SIAM 2022

Searchable Abstracts Document

**ACM-SIAM Symposium on
Discrete Algorithms (SODA22)**
January 9 – 12, 2022

**Symposium on
Algorithm Engineering and Experiments (ALENEX22)**
January 9 – 10, 2022

**Symposium on
Simplicity in Algorithms (SOSA22)**
January 10 – 11, 2022

**Symposium on
Algorithmic Principles of Computer Systems (APOCS22)**
January 12, 2022

This document was current as of January 3, 2022. Abstracts appear as submitted.

IP1

Geometry of Polynomials and Applications to Optimization and Counting

I will discuss the fruitful paradigm of encoding discrete phenomena in complex multivariate polynomials, and understanding them via the interplay of the coefficients, zeros, and function values of these polynomials. For example, the states of a hard-core lattice gas model, the forests or matchings in a graph, and the bases of a matroid can all be viewed in these terms. Over the last decade, this perspective has led to several breakthroughs in computer science, and an unexpected bridge between distant scientific areas including combinatorics, probability, and statistical physics has been built. In this talk, I will discuss several classes of these polynomials and their applications to the traveling salesperson problem, and to the analysis of Markov chains.

Shayan Oveis Gharan
University of Washington
shayan@cs.washington.edu

IP2

Property Testing and Regularity Methods

The goal of property testing is to be able to quickly distinguish between objects that satisfy a property from those that are far from satisfying that property. Very general results in different settings based on regularity methods allow one to give randomized algorithms that run using a "constant" number of queries (independent of the object size). However, due to the use of regularity methods, these general results are impractical, with the number of queries having a tower-type (or worse) dependence on the farness parameter. This talk is on recent progress toward practical algorithms in the graph, arithmetic, and permutation settings.

Jacob Fox
Stanford University
jacobfox@stanford.edu

IP3

Imitation Games

One of Alan Turing's most influential papers is his 1950 Computing machinery and intelligence, in which he introduces the famous "Turing test" for probing the nature of intelligence by evaluating the abilities of machines to behave as humans. In this test, which he calls the Imitation Game, a (human) referee has to distinguish between two (remote and separate) entities, a human and a computer, only by observing answers to a sequence of arbitrary questions to each entity. The set-up of an imitation game is extremely versatile and powerful, and variations of it have been used throughout the theory of computation and discrete mathematics to (re)define and understand basic notions, leading to new theories and applications. I will survey several such imitation games, central to cryptography, pseudorandomness, additive combinatorics and differential privacy.

Avi Wigderson
Institute for Advanced Studies
avi@ias.edu

IP4

Approximately Optimal Mechanism Design and Item Pricing

The quintessential problem in economic mechanism design is to allocate a set of items or resources to agents with heterogeneous preferences with the goal of optimizing the economic efficiency of the allocation, or the seller's revenue, etc. Optimal mechanisms in these settings can be very complicated and computationally intractable. A rich literature has emerged in the last few years around the theme of simple approximations. Surprisingly many of these approximation results employ the same simple mechanism – item pricing – where every item receives a (distinct) price and buyers can purchase sets of their choice at the sum of the constituent prices. In this talk I will describe some of these approximation results and the rich set of algorithmic techniques they employ.

Shuchi Chawla
UT-Austin
shuchi@cs.utexas.edu

CP1

Learning-Augmented Weighted Paging

We consider a natural semi-online model for weighted paging, where at any time the algorithm is given predictions, possibly with errors, about the next arrival of each page. The model is inspired by Beladys classic optimal offline algorithm for unweighted paging, and extends the recently studied model for learning-augmented paging to the weighted setting. For the case of perfect predictions, we provide an ℓ-competitive deterministic and an $O(\log \ell)$-competitive randomized algorithm, where ℓ is the number of distinct weight classes. Both these bounds are tight, and imply an $O(\log W)$ and $O(\log \log W)$ competitive ratio, respectively, when the page weights lie between 1 and W. Previously, it was not known how to use these predictions in the weighted setting and only bounds of k and $O(\log k)$ were known, where k is the cache size. Our results also generalize to the interleaved paging setting and to the case of imperfect predictions, with the competitive ratios degrading smoothly from $O(\ell)$ and $O(\log \ell)$ to $O(k)$ and $O(\log k)$, respectively, as the prediction error increases. Our results are based on several insights on structural properties of Beladys algorithm and the sequence of page arrival predictions, and novel potential functions that incorporate these predictions. For the case of unweighted paging, the results imply a very simple potential function based proof of the optimality of Beladys algorithm, which may be of independent interest.

