$-dense. Our reduction is inspired by recent applications of the "birthday repetition" technique [AIM14, BKW15]. Our analysis relies on information theoretical machinery and is similar in spirit to analyzing a parallel repetition of two-prover games in which the provers may choose to answer some challenges multiple times, while completely ignoring other challenges.

.

[112] Frédéric Meunier, Wolfgang Mulzer, Pauline Sarrabezolles and Yannik Stein. The Rainbow at the End of the Line - A PPAD Formulation of the Colorful Carathéodory Theorem with Applications. **SODA 6C** (Nelva). January 17th, 2017, 16:50-17:10

Let $C_1, ..., C_{d+1}$ be $d+1$ point sets in $\mathbb{R}^d$, each containing the origin in its convex hull. A subset C of $\bigcup_{i=1}^{d+1} C_i$ is called a colorful choice (or rainbow) for $C_1, \ldots, C_{d+1}$, if it contains exactly one point from each set C_i. The colorful Caratheodory theorem states that there always exists a colorful choice for $C_1, \ldots, C_{d+1}$ that has the origin in its convex hull. This theorem is very general and can be used to prove several other existence theorems in high-dimensional discrete geometry, such as the centerpoint theorem or Tverberg's theorem. The colorful Caratheodory problem (CCP) is the computational problem of finding such a colorful choice. Despite several efforts in the past, the computational complexity of CCP in arbitrary dimension is still open. We show that CCP lies in the intersection of the complexity classes PPAD and PLS. This makes it one of the few geometric problems in PPAD and PLS that are not known to be solvable in polynomial time. Moreover, it implies that the problem of computing centerpoints, computing Tverberg partitions, and computing points with large simplicial depth is contained in PPAD $\cap$ PLS. This is the first nontrivial upper bound on the complexity of these problems. Finally, we show that our PPAD formulation leads to a polynomial-time algorithm for a special case of CCP in which we have only two color classes C_1 and C_2 in d dimensions, each with the origin in its convex hull, and we would like to find a set with half the points from each color class that contains the origin in its convex hull.

.

[113] Pavel Hubáček and Eylon Yogev. Hardness of Continuous Local Search: Query Complexity and Cryptographic Lower Bounds. **SODA 6C** (Nelva). January 17th, 2017, 17:15-17:35

Local search proved to be an extremely useful tool when facing hard optimization problems (e.g., via the simplex algorithm, simulated annealing, or genetic algorithms). Although powerful, it has its limitations: there are functions for which exponentially many queries are needed to find a local optimum. In many contexts the optimization problem is defined by a continuous function, which might offer an advantage when performing the local search. This leads us to study the following natural question: How hard is *continuous* local search? The computational complexity of such search problems is captured by the complexity class $\mathbf{CLS}$ (Daskalakis and Papadimitriou SODA'11) which is contained in the intersection of $\mathbf{PLS}$ and $\mathbf{PPAD}$, two important subclasses of $\mathbf{TFNP}$ (the class of $\mathbf{NP}$ search problems with a guaranteed solution). In this work, we show the first hardness results for $\mathbf{CLS}$ (the smallest non-trivial class among the currently defined subclasses of $\mathbf{TFNP}$). Our hardness results are in terms of black-box (where only oracle access to the function is given) and white-box (where the function is represented succinctly by a circuit). In the black-box case, we show instances for which any (computationally unbounded) randomized algorithm must perform exponentially many queries in order to find a local optimum. In the white-box case, we show hardness for computationally bounded algorithms under cryptographic assumptions. Our results demonstrate a strong conceptual barrier precluding design of efficient algorithms for solving local search problems even over continuous domains. As our main technical contribution we introduce a new total search problem which we call End-Of-Metered-Line. The special structure of End-Of-Metered-Line enables us to: (1) show that it is contained in $\mathbf{CLS}$, and (2) prove hardness for it both in the black-box and the white-box setting.

.

[114] Daniel Průša and Tomáš Werner. LP Relaxations of Some NP-Hard Problems Are as Hard as Any LP. **SODA 6C** (Nelva). January 17th, 2017, 17:40-18:00

We show that solving linear programming (LP) relaxations of many classical NP-hard combinatorial optimization problems is as hard as solving the general LP problem. Precisely, the general LP can be reduced in linear time to the LP relaxation of each of these problems. This result poses a fundamental limitation for designing efficient algorithms to solve the LP relaxations, because finding such an algorithm might improve the complexity of best known algorithms for the general LP. Besides linear-time reductions, we show that the LP relaxations of the considered problems are P-complete under log-space reduction, therefore also hard to parallelize.

.

[115] Johannes Fischer, Florian Kurpicz and Peter Sanders. Engineering a Distributed Full-Text Index. **ALENEX 3** (Las Arenas Foyer). January 18th, 2017, 09:20-09:40

We present a distributed full-text index for big data applications in a distributed environment. Our index can answer different types of pattern matching queries (existential, counting and enumeration). We perform experiments on inputs up to 100 GiB using up to 512 processors, and compare our index with the distributed suffix array by Arroyuelo et al. [Parall. Comput. 40(9): 471–495, 2014]. The result is that our index answers counting queries up to 5.5 times faster than the distributed suffix array, while using about the same space. We also provide a succinct variant of our index that uses only one third of the memory compared with our non-succinct variant, at the expense of only 20% slower query times.

.

[116] Julian Labeit and Simon Gog. Elias-Fano meets Single-Term Top-k Document Retrieval. **ALENEX 3** (Saray). January 18th, 2017, 09:45-10:05

A fundamental problem in Information Retrieval is to determine the k most relevant documents of a collection for a given query word or phrase P. In a recent result, Navarro and Nekrich [SODA 2012] showed that this problem can be solved in optimal time complexity of $O(|P| + k)$ with a precomputed linear-space index. The size of this optimal-time index was estimated to be 80 times the collection size, rendering it not to be practical. In subsequent work, Navarro and Konow [DCC 2013] and Gog and Navarro [ALENEX 2015] created a practical version with slightly worse query time guarantees but reduced the space to $2.5 - 3$ times the collection size. The index is conceptually simple and is divided in five components. In this paper we show how the $n \log N$ bits required by the usually largest component – the so called *repetition array* – can be reduced to $n \log \log n + O(n)$, where n is the size of the collection and N the number of documents. As the overall query time complexity matches the one of the old index, we achieve a theoretically superior time-space trade-off. We explore the practical properties of the improved index in a detailed experimental study and compare to the previously established baseline. Index sizes are now between $1.5 - 2$ times the collection size while query speed is comparable to the larger indexes. We also show that the new approach automatically adapts to highly repetitive text collections, which are for instance produced by version control systems.

.

[117] Preethi Lahoti, Patrick K. Nicholson and Bilyana Taneva. Efficient Set Intersection Counting Algorithm for Text Similarity Measures. **ALENEX 3** (Saray). January 18th, 2017, 10:10-10:30

The problem of set intersection counting appears as a subroutine in many techniques used in natural language processing, in which similarity is often measured as a function of document cooccurence counts between pairs of noun phrases or entities. Such techniques include clustering of text phrases and named entities, topic labeling, entity disambiguation, sentiment analysis, and search for synonyms. These techniques can have real-time constraints that require very fast computation of thousands of set intersection counting queries with little space overhead and minimal error. On one hand, while sketching techniques for approximating intersection counting exist and have very fast query time, many have issues with accuracy, especially for lists that have low Jaccard similarity. On the other hand, space-efficient computation of *exact* intersection sizes is particularly challenging in real-time. In this paper, we show how an efficient space-time trade-off can be achieved for set intersection counting, by combining state-of-the-art algorithms with precomputation and judicious use of compression. In addition, we show that the performance can be further improved by combining the best aspects of these algorithms. We present experimental evidence that real-time computation of exact intersection sizes is feasible with low memory overhead: we improve the mean query time of baseline approaches by over a factor of 100 using a data structure that takes merely twice the size of an inverted index. Overall, in our experiments, we achieve running times within the same order of magnitude as well-known approximation techniques.

.

[118] Bart M.P. Jansen and Marcin Pilipczuk. Approximation and Kernelization for Chordal Vertex Deletion. **SODA 7A** (Las Arenas II-IV). January 18th, 2017, 09:20-09:40

The Chordal Vertex Deletion (ChVD) problem asks to delete a minimum number of vertices from an input graph to obtain a chordal graph. In this paper we develop a polynomial kernel for ChVD under the parameterization by the solution size. Using a new Erdős-Pósa type packing/covering duality for holes in nearly-chordal graphs, we present a polynomial-time algorithm that reduces any instance (G, k) of ChVD to an equivalent instance with poly(k) vertices. The existence of a polynomial kernel answers an open problem of Marx from 2006 [Proceedings of the 32nd International Workshop on Graph-Theoretic Concepts in Computer Science, LNCS 4271, 37–48, 2006]. To obtain the kernelization, we develop the first poly(OPT)-approximation algorithm for ChVD, which is of independent interest. In polynomial time, it either decides that G has no chordal deletion set of size k, or outputs a solution of size $O(k^4 \log^2 k)$.

.

[119] Akanksha Agrawal, Daniel Lokshtanov, Pranabendu Misra, Saket Saurabh and Meirav Zehavi. Feedback Vertex Set Inspired Kernel for Chordal Vertex Deletion. **SODA 7A** (Las Arenas). January 18th, 2017, 09:20-09:40

Given a graph G and a parameter k, the CHORDAL VERTEX DELETION (CVD) problem asks whether there exists a subset $U \subseteq V(G)$ of size at most k that hits all induced cycles of size at least 4. The existence of a polynomial kernel for CVD was a well-known open problem in the field of Parameterized Complexity. Recently, Jansen and Pilipczuk resolved this question affirmatively by designing a polynomial kernel for CVD of size $\mathcal{O}(k^{161} \log^{58} k)$, and asked whether one can design a kernel of size $\mathcal{O}(k^{10})$. While we do not completely resolve this question, we design a significantly smaller kernel of size $\mathcal{O}(k^{25} log^{14} k)$, inspired by the $\mathcal{O}(k^2)$-size kernel for FEEDBACK VERTEX SET. To obtain this result, we first design an $\mathcal{O}(\text{opt} \cdot \log^2 n)$-factor approximation algorithm for CVD, which is central to our kernelization procedure. Thus, we improve upon both the kernelization algorithm and the approximation algorithm of Jansen and Pilipczuk. Next, we introduce the notion of the independence degree of a vertex, which is our main conceptual contribution. We believe that this notion could be useful in designing kernels for other problems.

.

[120] Fedor V. Fomin, Daniel Lokshtanov, Michał Pilipczuk, Saket Saurabh and Marcin Wrochna. Fully polynomial-time parameterized computations for graphs and matrices of low treewidth. **SODA 7A** (Las Arenas II-IV). January 18th, 2017, 09:45-10:05

We investigate the complexity of several fundamental polynomial-time solvable problems on graphs and on matrices, when the given instance has low treewidth; in the case of matrices, we consider the treewidth of the graph formed by non-zero entries. In each of the considered cases, the best known algorithms working on general graphs run in polynomial, but far from linear, time. Thus, our goal is to construct algorithms with running time of the form poly$(k) \cdot n$ or poly$(k) \cdot n \log n$, where k is the width of the tree decomposition given on the input. Such procedures would outperform the best known algorithms for the considered problems already for moderate values of the treewidth, like $O(n^{1/c})$ for some small constant c. Our results include:

- an algorithm for computing the determinant and the rank of an $n \times n$ matrix using $O(k^3 \cdot n)$ time and arithmetic operations;
- an algorithm for solving a system of linear equations using $O(k^3 \cdot n)$ time and arithmetic operations;
- an $O(k^3 \cdot n \log n)$-time randomized algorithm for finding the cardinality of a maximum matching in a graph;
- an $O(k^4 \cdot n \log^2 n)$-time randomized algorithm for constructing a maximum matching in a graph;
- an $O(k^2 \cdot n \log n)$-time algorithm for finding a maximum vertex flow in a directed graph.

Moreover, we provide an approximation algorithm for treewidth with time complexity suited to the running times as above. Namely, the algorithm, when given a graph G and integer k, runs in time $O(k^7 \cdot n \log n)$ and either correctly reports that the treewidth of G is larger than k, or constructs a tree decomposition of G of width $O(k^2)$. The above results stand in contrast with the recent work of (A. Abboud, V. V. Williams, and J. R. Wang, *Approximation and fixed parameter subquadratic algorithms for radius and diameter in sparse graphs*, In SODA 2016, 377–39, SIAM, 2016), which shows that the existence of algorithms with similar running times is unlikely for the problems of finding the diameter and the radius of a graph of low treewidth.

.

[121] Fedor V. Fomin, Petr A. Golovach, Daniel Lokshtanov and Saket Saurabh. Spanning Circuits in Regular Matroids. **SODA 7A** (Las Arenas II-IV). January 18th, 2017, 10:10-10:30

We consider the fundamental Matroid Theory problem of finding a circuit in a matroid spanning a set T of given terminal elements. For graphic matroids this corresponds to the problem of finding a simple cycle passing through a set of given terminal edges in a graph. The algorithmic study of the problem on regular matroids, a superclass of graphic matroids, was initiated by Gavenčiak, Král', and Oum [ICALP'12], who proved that the case of the problem with $|T| = 2$ is fixed-parameter tractable (FPT) when parameterized by the length of the circuit. We extend the result of Gavenčiak, Král', and Oum by showing that for regular matroids

- the MINIMUM SPANNING CIRCUIT problem, deciding whether there is a circuit with at most ℓ elements containing T, is FPT parameterized by $k = \ell - |T|$;
- the SPANNING CIRCUIT problem, deciding whether there is a circuit containing T, is FPT parameterized by $|T|$.

We note that extending our algorithmic findings to binary matroids, a superclass of regular matroids, is highly unlikely: MINIMUM SPANNING CIRCUIT parameterized by ℓ is W[1]-hard on binary matroids even when $|T| = 1$. We also show a limit to how far our results can be strengthened by considering a smaller parameter. More precisely, we prove that MINIMUM SPANNING CIRCUIT parameterized by $|T|$ is W[1]-hard even on cographic matroids, a proper subclass of regular matroids.

.

[122] Haris Angelidakis, Yury Makarychev and Vsevolod Oparin. Algorithmic and Hardness Results for the Hub Labeling Problem. **SODA 7B** (Las Arenas I). January 18th, 2017, 09:20-09:40

There has been significant success in designing highly efficient algorithms for distance and shortest-path queries in recent years; many of the state-of-the-art algorithms use the hub labeling framework. In this paper, we study the approximability of the Hub Labeling problem. We prove a hardness of $\Omega(\log n)$ for Hub Labeling, matching known approximation guarantees. The hardness result applies to graphs that have multiple shortest paths between some pairs of vertices. No hardness of approximation results were known previously. Then, we focus on graphs that have a unique shortest path between each pair of vertices. This is a very natural family of graphs, and much research on the Hub Labeling problem has studied such graphs. We give an $O(\log D)$ approximation algorithm for graphs of shortest-path diameter D with unique shortest paths. In particular, we get an $O(\log \log n)$ approximation for graphs of polylogarithmic diameter, while previously known algorithms gave an $O(\log n)$ approximation. Finally, we present a polynomial-time approximation scheme (PTAS) and quasi-polynomial-time algorithms for Hub Labeling on trees; additionally, we analyze a simple combinatorial heuristic for Hub Labeling on trees, proposed by Peleg in 2000. We show that this heuristic gives an approximation factor of 2.

.

[123] Adrian Kosowski and Laurent Viennot. Beyond Highway Dimension: Small Distance Labels Using Tree Skeletons. **SODA 7B** (Las Arenas I). January 18th, 2017, 09:45-10:05

The goal of a hub-based distance labeling scheme for a network $G = (V, E)$ is to assign a small subset $S(u) \subseteq S$ to each node $u \in V$, in such a way that for any pair of nodes u, v, the intersection of hub sets $S(u) \cap S(v)$ contains a node on the shortest uv-path. The existence of small hub sets, and consequently efficient shortest path processing algorithms, for road networks is an empirical observation. A theoretical explanation for this phenomenon was proposed by Abraham et al. (SODA 2010) through a network parameter they called *highway dimension* h, which captures the size of a hitting set for a collection of shortest paths of length at least r intersecting a given ball of radius $2r$. In this work, we revisit this explanation, introducing a more tractable (and directly comparable) parameter based solely on the structure of shortest-path spanning trees, which we call *skeleton dimension* k. We show that skeleton dimension admits an intuitive definition for both directed and undirected graphs, provides a way of computing labels more efficiently than by using highway dimension, and leads to comparable or stronger theoretical bounds on hub set size.

.

[124] Shiri Chechik, Sarel Cohen, Amos Fiat and Haim Kaplan. $(1 + \epsilon)$-Approximate f-Sensitive Distance Oracles. **SODA 7B** (Las Arenas I). January 18th, 2017, 10:10-10:30

An f-Sensitive Distance Oracle with stretch α preprocesses a graph $G(V, E)$ and produces a small data structure that is used to answer subsequent queries. A query is a triple consisting of a set $F \subset E$ of at most f edges, and vertices s and t. The oracle answers a query (F, s, t) by returning a value $\tilde{d}$ which is equal to the length of some path between s and t in the graph $G \setminus F$ (the graph obtained from G by discarding all edges in F). Moreover, $\tilde{d}$ is at most α times the length of the shortest path between s and t in $G \setminus F$. The oracle can also construct a path between s and t in $G \setminus F$ of length $\tilde{d}$. To the best of our knowledge we give the first nontrivial f-sensitive distance oracle with fast query time and small stretch capable of handling multiple edge failures. Specifically, for any $f = o(\frac{\log n}{\log \log n})$ and a fixed $\epsilon > 0$ our oracle answers queries (F, s, t) in time $\widetilde{O}(1)$ with $(1 + \epsilon)$ stretch using a data structure of size $n^{2+o(1)}$. For comparison, the naïve alternative requires $m^f n^2$ space for sublinear query time.

.

[125] Alan Frieze and Tony Johansson. On the insertion time of random walk cuckoo hashing. **SODA 7C** (Nelva). January 18th, 2017, 09:20-09:40

Cuckoo Hashing is a hashing scheme invented by Pagh and Rodler. It uses $d \geq 2$ distinct hash functions to insert items into the hash table. It has been an open question for some time as to the expected time for Random Walk Insertion to add items. We show that if the number of hash functions $d = O(1)$ is sufficiently large, then the expected insertion time is $O(1)$ per item.

.

[126] Michael A. Bender, Jeremy T. Fineman, Seth Gilbert, Tsvi Kopelowitz and Pablo Montes. File Maintenance: When in Doubt, Change the Layout! **SODA 7C** (Nelva). January 18th, 2017, 09:45-10:05

This paper gives a new deamortized solution to the sequential-file-maintenance problem. The data structure uses several new tools for solving this historically complicated problem. These tools include an unbalanced ternary-tree layout embedded in the sparse table, one-way rebalancing, and extra structural properties to keep interaction among rebalances to a minimum.

.

[127] Peyman Afshani, Michael A. Bender, Martín Farach-Colton, Jeremy T. Fineman, Mayank Goswami and Meng-Tsung Tsai. Cross-Referenced Dictionaries and the Limits of Write Optimization. **SODA 7C** (Nelva). January 18th, 2017, 10:10-10:30

Dictionaries remain the most well studied class of data structures. A dictionary supports insertions, deletions, membership queries, and usually successor, predecessor, and extract-min. In a RAM, all such operations take $O(\log N)$ time on N elements. Dictionaries are often cross-referenced as follows. Consider a set of tuples $\{\langle a_i, b_i, c_i \ldots \rangle\}$. A database might include more than one dictionary on such a set, for example, one indexed on the a's, another on the b's, and so on. Once again, in a RAM, inserting into a set of L cross-referenced dictionaries takes $O(L \log N)$ time, as does deleting. The situation is more interesting in external memory. On a Disk Access Machine (DAM), B-trees achieve $O(\log_B N)$ I/Os for insertions and deletions on a single dictionary and K-element range queries take optimal $O(\log_B N + K/B)$ I/Os. These bounds are also achievable by a B-tree on cross-referenced dictionaries, with a slowdown of an L factor on insertion and deletions. In recent years, both the theory and practice of external-memory dictionaries has been revolutionized by ***write-optimization*** techniques. A dictionary is write optimized if it is close to a B-tree for query time while beating B-trees on insertions. The best (and optimal) dictionaries achieve a substantially improved insertion and deletion cost of $O(\frac{\log_{1+B^\epsilon} N}{B^{1-\epsilon}})$, $0 \leq \epsilon \leq 1$, amortized I/Os on a single dictionary while maintaining optimal $O(\log_{1+B^\epsilon} N + K/B)$-I/O range queries. Although write optimization still helps for insertions into cross-referenced dictionaries, its value for deletions would seem to be greatly reduced. A deletion into a cross-referenced dictionary only specifies a key a. It seems to be necessary to look up the associated values $b, c \ldots$ in order to delete them from the other dictionaries. This takes $\Omega(\log_B N)$ I/Os, well above the per-dictionary write-optimization budget of $O(\frac{\log_{1+B^\epsilon} N}{B^{1-\epsilon}})$ I/Os. So the total deletion cost is $O(\log_B N + L\frac{\log_{1+B^\epsilon} N}{B^{1-\epsilon}})$ I/Os. In short, for deletions, write optimization offers an advantage over B-trees in that L multiplies a lower order term, but when $L = 2$, write optimization seems to offer no asymptotic advantage over B-trees. That is, no known query-optimal solution for pairs of cross-referenced dictionaries seem to beat B-trees for deletions. In this paper, we show a lower bound establishing that a pair of cross-referenced dictionaries that are optimal for range queries and that supports deletions cannot match the write optimization bound available to insert-only dictionaries. This result thus establishes a limit to the applicability of write-optimization techniques on which many new databases and file systems are based.