Nikhil Bansal
University of Michigan
bansaln@umich.edu

Christian Coester
Tel Aviv University
christian.coester@gmail.com

Ravi Kumar
Google Inc.
ravi.k53@gmail.com

Manish Purohit
Google

mpurohit@google.com

Erik Vee
Google
Mountain View, CA
erikvee@google.com

CP1

Online Nash Social Welfare Maximization with Predictions

We consider the problem of allocating a set of divisible goods to N agents in an online manner, aiming to maximize the Nash social welfare, a widely studied objective which provides a balance between fairness and efficiency. The goods arrive in a sequence of T periods and the value of each agent for a good is adversarially chosen when the good arrives. We first observe that no online algorithm can achieve a competitive ratio better than the trivial $O(N)$, unless it is given additional information about the agents' values. We then consider a setting where for each agent, the algorithm is only given a prediction of her monopolist utility, i.e., her utility if all goods were given to her alone (corresponding to the sum of her values over the T periods). Our main result is an online algorithm whose competitive ratio is parameterized by the multiplicative errors in these predictions. The algorithm achieves a competitive ratio of $\min\{O(\log N), \log T\}$ if the predictions are perfectly accurate. Moreover, the competitive ratio degrades smoothly with the errors in the predictions, and is surprisingly robust: the logarithmic competitive ratio holds even if the predictions are very inaccurate. We complement this positive result by showing that our bounds are essentially tight: no online algorithm, even if provided with perfect predictions, can achieve a competitive ratio of $O(\log^{1-\epsilon} N)$ or $O(\log^{1-\epsilon} T)$ for any $\epsilon > 0$.

Billy Jin, Siddhartha Banerjee
Cornell University
bzj3@cornell.edu, sbanerjee@cornell.edu

Vasilis Gkatzelis
Drexel University, Computer Science department
gkatz@drexel.edu

Artur Gorokh
Facebook
arturgorokh@fb.com

CP1

Robust Load Balancing with Machine Learned Advice

Motivated by the exploding growth of web-based services and the importance of efficiently managing the computational resources of such systems, we introduce and study a theoretical model for load balancing of very large databases such as commercial search engines. Our model is a more realistic version of the well-received balls-into-bins model with an additional constraint that limits the number of servers that carry each piece of the data. This additional constraint is necessary when, on one hand, the data is so large that we can not copy the whole data on each server. On the other hand, the query response time is so limited that we can not ignore the fact that the number of queries for each piece of the data changes over time, and hence we can not simply split the data over different machines In this paper, we develop an almost optimal load balanc-

ing algorithm that works given an estimate of the load of each piece of the data. Our algorithm is almost perfectly robust to wrong estimates, to the extent that even when all of the loads are adversarially chosen the performance of our algorithm is $1 - 1/e$, which is provably optimal. Along the way, we develop various techniques for analyzing the balls-into-bins process under certain correlations and build a novel connection with the multiplicative weights update scheme.

Binghui Peng
Tsinghua University
bp2601@columbia.edu

Sara Ahmadian, Hossein Esfandiari
Google Research
sahmadian@google.com, esfandiari@google.com

Vahab Mirrokni
Google Inc.
mirrokni@google.com

CP1

Online Discrepancy with Recourse for Vectors and Graphs

In the vector-balancing problem we are given T vectors in $[-1, 1]^n$, and the goal is to find a signing $\sigma(a) \in \{\pm 1\}$ of each vector a to minimize the discrepancy $\|\sum_a \sigma(a) \cdot a\|_\infty$. In this paper we initiate its study in the fully-dynamic setting with recourse: the algorithm sees a stream of T insertions and deletions of vectors, and at each time must maintain a low-discrepancy signing, while also minimizing the amortized recourse (the number of times any vector changes its sign) per update. For general vectors, we show algorithms which almost match Spencer's $O(\sqrt{n})$ offline discrepancy bound, with $O(n \cdot poly\log T)$ amortized recourse per update. The crucial idea behind our algorithm is to compute a basic feasible solution to the linear relaxation in a distributed and recursive manner. We then give algorithms for low-discrepancy edge orientation, where we dynamically maintain signings for 2-sparse vectors. Alternatively, this can be seen as orienting a dynamic set of edges of a graph to minimize the absolute difference between in- and out-degrees at any vertex. We present a deterministic algorithm with $O(poly\log n)$ discrepancy and $O(poly\log n)$ amortized recourse. The core ideas are to dynamically maintain an expander-decomposition with low recourse, and then to show that, as the expanders change over time, a natural local-search algorithm converges quickly.

Anupam Gupta
Carnegie Mellon University
anupamg@cs.cmu.edu

Vijaykrishna Gurunathan
Stanford University
krishnavijay1999@gmail.com

Ravishankar Krishnaswamy
Microsoft Research India
rakri@microsoft.com

Amit Kumar
IIT Delhi
amitk@cse.iitd.ac.in

Sahil Singla

Georgia Tech
ssingla@gatech.edu

CP1
Online Graph Algorithms with Predictions

In this paper, we study online graph problems with predictions. Our contributions are the following: * We give a novel definition of prediction error for graph problems called metric error with outliers which captures both inaccuracies and erroneous omissions/inclusions in the prediction. * We give a general framework for obtaining online algorithms with predictions that combines, in a "black box' fashion, existing online and offline algorithms. * Using our framework, we obtain tight bounds on the competitive ratio of several classical graph problems in terms of metric error with outliers: Steiner tree, Steiner forest, priority Steiner tree/forest, and uncapacitated/capacitated facility location. Both the definition of metric error with outliers and the general framework for combining offline and online algorithms are not specific to the problems that we consider in this paper. We hope that these will be useful for future work on other problems in this domain.

Noam Touitou
Tel Aviv University
noamtwx@gmail.com

Yossi Azar
Tel-Aviv University
azar@tau.ac.il

Debmalya Panigrahi
Duke University
debmalya@cs.duke.edu

CP2
Nested Dissection Meets IPMs: Planar Min-Cost Flow in Nearly-Linear Time

We present a nearly-linear time algorithm for finding a minimum-cost flow in planar graphs with polynomially bounded integer costs and capacities. The previous fastest algorithm for this problem was based on interior point methods (IPMs) and worked for general sparse graphs in $O(n^{1.5}\text{poly}(\log n))$ time [Daitch-Spielman, STOC'08]. Intuitively, $\Omega(n^{1.5})$ is a natural runtime barrier for IPM based methods, since they require $\sqrt{n}$ iterations, each routing a possibly-dense electrical flow. To break this barrier, we develop a new implicit representation for flows based on generalized nested-dissection [Lipton-Rose-Tarjan, JSTOR'79] and approximate Schur complements [Kyng-Sachdeva, FOCS'16]. This implicit representation permits us to design a data structure to route an electrical flow with sparse demands in roughly $\sqrt{n}$ update time, resulting in a total running time of $O(n \cdot \text{poly}(\log n))$. Our results immediately generalize to all families of separable graphs.

Sally Dong
University of Washington
sallyqd@uw.edu

Yu Gao
Georgia Institute of Technology
ygao380@gatech.edu

Gramoz Goranci

University of Glasgow
gramoz.goranci@glasgow.ac.uk

Yin Tat Lee
University of Washington
yintat@uw.edu

Richard Peng
Georgia Tech
rpeng@cc.gatech.edu

Sushant Sachdeva
University of Toronto
sachdeva@cs.toronto.edu

Guanghao Ye
University of Washington
ghye@uw.edu

CP2
Improved Strongly Polynomial Algorithms for Deterministic MDPs, 2VPI Feasibility, and Discounted All-Pairs Shortest Paths

We revisit the problem of finding optimal strategies for deterministic Markov Decision Processes (DMDPs), and a closely related problem of testing feasibility of systems of m linear inequalities on n real variables with at most two variables per inequality (2VPI). We give a randomized trade-off algorithm solving both problems and running in $\tilde{O}(nmh+(n/h)^3)$ time using $\tilde{O}(n^2/h+m)$ space for any parameter $h \in [1, n]$. In particular, using subquadratic space we get $\tilde{O}(nm+n^{3/2}m^{3/4})$ running time, which improves by a polynomial factor upon all the known upper bounds for non-dense instances with $m = O(n^{2-\epsilon})$. Moreover, using linear space we match the randomized $\tilde{O}(nm + n^3)$ time bound of Cohen and Megiddo [SICOMP'94] that required $\tilde{\Theta}(n^2 + m)$ space. Additionally, we show a new algorithm for the Discounted All-Pairs Shortest Paths problem, introduced by Madani et al. [TALG'10], that extends the DMDPs with optional end vertices. For the case of uniform discount factors, we give a deterministic algorithm running in $\tilde{O}(n^{3/2}m^{3/4})$ time, which improves significantly upon the randomized bound $\tilde{O}(n^2\sqrt{m})$ of Madani et al.

Adam Karczmarz
University of Warsaw
a.karczmarz@mimuw.edu.pl

CP2
The Popular Assignment Problem: When Cardinality Is More Important Than Popularity

We consider a matching problem in a bipartite graph $G = (A \cup B, E)$ where each node in A is an agent having preferences in partial order over her neighbors, while nodes in B are objects with no preferences. The size of our matching is more important than node preferences; we are thus interested in maximum matchings only. Any pair of maximum matchings in G (equivalently, perfect matchings or *assignments*) can be compared by holding a head-to-head election between them where agents are voters. The goal is to compute an assignment for which no assignment is "more popular". This is the popular assignment problem and it generalizes the well-studied popular matching problem. We show a polynomial-time algorithm that computes a popular assignment if one exists. We also consider

the problem of finding an almost popular assignment, i.e., an assignment with minimum unpopularity margin. We give an $O^*(|E|^k)$ time algorithm for deciding if there exists an assignment with unpopularity margin at most k. We prove that this problem is $\mathsf{W}_l[1]$-hard with parameter k, so our algorithm is essentially optimal. We also consider the minimum-cost popular assignment problem when there are edge costs, and show its NP-hardness even when all edge costs are in $\{0, 1\}$. By contrast, we propose a polynomial-time algorithm for deciding if a popular assignment with a given set of forced/forbidden edges exists. Our algorithms are combinatorial and based on LP duality.

Telikepalli Kavitha
Tata Institute of Fundamental Research
Mumbai, India
kavitha@tifr.res.in

Tamás Király
Eotvos Lorand University
Budapest, Hungary
tamas.kiraly@ttk.elte.hu

Jannik Matuschke
Research Center for Operations Management, KU Leuven
Leuven, Belgium
jannik.matuschke@kuleuven.be

Ildikó Schlotter
Centre for Economic and Regional Studies
schlotter.ildiko@krtk.hu

Ulrike Schmidt-Kraepelin
Technische Universitat Berlin
Berlin, Germany
u.schmidt-kraepelin@tu-berlin.de

CP2

A Faster Algorithm for Quickest Transshipments via an Extended Discrete Newton Method

The Quickest Transshipment Problem is to route flow as quickly as possible from sources with supplies to sinks with demands in a network with capacities and transit times on the arcs. It is of fundamental importance for numerous applications in areas such as logistics, production, traffic, evacuation, and finance. More than 25 years ago, Hoppe and Tardos presented the first (strongly) polynomial-time algorithm for this problem. Their approach, as well as subsequently derived algorithms with strongly polynomial running time, are hardly practical as they rely on parametric submodular function minimization via Megiddo's method of parametric search. The main contribution of this paper is a considerably faster algorithm for the Quickest Transshipment Problem that instead employs a subtle extension of the Discrete Newton Method. This improves the previously best known running time of $\tilde{O}(m^4 k^{14})$ to $\tilde{O}(m^2 k^5 + m^3 k^3 + m^3 n)$, where n is the number of nodes, m the number of arcs, and k the number of sources and sinks.