.

[128] Jan Dean Catarata, Scott Corbett, Harry Stern, Mario Szegedy, Tomas Vyskocil and Zheng Zhang. The Moser-Tardos Resample algorithm: Where is the limit? (an experimental inquiry). **ALENEX 4** (Saray). January 18th, 2017, 10:55-11:15

The celebrated Lovász Local Lemma (LLL) guarantees that locally sparse systems always have solutions. The Moser-Tardos RESAMPLE algorithm does not only find such a solution in linear time, but its beautiful analysis has greatly enhanced LLL related research. Nevertheless two major questions remain open.

1. How far *beyond* Lovász's condition can we expect that RESAMPLE still performs in polynomial (linear) expected running time?
2. In RESAMPLE we have a choice between different constraint-selection strategies. How much does this choice matter?

To state the first question correctly is a challenge already. For a solvable fixed instance RESAMPLE always comes up with a solution, but the catch is that the number of steps may be very large. We have therefore looked at parameterized instance families and tried to identify phase transitions in terms of these parameters. Perhaps the biggest lesson we have learned is that if we want to see phase transition thresholds, i.e. identify parameter values where RESAMPLE "stops working," we need to understand what happens when RESAMPLE *does not work*. We have noticed that in this case the algorithm settles at a *metastable equilibrium* (at least for the homogenous instances we have considered), a phenomenon mostly studied for physical systems. Concerning the policies for picking the violated constraints (such as first violated, random violated, recursive fix, etc.), in the context of the grid-coloring problem the methods worked exactly for the same parameter range the number of resample steps differed by no more than 20 percent. All results are experimental, although we discuss a possible reason behind some phenomena.

.

[129] Christopher Musco, Maxim Sviridenko and Justin Thaler. Determining Tournament Payout Structures for Daily Fantasy Sports. **ALENEX 4** (Saray). January 18th, 2017, 11:20-11:40

With an exploding global market and the recent introduction of online cash prize tournaments, fantasy sports contests are quickly becoming a central part of the social gaming and sports industries. For sports fans and online media companies, fantasy sports contests are an opportunity for large financial gains. However, they present a host of technical challenges that arise from the complexities involved in running a web-scale, prize driven fantasy sports platform. We initiate the study of these challenges by examining one concrete problem in particular: how to algorithmically generate contest payout structures that are 1) economically motivating and appealing to contestants and 2) reasonably structured and succinctly representable. We formalize this problem and present a general two-staged approach for producing satisfying payout structures given constraints on contest size, entry fee, prize bucketing, etc. We then propose and evaluate several potential algorithms for solving the payout problem efficiently, including methods based on dynamic programming, integer programming, and heuristic techniques. Experimental results show that a carefully designed heuristic scales very well, even to contests with over 100,000 prize winners. Our approach extends beyond fantasy sports – it is suitable for generating engaging payout structures for any contest with a large number of entrants and a large number of prize winners, including other massive online games, poker tournaments, and real-life sports tournaments.

.

[130] Stefan Funke, Thomas Mendel, Alexander Miller, Sabine Storandt and Maria Wiebe. Map Simplification with Topology Constraints: Exactly and in Practice. **ALENEX 4** (Saray). January 18th, 2017, 11:45-12:05

We consider the classical line simplification problem subject to a given error bound ϵ but with additional topology constraints as they arise for example in the map rendering domain. While theoretically inapproximability has been proven for these problem variants, we show that in practice one can solve medium sized instances optimally using an integer linear programming approach and larger instances using an heuristic approach which for medium-sized real-world instances yields close-to-optimal results. Our approaches are evaluated on data sets which are synthetically generated, stem from the OpenStreetMap project and the recent GISCup competition.

.

[131] Saverio Basso and Alberto Ceselli. Asynchronous Column Generation. **ALENEX 4** (Saray). January 18th, 2017, 12:10-12:30

In this paper we face a very fundamental problem in Operations Research: to find good dual bounds to generic mixed integer mathematical programs (MIPs) as quickly as possible. In particular, we focus on the scenario where large scale data needs to be considered, multi-core CPU architectures are available, and massive parallelism can be exploited by means of decomposition methods. We consider column generation techniques to solve extended formulations obtained by means of Dantzig-Wolfe decomposition for MIPs. We propose a concurrent algorithm, that relaxes the synchronized behavior of classical column generation. Our approach relies on simple data structures and efficient synchronization, still providing the same global convergence properties of classical sequential column generation methods. We present and discuss the results of an extensive experimental campaign, comparing our concurrent algorithm to both a naive parallelization of column generation and the cutting planes algorithm implemented in state-of-the-art commercial optimization packages, considering large scale datasets of a hard packing problem from the literature as representative benchmark. Our approach turns out to be on average one order of magnitude faster than competitors, attaining almost linear speedups as the number of available CPU cores increases.

.

[132] Ilan Newman, Yuri Rabinovich, Deepak Rajendraprasad and Christian Sohler. Testing for Forbidden Order Patterns in an Array. **SODA 8A** (Las Arenas II-IV). January 18th, 2017, 10:55-11:15

In this paper, we study testing of sequence properties that are defined by forbidden order patterns. A sequence $f : \{1, \ldots, n\} \to \mathbb{R}$ of length n contains a pattern $\pi \in \mathfrak{S}_k$ ($\mathfrak{S}_k$ is the group of permutations of k elements), iff there are indices $i_1 < i_2 < \cdots < i_k$, such that $f(i_x) > f(i_y)$ whenever $\pi(x) > \pi(y)$. If f does not contain π, we say f is π-free. For example, for $\pi = (2, 1)$, the property of being π-free is equivalent to being non-decreasing, i.e. monotone. The property of being $(k, k-1, \ldots, 1)$-free is equivalent to the property of having a partition into at most $k-1$ non-decreasing subsequences.

Let $\pi \in \mathfrak{S}_k$, k constant, be a (forbidden) pattern. Assuming f is stored in an array, we consider the property testing problem of distinguishing the case that f is π-free from the case that f differs in more than ϵn places from any π-free sequence. We show the following results: There is a clear dichotomy between the monotone patterns and the non-monotone ones:

- For monotone patterns of length k, i.e., $(k, k-1, \ldots, 1)$ and $(1, 2, \ldots, k)$, we design *non-adaptive* one-sided error ϵ-tests of $(\epsilon^{-1} \log n)^{O(k^2)}$ query complexity.
- For non-monotone patterns, we show that for any size-k non-monotone π, *any non-adaptive* one-sided error ϵ-test requires at least $\Omega(\sqrt{n})$ queries. This general lower bound can be further strengthened for specific non-monotone k-length patterns to $\Omega(n^{1-2/(k+1)})$. On the other hand, there always exists a *non-adaptive* one-sided error ϵ-test for $\pi \in \mathfrak{S}_k$ with $O(\epsilon^{-1/k} n^{1-1/k})$ query complexity. Again, this general upper bound can be further strengthened for specific non-monotone patterns. E.g., for $\pi = (1, 3, 2)$, we describe an ϵ-test with (almost tight) query complexity of $\widetilde{O}(\sqrt{n})$.

Finally, we show that adaptivity can make a big difference in testing non-monotone patterns, and develop an *adaptive* algorithm that for any $\pi \in \mathfrak{S}_3$, tests π-freeness by making $(\epsilon^{-1} \log n)^{O(1)}$ queries. For all algorithms presented here, the running times are linear in their query complexity.

.

[133] Aram W. Harrow, Cedric Yen-Yu Lin and Ashley Montanaro. Sequential measurements, disturbance and property testing. **SODA 8A** (Las Arenas II-IV). January 18th, 2017, 11:20-11:40

We describe two procedures which, given access to one copy of a quantum state and a sequence of two-outcome measurements, can distinguish between the case that at least one of the measurements accepts the state with high probability, and the case that all of the measurements have low probability of acceptance. The measurements cannot simply be tried in sequence, because early measurements may disturb the state being tested. One procedure is based on a variant of Marriott-Watrous amplification. The other procedure is based on the use of a test for this disturbance, which is applied with low probability. We find a number of applications:

- Quantum query complexity separations in the property testing model for testing isomorphism of functions under group actions. We give quantum algorithms for testing isomorphism, linear isomorphism and affine isomorphism of boolean functions which use exponentially fewer queries than is possible classically, and a quantum algorithm for testing graph isomorphism which uses polynomially fewer queries than the best algorithm known.
- Testing properties of quantum states and operations. We show that any finite property of quantum states can be tested using a number of copies of the state which is logarithmic in the size of the property, and give a test for genuine multipartite entanglement of states of n qubits that uses $O(n)$ copies of the state. We also show that equivalence of two unitary operations under conjugation by a unitary picked from a fixed set can be tested efficiently. This is a natural quantum generalisation of testing isomorphism of boolean functions.
- Correcting an error in a result of Aaronson on de-Merlinizing quantum protocols. This result claimed that, in any one-way quantum communication protocol where two parties are assisted by an all-powerful but untrusted third party, the third party can be removed with only a modest increase in the communication cost. We give a corrected proof of a key technical lemma required for Aaronson's result.

.

[134] László Miklós Lovász and Jacob Fox. A tight bound for Green's arithmetic triangle removal lemma in vector spaces. **SODA 8A** (Las Arenas II-IV). January 18th, 2017, 11:45-12:05

Let p be a fixed prime. A triangle in $\mathbb{F}_p^n$ is an ordered triple (x,y,z) of points satisfying $x+y+z=0$. Let $N=p^n=|\mathbb{F}_p^n|$. Green proved an arithmetic triangle removal lemma which says that for every $\epsilon>0$ and prime p, there is a $\delta>0$ such that if $X,Y,Z\subset\mathbb{F}_p^n$ and the number of triangles in $X\times Y\times Z$ is at most δN^2, then we can delete ϵN elements from X, Y, and Z and remove all triangles. Green posed the problem of improving the quantitative bounds on the arithmetic triangle removal lemma, and, in particular, asked whether a polynomial bound holds. Despite considerable attention, prior to this paper, the best known bound, due to the first author, showed that $1/\delta$ can be taken to be an exponential tower of twos of height logarithmic in $1/\epsilon$. We solve Green's problem, proving an essentially tight bound for Green's arithmetic triangle removal lemma in $\mathbb{F}_p^n$. We show that a polynomial bound holds, and further determine the best possible exponent. Namely, there is a computable number C_p such that we may take $\delta=(\epsilon/3)^{C_p}$, and we must have $\delta\le\epsilon^{C_p-o(1)}$. In particular, $C_2=1+1/(5/3-\log_2 3)\approx 13.239$, and $C_3=1+1/c_3$ with $c_3=1-\frac{\log b}{\log 3}$, $b=a^{-2/3}+a^{1/3}+a^{4/3}$, and $a=\frac{\sqrt{33}-1}{8}$, which gives $C_3\approx 13.901$. The proof uses Kleinberg, Sawin, and Speyer's essentially sharp bound on multicolored sum-free sets, which builds on the recent breakthrough on the cap set problem by Croot-Lev-Pach, and the subsequent work by Ellenberg-Gijswijt, Blasiak-Church-Cohn-Grochow-Naslund-Sawin-Umans, and Alon.

.

[135] Jacob Fox and Fan Wei. Fast Permutation Property Testing. **SODA 8A** (Las Arenas II-IV). January 18th, 2017, 12:10-12:30

The goal of property testing is to quickly distinguish between objects which satisfy a property and objects that are ϵ-far from satisfying the property. There are now several general results in this area which show that natural properties of combinatorial objects can be tested with "constant" query complexity, depending only on ϵ and the property, and not on the size of the object being tested. The upper bound on the query complexity coming from the proof techniques are often enormous and impractical. It remains a major open problem if better bounds hold. Hoppen, Kohayakawa, Moreira, and Sampaio conjectured and Klimošová and Král' proved that hereditary permutation properties are strongly testable, i.e., can be tested with respect to Kendall's tau distance. The query complexity bound coming from this proof is huge. Even for testing a single forbidden subpermutation it is of Ackermann-type in $1/\epsilon$. We give a new proof which gives a polynomial bound in $1/\epsilon$ for testing a single forbidden subpermutation. Maybe surprisingly, for testing with respect to the rectangular distance, we prove there is a universal (not depending on the property), polynomial in $1/\epsilon$ query complexity bound for two-sided testing hereditary properties of sufficiently large permutations. We further give a nearly linear bound with respect to a closely related metric which also depends on the smallest forbidden subpermutation for the property. Finally, we show that several different permutation metrics of interest are related to the rectangular distance, yielding similar results for testing with respect to these metrics.

.

[136] Stephan Kreutzer, Roman Rabinovich and Sebastian Siebertz. Polynomial Kernels and Wideness Properties of Nowhere Dense Graph Classes. **SODA 8B** (Las Arenas I). January 18th, 2017, 10:55-11:15

Nowhere dense classes of graphs (J. Nešetřil and P. Ossona de Mendez, *First order properties on nowhere dense structures*, The Journal of Symbolic Logic, Vol. 75(03), pages 868–887, 2010.), (J. Nešetřil and P. Ossona de Mendez, *On nowhere dense graphs*, European Journal of Combinatorics, Vol. 32(4), pages 600–617, 2011.) are very general classes of uniformly sparse graphs with several seemingly unrelated characterisations. From an algorithmic perspective, a characterisation of these classes in terms of *uniform quasi-wideness*, a concept originating in finite model theory, has proved to be particularly useful. Uniform quasi-wideness is used in many fpt-algorithms on nowhere dense classes. However, the existing constructions showing the equivalence of nowhere denseness and uniform quasi-wideness imply a non-elementary blow up in the parameter dependence of the fpt-algorithms, making them infeasible in practice. As a first main result of this paper, we use tools from logic, in particular from a sub-field of model theory known as stability theory, to establish polynomial bounds for the equivalence of nowhere denseness and uniform quasi-wideness. As an algorithmic application of our new methods, we obtain for every fixed value of $r\in\mathbb{N}$ a polynomial kernel for the distance-r dominating set problem on nowhere dense classes of graphs. This is particularly interesting, as it implies that for every subgraph-closed class $\mathcal{C}$, the distance-r dominating set problem admits a kernel on $\mathcal{C}$ for every value of r if, and only if, it admits a polynomial kernel for every value of r (under the standard assumption of parameterized complexity theory that FPT $\neq$ W[2]). Finally, we demonstrate how to use the new methods to improve the parameter dependence of many fixed-parameter algorithms. As an example we provide a single exponential parameterized algorithm for the Connected Dominating Set problem on nowhere dense graph classes.

.

[137] Euiwoong Lee. Partitioning a Graph into Small Pieces with Applications to Path Transversal. **SODA 8B** (Las Arenas I). January 18th, 2017, 11:20-11:40

Given a graph $G = (V, E)$ and an integer $k \in \mathbb{N}$, we study *k-Vertex Separator* (resp. *k-Edge Separator*), where the goal is to remove the minimum number of vertices (resp. edges) such that each connected component in the resulting graph has at most k vertices. Our primary focus is on the case where k is either a constant or a slowly growing function of n (e.g. $O(\log n)$ or $n^{o(1)}$). Our problems can be interpreted as a special case of three general classes of problems that have been studied separately (balanced graph partitioning, Hypergraph Vertex Cover (HVC), and fixed parameter tractability (FPT)). Our main result is an $O(\log k)$-approximation algorithm for k-Vertex Separator that runs in time $2^{O(k)}n^{O(1)}$, and an $O(\log k)$-approximation algorithm for k-Edge Separator that runs in time $n^{O(1)}$. Our result on k-Edge Separator improves the best previous graph partitioning algorithm for small k. Our result on k-Vertex Separator improves the simple $(k+1)$-approximation from HVC. When OPT $> k$, the running time $2^{O(k)}n^{O(1)}$ is faster than the lower bound $k^{\Omega(\mathsf{OPT})}n^{\Omega(1)}$ for exact algorithms assuming the Exponential Time Hypothesis. While the running time of $2^{O(k)}n^{O(1)}$ for k-Vertex Separator seems unsatisfactory, we show that the superpolynomial dependence on k may be needed to achieve a polylogarithmic approximation ratio, based on hardness of *Densest k-Subgraph*. We also study k-Path Transversal, where the goal is to remove the minimum number of vertices such that there is no simple path of length k. With additional ideas from FPT algorithms and graph theory, we present an $O(\log k)$-approximation algorithm for k-Path Transversal that runs in time $2^{O(k^3 \log k)}n^{O(1)}$. Previously, the existence of even $(1-\delta)k$-approximation algorithm for fixed $\delta > 0$ was open.

.

[138] Magnus Wahlström. LP-branching algorithms based on biased graphs. **SODA 8B** (Las Arenas I). January 18th, 2017, 11:45-12:05

We give a combinatorial condition for the existence of efficient, LP-based FPT algorithms for a broad class of graph-theoretical optimisation problems. Our condition is based on the notion of biased graphs known from matroid theory. Specifically, we show that given a biased graph $\Psi = (G, \mathcal{B})$, where $\mathcal{B}$ is a class of balanced cycles in G, the problem of finding a set X of at most k vertices in G which intersects every unbalanced cycle in G admits an FPT algorithm using an LP-branching approach, similar to those previously seen for VCSP problems (Wahlström, SODA 2014). Our algorithm has two parts. First we define a *local problem*, where we are additionally given a root vertex $v_0 \in V$ and asked only to delete vertices X (excluding v_0) so that the connected component of v_0 in $G - X$ contains no unbalanced cycle. We show that this local problem admits a persistent, half-integral LP-relaxation with a polynomial-time solvable separation oracle, and can therefore be solved in FPT time via LP-branching, assuming only oracle membership queries for the class of balanced cycles in G. We then show that solutions to this local problem can be used to tile the graph, producing an optimal solution to the original, global problem as well. This framework captures many of the problems previously solved via the VCSP approach to LP-branching, as well as new generalisations, such as Group Feedback Vertex Set for infinite groups (e.g., for graphs whose edges are labelled by matrices). A major advantage compared to previous work is that it is immediate to check the applicability of the result for a given problem, whereas testing applicability of the VCSP approach for a specific VCSP, requires determining the existence of an embedding language with certain algebraically defined properties, which is not known to be decidable in general.

.

[139] Klaus Jansen and Kim-Manuel Klein. About the Structure of the Integer Cone and its Application to Bin Packing. **SODA 8B** (Las Arenas I). January 18th, 2017, 12:10-12:30

We consider the bin packing problem with d different item sizes and revisit the structure theorem given by Goemans and Rothvoß (*Polynomiality for Bin Packing with a Constant Number of Item Types*, SODA 2014) about solutions of the integer cone. We present new techniques on how solutions can be modified and give a new structure theorem that relies on the set of vertices of the underlying integer polytope. As a result of our new structure theorem, we obtain an algorithm for the bin packing problem with running time $|V|^{2^{O(d)}} enc(I)$, where V is the set of vertices of the integer knapsack polytope and $enc(I)$ is the encoding length of the bin packing instance. The algorithm is fixed parameter tractable, parameterized by the number of vertices of the integer knapsack polytope $|V|$. This shows that the bin packing problem can be solved efficiently when the underlying integer knapsack polytope has an easy structure, i.e. has a small number of vertices. Furthermore, we show that the presented bounds of the structure theorem are asymptotically tight. We give a construction of bin packing instances using new structural insights and classical number theoretical theorems which yield the desired lower bound.

.

[140] Marco Molinaro. Online and Random-order Load Balancing Simultaneously. **SODA 8C** (Nelva). January 18th, 2017, 10:55-11:15

We consider the problem of online load balancing under ℓ_p-norms: sequential jobs need to be assigned to one of the machines and the goal is to minimize the ℓ_p-norm of the machine loads. This generalizes the classical problem of scheduling for makespan minimization (case ℓ_∞) and has been thoroughly studied. However, despite the recent push for beyond worst-case analyses, no such results are known for this problem.

In this paper we provide algorithms with **simultaneous** guarantees for the worst-case model as well as for the random-order (i.e. secretary) model, where an arbitrary set of jobs comes in random order. First, we show that the greedy algorithm (with restart), known to have optimal $O(p)$ worst-case guarantee, also has a (typically) improved random-order guarantee. However, the behavior of this algorithm in the random-order model degrades with p. We then propose algorithm SimultaneousLB that has **simultaneously optimal guarantees** (within constants) in both worst-case and random-order models. In particular, the random-order guarantee of SimultaneousLB improves as p increases.

One of the main components is a new algorithm with improved regret for Online Linear Optimization (OLO) over the non-negative vectors in the ℓ_q ball. Interestingly, this OLO algorithm is also used to prove a purely probabilistic inequality that controls the correlations arising in the random-order model, a common source of difficulty for the analysis. Another important component used in both SimultaneousLB and our OLO algorithm is a smoothing of the ℓ_p-norm that may be of independent interest. This smoothness property allows us to see algorithm SimultaneousLB as essentially a greedy one in the worst-case model and as a primal-dual one in the random-order model, which is instrumental for its simultaneous guarantees.

.

[141] Moran Feldman and Rani Izsak. Building a Good Team: Secretary Problems and the Supermodular Degree. **SODA 8C** (Nelva). January 18th, 2017, 11:20-11:40

In the (classical) SECRETARY PROBLEM, one has to hire the best among n candidates. The candidates are interviewed, one at a time, at a uniformly random order, and one has to decide on the spot, whether to hire a candidate or continue interviewing. It is well known that the best candidate can be hired with a probability of $1/e$ (Dynkin, 1963). Recent works extend this problem to settings in which multiple candidates can be hired, subject to some constraint. Here, one wishes to hire a set of candidates maximizing a given objective set function. Almost all extensions considered in the literature assume the objective set function is either linear or submodular. Unfortunately, real world functions might not have either of these properties. Consider, for example, a scenario where one hires researchers for a project. Indeed, it can be that some researchers can substitute others for that matter. However, it can also be that some combinations of researchers result in synergy (see, e.g., Woolley et al., Science 2010, for a study on collective intelligence). The first phenomenon can be modeled by a submoudlar set function, while the latter cannot. In this work, we study the secretary problem with an *arbitrary* non-negative monotone valuation function, subject to a general matroid constraint. One can prove that, generally, only very poor results can be obtained for this class of objective functions. We tackle this hardness by combining the following: (1) Parametrizing our algorithms by the *supermodular degree* of the objective function (defined by Feige and Izsak, ITCS 2013), which, roughly speaking, measures the distance of a function from being submodular. (2) Suggesting an (arguably) natural model that permits approximation guarantees that are *polynomial* in the supermodular degree (as opposed to the standard model which allows only *exponential* guarantees). Our algorithms learn the input by running a non-trivial estimation algorithm on a portion of it whose size depends on the supermodular degree. We also provide better approximation guarantees for the special case of a uniform matroid constraint. To the best of our knowledge, our results represent the first algorithms for a secretary problem handling arbitrary non-negative monotone valuation functions.

.

[142] Aviad Rubinstein and Sahil Singla. Combinatorial Prophet Inequalities. **SODA 8C** (Nelva). January 18th, 2017, 11:45-12:05

We introduce a novel framework of Prophet Inequalities for combinatorial valuation functions. For a (non-monotone) submodular objective function over an arbitrary matroid feasibility constraint, we give an $O(1)$-competitive algorithm. For a monotone subadditive objective function over an arbitrary downward-closed feasibility constraint, we give an $O(\log n \log^2 r)$-competitive algorithm (where r is the cardinality of the largest feasible subset). Inspired by the proof of our subadditive prophet inequality, we also obtain an $O(\log n \cdot \log^2 r)$-competitive algorithm for the Secretary Problem with a monotone subadditive objective function subject to an arbitrary downward-closed feasibility constraint. Even for the special case of a cardinality feasibility constraint, our algorithm circumvents an $\Omega(\sqrt{n})$ lower bound by Bateni, Hajiaghayi, and Zadimoghaddam in a restricted query model. En route to our submodular prophet inequality, we prove a technical result of independent interest: we show a variant of the Correlation Gap Lemma for non-monotone submodular functions.

.

[143] Anupam Gupta, Viswanath Nagarajan and Sahil Singla. Adaptivity Gaps for Stochastic Probing: Submodular and XOS Functions. **SODA 8C** (Nelva). January 18th, 2017, 12:10-12:30

Suppose we are given a submodular function f over a set of elements, and we want to maximize its value subject to certain constraints. Good approximation algorithms are known for such problems under both monotone and non-monotone submodular functions. We consider these problems in a stochastic setting, where elements are not all active and we can only get value from active elements. Each element e is active independently with some known probability p_e, but we don't know the element's status *a priori*. We find it out only when we *probe* the element e—probing reveals whether it's active or not, whereafter we can use this information to decide which other elements to probe. Eventually, if we have a probed set S and a subset $active(S)$ of active elements in S, we can pick any $T \subseteq active(S)$ and get value $f(T)$. Moreover, the sequence of elements we probe must satisfy a given *prefix-closed constraint*—e.g., these may be given by a matroid, or an orienteering constraint, or deadline, or precedence constraint, or an arbitrary downward-closed constraint—if we can probe some sequence of elements we can probe any prefix of it. What is a good strategy to probe elements to maximize the expected value? In this paper we study the gap between adaptive and non-adaptive strategies for f being a submodular or a fractionally subadditive (XOS) function. If this gap is small, we can focus on finding good non-adaptive strategies instead, which are easier to find as well as to represent. We show that the adaptivity gap is a constant for monotone and non-monotone submodular functions, and logarithmic for XOS functions of small *width*. These bounds are nearly tight. Our techniques show new ways of arguing about the optimal adaptive decision tree for stochastic problems.

.

[144] Daniel Funke and Peter Sanders. Parallel d-D Delaunay Triangulations in Shared and Distributed Memory. **ALENEX 5** (Saray). January 18th, 2017, 14:00-14:20

Computing the Delaunay triangulation (DT) of a given point set in $\mathbb{R}^D$ is one of the fundamental operations in computational geometry. In this paper we present a novel divide-and-conquer (DAC) algorithm that lends itself equally well to shared and distributed memory parallelism. While previous DAC algorithms generally suffer from a complex – often sequential – merge or divide step, we reduce the merging of two partial triangulations to re-triangulating a small subset of their vertices using the same parallel algorithm and combining the three triangulations via parallel hash table lookups. In experiments we achieve a reasonable speedup on shared memory machines and compare favorably to CGAL's three-dimensional parallel DT implementation on some inputs. In the distributed memory setting we show that our approach scales to 2048 processing elements, which allows us to compute 3-D DTs for inputs with billions of points.

.

[145] Peyman Afshani, Mark de Berg, Henri Casanova, Benjamin Karsin, Colin Lambrechts, Nodari Sitchinava and Constantinos Tsirogiannis. An Efficient Algorithm for the 1D Total Visibility-Index Problem. **ALENEX 5** (Saray). January 18th, 2017, 14:25-14:45

Let T be a terrain, and let P be a set of points (locations) on its surface. An important problem in Geographic Information Science (GIS) is computing the *visibility index* of a point p on P, that is, the number of points in P that are visible from p. The *total visibility-index* problem asks for computing the visibility index of *every* point in P. Most applications of this problem involve 2-dimensional terrains represented by a grid of $n \times n$ square cells, where each cell is associated with an elevation value, and P consists of the center-points of these cells. Current approaches for computing the total visibility-index on such a terrain take at least quadratic time with respect to the number of the terrain cells. While finding a subquadratic solution to this *2D total visibility-index* problem is an open problem, surprisingly, no subquadratic solution has been proposed for the one-dimensional (1D) version of the problem; in the 1D problem, the terrain is an x-monotone polyline, and P is the set of the polyline vertices. We present an $O(n \log^2 n)$ algorithm that solves the 1D total visibility-index problem in the RAM model. Our algorithm is based on a geometric dualization technique, which reduces the problem into a set of instances of the red-blue line segment intersection counting problem. We also present a parallel version of this algorithm, which requires $O(\log^2 n)$ time and $O(n \log^2 n)$ work in the CREW PRAM model. We implement a naive $O(n^2)$ approach and three variations of our algorithm: one employing an existing red-blue line segment intersection algorithm and two new approaches that perform the intersection counting by leveraging features specific to our problem. We present experimental results for both serial and parallel implementations on large synthetic and real-world datasets, using two distinct hardware platforms. Results show that all variants of our algorithm outperform the naive approach by several orders of magnitude on large datasets. Furthermore, we show that our new intersection counting implementations achieve more than 8 times speedup over the existing red-blue line segment intersection algorithm. Our parallel implementation is able to process a terrain of 2^{24} vertices in under 1 minute using 16 cores, achieving more than 7 times speedup over serial execution.

.

[146] Frank Staals and Constantinos Tsirogiannis. Computing the Expected Value and Variance of Geometric Measures. **ALENEX 5** (Saray). January 18th, 2017, 14:50-15:10

Let P be a set of points in $\mathbb{R}^d$, and let M be a function that maps any subset of P to a positive real number. We examine the problem of computing the exact mean and variance of M when a subset of points in P is selected according to a well-defined random distribution. We consider two distributions; in the first distribution (which we call the *Bernoulli* distribution), each point $p \in P$ is included in the random subset independently, with probability $\pi(p)$. In the second distribution (the *fixed-size* distribution), a subset of exactly s points is selected uniformly at random among all possible subsets of s points in P. This problem is a crucial part of modern ecological analyses; each point in P represents a species in d-dimensional trait space, and the goal is to compute the statistics of a geometric measure on this trait space, when subsets of species are selected under random processes. We present efficient exact algorithms for computing the mean and variance of several geometric measures when point sets are selected under one of the described random distributions. More specifically, we provide algorithms for the following measures: the bounding box volume, the convex hull volume, the mean pairwise distance (MPD), the squared Euclidean distance from the centroid, and the diameter of the minimum enclosing disk. We also describe an efficient $(1-\varepsilon)$-approximation algorithm for computing the mean and variance of the mean pairwise distance. We implemented three of our algorithms: an algorithm that computes the exact mean volume of the 2D bounding box in the Bernoulli distribution, an algorithm that computes the exact mean and variance of the MPD for d-dimensional point sets in the fixed-size distribution, and an $(1-\varepsilon)$-approximation algorithm for the same measure. We conducted experiments where we compared the performance of our implementations with a standard heuristic approach used in ecological applications. We show that our implementations can provide major speedups compared to the standard approach, and they produce results of higher precision, especially for the calculation of the variance. We also compared the implementation of our exact MPD algorithm with the corresponding $(1-\varepsilon)$-approximation method; we show that the approximation method performs faster in certain cases, while also providing high-precision approximations. We thus demonstrate that, as an alternative to the exact algorithm, this method can also be used as a reliable tool for ecological analysis.

.

[147] Daniel Bahrdt, Michael Becher, Stefan Funke, Filip Krumpe, Andre Nusser, Martin Seybold and Sabine Storandt. Growing Balls in $\mathbb{R}^d$. **ALENEX 5** (Saray). January 18th, 2017, 15:15-15:35

Given a set of prioritized balls with fixed centers in R^d whose radii grow linearly over time, we want to compute the *elimination order* of these balls assuming that when two balls touch, the one with lower priority is 'crushed'. A straightforward algorithm has running time $O(n^2 \log n)$ which we improve to expected $O(\Delta^d n(\log n + \Delta^d))$ where $\Delta = r_{max}/r_{min}$ is the ratio between largest and smallest radius amongst the balls. For a natural application of this problem, namely drawing labels on the globe, we have $\Delta = O(1)$. An efficient implementation based on a spherical Delaunay triangulation allows to compute the elimination order for millions of labels on commodity Desktop hardware. Dealing with rounding error induced robustness issues turned out to be one of the major challenges in the implementation.

.

[148] Lars Arge, Mathias Rav, Sarfraz Raza and Morten Revsbæk. I/O-Efficient Event Based Depression Flood Risk. **ALENEX 5** (Saray). January 18th, 2017, 15:40-16:00

An important problem in terrain analysis is modeling how water flows across a terrain and creates floods by filling up depressions. The accuracy of such modeling depends critically on the precision of the terrain data, and available high-resolution terrain models of even fairly small geographic regions often exceed the size of a computer's main memory. In such cases movement of data between main memory and external memory (such as disk) is often the bottleneck in the computation. Thus it is important to develop I/O-efficient modeling algorithms, that is, algorithms that minimize the movement of blocks of data between main memory and disk. In this paper we develop practically I/O-efficient algorithms for the problem of computing the areas of a terrain that are flooded in a given flash flood event due to water collecting in depressions. Previous work only considered events where rain falls at a constant uniform rate on the entire terrain. In reality, local extreme flash floods can affect downstream areas that do not receive heavy rainfall directly, so it is important to model such non-uniform events. Our main algorithm uses $O(\mathrm{Sort}(N) + \mathrm{Scan}(H \cdot X))$ I/Os, where N is the size of the terrain, $\mathrm{Sort}(N)$ and $\mathrm{Scan}(N)$ are the number of I/Os required to sort and read N elements in the standard two-level I/O-model, respectively, X is the number of sinks in the terrain and H the height of the so-called merge-tree, which is a hierarchical representation of the depressions of the terrain. Under practically realistic assumptions about the main memory size compared to X and H, we also develop $O(\mathrm{Sort}(N))$ I/O-algorithms. One of these algorithms can handle an event in optimal $O(\mathrm{Scan}(N))$ I/Os after using $O(\mathrm{Sort}(N))$ I/Os on preprocessing the terrain. We have implemented our algorithms and show that they work very well in practice.

.

[149] Michael Kapralov, Sanjeev Khanna, Madhu Sudan and Ameya Velingker. $(1+\Omega(1))$-Approximation to MAX-CUT Requires Linear Space. **SODA 9A** (Las Arenas II-IV). January 18th, 2017, 14:00-14:20

We consider the problem of estimating the value of MAX-CUT in a graph in the streaming model of computation. We show that there exists a constant $\epsilon_* > 0$ such that any randomized streaming algorithm that computes a $(1+\epsilon_*)$-approximation to MAX-CUT requires $\Omega(n)$ space on an n vertex graph. By contrast, there are algorithms that produce a $(1+\epsilon)$-approximation in space $O(n/\epsilon^2)$ for every $\epsilon > 0$. Our result is the first linear space lower bound for the task of approximating the max cut value and partially answers an open question from the literature. The prior state of the art ruled out $(2-\epsilon)$-approximation in $\tilde{O}(\sqrt{n})$ space or $(1+\epsilon)$-approximation in $n^{1-O(\epsilon)}$ space, for any $\epsilon > 0$. Previous lower bounds for the MAX-CUT problem relied, in essence, on a lower bound on the communication complexity of the following task: Several players are each given some edges of a graph and they wish to determine if the union of these edges is ϵ-close to forming a bipartite graph, using one-way communication. The previous works proved a lower bound of $\Omega(\sqrt{n})$ for this task when $\epsilon = 1/2$, and $n^{1-O(\epsilon)}$ for every $\epsilon > 0$, even when one of the players is given a candidate bipartition of the graph and the graph is promised to be bipartite with respect to this partition or ϵ-far from bipartite. This added information was essential in enabling the previous analyses but also yields a weak bound since, with this extra information, there is an $n^{1-O(\epsilon)}$ communication protocol for this problem. In this work, we give an $\Omega(n)$ lower bound on the communication complexity of the original problem (without the extra information) for $\epsilon = \Omega(1)$ in the three-player setting. Obtaining this $\Omega(n)$ lower bound on the communication complexity is the main technical result in this paper. We achieve it by a delicate choice of distributions on instances as well as a novel use of the convolution theorem from Fourier analysis combined with graph-theoretic considerations to analyze the communication complexity.

.

[150] Sepehr Assadi, Sanjeev Khanna and Yang Li. On Estimating Maximum Matching Size in Graph Streams. **SODA 9A** (Las Arenas II-IV). January 18th, 2017, 14:25-14:45

We study the problem of estimating the maximum matching *size* in graphs whose edges are revealed in a streaming manner. We consider both *insertion-only* streams, which only contain edge insertions, and *dynamic* streams that allow both insertions and deletions of the edges, and present new upper and lower bound results for both cases.

On the upper bound front, we show that an α-approximate estimate of the matching size can be computed in dynamic streams using $\tilde{O}(n^2/\alpha^4)$ space, and in insertion-only streams using $\tilde{O}(n/\alpha^2)$-space. These bounds respectively shave off a factor of α from the space necessary to compute an α-approximate matching (as opposed to only size), thus proving a non-trivial separation between approximate estimation and approximate computation of matchings in data streams.

On the lower bound front, we prove that any α-approximation algorithm for estimating matching size in dynamic graph streams requires $\Omega(\sqrt{n}/\alpha^{2.5})$ bits of space, *even* if the underlying graph is both *sparse* and has *arboricity* bounded by $O(\alpha)$. We further improve our lower bound to $\Omega(n/\alpha^2)$ in the case of *dense* graphs. These results establish the first non-trivial streaming lower bounds for *super-constant* approximation of matching size.

Furthermore, we present the first *super-linear* space lower bound for computing a $(1+\epsilon)$-approximation of matching size *even* in insertion-only streams. In particular, we prove that a $(1+\epsilon)$-approximation to matching size requires $\mathsf{RS}(n)\cdot n^{1-O(\epsilon)}$ space; here, $\mathsf{RS}(n)$ denotes the maximum number of edge-disjoint *induced matchings* of size $\Theta(n)$ in an n-vertex graph. It is a major open problem with far-reaching implications to determine the value of $\mathsf{RS}(n)$, and current results leave open the possibility that $\mathsf{RS}(n)$ may be as large as $n/\log n$. Moreover, using the best known lower bounds for $\mathsf{RS}(n)$, our result already rules out any $O(n\cdot \mathrm{poly}(\log n/\epsilon))$-space algorithm for $(1+\epsilon)$-approximation of matchings. We also show how to avoid the dependency on the parameter $\mathsf{RS}(n)$ in proving lower bound for dynamic streams and present a near-optimal lower bound of $n^{2-O(\epsilon)}$ for $(1+\epsilon)$-approximation in this model. Using a well-known connection between matching size and *matrix rank*, all our lower bounds also hold for the problem of estimating matrix rank. In particular our results imply a near-optimal $n^{2-O(\epsilon)}$ bit lower bound for $(1+\epsilon)$-approximation of matrix ranks for dense matrices in dynamic streams, answering an open question of Li and Woodruff (STOC 2016).

.

[151] Themistoklis Gouleakis, Christos Tzamos and Manolis Zampetakis. Faster Sublinear Algorithms via Conditional Sampling. **SODA 9A** (Las Arenas II-IV). January 18th, 2017, 14:50-15:10

A conditional sampling oracle for a probability distribution D returns samples from the conditional distribution of D restricted to a specified subset of the domain. A recent line of work [CFGM13, CRS14] has shown that having access to such a conditional sampling oracle requires only polylogarithmic or even constant number of samples to solve distribution testing problems like identity and uniformity. This significantly improves over the standard sampling model where polynomially many samples are necessary. Inspired by these results, we introduce a computational model based on conditional sampling to develop sublinear algorithms with exponentially faster runtimes compared to standard sublinear algorithms. We focus on geometric optimization problems over points in high dimensional Euclidean space. Access to these points is provided via a conditional sampling oracle that takes as input a succinct representation of a subset of the domain and outputs a uniformly random point in that subset. We study two well studied problems: k-means clustering and estimating the weight of the minimum spanning tree. In contrast to prior algorithms for the classic model, our algorithms have time, space and sample complexity that is polynomial in the dimension and polylogarithmic in the number of points. Finally, we comment on the applicability of the model and compare with existing ones like streaming, parallel and distributed computational models.

.

[152] Michael B. Cohen, Cameron Musco and Christopher Musco. Input Sparsity Time Low-Rank Approximation via Ridge Leverage Score Sampling. **SODA 9A** (Las Arenas II-IV). January 18th, 2017, 15:15-15:35

We present a new algorithm for finding a near optimal low-rank approximation of a matrix A in $O(nnz(A))$ time. Our method is based on a recursive sampling scheme for computing a representative subset of A's columns, which is then used to find a low-rank approximation. This approach differs substantially from prior $O(nnz(A))$ time algorithms, which are all based on fast Johnson-Lindenstrauss random projections. It matches the guarantees of these methods while offering a number of advantages. Not only are sampling algorithms faster for sparse and structured data, but they can also be applied in settings where random projections cannot. For example, we give new single-pass streaming algorithms for the column subset selection and projection-cost preserving sample problems. Our method has also been used to give the fastest algorithms for provably approximating kernel matrices.

.

[153] John Kallaugher and Eric Price. A Hybrid Sampling Scheme for Triangle Counting. **SODA 9A** (Las Arenas II-IV). January 18th, 2017, 15:40-16:00

We study the problem of estimating the number of triangles in a graph stream. No streaming algorithm can get sublinear space on all graphs, so methods in this area bound the space in terms of parameters of the input graph such as the maximum number of triangles sharing a single edge. We give a sampling algorithm that is additionally parameterized by the maximum number of triangles sharing a single vertex. Our bound matches the best known turnstile results in all graphs, and gets better performance on simple graphs like $G(n, p)$ or a set of independent triangles. We complement the upper bound with a lower bound showing that no sampling algorithm can do better on those graphs by more than a log factor. In particular, any insertion stream algorithm must use $\sqrt{T}$ space when all the triangles share a common vertex, and any sampling algorithm must take $T^{1/3}$ samples when all the triangles are independent. We add another lower bound, also matching our algorithm's performance, which applies to *all* graph classes. This lower bound covers "triangle-dependent" sampling algorithms, a subclass that includes our algorithm and all previous sampling algorithms for the problem.