Miriam Schloeter
ETH Zurich
miriam.schloeter@ifor.math.ethz.ch

Martin Skutella, Khai Van Tran
TU Berlin

skutella@math.tu-berlin.de, kvtran@math.tu-berlin.de

CP3

Computational Topology in a Collapsing Universe: Laplacians, Homology, Cohomology

The dth combinatorial Laplacian of a simplicial complex K is defined as $L_d = \partial_{d+1}\partial_{d+1}^T + \partial_d^T \partial_d$, with the zeroth Laplacian being the well-studied graph Laplacian. In this paper, we describe a solver for the linear systems $L_1 x = b$, where L_1 is the 1-Laplacian of a simplicial complex K under the following assumptions: $\dim H_1(K) = 0$ and $K \subset X$ where X is a collapsible simplicial complex embedded in $\mathbb{R}^3$ with a known collapsing sequence. Our algorithm runs in $\tilde{O}(n \log^2(n\kappa/\varepsilon))$ time, where n is the total number of vertices, edges, and triangles in X, κ is the largest condition number of the two parts of the Laplacian, and ε quantifies the approximation quality. This result is a generalization of Cohen et al. [SODA 2014]. In addition, we describe faster algorithms for testing null-homology of $(d-1)$-cycles, testing null-cohomology of d-cocycles, and computing a $(d-1)$-cohomology basis from a given $(d-1)$-homology basis in a d-simplicial complex K provided $K \subset X$ for a collapsible X embedded in $\mathbb{R}^{d+1}$ with a known collapsing sequence. For all of the problems above and when $d = 2$, that is $K \subset \mathbb{R}^3$, if the collapsible supercomplex X is not provided, we can expand K into a convex ball of possibly quadratic complexity, which is known to be collapsible, resulting in nearly quadratic time algorithms.

Amir Nayyeri
School of Electrical Eng and Computer Science
Oregon State University
nayyeria@eecs.oregonstate.edu

Mitchell Black, William Maxwell, Eli Winkelman
Oregon State University
blackmit@oregonstate.edu, maxwellw@oregonstate.edu, winkelme@oregonstate.edu

CP3

Untangling Planar Graphs and Curves by Staying Positive

Any generic planar closed curve with n crossings can be turned into a simple closed curve by applying $O(n^{3/2})$ homotopy moves without ever increasing the number of self-crossings; this improves over the $O(n^2)$ upper bound from Steinitz [Ency. Math. Wiss. III 1916], and matches the best lower bound. We prove the existence of a positive move that decreases the depth-sum potential at every step. Using similar techniques, we show that any 2-terminal plane graph with n vertices can be reduced to a single edge between the terminals using $O(n^{3/2})$ electrical transformations, consisting of degree-1 reductions, series-parallel reductions, and ΔY-transformations; this proves a conjecture of Feo and Provan that was open for more than 30 years.

Dylan Fridman
Department of Computer Science, Dartmouth College
dylan.fridman.24@dartmouth.edu

Hsien-Chih Chang
Department of Computer Science
Dartmouth College
hsien-chih.chang@dartmouth.edu

Santiago Aranguri
Department of Mathematics, Stanford University
aranguri@stanford.edu

CP3

Cubic Upper and Lower Bounds for Subtrajectory Clustering under the Continuous Fréchet Distance

Detecting commuting patterns or migration patterns in movement data is an important problem in computational movement analysis. Given a trajectory, or set of trajectories, this corresponds to clustering similar subtrajectories. We study subtrajectory clustering under the continuous and discrete Fréchet distances. The most relevant theoretical result is by Buchin et al. (2011). They provide, in the continuous case, an $O(n^5)$ time algorithm and a 3SUM-hardness lower bound, and in the discrete case, an $O(n^3)$ time algorithm. We show, in the continuous case, an $O(n^3 \log^2 n)$ time algorithm and a 3OV-hardness lower bound, and in the discrete case, an $O(n^2 \log n)$ time algorithm and a quadratic lower bound. Our bounds are almost tight unless SETH fails.