.

[154] Pinyan Lu, Kuan Yang, Chihao Zhang and Minshen Zhu. An FPTAS for Counting Proper Four-Colorings on Cubic Graphs. **SODA 9B** (Las Arenas I). January 18th, 2017, 14:00-14:20

Graph coloring is arguably the most exhaustively studied problem in the area of approximate counting. It is conjectured that there is a fully polynomial-time (randomized) approximation scheme (FPTAS/FPRAS) for counting the number of proper colorings as long as $q \geq \Delta + 1$, where q is the number of colors and Δ is the maximum degree of the graph. The bound of $q = \Delta + 1$ is the uniqueness threshold for Gibbs measure on Δ-regular infinite trees. However, the conjecture remained open even for any fixed $\Delta \geq 3$ (The cases of $\Delta = 1, 2$ are trivial). In this paper, we design an FPTAS for counting the number of proper 4-colorings on graphs with maximum degree 3 and thus confirm the conjecture in the case of $\Delta = 3$. This is the first time to achieve this optimal bound of $q = \Delta + 1$. Previously, the best FPRAS requires $q > \frac{11}{6}\Delta$ and the best deterministic FPTAS requires $q > 2.581\Delta + 1$ for general graphs. In the case of $\Delta = 3$, the best previous result is an FPRAS for counting proper 5-colorings. We note that there is a barrier to go beyond $q = \Delta + 2$ for single-site Glauber dynamics based FPRAS and we overcome this by correlation decay approach. Moreover, we develop a number of new techniques for the correlation decay approach which can find applications in other approximate counting problems.

.

[155] Heng Guo and Mark Jerrum. Random cluster dynamics for the Ising model is rapidly mixing. **SODA 9B** (Las Arenas I). January 18th, 2017, 14:25-14:45

We show for the first time that the mixing time of Glauber (single edge update) dynamics for the random cluster model at $q = 2$ is bounded by a polynomial in the size of the underlying graph. As a consequence, the Swendsen-Wang algorithm for the ferromagnetic Ising model at any temperature has the same polynomial mixing time bound.

.

[156] Prateek Bhakta, Ben Cousins, Matthew Fahrbach and Dana Randall. Approximately Sampling Elements with Fixed Rank in Graded Posets. **SODA 9B** (Las Arenas I). January 18th, 2017, 14:50-15:10

Graded posets frequently arise throughout combinatorics, where it is natural to try to count the number of elements of a fixed rank. These counting problems are often #**P**-complete, so we consider approximation algorithms for counting and uniform sampling. We show that for certain classes of posets, biased Markov chains that walk along edges of their Hasse diagrams allow us to approximately generate samples with any fixed rank in expected polynomial time. Our arguments do not rely on the typical proofs of log-concavity, which are used to construct a stationary distribution with a specific mode in order to give a lower bound on the probability of outputting an element of the desired rank. Instead, we infer this directly from bounds on the mixing time of the chains through a method we call *balanced bias*. A noteworthy application of our method is sampling restricted classes of integer partitions of n. We give the first provably efficient Markov chain algorithm to uniformly sample integer partitions of n from general restricted classes. Several observations allow us to improve the efficiency of this chain to require $O(n^{1/2}\log(n))$ space, and for unrestricted integer partitions, expected $O(n^{9/4})$ time. Related applications include sampling permutations with a fixed number of inversions and lozenge tilings on the triangular lattice with a fixed average height.

.

[157] Roee David and Uriel Feige. Random walks with the minimum degree local rule have $O(n^2)$ cover time. **SODA 9B** (Las Arenas I). January 18th, 2017, 15:15-15:35

For a simple (unbiased) random walk on a connected graph with n vertices, the cover time (the expected number of steps it takes to visit all vertices) is at most $O(n^3)$. We consider locally biased random walks, in which the probability of traversing an edge depends on the degrees of its endpoints. We confirm a conjecture of Abdullah, Cooper and Draief [2015] that the min-degree local bias rule ensures a cover time of $O(n^2)$. For this we formulate and prove the following lemma about spanning trees. Let $R(e)$ denote for edge e the minimum degree among its two endpoints. We say that a weight function W for the edges is feasible if it is nonnegative, dominated by R (for every edge $W(e) \leq R(e)$) and the sum over all edges of the ratios $W(e)/R(e)$ equals $n - 1$. For example, in trees $W(e) = R(e)$, and in regular graphs the sum of edge weights is $d(n - 1)$. **Lemma:** for every feasible W, the minimum weight spanning tree has total weight $O(n)$. For regular graphs, a similar lemma was proved by Kahn, Linial, Nisan and Saks [1989].

.

[158] Siu On Chan, Tsz Chiu Kwok and Lap Chi Lau. Random Walks and Evolving Sets: Faster Convergences and Limitations. **SODA 9B** (Las Arenas I). January 18th, 2017, 15:40-16:00

Analyzing the mixing time of random walks is a well-studied problem with applications in random sampling and more recently in graph partitioning. In this work, we present new analysis of random walks and evolving sets using more combinatorial graph structures, and show some implications in approximating small-set expansion. On the other hand, we provide examples showing the limitations of using random walks and evolving sets in disproving the small-set expansion hypothesis.

1. We define a combinatorial analog of the spectral gap, and use it to prove the convergence of non-lazy random walks. A corollary is a tight lower bound on the small-set expansion of graph powers for any graph.
2. We prove that random walks converge faster when the robust vertex expansion of the graph is larger. This provides an improved analysis of the local graph partitioning algorithm using the evolving set process, and also derives an alternative proof of an improved Cheeger's inequality.
3. We give an example showing that the evolving set process fails to disprove the small-set expansion hypothesis. This refutes a conjecture of Oveis Gharan and shows the limitations of all existing local graph partitioning algorithms in approximating small-set expansion.

.

[159] James B. Orlin and Antonio Sedeño-Noda. An $O(nm)$ time algorithm for finding the min length directed cycle in a graph. **SODA 9C** (Nelva). January 18th, 2017, 14:00-14:20

In this paper, we introduce an $O(nm)$ time algorithm to determine the minimum length directed cycle (also called the "minimum weight directed cycle") in a directed network with n nodes and m arcs and with no negative length directed cycles. This result improves upon the previous best time bound of $O(nm + n^2 \log\log n)$. Our algorithm first determines the cycle with minimum mean length λ^* in $O(nm)$ time. Subsequently, it chooses node potentials so that all reduced costs are λ^* or greater. It then solves the all pairs shortest path problem, but restricts attention to paths of length at most $n\lambda^*$. We speed up the shortest path calculations to $O(m)$ per source node, leading to an $O(nm)$ running time in total. We also carry out computational experiments comparing the performance of the proposed methods and other state-of-the-art methods. Experiments confirmed that it is advantageous to solve the minimum mean cycle problem prior to solving shortest path problems. Analysis of our experiments suggest that the running time to solve the minimum length directed cycle problem was much faster than $O(n^2)$ on average.

.

[160] Loukas Georgiadis, Giuseppe F. Italiano and Nikos Parotsidis. Strong Connectivity in Directed Graphs under Failures, with Applications. **SODA 9C** (Nelva). January 18th, 2017, 14:25-14:45

Let G be a directed graph (digraph) with m edges and n vertices, and let $G \setminus e$ (resp., $G \setminus v$) be the digraph obtained after deleting edge e (resp., vertex v) from G. We show how to compute in $O(m + n)$ worst-case time:

- The total number of strongly connected components in $G \setminus e$ (resp., $G \setminus v$), *for all edges* e (resp., *for all vertices* v) in G.
- The size of the largest and of the smallest strongly connected components in $G \setminus e$ (resp., $G \setminus v$), *for all edges* e (resp., *for all vertices* v) in G.

Let G be strongly connected. We say that edge e (resp., vertex v) separates two vertices x and y, if x and y are no longer strongly connected in $G \setminus e$ (resp., $G \setminus v$). We also show how to build in $O(m + n)$ time $O(n)$-space data structures that can answer in optimal time the following basic connectivity queries on digraphs:

- Report in $O(n)$ worst-case time all the strongly connected components of $G \setminus e$ (resp., $G \setminus v$), for a query edge e (resp., vertex v).
- Test whether an edge or a vertex separates two query vertices in $O(1)$ worst-case time.
- Report all edges (resp., vertices) that separate two query vertices in optimal worst-case time, i.e., in time $O(k)$, where k is the number of separating edges (resp., separating vertices). (For $k = 0$, the time is $O(1)$).

All our bounds are tight and are obtained with a common algorithmic framework, based on a novel compact representation of the decompositions induced by 1-edge and 1-vertex cuts in digraphs, which might be of independent interest. With the help of our data structures we can design efficient algorithms for several other connectivity problems on digraphs and we can also obtain in linear time a strongly connected spanning subgraph of G with $O(n)$ edges that maintains the 1-connectivity cuts of G and the decompositions induced by those cuts.

.

[161] Shiri Chechik, Thomas Dueholm Hansen, Giuseppe F. Italiano, Veronika Loitzenbauer and Nikos Parotsidis. Faster Algorithms for Computing Maximal 2-Connected Subgraphs in Sparse Directed Graphs. **SODA 9C** (Nelva). January 18th, 2017, 14:50-15:10

Connectivity related concepts are of fundamental interest in graph theory. The area has received extensive attention over four decades, but many problems remain unsolved, especially for directed graphs. A directed graph is 2-edge-connected (resp., 2-vertex-connected) if the removal of any edge (resp., vertex) leaves the graph strongly connected. In this paper we present improved algorithms for computing the maximal 2-edge- and 2-vertex-connected subgraphs of a given directed graph. These problems were first studied more than 35 years ago, with $\widetilde{O}(mn)$ time algorithms for graphs with m edges and n vertices being known since the late 1980s. In contrast, the same problems for undirected graphs are known to be solvable in linear time. Henzinger et al. [ICALP 2015] recently introduced $O(n^2)$ time algorithms for the directed case, thus improving the running times for dense graphs. Our new algorithms run in time $O(m^{3/2})$, which further improves the running times for sparse graphs. The notion of 2-connectivity naturally generalizes to k-connectivity for $k > 2$. For constant values of k, we extend one of our algorithms to compute the maximal k-edge-connected in time $O(m^{3/2} \log n)$, improving again for sparse graphs the best known algorithm by Henzinger et al. [ICALP 2015] that runs in $O(n^2 \log n)$ time.

.

[162] Monika Henzinger, Satish Rao and Di Wang. Local Flow Partitioning for Faster Edge Connectivity. **SODA 9C** (Nelva). January 18th, 2017, 15:15-15:35

We study the problem of computing a minimum cut in a simple, undirected graph and give a deterministic $O(m \log^2 n \log\log^2 n)$ time algorithm. This improves both on the best previously known deterministic running time of $O(m \log^{12} n)$ (Kawarabayashi and Thorup STOC'2015) and the best previously known randomized running time of $O(m \log^3 n)$ (Karger STOC'1996) for this problem, though Karger's algorithm can be further applied to weighted graphs. Our approach is using the Kawarabayashi and Thorup graph compression technique, which repeatedly finds low-conductance cuts. To find these cuts they use a diffusion-based local algorithm. We use instead a flow-based local algorithm and suitably adjust their framework to work with our flow-based subroutine. Both flow and diffusion based methods have a long history of being applied to finding low conductance cuts. Diffusion algorithms have several variants that are naturally local while it is more complicated to make flow methods local. Some prior work has proven nice properties for local flow based algorithms with respect to improving or cleaning up low conductance cuts. Our flow subroutine, however, is the first that is both local and produces low conductance cuts. Thus, it may be of independent interest.

.

[163] Saleh Soltan, Mihalis Yannakakis and Gil Zussman. Doubly Balanced Connected Graph Partitioning. **SODA 9C** (Nelva). January 18th, 2017, 15:40-16:00

We introduce and study the Doubly Balanced Connected graph Partitioning (DBCP) problem: Let $G = (V, E)$ be a connected graph with a weight (supply/demand) function $p : V \to \{-1, +1\}$ satisfying $p(V) = \sum_{j \in V} p(j) = 0$. The objective is to partition G into (V_1, V_2) such that $G[V_1]$ and $G[V_2]$ are connected, $|p(V_1)|, |p(V_2)| \leq c_p$, and $\max\{\frac{|V_1|}{|V_2|}, \frac{|V_2|}{|V_1|}\} \leq c_s$, for some constants c_p and c_s. When G is 2-connected, we show that a solution with $c_p = 1$ and $c_s = 3$ always exists and can be found in polynomial time. Moreover, when G is 3-connected, we show that there is always a 'perfect' solution (a partition with $p(V_1) = p(V_2) = 0$ and $|V_1| = |V_2|$, if $|V| \equiv 0 (\bmod 4)$), and it can be found in polynomial time. Our techniques can be extended, with similar results, to the case in which the weights are arbitrary (not necessarily ± 1), and to the case that $p(V) \neq 0$ and the excess supply/demand should be split evenly. They also apply to the problem of partitioning a graph with two types of nodes into two large connected subgraphs that preserve approximately the proportion of the two types.

.

[164] Zachary Abel, Victor Alvarez, Erik D. Demaine, Sándor P. Fekete, Aman Gour, Adam Hesterberg, Phillip Keldenich and Christian Scheffer. Three Colors Suffice: Conflict-Free Coloring of Planar Graphs. **SODA 10A** (Las Arenas II-IV). January 18th, 2017, 16:25-16:45

A *conflict-free k-coloring* of a graph assigns one of k different colors to some of the vertices such that, for every vertex v, there is a color that is assigned to exactly one vertex among v and v's neighbors. Such colorings have applications in wireless networking, robotics, and geometry, and are well-studied in graph theory. Here we study the natural problem of the *conflict-free chromatic number* $\chi_{CF}(G)$ (the smallest k for which conflict-free k-colorings exist), with a focus on planar graphs. For general graphs, we prove the conflict-free variant of the famous Hadwiger Conjecture: If G does not contain K_{k+1} as a minor, then $\chi_{CF}(G) \leq k$. For planar graphs, we obtain a tight worst-case bound: three colors are sometimes necessary and always sufficient. In addition, we give a complete characterization of the algorithmic/computational complexity of conflict-free coloring. It is NP-complete to decide whether a planar graph has a conflict-free coloring with *one* color, while for outerplanar graphs, this can be decided in polynomial time. Furthermore, it is NP-complete to decide whether a planar graph has a conflict-free coloring with *two* colors, while for outerplanar graphs, two colors always suffice. For the *bicriteria* problem of minimizing the number of colored vertices subject to a given bound k on the number of colors, we give a full algorithmic characterization in terms of complexity and approximation for outerplanar and planar graphs.

.

[165] Nikhil Bansal, Daniel Reichman and Seeun William Umboh. LP-Based Robust Algorithms for Noisy Minor-Free and Bounded Treewidth Graphs. **SODA 10A** (Las Arenas II-IV). January 18th, 2017, 16:50-17:10

We give a general approach for solving optimization problems on noisy minor free and bounded treewidth graphs, where a fraction of edges are adversarially corrupted. The noisy setting was first considered by Magen and Moharrami and they gave a $(1+\epsilon)$-estimation algorithm for the independent set problem. Later, Chan and Har-Peled designed a local search algorithm that finds a $(1+\epsilon)$-approximate independent set. However, nothing was known regarding other problems in the noisy setting. Our main contribution is a general LP-based framework that yields $(1+\epsilon)$-approximation algorithms for noisy MAX-k-CSPs.

.

[166] Fabrizio Frati, Maurizio Patrignani and Vincenzo Roselli. LR-Drawings of Ordered Rooted Binary Trees and Near-Linear Area Drawings of Outerplanar Graphs. **SODA 10A** (Las Arenas II-IV). January 18th, 2017, 17:15-17:35

We study a family of algorithms, introduced by Chan *[SODA 1999]*, for drawing ordered rooted binary trees. Any algorithm in this family (which we name an *LR-algorithm*) takes in input an ordered rooted binary tree T with a root r_T, and recursively constructs drawings Γ_L of the left subtree L of r_T and Γ_R of the right subtree R of r_T; then either it applies the *left rule*, i.e., it places Γ_L one unit below and to the left of r_T, and Γ_R one unit below Γ_L with the root of R vertically aligned with r_T, or it applies the *right rule*, i.e., it places Γ_R one unit below and to the right of r_T, and Γ_L one unit below Γ_R with the root of L vertically aligned with r_T. In both cases, the edges between r_T and its children are represented by straight-line segments. Different LR-algorithms result from different choices on whether the left or right rule is applied at any node of T. We are interested in constructing *LR-drawings* (that are drawings obtained via LR-algorithms) with small width. Chan showed three LR-algorithms that achieve, for an n-node ordered rooted binary tree, width $O(n^{0.695})$, width $O(n^{0.5})$, and width $O(n^{0.48})$.

We prove that, for every n-node ordered rooted binary tree, an LR-drawing with minimum width can be constructed in $O(n^{1.48})$ time. Further, we show an infinite family of n-node ordered rooted binary trees requiring $\Omega(n^{0.418})$ width in any LR-drawing; no lower bound better than $\Omega(\log n)$ was previously known. Finally, we present the results of an experimental evaluation that allowed us to determine the minimum width of all the ordered rooted binary trees with up to 455 nodes.

Our interest in LR-drawings is mainly motivated by a result of Di Battista and Frati *[Algorithmica 2009]*, who proved that n-vertex outerplanar graphs have outerplanar straight-line drawings in $O(n^{1.48})$ area by means of a drawing algorithm which resembles an LR-algorithm.

We deepen the connection between LR-drawings and outerplanar drawings by proving that, if n-node ordered rooted binary trees have LR-drawings with $f(n)$ width, for any function $f(n)$, then n-vertex outerplanar graphs have outerplanar straight-line drawings in $O(f(n))$ area.

Finally, we exploit a structural decomposition for ordered rooted binary trees introduced by Chan in order to prove that every n-vertex outerplanar graph has an outerplanar straight-line drawing in $O(n \cdot 2^{\sqrt{2\log_2 n}}\sqrt{\log n})$ area.

.

[167] Guido Brückner and Ignaz Rutter. Partial and Constrained Level Planarity. **SODA 10A** (Las Arenas II-IV). January 18th, 2017, 17:40-18:00

Let $G=(V,E)$ be a directed graph and $\ell\colon V \to [k] := \{1,\dots,k\}$ a level assignment such that $\ell(u) < \ell(v)$ for all directed edges $(u,v) \in E$. A level planar drawing of G is a drawing of G where each vertex v is mapped to a unique point on the horizontal line ℓ_i with y-coordinate $\ell(v)$, and each edge is drawn as a y-monotone curve between its endpoints such that no two curves cross in their interior. In the problem Constrained Level Planarity (CLP for short), we are further given a partial ordering $\prec_i$ of $V_i := \ell^{-1}(i)$ for $i \in [k]$, and we seek a level planar drawing where the order of the vertices on ℓ_i is a linear extension of $\prec_i$. A special case of this is the problem Partial Level Planarity (PLP for short), where we are asked to extend a given level-planar drawing $\mathcal{H}$ of a subgraph $H \subseteq G$ to a complete drawing $\mathcal{G}$ of G without modifying the given drawing, i.e., the restriction of $\mathcal{G}$ to H must coincide with $\mathcal{H}$. We give a simple polynomial-time algorithm with running time $\mathcal{O}(n^5)$ for CLP of single-source graphs that is based on a simplified version of an existing level-planarity testing algorithm for single-source graphs. We introduce a modified type of PQ-tree data structure that is capable of efficiently handling the arising constraints to improve the running time to $O(n+k\ell)$, where ℓ is the size of the constraints. We complement this result by showing that PLP is NP-complete even in very restricted cases. In particular, PLP remains NP-complete even when G is a subdivision of a triconnected planar graph with bounded degree.

.

[168] Timothy Naumovitz, Michael Saks and C. Ceshadhri. Accurate and Nearly Optimal Sublinear Approximations to Ulam Distance. **SODA 10B** (Las Arenas I). January 18th, 2017, 16:25-16:45

The *Ulam distance* between two permutations of length n is the minimum number of insertions and deletions needed to transform one sequence into the other. Equivalently, the Ulam distance d is n minus the length of the longest common subsequence (LCS) between the permutations. Our main result is an algorithm, that for any fixed $\varepsilon > 0$, provides a $(1+\varepsilon)$-multiplicative approximation for d in $\tilde{O}_\varepsilon(n/d+\sqrt{n})$ time, which has been shown to be optimal up to polylogarithmic factors. This is the first sublinear time algorithm (provided that $d=(\log n)^{\omega(1)}$) that obtains arbitrarily good multiplicative approximations to the Ulam distance. The previous best bound is an $O(1)$-approximation (with a large constant) by (A.

Andoni and H. L. Nguyen, *Near-Optimal Sublinear Time Algorithms for Ulam Distance*, Proceedings of the 21st Symposium on Discrete Algorithms (SODA), 76–86, 2010) with the same running time bound (ignoring polylogarithmic factors).

The improvement in the approximation factor from $O(1)$ to $(1+\varepsilon)$ allows for significantly more powerful sublinear algorithms. For example, for any fixed $\delta > 0$, we can get additive δn approximations for the LCS between permutations in $\tilde{O}_\delta(\sqrt{n})$ time. Previous sublinear algorithms require δ to be at least $1-1/C$, where C is the approximation factor, which is close to 1 when C is large. Our algorithm is obtained by abstracting the basic algorithmic framework of (A. Andoni and H.L. Nguyen, ibid.) and combining it with the sublinear approximations for the longest increasing subsequence by (M. Saks and C. Seshadhri, *Estimating the longest increasing sequence in polylogarithmic time*, Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS), 458–467, 2010).

.

[169] Rasmus Kyng, Jakub Pachocki, Richard Peng and Sushant Sachdeva. A framework for analyzing resparsification algorithms. **SODA 10B** (Las Arenas I). January 18th, 2017, 16:50-17:10

A spectral sparsifier of a graph G is a sparser graph H that approximately preserves the quadratic form of G, i.e., for all vectors x, $x^T L_G x \approx x^T L_H x$, where L_G, L_H denote the respective graph Laplacians. Spectral sparsifiers generalize cut sparsifiers, and have found several applications in designing graph algorithms. In recent years, there has been interest in computing spectral sparsifiers in the semi-streaming and dynamic settings. Natural algorithms in these settings involve repeated sparsification of a dynamic graph. We present a framework for analyzing algorithms for graph sparsification that perform repeated sparsifications. The framework yields analyses that avoid a worst-case error accumulation across various resparsification steps, and only incur the error corresponding to a single sparsification step, leading to better results. As an application, we show how to maintain a spectral sparsifier of a graph, with $O(n \log n)$ edges in a semi-streaming setting: We present a simple algorithm that, for a graph G on n vertices and m edges, computes a spectral sparsifier of G with $O(n \log n)$ edges in a single pass over G, using only $O(n \log n)$ space, and $O(m \log^2 n)$ total time. This improves on previous best constructions in the semi-streaming setting for both spectral and cut sparsifiers. The algorithm also extends to semi-streaming row sampling for general PSD matrices. As another example, we use this framework to give a parallel algorithm that achieves the best combinatorial construction of spectral graph sparsifiers, combining an algorithm due to Koutis with improved spanner construction.

.

[170] Santosh Vempala and David P. Woodruff. Adaptive Matrix Vector Product. **SODA 10B** (Las Arenas I). January 18th, 2017, 17:15-17:35

We consider the following streaming problem: given a hardwired $m \times n$ matrix A together with a $\mathrm{poly}(mn)$-bit hardwired string of advice that may depend on A, for any x with coordinates $x_1, \ldots x_n$ presented in order, output the coordinates of $A \cdot x$ in order. Our focus is on using as little memory as possible while computing $A \cdot x$; we do not count the size of the output tape on which the coordinates of $A \cdot x$ are written; for some matrices A such as the identity matrix, a constant number of words of space is achievable. Such an algorithm has to adapt its memory contents to the changing structure of A and exploit it on the fly. We give a nearly tight characterization, for any number of passes over the coordinates of x, of the space complexity of such a streaming algorithm. Our characterization is constructive, in that we provide an efficient algorithm matching our lower bound on the space complexity. The essential parameters, *streaming rank* and *multi-pass streaming rank* of A, might be of independent interest, and we show they can be computed efficiently. We give several applications of our results to computing Johnson-Lindenstrauss transforms. Finally, we note that we can characterize the optimal space complexity when the coordinates of $A \cdot x$ can be written in any order.

.

[171] Kenneth L. Clarkson and David P. Woodruff. Low-Rank PSD Approximation in Input-Sparsity Time. **SODA 10B** (Las Arenas I). January 18th, 2017, 17:40-18:00

We give algorithms for approximation by low-rank positive semidefinite (PSD) matrices. For symmetric input matrix $A \in \mathbb{R}^{n \times n}$, target rank k, and error parameter $\epsilon > 0$, one algorithm finds with constant probability a PSD matrix Y of rank k such that $\|A - Y\|_F^2 \le (1+\epsilon)\|A - A_{k,+}\|_F^2$, where $A_{k,+}$ denotes the best rank-k PSD approximation to A, and the norm is Frobenius. The algorithm takes time $O(\mathrm{nnz}(A) \log n) + n\mathrm{poly}((\log n)k/\epsilon) + \mathrm{poly}(k/\epsilon)$, where $\mathrm{nnz}(A)$ denotes the number of nonzero entries of A, and $\mathrm{poly}(k/\epsilon)$ denotes a polynomial in k/ϵ. (There are two different polynomials in the time bound.) Here the output matrix Y has the form $CUC^\top$, where the $O(k/\epsilon)$ columns of C are columns of A. In contrast to prior work, we do not require the input matrix A to be PSD, our output is rank k (not larger), and our running time is $O(\mathrm{nnz}(A) \log n)$ provided this is larger than $n\mathrm{poly}((\log n)k/\epsilon)$. We give a similar algorithm that is faster and simpler, but whose rank-k PSD output does not involve columns of A, and does not require A to be symmetric. We give similar algorithms for best rank-k approximation subject to the constraint of symmetry. We also show that there are asymmetric input matrices that cannot have good symmetric column-selected approximations.

.

[172] Sivakanth Gopi, Swastik Kopparty, Rafael Oliveira, Noga Ron-Zewi and Shubhangi Saraf. Locally testable and locally correctable codes approaching the Gilbert-Varshamov bound. **SODA 10C** (Nelva). January 18th, 2017, 16:25-16:45

One of the most important open problems in the theory of error-correcting codes is to determine the tradeoff between the rate R and minimum distance δ of a binary code. The best known tradeoff is the Gilbert-Varshamov bound, and says that for every $\delta \in (0, 1/2)$, there are codes with minimum distance δ and rate $R = R_{\mathsf{GV}}(\delta) > 0$ (for a certain simple function $R_{\mathsf{GV}}(\cdot)$). In this paper we show that the Gilbert-Varshamov bound can be achieved by codes which support *local* error-detection and error-correction algorithms. Specifically, we show the following results.

1. **Local Testing:** For all $\delta \in (0, 1/2)$ and all $R < R_{\mathsf{GV}}(\delta)$, there exist codes with length n, rate R and minimum distance δ that are locally testable with quasipolylog(n) query complexity.
2. **Local Correction:** For all $\epsilon > 0$, for all $\delta < 1/2$ sufficiently large, and all $R < (1-\epsilon)R_{\mathsf{GV}}(\delta)$, there exist codes with length n, rate R and minimum distance δ that are locally correctable from $\frac{\delta}{2} - o(1)$ fraction errors with $O(n^{\epsilon})$ query complexity.

Furthermore, these codes have an efficient randomized construction, and the local testing and local correction algorithms can be made to run in time polynomial in the query complexity. Our results on locally correctable codes also immediately give locally decodable codes with the same parameters. Our local testing result is obtained by combining Thommesen's random concatenation technique and the best known locally testable codes from Kopparty et al. (2016). Our local correction result, which is significantly more involved, also uses random concatenation, along with a number of further ideas: the Guruswami-Sudan-Indyk list decoding strategy for concatenated codes, Alon-Edmonds-Luby distance amplification, and the local list-decodability, local list-recoverability and local testability of Reed-Muller codes. Curiously, our final local correction algorithms go via local list-decoding and local testing algorithms; this seems to be the first time local testability is used in the construction of a locally correctable code.

.

[173] Parikshit Gopalan, Guangda Hu, Swastik Kopparty, Shubhangi Saraf, Carol Wang and Sergey Yekhanin. Maximally Recoverable Codes for Grid-like Topologies. **SODA 10C** (Nelva). January 18th, 2017, 16:50-17:10

The explosion in the volumes of data being stored online has resulted in distributed storage systems transitioning to erasure coding based schemes. Yet, the codes being deployed in practice are fairly short. In this work, we address what we view as the main coding theoretic barrier to deploying longer codes in storage: at large lengths, failures are not independent and correlated failures are inevitable. This motivates designing codes that allow quick data recovery even after large correlated failures, and which have efficient encoding and decoding. We propose that code design for distributed storage be viewed as a two step process. The first step is choose a *topology* of the code, which incorporates knowledge about the correlate d failures that need to be handled, and ensures local recovery from such failures. In the second step one specifies a code with the chosen topology by choosing coefficients from a finite field $\mathbb{F}_q$. In this step, one tries to balance reliability (which is better over larger fields) with encoding and decoding efficiency (which is better over smaller fields). This work initiates an in-depth study of this reliability/efficiency tradeoff. We consider the field-size needed for achieving *maximal recoverability*: the strongest reliability possible with a given topology. We propose a family of topologies called grid-like topologies which unify a number of topologies considered both in theory and practice, and prove the following results about codes for such topologies:

- The first super-polynomial lower bound on the field size needed for achieving maximal recoverability in a simple grid-like topology. To our knowledge, there was no super-linear lower bound known before, for any topology.
- A combinatorial characterization of erasure patterns correctable by Maximally Recoverable codes for a topology which corresponds to tensoring MDS codes with a parity check code. This topology is used in practice (for instance see [MLRH14]). We conjecture a similar characterization for Maximally Recoverable codes instantiating arbitrary tensor product topologies.

.

[174] Venkatesan Guruswami and Ankit Singh Rawat. MDS Code Constructions with Small Sub-packetization and Near-optimal Repair Bandwidth. **SODA 10C** (Nelva). January 18th, 2017, 17:15-17:35

An (n, M) vector code $\mathcal{C} \subseteq \mathbb{F}^n$ is a collection of M codewords where n elements (from the field $\mathbb{F}$) in each of the codewords are referred to as code blocks. Assuming that $\mathbb{F} \cong \mathbb{B}^{\ell}$, the code blocks are treated as ℓ-length vectors over the base field $\mathbb{B}$. Equivalently, the code is said to have the sub-packetization level ℓ. This paper addresses the problem of constructing MDS vector codes which enable exact reconstruction of each code block by downloading small amount of information from the remaining code blocks. The repair bandwidth of a code measures the information flow from the remaining code blocks during the reconstruction of a single code block. This problem naturally arises in the context of distributed storage systems as the node repair problem [Dimakis et al.]. Assuming that $M = |\mathbb{B}|^{k\ell}$, the repair bandwidth of an MDS vector code is lower bounded by $\big((n-1)/(n-k)\big) \cdot \ell$ symbols (over the base field $\mathbb{B}$) which is also referred to as the cut-set bound [Dimakis et al.]. For all values of n and k, the MDS vector codes that attain the cut-set bound with the sub-packetization level $\ell = (n-k)^{\lceil n/(n-k) \rceil}$ are known in the literature [Sasidharan et al., Ye and Barg]. This paper presents a construction for MDS vector codes which simultaneously ensures both small repair bandwidth and small sub-packetization level. The obtained codes have the smallest possible sub-packetization level $\ell = O(n-k)$ for an MDS vector code and the repair bandwidth which is at most twice the cut-set bound. The paper then generalizes this code construction so that the repair bandwidth of the obtained codes approach the cut-set bound at the cost of increased sub-packetization level. The constructions presented in this paper give MDS vector codes which are linear over the base field $\mathbb{B}$.

.

[175] Bernhard Haeupler and Ameya Velingker. Bridging the Capacity Gap Between Interactive and One-Way Communication. **SODA 10C** (Nelva). January 18th, 2017, 17:40-18:00

We study the communication rate of coding schemes for interactive communication that transform any two-party interactive protocol into a protocol that is robust to noise. Recently, Haeupler showed that if an $\epsilon > 0$ fraction of transmissions are corrupted, adversarially or randomly, then it is possible to achieve a communication rate of $1-\widetilde{O}(\sqrt{\epsilon})$. Furthermore, Haeupler conjectured that this rate is optimal for general input protocols. This stands in contrast to the classical setting of one-way communication in which error-correcting codes are known to achieve an optimal communication rate of $1 - \Theta(H(\epsilon)) = 1 - \widetilde{\Theta}(\epsilon)$. In this work, we show that the quadratically smaller rate loss of the one-way setting can also be achieved in interactive coding schemes for a very natural class of input protocols. We introduce the notion of *average message length*, or the average number of bits a party sends before receiving a reply, as a natural parameter for measuring the level of interactivity in a protocol. Moreover, we show that any protocol with average message length $\ell = \Omega(\text{poly}(1/\epsilon))$ can be simulated by a protocol with optimal communication rate $1 - \Theta(H(\epsilon))$ over an oblivious adversarial channel with error fraction ϵ. Furthermore, under the additional assumption of access to public shared randomness, the optimal communication rate is achieved *ratelessly*, i.e., the communication rate adapts automatically to the actual error rate ϵ without having to specify it in advance.

This shows that the capacity gap between one-way and interactive communication can be bridged even for very small (constant in ϵ) average message lengths, which are likely to be found in many applications.

.

[176] Sergio Cabello. Subquadratic Algorithms for the Diameter and the Sum of Pairwise Distances in Planar Graphs. **SODA Awards** (Las Arenas II-IV). January 18th, 2017, 18:00-18:20

We show how to compute in $O(n^{11/6} \operatorname{polylog}(n))$ expected time the diameter and the sum of the pairwise distances in an undirected planar graph with n vertices and positive edge weights. These are the first algorithms for these problems using time $O(n^c)$ for some constant $c < 2$.

.

[177] Ami Paz and Gregory Schwartzman. A $(2 + \epsilon)$-Approximation for Maximum Weight Matching in the Semi-Streaming Model. **SODA Awards** (Las Arenas II-IV). January 18th, 2017, 18:25-18:45

We present a simple deterministic single-pass $(2+\epsilon)$-approximation algorithm for the maximum weight matching problem in the semi-streaming model. This improves upon the currently best known approximation ratio of $(3.5+\epsilon)$. Our algorithm uses $O(n \log^2 n)$ space for constant values of ϵ. It relies on a variation of the local-ratio theorem, which may be of independent interest in the semi-streaming model.

.

[178] Nikhil Bansal. Uncertainty and Optimization. **Invited Talk 4** (Las Arenas II-IV). January 19th, 2017, 09:00-10:00

In many settings the input arrives over time and an algorithm must make its current decisions without a precise knowledge of the future. Traditionally, these problems have been studied using mostly disjoint approaches such as competitive anlaysis, regret minimization and under stochastic assumptions. But in recent years, several interesting connections have emerged, often based on methods from optimization and duality. In this talk, we will outline some of these developments as well as some future research directions.

.

[179] Jiawei Gao, Russell Impagliazzo, Antonina Kolokolova and Ryan Williams. Completeness for First-Order Properties on Sparse Structures with Algorithmic Applications. **SODA 11A** (Las Arenas II-IV). January 19th, 2017, 10:30-10:50

Properties definable in first-order logic are algorithmically interesting for both theoretical and pragmatic reasons. Many of the most studied algorithmic problems, such as Hitting Set and Orthogonal Vectors, are first-order, and the first-order properties naturally arise as relational database queries. A relatively straightforward algorithm for evaluating a property with $k + 1$ quantifiers takes time $O(m^k)$ and, assuming the Strong Exponential Time Hypothesis (SETH), some such properties require $O(m^{k-\epsilon})$ time for any $\epsilon > 0$. (Here, m represents the size of the input structure, i.e. the number of tuples in all relations.) We give algorithms for every first-order property that improves this upper bound to $m^k/2^{\Theta(\sqrt{\log n})}$, i.e., an improvement by a factor more than any poly-log, but less than the polynomial required to refute SETH. Moreover, we show that further improvement is *equivalent* to improving algorithms for sparse instances of the well-studied Orthogonal Vectors problem. Surprisingly, both results are obtained by showing completeness of the Sparse Orthogonal Vectors problem for the class of first-order properties under fine-grained reductions. To obtain improved algorithms, we apply the fast Orthogonal Vectors algorithm of (A. Abboud, R. Williams, and H. Yu. *More applications of the polynomial method to algorithm design.* In Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 218-230. SIAM, 2015), (T. M. Chan and R. Williams. *Deterministic APSP, Orthogonal Vectors, and More: Quickly derandomizing Razborov-Smolensky.*

In Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, pages 1246-1255. SIAM, 2016). While fine-grained reductions (reductions that closely preserve the conjectured complexities of problems) have been used to relate the hardness of disparate specific problems both within P and beyond, this is the first such completeness result for a standard complexity class.

.

[180] Kasper Green Larsen and Ryan Williams. Faster Online Matrix-Vector Multiplication. **SODA 11A** (Las Arenas II-IV). January 19th, 2017, 10:55-11:15

We consider the *Online Boolean Matrix-Vector Multiplication* (OMV) problem studied by Henzinger *et al.* [STOC'15]: given an $n \times n$ Boolean matrix M, we receive n Boolean vectors $v_1, \ldots, v_n$ one at a time, and are required to output Mv_i (over the Boolean semiring) before seeing the vector v_{i+1}, for all i. Previous known algorithms for this problem are combinatorial, running in $O(n^3/\log^2 n)$ time. Henzinger *et al.* conjecture there is no $O(n^{3-\varepsilon})$ time algorithm for OMV, for all $\varepsilon > 0$; their OMV conjecture is shown to imply strong hardness results for many basic dynamic problems. We give a substantially faster method for computing OMV, running in $n^3/2^{\Omega(\sqrt{\log n})}$ randomized time. In fact, after seeing $2^{\omega(\sqrt{\log n})}$ vectors, we already achieve $n^2/2^{\Omega(\sqrt{\log n})}$ amortized time for matrix-vector multiplication. Our approach gives a way to reduce matrix-vector multiplication to solving a version of the Orthogonal Vectors problem, which in turn reduces to "small" algebraic matrix-matrix multiplication. Applications include faster independent set detection, partial match retrieval, and 2-CNF evaluation. We also show how a modification of our method gives a cell probe data structure for OMV with worst case $O(n^{7/4}/\sqrt{w})$ time per query vector, where w is the word size. This result rules out an unconditional proof of the OMV conjecture using purely information-theoretic arguments.

.

[181] Daniel Lokshtanov, Ramamohan Paturi, Suguru Tamaki, Ryan Williams and Huacheng Yu. Beating Brute Force for Systems of Polynomial Equations over Finite Fields. **SODA 11A** (Las Arenas II-IV). January 19th, 2017, 11:20-11:40

We consider the problem of solving *systems of multivariate polynomial equations of degree* k over a finite field. For every integer $k \geq 2$ and finite field $\mathbb{F}_q$ where $q = p^d$ for a prime p, we give, to the best of our knowledge, the first algorithms that achieve an exponential speedup over the brute force $O(q^n)$ time algorithm in the worst case. We present two algorithms, a randomized algorithm with running time $q^{n+o(n)} \cdot q^{-n/O(k)}$ time if $q \leq 2^{4ekd}$, and $q^{n+o(n)} \cdot (\frac{\log q}{dek})^{-dn}$ otherwise, where $e = 2.718\ldots$ is Napier's constant, and a deterministic algorithm for *counting* solutions with running time $q^{n+o(n)} \cdot q^{-n/O(kq^{6/7d})}$. For the important special case of quadratic equations in $\mathbb{F}_2$, our randomized algorithm has running time $O(2^{0.8765n})$. For systems over $GF(2)$ we also consider the case where the input polynomials do not have bounded degree, but instead can be efficiently represented as a $\Sigma\Pi\Sigma$ circuit, i.e., a sum of products of sums of variables. For this case we present a deterministic algorithm running in time $2^{n-\delta n}$ for $\delta = 1/O(\log(s/n))$ for instances with s product gates in total and n variables. Our algorithms adapt several techniques recently developed via the polynomial method from circuit complexity. The algorithm for systems of $\Sigma\Pi\Sigma$ polynomials also introduces a new *degree reduction* method that takes an instance of the problem and outputs a subexponential-sized set of instances, in such a way that feasibility is preserved and every polynomial among the output instances has degree $O(\log(s/n))$.

.

[182] Massimo Cairo and Romeo Rizzi. The Complexity of Simulation and Matrix Multiplication. **SODA 11A** (Las Arenas II-IV). January 19th, 2017, 11:45-12:05

Computing the simulation preorder of a given Kripke structure (i.e., a directed graph with n labeled vertices) has crucial applications in model checking of temporal logic. It amounts to solving a specific two-players reachability game, called simulation game. We offer the first conditional lower bounds for this problem, and we relate its complexity (for computation, verification, and certification) to some variants of $n \times n$ matrix multiplication. We show that any $O(n^\alpha)$-time algorithm for simulation games, even restricting to acyclic games/structures, can be used to compute $n \times n$ boolean matrix multiplication (BMM) in $O(n^\alpha)$ time. In the acyclic case, we match this bound by presenting the first subcubic algorithm, based on fast BMM, and running in $n^{\omega+o(1)}$ time (where $\omega < 2.376$ is the exponent of matrix multiplication). For both acyclic and cyclic structures, we point out the existence of natural and canonical $O(n^2)$-size certificates, that can be verified in truly subcubic time by means of matrix multiplication. In the acyclic case, $O(n^2)$ time is sufficient, employing standard $(+, \times)$-matrix product verification. In the cyclic case, a $\min$-edge witness matrix multiplication (EWMM) is used, i.e., a matrix multiplication on the semi-ring $(\max, \times)$ where one matrix contains only 0's and 1's, which is computable in truly subcubic $n^{(3+\omega)/2+o(1)}$ time. Finally, we show a reduction from EWMM to cyclic simulation games which implies a separation between the cyclic and the acyclic cases, unless EWMM can be verified in $n^{\omega+o(1)}$ time.