Sampson Wong, Joachim Gudmundsson
University of Sydney
sampson.wong123@gmail.com,
joachim.gudmundsson@gmail.com

CP3

Hopcroft's Problem, Log-Star Shaving, 2D Fractional Cascading, and Decision Trees

We revisit Hopcroft's problem and related fundamental problems about geometric range searching. Given n points and n lines in the plane, we show how to count the number of point-line incidence pairs or the number of point-above-line pairs in $O(n^{4/3})$ time, which matches the conjectured lower bound and improves the best previous time bound of $n^{4/3}2^{O(\log^* n)}$ obtained almost 30 years ago by Matousek. We describe two interesting and different ways to achieve the result: the first is randomized and uses a new 2D version of fractional cascading for arrangements of lines; the second is deterministic and uses decision trees in a manner inspired by the sorting technique of Fredman (1976). The second approach extends to any constant dimension. Many consequences follow from these new ideas: for example, we obtain an $O(n^{4/3})$-time algorithm for line segment intersection counting in the plane, $O(n^{4/3})$-time randomized algorithms for bichromatic closest pair and Euclidean minimum spanning tree in three or four dimensions, and a randomized data structure for halfplane range counting in the plane with $O(n^{4/3})$ preprocessing time and space and $O(n^{1/3})$ query time.

Timothy M. Chan
University of Illinois at Urbana-Champaign
tmc@illinois.edu

Da Wei Zheng
Department of Computer Science
University of Illinois at Urbana Champaign
dwzheng2@illinois.edu

CP4

Engineering Uniform Sampling of Graphs with a

Prescribed Power-Law Degree Sequence

We consider the following common network analysis problem: given a degree sequence $\mathbf{d} = (d_1, \ldots, d_n) \in \mathbb{N}^n$ return a uniform sample from the ensemble of all simple graphs with matching degrees. In practice, the problem is typically solved using Markov Chain Monte Carlo approaches, such as Edge-Switching or Curveball, even if no practical useful rigorous bounds are known on their mixing times. In contrast, Arman et al. sketch Inc-Powerlaw, a novel and much more involved algorithm capable of generating graphs for power-law bounded degree sequences with $\gamma \gtrapprox 2.88$ in expected linear time. For the first time, we give a complete description of the algorithm and add novel switchings. To the best of our knowledge, our open-source implementation of Inc-Powerlaw is the first practical generator with rigorous uniformity guarantees for the aforementioned degree sequences. In an empirical investigation, we find that for small average-degrees Inc-Powerlaw is very efficient and generates graphs with one million nodes in less than a second. For larger average-degrees, parallelism can partially mitigate the increased running-time.

Daniel Allendorf, Ulrich Meyer, Manuel Penschuck, Hung Tran
Goethe University Frankfurt
dallendorf@ae.cs.uni-frankfurt.de,
umeyer@ae.cs.uni-frankfurt.de,
manuel@ae.cs.uni-frankfurt.de,
htran@ae.cs.uni-frankfurt.de

Nick Wormald
Monash University
nicholas.wormald@monash.edu

CP4

McSparse: Exact Solutions of Sparse Maximum Cut and Sparse Unconstrained Binary Quadratic Optimization Problems

While the Maximum Cut Problem and Unconstrained Binary Quadratic Optimization are of high interest in the scientific community and gain increasing importance, state-of-the-art solvers for these problems are publicly available, yet not for sparse instances of larger scale. We present the novel solver McSparse to fill this gap. It is installed as an internet service similar to the well-known services Biq Mac and BiqCrunch. We explain details of the algorithmic innovations based on integer linear programming and polyhedral combinatorics leading to the branch-and-cut algorithm implemented in McSparse. Substantial improvements with respect to former such approaches are demonstrated and the sustained performance is compared to those of other state-of-the-art methods using a broad set of benchmark instances.

Sven Mallach, Jonas Charfreitag, Petra Mutzel
University of Bonn
smallach@cs.uni-bonn.de, jcharfreitag@cs.uni-bonn.de,
petra.mutzel@cs.uni-bonn.de

Michael Jünger
University of Cologne
juenger-sfb@informatik.uni-koeln.de

CP4

Practical Fully Dynamic Minimum Cut Algorithms

We present a practically efficient algorithm for maintain-

ing a global minimum cut in large dynamic graphs under both edge insertions and deletions. While there has been theoretical work on this problem, our algorithm is the first implementation of a fully-dynamic algorithm. The algorithm uses the theoretical foundation and combines it with efficient and finely-tuned implementations to give an algorithm that can maintain the global minimum cut of a graph with rapid update times. We show that our algorithm gives up to multiple orders of magnitude speedup compared to static approaches both on edge insertions and deletions.

Monika Henzinger
University of Vienna, Austria
monika.henzinger@univie.ac.at

Alexander Noe
University of Vienna
alexander.noe@univie.ac.at

Christian Schulz
Heidelberg University, Germany
christian.schulz@informatik.uni-heidelberg.de

CP4
Distance Closures: Unifying Search and Lookup-Based Shortest Path Speedup Techniques

Most popular speed-up techniques for shortest path queries in road networks are based either on pruned *graph search* or clever *lookup schemes* and allow for answering of shortest path distance queries within continent-sized road networks in less than a *milli-* (search-based) or even *microsecond* (lookup-based) compared to several seconds of a normal Dijkstra run. While both paradigms previously have been considered mostly separately, we present a framework that unifies these seemingly different views on shortest path computations. Apart from the conceptual novelty, this allows for new and (practically very attractive) space-time tradeoffs for shortest-path computation. To our knowledge we are also the first to report on computational results for the largest connected component of the OpenStreetMap planet road network with more than half a billion nodes.

Daniel Bahrdt, Stefan Funke, Sokol Makolli,
Claudius Proissl
University of Stuttgart
daniel.bahrdt@fmi.uni-stuttgart.de, stefan.funke@fmi.uni-stuttgart.de, st116214@stud.uni-stuttgart.de, claudius.proissl@fmi.uni-stuttgart.de

CP4
Effective Data Reduction for the Vertex Clique Cover Problem

The vertex clique cover (VCC) problem—the problem of computing a minimum cardinality set of cliques covering all vertices of a graph—is a classic NP-hard problem. Despite recent advances in parameterized algorithms that have been used to solve NP-hard problems in practice, the VCC problem has been almost completely unexplored. In particular, data reduction rules for the VCC problem, which transform the input graph to a smaller equivalent instance, are nearly nonexistent: instead, the complementary graph coloring problem has received the lion's share of attention, and the available rules for that problem are either theoretical or they do not translate to effective rules for solving the VCC problem on sparse graphs. In this paper, we introduce a large suite of data reduction rules for the VCC problem. These rules enable us to solve large,

sparse, real-world graphs significantly faster than the state of the art. Of the 52 graphs tested, without any additional techniques, our reduction rules completely solve 14 graphs with up to 326K vertices in a few milliseconds. Furthermore, applying our rules as a preprocessing step accelerates the state-of-the-art iterated greedy (IG) approach due to Chalupa, enabling us to find higher-quality solutions up to multiple orders of magnitude faster than previously possible. Furthermore, we integrate our data reductions into the branch-and-reduce framework, exactly solving instances on up to millions of vertices.

Darren Strash, Louise Thompson
Department of Computer Science
Hamilton College
dstrash@hamilton.edu, lmthomps@hamilton.edu

CP5
Tight Running Times for Minimum ℓ_q-Norm Load Balancing: Beyond Exponential Dependencies on $1/\epsilon$

We consider a classical scheduling problem on m identical machines. For an arbitrary constant $q > 1$, the aim is to assign jobs to machines such that $\sum_{i=1}^{m} C_i^q$ is minimized, where C_i is the total processing time of jobs assigned to machine i. It is well known that this problem is strongly NP-hard. Under mild assumptions, the running time of an $(1 + \epsilon)$-approximation algorithm for a strongly NP-hard problem cannot be polynomial on $1/\epsilon$, unless P = NP. For most problems in the literature, this translates into algorithms with running time at least as large as $2^{\Omega(1/\varepsilon)} + n^{O(1)}$. For the natural scheduling problem above, we establish the existence of an algorithm which violates this threshold. More precisely, we design a PTAS that runs in $2^{\tilde{O}(\sqrt{1/\epsilon})} + n^{O(1)}$ time. This result is in sharp contrast to the closely related minimum makespan variant, where an exponential lower bound is known under the exponential time hypothesis (ETH). We complement our result with an essentially matching lower bound on the running time, showing that our algorithm is best-possible under ETH. The lower bound proof exploits new number-theoretical constructions for variants of progression-free sets, which might be of independent interest.

Lin Chen
University of Houston
chenlin198662@gmail.com

Liangde Tao
Zhejiang University
vast.tld@gmail.com

Jose Verschae
Pontificia Universidad Católica de Chile
jverschae@uc.cl

CP5
On the Hardness of Scheduling with Non-Uniform Communication Delays

In the problem of scheduling with non-uniform communication delays, the input is a set of jobs with precedence constraints. Associated with every precedence constraint between a pair of jobs is a communication delay, the time duration the scheduler has to wait between the two jobs if they are scheduled on different machines. The objective is to assign the jobs to machines to minimize the makespan

of the schedule. Despite being a fundamental problem in theory and a consequential problem in practice, the approximability of scheduling problems with communication delays is not very well understood. One of the top ten open problems in scheduling theory, in the influential list by Schuurman and Woeginger and its latest update by Bansal, asks if the problem admits a constant-factor approximation algorithm. In this paper, we answer this question in the negative by proving a logarithmic hardness for the problem under the standard complexity theory assumption that NP-complete problems do not admit quasi-polynomial-time algorithms. Our hardness result is obtained using a surprisingly simple reduction from a problem that we call Unique Machine Precedence constraints Scheduling (UMPS). We believe that this problem is of central importance in understanding the hardness of many scheduling problems and we conjecture that it is very hard to approximate. Among other things, our conjecture implies a logarithmic hardness of related machine scheduling with precedences.

Sami Davies
University of Washington
daviess@uw.edu

Janardhan Kulkarni
Microsoft Research, Redmond
jakul@microsoft.com

Thomas Rothvoss
University of Washington
rothvoss@uw.edu

Sai Sandeep
Carnegie Mellon University
spallerl@andrew.cmu.edu

Jakub Tarnawski
Microsoft Research
jatarnaw@microsoft.com

Yihao Zhang
University of Washington
yihaoz93@uw.edu

CP5
Distortion-Oblivious Algorithms for Minimizing Flow Time

We consider the classic online problem of scheduling on a single machine to minimize total flow time. In STOC 2021, the concept of robustness to distortion in processing times was introduced: for every distortion factor μ, an $O(\mu^2)$-competitive algorithm ALG_μ which handles distortions up to μ was presented. However, using that result requires one to know the distortion of the input in advance, which is impractical. We present the first *distortion-oblivious* algorithms: algorithms which are competitive for *every* input of *every* distortion, and thus do not require knowledge of the distortion in advance. Moreover, the competitive ratios of our algorithms are $\tilde{O}(\mu)$, which is a quadratic improvement over the algorithm from STOC 2021, and is nearly optimal (we show a randomized lower bound of $\Omega(\mu)$ on competitiveness).