.

[183] Arturs Backurs, Piotr Indyk and Ludwig Schmidt. Better Approximations for Tree Sparsity in Nearly-Linear Time. **SODA 11A** (Las Arenas II-IV). January 19th, 2017, 12:10-12:30

The Tree Sparsity problem is defined as follows: given a node-weighted tree of size n and an integer k, output a rooted subtree of size k with maximum weight. The best known algorithm solves this problem in time $O(kn)$, i.e., quadratic in the size of the input tree for $k = \Theta(n)$. In this work, we design $(1 + \varepsilon)$-approximation algorithms for the Tree Sparsity problem that run in nearly-linear time. Unlike prior algorithms for this problem, our results offer single criterion approximations, i.e., they do not increase the sparsity of the output solution, and work for arbitrary trees (not only balanced trees). We also provide further algorithms for this problem with different runtime vs approximation trade-offs. Finally, we show that if the exact version of the Tree Sparsity problem can be solved in strongly subquadratic time, then the $(\min, +)$ convolution problem can be solved in strongly subquadratic time as well. The latter is a well-studied problem for which no strongly subquadratic time algorithm is known.

.

[184] Marc Lelarge. Counting matchings in irregular bipartite graphs and random lifts. **SODA 11B** (Las Arenas I). January 19th, 2017, 10:30-10:50

We give a sharp lower bound on the number of matchings of a given size in a bipartite graph. When specialized to regular bipartite graphs, our results imply Schrijver's theorem and Friedland's Lower Matching Conjecture proven by Gurvits and Csikvári. Indeed, our work extends the recent work of Csikvári done for regular and bi-regular bipartite graphs. Moreover, our lower bounds are order optimal as they are attained for a sequence of 2-lifts of the original graph as well as for random n-lifts of the original graph when n tends to infinity. We then extend our results to permanents and subpermanents sums. For permanents, we are able to recover the lower bound of Schrijver recently proved by Gurvits using stable polynomials. Our proof is algorithmic and borrows ideas from the theory of local weak convergence of graphs, statistical physics and covers of graphs. We provide new lower bounds for subpermanents sums and obtain new results on the number of matchings in random n-lifts with some implications for the matching measure and the spectral measure of random n-lifts as well as for the spectral measure of infinite trees.

.

[185] Torsten Mütze and Jerri Nummenpalo. A constant-time algorithm for middle levels Gray codes. **SODA 11B** (Las Arenas I). January 19th, 2017, 10:55-11:15

For any integer $n \geq 1$ a *middle levels Gray code* is a cyclic listing of all n-element and $(n + 1)$-element subsets of $\{1, 2, \ldots, 2n+1\}$ such that any two consecutive subsets differ in adding or removing a single element. The question whether such a Gray code exists for any $n \geq 1$ has been the subject of intensive research during the last 30 years, and has been answered affirmatively only recently [T. Mütze. Proof of the middle levels conjecture. *Proc. London Math. Soc.*, 112(4):677–713, 2016]. In a follow-up paper [T. Mütze and J. Nummenpalo. An efficient algorithm for computing a middle levels Gray code. *Proc. ESA*, 2015] this existence proof was turned into an algorithm that computes each new set in the Gray code in time $\mathcal{O}(n)$ on average. In this work we complete this line of research by presenting an algorithm for computing a middle levels Gray code in optimal time and space: Each new set is generated in time $\mathcal{O}(1)$, and the required space is $\mathcal{O}(n)$.

.

[186] Chaya Keller, Shakhar Smorodinky and Gábor Tardos. On Max-Clique for intersection graphs of sets and the Hadwiger-Debrunner numbers. **SODA 11B** (Las Arenas I). January 19th, 2017, 11:20-11:40

Let $HD_d(p, q)$ denote the minimal size of a transversal that can always be guaranteed for a family of compact convex sets in $\mathbb{R}^d$ which satisfy the (p, q)-property ($p \geq q \geq d + 1$). In a celebrated proof of the Hadwiger-Debrunner conjecture, Alon and Kleitman proved that $HD_d(p, q)$ exists for all $p \geq q \geq d + 1$. Specifically, they prove that $HD_d(p, d + 1)$ is $\tilde{O}(p^{d^2+d})$. This paper has two parts. In the first part we present several improved bounds on $HD_d(p, q)$. In particular, we obtain the first near tight estimate of $HD_d(p, q)$ for an extended range of values of (p, q) since the 1957 Hadwiger-Debrunner theorem. In the second part we prove a $(p, 2)$-theorem for families in $\mathbb{R}^2$ with union complexity below a specific quadratic bound. Based on this, we introduce a polynomial time constant factor approximation algorithm for MAX-CLIQUE of intersection graphs of convex sets satisfying this property. It is not likely that our constant factor approximation can be improved to a PTAS as MAX-CLIQUE for intersection graphs of fat ellipses is known to be APX-HARD and fat ellipses have sub-quadratic union complexity.

.

[187] Victor Chepoi, Feodor Dragan and Yann Vaxès. Core congestion is inherent in hyperbolic networks. **SODA 11B** (Las Arenas I). January 19th, 2017, 11:45-12:05

We investigate the impact the negative curvature has on the traffic congestion in large-scale networks. We prove that every Gromov hyperbolic network G admits a core, thus answering in the positive a conjecture by Jonckheere, Lou, Bonahon, and Baryshnikov, Internet Mathematics, 7 (2011) which is based on the experimental observation by Narayan and Saniee, Physical Review E, 84 (2011) that real-world networks with small hyperbolicity have a core congestion. Namely, we prove that for every subset X of n vertices of a graph with δ-thin geodesic triangles (in particular, of a δ-hyperbolic graph) G there exists a vertex m of G such that the ball $B(m, 4\delta)$ of radius 4δ centered at m intercepts at least one half of the total flow between all pairs of vertices of X, where the flow between two vertices $x, y \in X$ is carried by geodesic (or quasi-geodesic) (x, y)-paths. Moreover, we prove a primal-dual result showing that, for any commodity graph R on X and any $r \geq 8\delta$, the size $\sigma_r(R)$ of the least r-multi-core (i.e., the number of balls of radius r) intercepting all pairs of R is upper bounded by the maximum number of pairwise $(2r - 5\delta)$-apart pairs of R and that an r-multi-core of size $\sigma_{r-5\delta}(R)$ can be computed in polynomial time for every finite set X. Our result about total r-multi-cores is based on a Helly-type theorem for quasiconvex sets in δ-hyperbolic graphs (this is our second main result). Namely, we show that for any finite collection $\mathcal{Q}$ of pairwise intersecting ϵ-quasiconvex sets of a δ-hyperbolic graph G there exists a single ball $B(c, 2\epsilon + 5\delta)$ intersecting all sets of $\mathcal{Q}$. More generally, we prove that if $\mathcal{Q}$ is a collection of $2r$-close (i.e., any two sets of $\mathcal{Q}$ are at distance $\leq 2r$) ϵ-quasiconvex sets of a δ-hyperbolic graph G, then there exists a ball $B(c, r^*)$ of radius $r^* := \max\{2\epsilon + 5\delta, r + \epsilon + 3\delta\}$ intersecting all sets of $\mathcal{Q}$. These kind of Helly-type results are also useful in geometric group theory. Using the Helly theorem for quasiconvex sets and a primal-dual approach, we show algorithmically that the minimum number of balls of radius $2\epsilon + 5\delta$ intersecting all sets of a family $\mathcal{Q}$ of ϵ-quasiconvex sets does not exceed the packing number of $\mathcal{Q}$ (maximum number of pairwise disjoint sets of $\mathcal{Q}$). We extend the covering and packing result to set-families ${}^{\kappa}\mathcal{Q}$ in which each set is a union of at most κ ϵ-quasiconvex sets of a δ-hyperbolic graph G. Namely, we show that if $r \geq \epsilon + 2\delta$ and $\pi_r({}^{\kappa}\mathcal{Q})$ is the maximum number of mutually $2r$-apart members of ${}^{\kappa}\mathcal{Q}$, then the minimum number of balls of radius $r + 2\epsilon + 6\delta$ intersecting all members of ${}^{\kappa}\mathcal{Q}$ is at most $2\kappa^2\pi_r({}^{\kappa}\mathcal{Q})$ and such a hitting set and a packing can be constructed in polynomial time for every finite ${}^{\kappa}\mathcal{Q}$ (this is our third main result). For set-families consisting of unions of κ balls in δ-hyperbolic graphs a similar result was obtained by Chepoi and Estellon (2007). In case of $\delta = 0$ (trees) and $\epsilon = r = 0$, (subtrees of a tree) we recover the result of Alon (2002) about the transversal and packing numbers of a set-family in which each set is a union of at most κ subtrees of a tree.

.

[188] Josef Cibulka and Jan Kynčl. Better upper bounds on the Füredi–Hajnal limits of permutations. **SODA 11B** (Las Arenas I). January 19th, 2017, 12:10-12:30

A *binary matrix* is a matrix with entries from the set $\{0, 1\}$. We say that a binary matrix A *contains* a binary matrix S if S can be obtained from A by removal of some rows, some columns, and changing some 1-entries to 0-entries. If A does not contain S, we say that A *avoids* S. A *k-permutation matrix* P is a binary $k \times k$ matrix with exactly one 1-entry in every row and one 1-entry in every column. The Füredi–Hajnal conjecture, proved by Marcus and Tardos, states that for every permutation matrix P, there is a constant c_P such that for every $n \in \mathbb{N}$, every $n \times n$ binary matrix A with at least $c_P n$ 1-entries contains P. We show that $c_P \leq 2^{O(k^{2/3} \log^{7/3} k / (\log\log k)^{1/3})}$ asymptotically almost surely for a random k-permutation matrix P. We also show that $c_P \leq 2^{(4+o(1))k}$ for every k-permutation matrix P, improving the constant in the exponent of a recent upper bound on c_P by Fox. We also consider a higher-dimensional generalization of the Stanley–Wilf conjecture about the number of d-dimensional n-permutation matrices avoiding a fixed d-dimensional k-permutation matrix, and prove almost matching upper and lower bounds of the form $(2^k)^{O(n)} \cdot (n!)^{d-1-1/(d-1)}$ and $n^{-O(k)} k^{\Omega(n)} \cdot (n!)^{d-1-1/(d-1)}$, respectively.

.

[189] Chien-Chung Huang and Telikepalli Kavitha. Popularity, Mixed Matchings, and Self-Duality. **SODA 11C** (Nelva). January 19th, 2017, 10:30-10:50

Our input instance is a bipartite graph $G = (A \cup B, E)$ where A is a set of applicants, B is a set of jobs, and each vertex $u \in A \cup B$ has a preference list ranking its neighbors in a strict order of preference. For any two matchings M and T in G, let $\phi(M, T)$ be the number of vertices that prefer M to T. A matching M is *popular* if $\phi(M, T) \geq \phi(T, M)$ for all matchings T in G. There is a utility function $w : E \to \mathbb{Q}$ and we consider the problem of matching applicants to jobs in a popular and utility-optimal manner. A popular *mixed matching* could have a much higher utility than all popular matchings, where a mixed matching is a probability distribution over matchings, i.e., a mixed matching $\Pi = \{(M_0, p_0), \ldots, (M_k, p_k)\}$ for some matchings $M_0, \ldots, M_k$ and $\sum_{i=0}^{k} p_i = 1$, $p_i \geq 0$ for all i. The function $\phi(\cdot, \cdot)$ easily extends to mixed matchings; a mixed matching Π is popular if $\phi(\Pi, \Lambda) \geq \phi(\Lambda, \Pi)$ for all mixed matchings Λ in G. Motivated by the fact that a popular mixed matching could have a much higher utility than all popular matchings, we study the popular fractional matching polytope $\mathcal{P}_G$. Our main result is that this polytope is half-integral and in the special case where a stable matching in G is a perfect matching, this polytope is integral. This implies that there is always a max-utility popular mixed matching Π such that $\Pi = \{(M_0, \frac{1}{2}), (M_1, \frac{1}{2})\}$ where M_0 and M_1 are matchings in G. As Π can be computed in polynomial time, an immediate consequence of our result is that in order to implement a max-utility popular mixed matching in G, we need just *a single* random bit. We analyze $\mathcal{P}_G$ whose description may have exponentially many constraints via an extended formulation with a linear number of constraints. The linear program that gives rise to this formulation has an unusual property: *self-duality*. In other words, this linear program is identical to its dual program. This is a rare case where an LP of a natural problem has such a property. The self-duality of this LP plays a crucial role in our proof of half-integrality of $\mathcal{P}_G$. We also show that our result carries over to the *roommates* problem, where the graph G need not be bipartite. The polytope of popular fractional matchings is still half-integral here and so we can compute a max-utility popular half-integral matching in G in polynomial time. To complement this result, we also show that the problem of computing a max-utility popular (integral) matching in a roommates instance is NP-hard.

.

[190] Shi Li. Constant Approximation Algorithm for Non-Uniform Capacitated Multi-Item Lot-Sizing via Strong Covering Inequalities. **SODA 11C** (Nelva). January 19th, 2017, 10:55-11:15

We study the non-uniform capacitated multi-item lot-sizing (CMILS) problem. In this problem, there is a set of demands over a planning horizon of T time periods and all demands must be satisfied on time. We can place an order at the beginning of each period s, incurring an ordering cost Ks. The total quantity of all products ordered at time s can not exceed a given capacity Cs. On the other hand, carrying inventory from time to time incurs inventory holding cost. The goal of the problem is to find a feasible solution that minimizes the sum of ordering and holding costs. Levi et al. (Levi, Lodi and Sviridenko, Math- matics of Operations Research 33(2), 2008) gave a 2-approximation for the problem when the capacities Cs are the same. In this paper, we extend their re- sult to the case of non-uniform capacities. That is, we give a constant approximation algorithm for the ca- pacitated multi-item lot-sizing problem with general capacities. The constant approximation is achieved by adding an exponentially large set of new covering in- equalities to the natural facility-location type linear programming relaxation for the problem. Along the way of our algorithm, we reduce the CMILS problem to two generalizations of the classic knapsack covering problem. We give LP-based constant approximation algorithms for both generalizations, via the iterative rounding technique.

.

[191] Abbas Bazzi, Samuel Fiorini, Sangxia Huang and Ola Svensson. Small Extended Formulation for Knapsack Cover Inequalities from Monotone Circuits. **SODA 11C** (Nelva). January 19th, 2017, 11:20-11:40

Initially developed for the min-knapsack problem, the knapsack cover inequalities are used in the current best relaxations for numerous combinatorial optimization problems of covering type. In spite of their widespread use, these inequalities yield linear programming (LP) relaxations of exponential size, over which it is not known how to optimize exactly in polynomial time. In this paper we address this issue and obtain LP relaxations of quasi-polynomial size that are at least as strong as that given by the knapsack cover inequalities. For the min-knapsack cover problem, our main result can be stated formally as follows: for any $\varepsilon > 0$, there is a $(1/\varepsilon)^{O(1)} n^{O(\log n)}$-size LP relaxation with an integrality gap of at most $2 + \varepsilon$, where n is the number of items. Prior to this work, there was no known relaxation of subexponential size with a constant upper bound on the integrality gap. Our construction is inspired by a connection between extended formulations and monotone circuit complexity via Karchmer-Wigderson games. In particular, our LP is based on $O(\log^2 n)$-depth monotone circuits with fan-in 2 for evaluating weighted threshold functions with n inputs, as constructed by Beimel and Weinreb. We believe that a further understanding of this connection may lead to more positive results complementing the numerous lower bounds recently proved for extended formulations.

.

[192] Robert Hildebrand, Robert Weismantel and Rico Zenklusen. Extension Complexity Lower Bounds for Mixed-Integer Extended Formulations. **SODA 11C** (Nelva). January 19th, 2017, 11:45-12:05

We prove that any mixed-integer linear extended formulation for the matching polytope of the complete graph on n vertices, with a polynomial number of constraints, requires $\Omega(\sqrt{\frac{n}{\log n}})$ many integer variables. By known reductions, this result extends to the traveling salesman polytope. This lower bound has various implications regarding the existence of small mixed-integer mathematical formulations of common problems in operations research. In particular, it shows that for many classic vehicle routing problems and problems involving matchings, any compact mixed-integer linear description of such a problem requires a large number of integer variables. This provides a first non-trivial lower bound on the number of integer variables needed in such settings.

.

[193] Avrim Blum, Ioannis Caragiannis, Nika Haghtalab, Ariel D. Procaccia, Eviatar B. Procaccia and Rohit Vaish. Opting Into Optimal Matchings. **SODA 11C** (Nelva). January 19th, 2017, 12:10-12:30

We revisit the problem of designing optimal, *individually rational* matching mechanisms (in a general sense, allowing for cycles in directed graphs), where each player—who is associated with a subset of vertices—matches as many of his own vertices when he opts into the matching mechanism as when he opts out. We offer a new perspective on this problem by considering an arbitrary graph, but assuming that vertices are associated with players at random. Our main result asserts that, under certain conditions, *any* fixed optimal matching is likely to be individually rational up to lower-order terms. We also show that a simple and practical mechanism is (fully) individually rational, and likely to be optimal up to lower-order terms. We discuss the implications of our results for market design in general, and kidney exchange in particular.

.

[194] David Adjiashvili, Andrea Baggio and Rico Zenklusen. Firefighting on Trees Beyond Integrality Gaps. **SODA 12A** (Las Arenas II-IV). January 19th, 2017, 14:00-14:20

The Firefighter problem and a variant of it, known as Resource Minimization for Fire Containment (RMFC), are natural models for optimal inhibition of harmful spreading processes. Despite considerable progress on several fronts, the approx- imability of these problems is still badly understood. This is the case even when the underlying graph is a tree, which is one of the most-studied graph structures in this context and the focus of this paper. In their simplest version, a fire spreads from one fixed vertex step by step from burning to adjacent non-burning vertices, and at each time step B many non-burning vertices can be protected from catching fire. The Firefighter problem asks, for a given B, to maximize the number of vertices that will not catch fire, whereas RMFC (on a tree) asks to find the smallest B that allows for saving all leaves of the tree. Prior to this work, the best known approximation ratios were an $O(1)$-approximation for the Firefighter problem and an $O(\log^* n)$-approximation for RMFC, both being LP-based and essentially matching the integrality gaps of two natural LP relaxations. We improve on both approximations by presenting a PTAS for the Firefighter problem and an O(1)-approximation for RMFC, both qualitatively matching the known hardness results. Our results are obtained through a combination of the known LPs with several new techniques, which allow for efficiently enumerating over super-constant size sets of constraints to strengthen the natural LPs.

.

[195] David Adjiashvili. Beating Approximation Factor Two for Weighted Tree Augmentation with Bounded Costs. **SODA 12A** (Las Arenas II-IV). January 19th, 2017, 14:25-14:45

The Weighted Tree Augmentation Problem (WTAP) is a fundamental well-studied problem in the field of network design. Given an undirected tree $G = (V, E)$, an additional set of edges $L \subseteq V \times V$ disjoint from E called *links* and a cost vector $c \in \mathbb{R}^L_{\geq 0}$, WTAP asks to find a minimum-cost set $F \subseteq L$ with the property that $(V, E \cup F)$ is 2-edge connected. The special case where $c_\ell = 1$ for all $\ell \in L$ is called the Tree Augmentation Problem (TAP). For the class of bounded cost vectors, we present a first improved approximation algorithm for WTAP since more than three decades. Concretely, for any $M \in \mathbb{R}_{\geq 1}$ and $\epsilon > 0$, we present an LP based $(\delta + \epsilon)$-approximation for WTAP restricted to cost vectors c in $[1, M]^L$ for $\delta \approx 1.96417$. More generally, our result is a $(\delta + \epsilon)$-approximation algorithm with running time $n^{r^{O(1)}}$, where $r = c_{\max}/c_{\min}$ is the ratio between the largest and the smallest cost of any link. For the special case of TAP we improve this factor to $\frac{5}{3} + \epsilon$. Our results rely on several new ideas, including a new LP relaxation of WTAP and a two-phase rounding algorithm. In the first phase, the algorithm uses the fractional LP solution to guide a simple decomposition method that breaks the tree into well-structured trees and equips each tree with a part of the fraction LP solution. In the second phase, the fractional solution in each part of the decomposition is rounded to an integral solution with two rounding procedures, and the best outcome is included in the solution. One rounding procedure exploits the constraints in the new LP, while the other one exploits a connection to the Edge Cover Problem. We show that both procedures can not have a bad approximation guarantee simultaneously to obtain the claimed approximation factor.

.