Yossi Azar
Tel-Aviv University
azar@tau.ac.il

Stefano Leonardi

Sapienza Università di Roma, Italy
leonardi@diag.uniroma1.it

Noam Touitou
Tel Aviv University
noamtwx@gmail.com

CP5
Competitive Strategies for Symmetric Rendezvous on the Line

In the Symmetric Rendezvous Search on the Line with Unknown Initial Distance, two identical agents are placed on the real line with their distance, the other's location, and their orientation unknown to them. Moving along the line at unit speed and executing the same randomized search strategy, the agents' goal is to meet up as early as possible. The expected meeting time depends on the unknown initial distance and orientations. The quality of a randomized search strategy is thus measured by its competitive ratio, i.e., the ratio of the expected meeting time and the earliest possible meeting time. We present a class of successively refined randomized search strategies with continuously improved competitive ratios. These strategies rely on the basic idea of performing an infinite sequence of steps of geometrically increasing size in random directions, always returning to the agent's initial position before the next step. Our more refined strategies use two novel ideas. First, remembering their past random choices, the agents choose the direction of the next step in a Markov-chain-like manner. Second, choosing the next few random directions in advance, each agent may combine consecutive steps in the same direction into one longer step. We show that this combination of looking into the past as well as into the future leads to a substantially improved competitive ratio of 13.93 compared to the previously best known bound of 24.85 (Ozsoyeller et al. 2013).

Khai Van Tran, Max Klimm, Guillaume Sagnol, Martin Skutella
TU Berlin
kvtran@math.tu-berlin.de,   klimm@math.tu-berlin.de, sagnol@math.tu-berlin.de, skutella@math.tu-berlin.de

CP6
Directed Tangle Tree-Decompositions and Applications

The tangle tree-decomposition theorem, proved by Robertson and Seymour in their seminal graph minors series, turns out to be an extremely valuable tool in structural and algorithmic graph theory. In this paper, we prove the analogous result for digraphs, the directed tangle tree-decomposition theorem. More precisely, we introduce directed tangles and provide a directed tree-decomposition of digraphs G that distinguishes all maximal directed tangles in G. Furthermore, for any integer k, we construct a directed tree-decomposition that distinguishes all directed tangles of order k. By relaxing the bound slightly, we can make the previous result algorithmic: for fixed k, we design a polynomial-time algorithm that finds a directed tree-decomposition distinguishing all directed tangles of order $6k-1$ separated by some separation of order less than k. As a direct application of the tangle tree-decomposition theorem, we prove that for every fixed k there is a polynomial-time algorithm which, on input G, and source and sink vertices $(s_1, t_1), \ldots, (s_k, t_k)$, either finds a family of paths $P_1, \ldots, P_k$ such that each P_i links s_i to t_i and every vertex of G is contained in at most two paths, or determines

that there is no set of pairwise vertex-disjoint paths each connecting s_i to t_i.

O-Joung Kwon
Incheon National University
ojoungkwon@gmail.com

Archontia Giannopoulou
Department of Informatics and Telecommunications
National and Kapodistrian University of Athens
archontia.giannopoulou@gmail.com

Ken-ichi Kawarabayashi
National Institute of Informatics, Japan
k_keniti@nii.ac.jp

Stephan Kreutzer
Technical University Berlin
stephan.kreutzer@tu-berlin.de

CP6
Sparsifying, Shrinking and Splicing for Minimum Path Cover in Parameterized Linear Time

A minimum path cover (MPC) of a directed acyclic graph (DAG) $G = (V, E)$ is a minimum-size set of paths that together cover all the vertices of the DAG. Computing an MPC is a basic polynomial problem, dating back to Dilworth's and Fulkerson's results in the 1950s. Since the size k of an MPC (also known as the width) can be small in practical applications, research has also studied algorithms whose running time is parameterized on k. We obtain two new MPC parameterized algorithms for DAGs running in time $O(k^2|V|\log|V| + |E|)$ and $O(k^3|V| + |E|)$. Our latter algorithm is the first solving the problem in parameterized linear time. Finally, we show that we can transform (in $O(k^2|V|)$ time) a given MPC into another MPC that uses less than $2|V|$ distinct edges, which we prove to be asymptotically tight. As such, we also obtain edge sparsification algorithms preserving the width of the DAG with the same running time as our MPC algorithms.