[196] Niv Buchbinder, Roy Schwartz and Baruch Weizman. Simplex Transformations and the Multiway Cut Problem. **SODA 12A** (Las Arenas II-IV). January 19th, 2017, 14:50-15:10

We consider `Multiway Cut`, a basic graph partitioning problem in which the goal is to find the minimum weight collection of edges disconnecting a given set of special vertices called terminals. `Multiway Cut` admits a well known simplex embedding relaxation, where rounding this embedding is equivalent to partitioning the simplex. Current best known solutions to the problem are comprised of a mix of several different ingredients, resulting in intricate algorithms. Moreover, the best of these algorithms is too complex to fully analyze analytically and its approximation factor was verified using a computer. We propose a new approach to simplex partitioning and the `Multiway Cut` problem based on general transformations of the simplex that allow dependencies between the different variables. Our approach admits much simpler algorithms, and in addition yields an approximation guarantee for the `Multiway Cut` problem that (roughly) matches the current best computer verified approximation factor.

.

[197] Fabrizio Grandoni, Tobias Mömke, Andreas Wiese and Hang Zhou. To Augment or Not to Augment: Solving Unsplittable Flow on a Path by Creating Slack. **SODA 12A** (Las Arenas II-IV). January 19th, 2017, 15:15-15:35

In the Unsplittable Flow on a Path problem (UFP) we are given a path with non-negative edge capacities and a set of tasks, each one characterized by a subpath, a demand, and a profit. Our goal is to select a subset of tasks of maximum total profit so that the total demand of the selected tasks on each edge does not exceed the respective edge capacity. UFP naturally captures several applications in bandwidth allocation, job scheduling, and caching. Following a sequence of improvements, the current best polynomial time approximation factor for UFP is $2+\epsilon$ (Anagnostopoulos et al. SODA'14). UFP also admits a QPTAS (Bansal et al. STOC'06, Batra et al. SODA'15), and finding a PTAS is considered a challenging open problem. In this paper we make progress in the direction of the mentioned open problem. Informally, we introduce a technique to obtain real PTASs from PTASs with *resource augmentation* where edge capacities can be violated by a $1 + \epsilon$ factor. While unfortunately we do not have a resource-augmentation PTAS for the general case of UFP, for many relevant special cases we have such an algorithm or we provide one in this paper. For example, our approach leads to a PTAS for the rooted case of UFP, where all tasks share a common edge. This is one of the simplest natural restrictions of UFP where the best-known approximation was $2 + \epsilon$ (like for the general case). At a high level, our technique is to sacrifice a few tasks in the optimal solution (with a small loss of profit) in order to create a sufficient amount of *slack capacity* on each edge. This slack turns out to be large enough to substitute the additional capacity we would gain from resource augmentation. Crucial for our approach is that we obtain slack from tasks with relatively small and relatively large demand simultaneously. In all prior polynomial time approximation algorithms the sacrificed tasks came from only one of these two groups.

.

[198] Fabrice Benhamouda, Tancrède Lepoint, Claire Mathieu and Hang Zhou. Optimization of Bootstrapping in Circuits. **SODA 12A** (Las Arenas II-IV). January 19th, 2017, 15:40-16:00

In 2009, Gentry proposed the first Fully Homomorphic Encryption (FHE) scheme, an extremely powerful cryptographic primitive that enables to perform computations, i.e., to evaluate circuits, on encrypted data without decrypting them first. This has many applications, particularly in cloud computing. In all currently known FHE schemes, encryptions are associated with some (non-negative integer) noise level. At each evaluation of an AND gate, this noise level increases. This increase is problematic because decryption succeeds only if the noise level stays below some maximum level L at every gate of the circuit. To ensure that property, it is possible to perform an operation called *bootstrapping* to reduce the noise level. Though critical, boostrapping is a time-consuming operation. This expense motivates a new problem in discrete optimization: minimizing the number of bootstrappings in a circuit while still controlling the noise level. In this paper, we (1) formally define the *bootstrap problem*, (2) design a polynomial-time L-approximation algorithm using a novel method of rounding of a linear program, and (3) show a matching hardness result: $(L-\epsilon)$-inapproximability for any $\epsilon > 0$.

.

[199] Mohammad Ali Abam, Mark de Berg and Mohammad Javad Rezaei Seraji. Geodesic Spanners for Points on a Polyhedral Terrain. **SODA 12B** (Las Arenas I). January 19th, 2017, 14:00-14:20

Let S be a set of n points on a polyhedral terrain $T \in \mathbb{R}^3$, and let $\varepsilon > 0$ be a fixed constant. We prove that S admits a $(2+\varepsilon)$-spanner with $O(n \log n)$ edges with respect to the geodesic distance. This is the first spanner with constant spanning ratio and a near-linear number of edges for points on a terrain.? On our way to this result, we prove that any set of n weighted points in $\mathbb{R}^d$ admits an additively weighted $(2+\varepsilon)$-spanner with $O(n)$ edges; this improves the previously best known bound on the spanning ratio (which was $5+\varepsilon$), and almost matches the lower bound.

.

[200] Bettina Speckmann, Kevin Buchin and Tim Ophelders. Computing the Fréchet Distance between Real-Valued Surfaces. **SODA 12B** (Las Arenas I). January 19th, 2017, 14:25-14:45

The Fréchet distance is a well-studied measure for the similarity of shapes. While efficient algorithms for computing the Fréchet distance between curves exist, there are only few results on the Fréchet distance between surfaces. Recent work has shown that the Fréchet distance is computable between piecewise linear functions f and $g\colon M \to \mathbb{R}^k$ with M a triangulated surface of genus zero. We focus on the case $k = 1$ and M being a topological sphere or disk with constant boundary. Intuitively, we measure the distance between terrains based solely on the height function. Our main result is that in this case computing the Fréchet distance between f and g is in NP. We additionally show that already for $k = 1$, computing a factor $2-\varepsilon$ approximation of the Fréchet distance is NP-hard, showing that this problem is in fact NP-complete. We also define an intermediate distance, between contour trees, which we also show to be NP-complete to compute. Finally, we discuss how our and other distance measures between contour trees relate to each other.

.

[201] Micha Sharir and Noam Solomon. Incidences with curves and surfaces in three dimensions, with applications to distinct and repeated distances. **SODA 12B** (Las Arenas I). January 19th, 2017, 14:50-15:10

We study a wide spectrum of incidence problems involving points and curves or points and surfaces in $\mathbb{R}^3$. The current (and in fact the only viable) approach to such problems, pioneered by Guth and Katz, requires a variety of tools from algebraic geometry, most notably (i) the polynomial partitioning technique, and (ii) the study of algebraic surfaces that are ruled by lines or, in more recent studies by Guth and Zahl, by algebraic curves of some constant degree. By exploiting and refining these tools, we obtain new and improved bounds for numerous incidence problems in $\mathbb{R}^3$. In broad terms, we consider two kinds of problems, those involving points and constant-degree algebraic *curves*, and those involving points and constant-degree algebraic *surfaces*. In some variants we assume that the points lie on some fixed constant-degree algebraic variety, and in others we consider arbitrary sets of points in 3-space. The case of points and curves has been considered in several previous studies, starting with Guth and Katz's work on points and lines. Our results, which are based on a recent work of Guth and Zahl concerning surfaces that are doubly ruled by curves, provide a grand generalization of all previous results. We reconstruct the bound for points and lines, and improve, in certain significant ways, recent bounds involving points and circles (by Sharir Sheffer and Zahl) and points and arbitrary constant-degree algebraic curves (by Sharir, Sheffer and Solomon). While in these latter instances the bounds are not known (and are strongly suspected not) to be tight, our bounds are, in a certain sense, the best that can be obtained with this approach, given the current state of knowledge. In the case of points and surfaces, the incidence graph between them can contain large complete bipartite graphs, each involving points on some curve and surfaces containing this curve (unlike earlier studies, we do not rule out this possibility, which makes our approach more general). Our bounds estimate the total size of the *vertex sets* in such a complete bipartite graph decomposition of the incidence graph. In favorable cases, our bounds translate into actual incidence bounds. Overall, here too our results can be regarded as providing a “grand generalization” of most of the previous studies of (special instances of) this problem. As applications of our point-surface incidence bounds, we consider the problems of distinct and repeated distances determined by a set of n points in $\mathbb{R}^3$, two of the most celebrated open problems in combinatorial geometry. We obtain new and improved bounds for two special cases, one in which the points lie on some algebraic variety of constant degree, and one involving distances between pairs in $P_1 \times P_2$, where P_1 is contained in a variety and P_2 is arbitrary.

.

[202] Boris Aronov, Edward Y. Miller and Micha Sharir. Eliminating Depth Cycles among Triangles in Three Dimensions. **SODA 12B** (Las Arenas I). January 19th, 2017, 15:15-15:35

Given n non-vertical pairwise disjoint triangles in 3-space, their vertical depth (above/below) relation may contain cycles. We show that, for any $\varepsilon > 0$, the triangles can be cut into $O(n^{3/2+\varepsilon})$ pieces, where each piece is a connected semi-algebraic set whose description complexity depends only on the choice of ε, such that the depth relation among these pieces is now a proper partial order. This bound is nearly tight in the worst case. We are not aware of any previous study of this problem with a subquadratic bound on the number of pieces. This work extends the recent study by two of the authors on eliminating depth cycles among lines in 3-space. Our approach is again algebraic, and makes use of a recent variant of the polynomial partitioning technique, due to Guth, which leads to a recursive procedure for cutting the triangles. In contrast to the case of lines, our analysis here is considerably more involved, due to the two-dimensional nature of the objects being cut, so additional tools, from topology and algebra, need to be brought to bear. Our result essentially settles a 35-year-old open problem in computational geometry, motivated by hidden-surface removal in computer graphics.

.

[203] Haim Kaplan, Wolfgang Mulzer, Liam Roditty, Paul Seiferth and Micha Sharir. Dynamic Planar Voronoi Diagrams for General Distance Functions and their Algorithmic Applications. **SODA 12B** (Las Arenas I). January 19th, 2017, 15:40-16:00

We describe a new data structure for dynamic nearest neighbor queries in the plane with respect to a general family of distance functions that includes L_p-norms and additively weighted Euclidean distances, and for general (convex, pairwise disjoint) sites that have constant description complexity (line segments, disks, etc.). Our data structure has a polylogarithmic update and query time, improving an earlier data structure of Agarwal, Efrat and Sharir that required $O(n^\varepsilon)$ time for an update and $O(\log n)$ time for a query. Our data structure has numerous applications, and in all of them it gives faster algorithms, typically reducing an $O(n^\varepsilon)$ factor in the bounds to polylogarithmic. To further demonstrate its effectiveness, we give here two new applications: an efficient construction of a spanner in a disk intersection graph, and a data structure for efficient connectivity queries in a dynamic disk graph. To obtain this data structure, we combine and extend various techniques and obtain several side results that are of independent interest. Our data structure depends on the existence and an efficient construction of *"vertical" shallow cuttings* in arrangements of bivariate algebraic functions. We prove that an appropriate level in an arrangement of a random sample of a suitable size provides such a cutting. To compute it efficiently, we develop a randomized incremental construction algorithm for finding the lowest k levels in an arrangement of bivariate algebraic functions (we mostly consider here collections of functions whose lower envelope has linear complexity, as is the case in the dynamic nearest-neighbor context). To analyze this algorithm, we improve a longstanding bound on the combinatorial complexity of the vertical decomposition of these levels. Finally, to obtain our structure, we plug our vertical shallow cutting construction into Chan's algorithm for efficiently maintaining the lower envelope of a dynamic set of planes in $\mathbb{R}^3$. While doing this, we also revisit Chan's technique and present a variant that uses a single binary counter, with a simpler analysis and an improved amortized deletion time.

.

[204] Mohsen Ghaffari and Hsin-Hao Su. Distributed Degree Splitting, Edge Coloring, and Orientations. **SODA 12C** (Nelva). January 19th, 2017, 14:00-14:20

We study a family of closely-related distributed graph problems, which we call *degree splitting*, where roughly speaking the objective is to partition (or orient) the edges such that each node's degree is split almost uniformly. Our findings lead to answers for a number of problems, a sampling of which includes: (1) We present a polylog$n$ round deterministic algorithm for $(2\Delta - 1) \cdot (1 + o(1))$ *edge-coloring*, where Δ denotes the maximum degree. Modulo the $1 + o(1)$ factor, this settles one of the long-standing open problems of the area from the 1990's (see e.g. Panconesi and Srinivasan [PODC'92]). Indeed, a weaker requirement of $(2\Delta - 1) \cdot$ polylog(Δ) edge-coloring in polylog n rounds was asked for in the 4th open question in the *Distributed Graph Coloring* book by Barenboim and Elkin. (2) We show that *sinkless orientation*—i.e., orienting edges such that each node has at least one outgoing edge—on Δ-regular graphs can be solved in $O(\log_\Delta \log n)$ rounds randomized and in $O(\log_\Delta n)$ rounds deterministically. These prove the corresponding lower bounds by Brandt et al. [STOC'16] and Chang, Kopelowitz, and Pettie [FOCS'16] to be tight. Moreover, these show that sinkless orientation exhibits an exponential separation between its randomized and deterministic complexities, akin to the results of Chang et al. for Δ-coloring Δ-regular trees. (3) We present a randomized $O(\log^4 n)$ round algorithm for orienting a-arboricity graphs with maximum out-degree $a(1+\epsilon)$. This can be also turned into a decomposition into $a(1+\epsilon)$ forests when $a = \Omega(\log n)$ and into $a(1+\epsilon)$ pseduo-forests when $a = o(\log n)$. Obtaining an efficient distributed decomposition into less than $2a$ forests was stated as the 10th open problem in the book by Barenboim and Elkin.

.

[205] Arkadev Chattopadhyay, Michael Langberg, Shi Li and Atri Rudra. Tight Network Topology Dependent Bounds on Rounds of Communication. **SODA 12C** (Nelva). January 19th, 2017, 14:25-14:45

We prove tight network topology dependent bounds on the round complexity of computing well studied k-party functions such as set-disjointness and element distinctness. Unlike the usual case in the CONGEST model in distributed computing, we fix the function and then vary the underlying network topology. This complements the recent such results on total communication that have received some attention. We also present some applications to distributed graph computation problems. Our main contribution is a proof technique that allows us to reduce the problem on a general graph topology to a relevant two-party communication complexity problem. However, unlike many previous works that also used the same high level strategy, we do *not* reason about a two-party communication problem that is induced by a cut in the graph. To 'stitch' back the various lower bounds from the two party communication problems, we use the notion of timed graph that has seen prior use in network coding. Our reductions use some tools from Steiner tree packing and multi-commodity flow problems that have a delay constraint.

.

[206] Lucas Boczkowski, Amos Korman and Emanuele Natale. Minimizing Message Size in Stochastic Communication Patterns: Fast Self-Stabilizing Protocols with 3 bits. **SODA 12C** (Nelva). January 19th, 2017, 14:50-15:10

This paper considers the basic $\mathcal{PULL}$ model of communication, in which in each round, each agent extracts information from few randomly chosen agents. We seek to identify the smallest amount of information revealed in each interaction (message size) that nevertheless allows for efficient and robust computations of fundamental information dissemination tasks.

We focus on the *Majority Bit Dissemination* problem that considers a population of n agents, with a designated subset of *source agents*. Each source agent holds an *input bit* and each agent holds an *output bit*. The goal is to let all agents converge their output bits on the most frequent input bit of the sources (the *majority bit*). Note that the particular case of a single source agent corresponds to the classical problem of *Broadcast* (also termed *Rumor Spreading*). We concentrate on the severe fault-tolerant context of *self-stabilization*, in which a correct configuration must be reached eventually, despite all agents starting the execution with arbitrary initial states. In particular, the specification of who is a source and what is its initial input bit may be set by an adversary. We first design a general compiler which can essentially transform any self-stabilizing algorithm with a certain property (called "the *bitwise-independence property*") that uses ℓ-bits messages to one that uses only $\log \ell$-bits messages, while paying only a small penalty in the running time. By applying this compiler recursively we then obtain a self-stabilizing *Clock Synchronization* protocol, in which agents synchronize their clocks modulo some given integer T, within $\tilde{O}(\log n \log T)$ rounds w.h.p., and using messages that contain 3 bits only.

We then employ the new Clock Synchronization tool to obtain a self-stabilizing Majority Bit Dissemination protocol which converges in $\tilde{O}(\log n)$ time, w.h.p., on every initial configuration, provided that the ratio of sources supporting the minority opinion is bounded away from half. Moreover, this protocol also uses only 3 bits per interaction.

.

[207] Dan Alistarh, James Aspnes, David Eisenstat, Rati Gelashvili and Ronald L. Rivest. Time-space Trade-offs in Population Protocols. **SODA 12C** (Nelva). January 19th, 2017, 15:15-15:35

Population protocols are a popular model of distributed computing, in which randomly-interacting agents with little computational power cooperate to jointly perform computational tasks. Inspired by developments in molecular computation, and in particular DNA computing, recent algorithmic work has focused on the complexity of solving simple yet fundamental tasks in the population model, such as *leader election* (which requires convergence to a single agent in a special "leader" state), and *majority* (in which agents must converge to a decision as to which of two possible initial states had higher initial count). Known results point towards an inherent trade-off between the *time complexity* of such algorithms, and the *space complexity*, i.e. size of the memory available to each agent. In this paper, we explore this trade-off and provide new upper and lower bounds for majority and leader election. First, we prove a unified lower bound, which relates the space available per node with the time complexity achievable by a protocol: for instance, our result implies that any protocol solving either of these tasks for n agents using $O(\log \log n)$ states must take $\Omega(n/\text{polylog} n)$ expected time. This is the first result to characterize time complexity for protocols which employ super-constant number of states per node, and proves that fast, poly-logarithmic running times require protocols to have relatively large space costs.

On the positive side, we give algorithms showing that fast, poly-logarithmic convergence time can be achieved using $O(\log^2 n)$ space per node, in the case of both tasks. Overall, our results highlight a time complexity separation between $O(\log \log n)$ and $\Theta(\log^2 n)$ state space size for both majority and leader election in population protocols, and introduce new techniques, which should be applicable more broadly.

.

[208] Niv Buchbinder, Iftach Haitner, Nissan Levi and Eliad Tsfadia. Fair Coin Flipping: Tighter Analysis and the Many-Party Case. **SODA 12C** (Nelva). January 19th, 2017, 15:40-16:00

In a multi-party *fair* coin-flipping protocol, the parties output a common (close to) unbiased bit, even when some corrupted parties try to bias the output. In this work we focus on the case of dishonest majority, i.e. at least half of the parties can be corrupted. Cleve [STOC 1986] has shown that in *any* m-round coin-flipping protocol the corrupted parties can bias the honest parties' common output bit by $\Theta(1/m)$. For more than two decades the best known coin-flipping protocols against majority was the protocol of Awerbuch, Blum, Chor, Goldwasser and Michali [Manuscript 1985], who presented a t-party, m-round protocol with bias $\Theta(t/\sqrt{m})$. This was changed by the breakthrough result of Moran, Naor, and Segev [TCC 2009], who constructed an m-round, *two*-party coin-flipping protocol with optimal bias $\Theta(1/m)$. Recently, Haitner, and Tsfadia [STOC 14] constructed an m-round, *three*-party coin-flipping protocol with bias $O(\log^3 m/m)$. Still for the case of more than three parties, against arbitrary number of corruptions, the best known protocol remained the $\Theta(t/\sqrt{m})$-bias protocol of Awerbuch et al. We make a step towards eliminating the above gap, presenting a t-party, m-round coin-flipping protocol, with bias

$O(\frac{t^3 \cdot 2^t \cdot \sqrt{\log m}}{m^{1/2+1/\left(2^{t-1}-2\right)}})$. This improves upon the $\Theta(t/\sqrt{m})$-bias protocol of Awerbuch et al. for any $t \leq 1/2 \cdot \log\log m$, and in particular for $t \in O(1)$, this yields an $1/m^{\frac{1}{2}+\Theta(1)}$-bias protocol. For the three-party case, this yields an $O(\sqrt{\log m}/m)$-bias protocol, improving over the the $O(\log^3 m/m)$-bias protocol of Haitner, and Tsfadia. Our protocol generalizes that of Haitner and Tsfadia, by presenting an appropriate "defense protocols" for the remaining parties to interact in, in the case that some parties abort or caught cheating Haitner, Tsfadia, only presented a two-party defense protocol, which limits their final protocol to handle three parties). We analyze our new protocols by presenting a new paradigm for analyzing fairness of coin-flipping protocols. We map the set of adversarial strategies that try to bias the honest parties outcome in the protocol to the set of the feasible solutions of a linear program. The gain each strategy achieves is the value of the corresponding solution. We then bound the the optimal value of the linear program by constructing a feasible solution to its dual.

.

[209] Sungjin Im and Benjamin Moseley. Fair Scheduling via Iterative Quasi-Uniform Sampling. **SODA 13A** (Las Arenas II-IV). January 19th, 2017, 16:30-16:50

In the paper we consider minimizing the ℓ_k-norms of flow time on a single machine offline using a preemptive scheduler for $k \geq 1$. We show the first $O(1)$-approximation for the problem, improving upon the previous best $O(\log\log P)$-approximation by Bansal and Pruhs (FOCS 09 and SICOMP 14) where P is the ratio of the maximum job size to the minimum. Our main technical ingredient is a novel combination of quasi-uniform sampling and iterative rounding, which is of interest in its own right.

.

[210] Rebecca Hoberg and Thomas Rothvoss. A Logarithmic Additive Integrality Gap for Bin Packing. **SODA 13A** (Las Arenas II-IV). January 19th, 2017, 16:55-17:15

For *bin packing*, the input consists of n items with sizes $s_1, \ldots, s_n \in [0,1]$ which have to be assigned to a minimum number of bins of size 1. Recently, the second author gave an LP-based polynomial time algorithm that employed techniques from *discrepancy theory* to find a solution using at most $OPT + O(\log OPT \cdot \log\log OPT)$ bins. In this paper, we build on the techniques of Rothvoss to present an approximation algorithm that has an additive gap of only $O(\log OPT)$ bins. This gap matches certain combinatorial lower bounds, and any further improvement would have to use more algebraic structure.

.

[211] Mong-Jen Kao. Iterative Partial Rounding for Vertex Cover with Hard Capacities. **SODA 13A** (Las Arenas II-IV). January 19th, 2017, 17:20-17:40

We provide a simple and novel algorithmic design technique, for which we call *iterative partial rounding*, that gives a tight rounding-based approximation for vertex cover with hard capacities (VC-HC). In particular, we obtain an f-approximation for VC-HC on hypergraphs, improving over a previous results of Cheung et al (SODA 2014) to the tight extent. This also closes the gap of approximation since it was posted by Chuzhoy and Naor in (FOCS 2002). We believe that our rounding technique is of independence interests when hard constraints are considered. Our main technical tool for establishing the approximation guarantee is a separation lemma that certifies the existence of a strong partition for solutions that are basic feasible in an extended version of the natural LP.

.

[212] Sam Chiu-wai Wong. Tight Algorithms for Vertex Cover with Hard Capacities on Multigraphs and Hypergraphs. **SODA 13A** (Las Arenas II-IV). January 19th, 2017, 17:20-17:40

In this paper we give a f-approximation algorithm for the minimum unweighted Vertex Cover problem with Hard Capacity constraints (VCHC) on f-hypergraphs. This problem generalizes standard vertex cover for which the best known approximation ratio is also f and cannot be improved assuming the unique game conjecture. Our result is therefore essentially the best possible. This improves over the previous 2.155 (for $f = 2$) and $2f$ approximation algorithms by Cheung, Goemans and Wong (CGW). At the heart of our approach is to apply iterative rounding to a natural LP relaxation that is slightly different from prior works which used (non-iterative) rounding. Our algorithm is significantly simpler and offers an intuitive explanation why f-approximation can be achieved for VCHC. We also present faster implementations of our method based on iteratively rounding the solution to certain CGW-style covering LPs.

.

[213] Christos Kalaitzis, Ola Svensson and Jakub Tarnawski. Unrelated Machine Scheduling of Jobs with Uniform Smith Ratios. **SODA 13A** (Las Arenas II-IV). January 19th, 2017, 17:45-18:05

We consider the classic problem of scheduling jobs on unrelated machines so as to minimize the weighted sum of completion times. Recently, for a small constant $\varepsilon > 0$, Bansal et al. gave a $(3/2 - \varepsilon)$-approximation algorithm improving upon the "natural" barrier of $3/2$ which follows from independent randomized rounding. In simplified terms, their result is obtained by an enhancement of independent randomized rounding via strong negative correlation properties. In this work, we take a different approach and propose to use the same elegant rounding scheme for the weighted completion time objective as devised by Shmoys and Tardos for optimizing a linear function subject to makespan constraints. Our main result is a 1.21-approximation algorithm for the natural special case where the weight of a job is proportional to its processing time (specifically, all jobs have the same Smith ratio), which expresses the notion that each unit of work has the same weight. In addition, as a direct consequence of the rounding, our algorithm also achieves a bi-criteria 2-approximation for the makespan objective. Our technical contribution is a tight analysis of the expected cost of the solution compared to the one given by the Configuration-LP relaxation – we reduce this task to that of understanding certain worst-case instances which are simple to analyze.

.

[214] Klaus Jansen and Lars Rohwedder. On the Configuration-LP of the Restricted Assignment Problem. **SODA 13A** (Las Arenas II-IV). January 19th, 2017, 18:10-18:30

We consider the classical problem of Scheduling on Unrelated Machines. In this problem a set of jobs is to be distributed among a set of machines and the maximum load (makespan) is to be minimized. The processing time p_{ij} of a job j depends on the machine i it is assigned to. Lenstra, Shmoys and Tardos gave a polynomial time 2-approximation for this problem (J. K. Lenstra, D. B. Shmoys and É. Tardos, *Approximation Algorithms for Scheduling Unrelated Parallel Machines*, Mathematical Programming, 46(3):259-271, 1990). In this paper we focus on a prominent special case, the Restricted Assignment problem, in which $p_{ij} \in \{p_j, \infty\}$. The configuration-LP is a linear programming relaxation for the Restricted Assignment problem. It was shown by Svensson that the multiplicative gap between integral and fractional solution, the integrality gap, is at most $2 - 1/17 \approx 1.9412$ (O. Svensson, *Santa Claus Schedules Jobs on Unrelated Machines*, 41(5):1318-1341, 2012). In this paper we significantly simplify his proof and achieve a bound of $2 - 1/6 \approx 1.8333$. As a direct consequence this provides a polynomial $(2-1/6+\epsilon)$-estimation algorithm for the Restricted Assignment problem by approximating the configuration-LP. The best lower bound known for the integrality gap is 1.5 and no estimation algorithm with a guarantee better than 1.5 exists unless $\mathrm{P} = \mathrm{NP}$.

.

[215] Eden Chlamtáč, Michael Dinitz, Guy Kortsarz and Bundit Laekhanukit. Approximating Spanners and Directed Steiner Forest: Upper and Lower Bounds. **SODA 13B** (Las Arenas I). January 19th, 2017, 16:30-16:50

It was recently found that there are very close connections between the existence of *additive spanners* (subgraphs where all distances are preserved up to an additive stretch), *distance preservers* (subgraphs in which demand pairs have their distance preserved exactly), and *pairwise spanners* (subgraphs in which demand pairs have their distance preserved up to a multiplicative or additive stretch) [Abboud-Bodwin SODA '16, Bodwin-Williams SODA '16]. We study these problems from an optimization point of view, where rather than studying the existence of extremal instances we are given an instance and are asked to find the sparsest possible spanner/preserver. We give an $O(n^{3/5+\varepsilon})$-approximation for distance preservers and pairwise spanners (for arbitrary constant $\varepsilon > 0$). This is the first nontrivial upper bound for either problem, both of which are known to be as hard to approximate as Label Cover. We also prove Label Cover hardness for approximating additive spanners, even for the cases of additive 1 stretch (where one might expect a polylogarithmic approximation, since the related multiplicative 2-spanner problem admits an $O(\log n)$-approximation) and additive polylogarithmic stretch (where the related multiplicative spanner problem has an $O(1)$-approximation). Interestingly, the techniques we use in our approximation algorithm extend beyond distance-based problem to pure connectivity network design problems. In particular, our techniques allow us to give an $O(n^{3/5+\varepsilon})$-approximation for the Directed Steiner Forest problem (for arbitrary constant $\epsilon > 0$) when all edges have uniform costs, improving the previous best $O(n^{2/3+\varepsilon})$-approximation due to Berman et al. [ICALP '11] (which holds for general edge costs).

.

[216] Amir Abboud, Greg Bodwin and Seth Pettie. A Hierarchy of Lower Bounds for Sublinear Additive Spanners. **SODA 13B** (Las Arenas I). January 19th, 2017, 16:55-17:15

Spanners, emulators, and approximate distance oracles can be viewed as *lossy* compression schemes that represent an unweighted graph metric in small space, say $\tilde{O}(n^{1+\delta})$ bits. There is an inherent tradeoff between the sparsity parameter δ and the *stretch function* f of the compression scheme, but the qualitative nature of this tradeoff has remained a persistent open problem. It has been known for some time that when $\delta \geq 1/3$ there are schemes with constant *additive* stretch (distance d is stretched to at most $f(d) = d + O(1)$), and recent results of Abboud and Bodwin show that when $\delta < 1/3$ there are no such schemes. Thus, to get practically efficient graph compression with $\delta \to 0$ we must pay super-constant additive stretch, but exactly how much do we have to pay? In this paper we show that the lower bound of Abboud and Bodwin is just the first step in a *hierarchy* of lower bounds that characterize the asymptotic behavior of the optimal stretch function f for sparsity parameter $\delta \in (0, 1/3)$. Specifically, for any integer $k \geq 2$, any compression scheme with size $O(n^{1+\frac{1}{2^k-1}-\epsilon})$ has a *sublinear additive stretch* function f:

$$f(d) = d + \Omega(d^{1-\frac{1}{k}}).$$

This lower bound matches Thorup and Zwick's (2006) construction of sublinear additive *emulators*. It also shows that Elkin and Peleg's $(1+\epsilon, \beta)$-spanners have an essentially optimal tradeoff between δ, ϵ, and β, and that the sublinear additive spanners of Pettie (2009) and Chechik (2013) are not too far from optimal. To complement these lower bounds we present a new construction of $(1+\epsilon, O(k/\epsilon)^{k-1})$-spanners with size $O((k/\epsilon)^{h_k} k n^{1+\frac{1}{2^{k+1}-1}})$, where $h_k < 3/4$. This size bound improves on the spanners of Elkin and Peleg (2004), Thorup and Zwick (2006), and Pettie (2009). According to our lower bounds neither the size nor stretch function can be substantially improved.

.

[217] Anupam Gupta, R. Ravi, Kunal Talwar and Seeun William Umboh. LAST but not Least: Buy-at-Bulk via Online Spanners. **SODA 13B** (Las Arenas I). January 19th, 2017, 17:20-17:40

The online (uniform) buy-at-bulk network design problem asks us to design a network, where the edge-costs exhibit economy-of-scale. Previous approaches to this problem used tree- embeddings, giving us randomized algorithms. Moreover, the optimal results with a logarithmic competitive ratio requires the metric on which the network is being built to be known up-front; the competitive ratios then depend on the size of this metric (which could be much larger than the number of terminals that arrive). We consider the buy-at-bulk problem in the least restrictive model where the metric is not known in advance, but revealed in parts along with the demand points seeking connectivity arriving online. For the single sink buy-at-bulk problem, we give a deterministic online algorithm with competitive ratio that is logarithmic in k, the number of terminals that have arrived, matching the lower bound known even for the online Steiner tree problem. In the oblivious case when the buy-at-bulk function used to compute the edge-costs of the network is not known in advance (but is the same across all edges), we give a deterministic algorithm with competitive ratio polylogarithmic in k, the number of terminals. At the heart of our algorithms are optimal constructions for online Light Approximate Shortest-path Trees (LASTs) and spanners, and their variants. We give constructions that have optimal trade-offs in terms of cost and stretch. We also define and give constructions for a new notion of LASTs where the set of roots (in addition to the points) expands over time. We expect these techniques will find applications in other online network-design problems.

.

[218] Greg Bodwin. Linear Size Distance Preservers. **SODA 13B** (Las Arenas I). January 19th, 2017, 17:45-18:05

The famous shortest path tree lemma states that, for any node s in a graph $G = (V, E)$, there is a subgraph on $O(n)$ edges that preserves all distances between node pairs in the set $\{s\} \times V$. A very basic question in distance sketching research, with applications to other problems in the field, is to categorize when *else* graphs admit sparse subgraphs that preserve distances between a set P of p node pairs, where P has some different structure than $\{s\} \times V$ or possibly no guaranteed structure at all. Trivial lower bounds of a path or a clique show that such a subgraph will need $\Omega(n + p)$ edges in the worst case. The question is then to determine when these trivial lower bounds are sharp; that is, when do graphs have *linear size distance preservers* on $O(n+p)$ edges? In this paper, we make the first new progress on this fundamental question in over ten years. We show:

1. All G, P has a distance preserver on $O(n)$ edges whenever $p = O(n^{1/3})$, even if G is directed and/or weighted. These are the first nontrivial preservers of size $O(n)$ known for directed graphs.
2. All G, P has a distance preserver on $O(p)$ edges whenever $p = \Omega\left(\frac{n^2}{rs(n)}\right)$, and G is undirected and unweighted. Here, $rs(n)$ is the Ruzsa-Szemerédi function from combinatoric graph theory. These are the first nontrivial preservers of size $O(p)$ known in any setting.
3. To preserve distances within a subset of s nodes in a graph, $\omega(s^2)$ edges are sometimes needed when $s = o\left(\frac{n^{2/3}}{2^{\Theta(\sqrt{\log n \cdot \log\log n})}}\right)$ even if G is undirected and unweighted. For weighted graphs, the range of this lower bound improves to $s = o(n^{2/3})$. This result reflects a polynomial improvement over lower bounds given by Coppersmith and Elkin (SODA '05).

An interesting technical contribution in this paper is a new method for "lazily" breaking ties between equally short paths in a graph, which allows us to draw our new connections between distance sketching and the Ruzsa-Szemerédi problem.

.

[219] Michael Elkin and Ofer Neiman. Efficient Algorithms for Constructing Very Sparse Spanners and Emulators. **SODA 13B** (Las Arenas I). January 19th, 2017, 18:10-18:30

Miller, Peng, Vladu and Xu, (Improved Parallel Algorithms for Spanners and Hopsets), SPAA'15, devised (implicitly) a distributed algorithm in the CONGEST model, that given a parameter $k = 1, 2, \ldots$, constructs an $O(k)$-spanner of an input unweighted n-vertex graph with $O(n^{1+1/k})$ expected edges in $O(k)$ rounds of communication. In this paper we improve the result of Miller et al., by showing a k-round distributed algorithm in the same model, that constructs a $(2k-1)$-spanner with $O(n^{1+1/k}/\epsilon)$ edges, with probability $1-\epsilon$, for any $\epsilon > 0$. Moreover, when $k = \omega(\log n)$, our algorithm produces (still in k rounds) *ultra-sparse* spanners, i.e., spanners of size $n(1+o(1))$, with probability $1-o(1)$. To our knowledge, this is the first distributed algorithm in the CONGEST or in the PRAM models that constructs spanners or skeletons (i.e., connected spanning subgraphs) that sparse. Our algorithm can also be implemented in linear time in the standard centralized model, and for large k, it provides spanners that are sparser than any other spanner given by a known (near-)linear time algorithm. We also devise improved bounds (and algorithms realizing these bounds) for $(1+\epsilon, \beta)$-spanners and emulators. In particular, we show that for any unweighted n-vertex graph and any $\epsilon > 0$, there exists a $(1+\epsilon, (\frac{\log\log n}{\epsilon})^{\log\log n})$-emulator with $O(n)$ edges. All previous constructions of $(1+\epsilon, \beta)$-spanners and emulators employ a super-linear number of edges, for all choices of parameters. Finally, we provide some applications of our results to approximate shortest paths' computation in unweighted graphs.

.

[220] Nicholas J. Cavanna, Kirk P. Gardner and Donald R. Sheehy. When and Why the Topological Coverage Criterion Works. **SODA 13C** (Nelva). January 19th, 2017, 16:30-16:50

In their seminal work on homological sensor networks, de Silva and Ghrist (*Algebraic & Geometric Topology*, 7:339–358, 2007) showed the surprising fact that it's possible to certify the coverage of a coordinate-free sensor network even with very minimal knowledge of the space to be covered. Here, coverage means that every point in the domain (except possibly those very near the boundary) has a nearby sensor. More generally, their algorithm takes a pair of nested neighborhood graphs along with a labeling of vertices as either boundary or interior and computes the relative homology of a simplicial complex induced by the graphs. This approach, called the Topological Coverage Criterion (TCC), requires some assumptions about the underlying geometric domain as well as some assumptions about the relationship of the input graphs to the domain. The goal of this paper is to generalize these assumptions and show how the TCC can be applied to both much more general domains as well as very weak assumptions on the input. We give a new, simpler proof of the de Silva-Ghrist Topological Coverage Criterion that eliminates any assumptions about the smoothness of the boundary of the underlying space, allowing the results to be applied to much more general problems. The new proof factors the geometric, topological, and combinatorial aspects, allowing us to provide a coverage condition that supports thick boundaries, k-coverage, and weighted coverage, in which sensors have varying radii.

.

[221] Éric Colin de Verdière and Salman Parsa. Deciding Contractibility of a Non-Simple Curve on the Boundary of a 3-Manifold. **SODA 13C** (Nelva). January 19th, 2017, 16:55-17:15

We present an algorithm for the following problem. Given a triangulated 3-manifold M and a (possibly non-simple) closed curve on the boundary of M, decide whether this curve is contractible in M. Our algorithm is combinatorial and runs in exponential time. This is the first algorithm that is specifically designed for this problem; its running time considerably improves upon the existing bounds implicit in the literature for the more general problem of contractibility of closed curves in a 3-manifold. The proof of the correctness of the algorithm relies on methods of 3-manifold topology and in particular on those used in the proof of the Loop Theorem.

.

[222] Jean-Daniel Boissonnat and Karthik C. S. An Efficient Representation for Filtrations of Simplicial Complexes. **SODA 13C** (Nelva). January 19th, 2017, 17:20-17:40

A filtration over a simplicial complex K is an ordering of the simplices of K such that all prefixes in the ordering are subcomplexes of K. Filtrations are at the core of Persistent Homology, a major tool in Topological Data Analysis. In order to represent the filtration of a simplicial complex, the entire filtration can be appended to any data structure that explicitly stores all the simplices of the complex such as the Hasse diagram or the recently introduced Simplex Tree [Algorithmica '14]. However, with the popularity of various computational methods that need to handle simplicial complexes, and with the rapidly increasing size of the complexes, the task of finding a compact data structure that can still support efficient queries is of great interest. This direction has been recently pursued for the case of maintaining simplicial complexes. For instance, Boissonnat et al. [SoCG '15] considered storing the simplices that are maximal for the inclusion and Attali et al. [IJCGA '12] considered storing the simplices that block the expansion of the complex. Nevertheless, so far there has been no data structure that compactly stores the *filtration* of a simplicial complex, while also allowing the efficient implementation of basic operations on the complex. In this paper, we propose a new data structure called the Critical Simplex Diagram (CSD) which is a variant of the Simplex Array List (SAL) [SoCG '15]. Our data structure allows to store in a compact way the filtration of a simplicial complex, and allows for the efficient implementation of a large range of basic operations. Moreover, we prove that our data structure is essentially optimal with respect to the requisite storage space. Next, we show that the CSD representation admits the following construction algorithms.

- A new *edge-deletion* algorithm for the fast construction of Flag complexes, which only depends on the number of critical simplices and the number of vertices.
- A new *matrix-parsing* algorithm to quickly construct relaxed Delaunay complexes, depending only on the number of witnesses and the dimension of the complex.

.

[223] Clément Maria and Jonathan Spreer. A polynomial time algorithm to compute quantum invariants of 3-manifolds with bounded first Betti number. **SODA 13C** (Nelva). January 19th, 2017, 17:45-18:05

In this article, we introduce a fixed parameter tractable algorithm for computing the Turaev-Viro invariants $\mathrm{TV}_{4,q}$, using the dimension of the first homology group of the manifold as parameter. This is, to our knowledge, the first parameterised algorithm in computational 3-manifold topology using a topological parameter. The computation of $\mathrm{TV}_{4,q}$ is known to be #P-hard in general; using a topological parameter provides an algorithm polynomial in the size of the input triangulation for the extremely large family of 3-manifolds with first homology group of bounded rank. Our algorithm is easy to implement and running times are comparable with running times to compute integral homology groups for standard libraries of triangulated 3-manifolds. The invariants we can compute this way are powerful: in combination with integral homology and using standard data sets we are able to roughly double the pairs of 3-manifolds we can distinguish. We hope this qualifies $\mathrm{TV}_{4,q}$ to be added to the short list of standard properties (such as orientability, connectedness, Betti numbers, etc.) that can be computed ad-hoc when first investigating an unknown triangulation.

.

[224] Tamal Dey, Zhe Dong and Yusu Wang. Parameter-free Topology Inference and Sparsification for Data on Manifolds. **SODA 13C** (Nelva). January 19th, 2017, 18:10-18:30

In topology inference from data, current approaches face two major problems. One concerns the selection of a correct parameter to build an appropriate complex on top of the data points; the other involves with the typical 'large' size of this complex. We address these two issues in the context of inferring homology from sample points of a smooth manifold of known dimension sitting in an Euclidean space $\mathbb{R}^k$. We show that, for a sample size of n points, we can identify a set of $O(n^2)$ points (as opposed to $O(n^{\lceil \frac{k}{2} \rceil})$ Voronoi vertices) approximating a subset of the medial axis that suffices to compute a distance sandwiched between the well known *local feature size* and the local *weak feature size* (in fact, the approximating set can be further reduced in size to $O(n)$). This distance, called the *lean feature size*, helps pruning the input set at least to the level of local feature size while making the data locally uniform. The local uniformity in turn helps in building a complex for homology inference on top of the sparsified data without requiring any user-supplied distance threshold. Unlike most topology inference results, ours does not require that the input is dense relative to a *global* feature such as *reach* or *weak feature ize*; instead it can be adaptive with respect to the local feature size. We present some empirical evidence in support of our theoretical claims.

.

Author Index

E

F

G

H

I

J

K

Venue Floor Plan