Manuel Caceres
University of Helsinki
manuel.caceresreyes@helsinki.fi

Massimo Cairo
Scuola Normale Superiore
cairomassimo@gmail.com

Brendan Mumey
Montana State University
brendan.mumey@montana.edu

Romeo Rizzi
Universita di Verona
romeo.rizzi@univr.it

Alexandru Tomescu
University of Helsinki
alexandru.tomescu@helsinki.fi

CP6
Collapsing the Tower - On the Complexity of Multistage Stochastic IPs

In this paper we study the computational complexity of solving a class of block structured integer programs (IPs) - so called multistage stochastic IPs. A multistage stochastic IP is an IP of the form $\max\{c^T x \mid \mathcal{A}x = b, l \leq x \leq u, x \text{ integral}\}$ where the constraint matrix $\mathcal{A}$ consists of small block matrices ordered on the diagonal line and for each stage there are larger blocks with few columns connecting the blocks in a tree like fashion. One of the major gaps that remained however was the parameter dependency in the running time for an algorithm solving multistage stochastic IPs. Previous algorithms require a tower of t exponentials, where t is the number of stages, while only a double exponential lower bound was known. In this paper we show that the tower of t exponentials is actually not necessary. We can show an improved running time for the algorithm solving multistage stochastic IPs with a running time of $2^{(d||A||_\infty)^{\mathcal{O}(d^{3t+1})}} \cdot poly(d, n)$, where d is the sum of columns in the connecting blocks and n is the number of blocks on the lowest stage. Hence, we obtain the first bound by an elementary function for the running time of an algorithm solving multistage stochastic IPs. In contrast to previous works, our algorithm has only a triple exponential dependency on the parameters and only doubly exponential for every constant t.

Kim-Manuel Klein, Janina Reuter
Kiel University
kmk@informatik.uni-kiel.de,　　janina.reuter@email.uni-kiel.de

CP6

Algorithmic Extensions of Diracs Theorem

In 1952, Dirac proved the following theorem about long cycles in graphs with large minimum vertex degrees: Every n-vertex 2-connected graph G with minimum vertex degree $\delta \geq 2$ contains a cycle with at least $\min\{2\delta, n\}$ vertices. In particular, if $\delta \geq n/2$, then G is Hamiltonian. The proof of Dirac's theorem is constructive, and it yields an algorithm computing the corresponding cycle in polynomial time. The combinatorial bound of Dirac's theorem is tight in the following sense. There are 2-connected graphs that do not contain cycles of length more than $2\delta+1$. Also, there are non-Hamiltonian graphs with all vertices but one of degree at least $n/2$. This prompts naturally to the following algorithmic questions. For $k \geq 1$, (A) How difficult is to decide whether a 2-connected graph contains a cycle of length at least $\min\{2\delta + k, n\}$? (B) How difficult is to decide whether a graph G is Hamiltonian, when at least $n - k$ vertices of G are of degrees at least $n/2 - k$? We resolve both questions by proving the following algorithmic generalization of Dirac's theorem: If all but k vertices of a 2-connected graph G are of degree at least δ, then deciding whether G has a cycle of length at least $\min\{2\delta + k, n\}$ can be done in time $2^{\mathcal{O}(k)} \cdot n^{\mathcal{O}(1)}$. The proof builds on new graph-theoretical results that are interesting on their own.

Danil Sagunov
St. Petersburg Department of V.A. Steklov Institute of Mathematics
danilka.pro@gmail.com

Fedor Fomin
Dep. of Informatics
University of Bergen
fedor.fomin@uib.no

Petr Golovach, Kirill Simonov
University of Bergen, Norway

petr.golovach@uib.no, kirillsimonov@gmail.com

CP7

Counting List Homomorphisms from Graphs of Bounded Treewidth: Tight Complexity Bounds

The goal of this work is to give precise bounds on the counting complexity of a family of generalized coloring problems (list homomorphisms) on bounded-treewidth graphs. Standard techniques show that if a graph G is given with a tree decomposition of width t, then the number of list homomorphisms to a graph H can be counted in time $|V(H)|^t \cdot n^{\mathcal{O}(1)}$. Our main result is determining, for every fixed graph H, how much the base $|V(H)|$ in the running time can be improved. For a certain graph invariant $\mathrm{irr}(H)$, earlier results show that if $\mathrm{irr}(H) = 1$, then the problem of counting list homomorphisms to H is polynomial-time solvable, and otherwise it is #P-hard. We show that, for every fixed graph H, the number of list homomorphisms from (G, L) to H - can be counted in time $\mathrm{irr}(H)^t \cdot n^{\mathcal{O}(1)}$ if a tree decomposition of G having width at most t is given in the input, and - given that $\mathrm{irr}(H) \geq 2$, cannot be counted in time $(\mathrm{irr}(H) - \epsilon)^t \cdot n^{\mathcal{O}(1)}$ for any $\epsilon > 0$, even if a tree decomposition of G having width at most t is given in the input, unless the Counting Strong Exponential-Time Hypothesis (#SETH) fails. Thereby we give a precise and complete complexity classification featuring matching upper and lower bounds for all target graphs with or without loops.

Jacob Focke, Dániel Marx
CISPA Helmholtz Center for Information Security
jacob.focke@cispa.de, marx@cispa.saarland

Pawel Rzazewski
Warsaw University of Technology, Poland
p.rzazewski@mini.pw.edu.pl

CP7

Counting Homomorphic Cycles in Degenerate Graphs

Since counting subgraphs in general graphs is a computationally demanding problem, it is natural to try and design fast algorithms for restricted families of graphs. One such family that has been extensively studied is that of graphs of bounded degeneracy. This line of work, which started in the early 80's, culminated in a recent work of Gishboliner et al., which highlighted the importance of the task of counting homomorphic copies of cycles in graphs of bounded degeneracy. Our main result in this paper is a surprisingly tight relation between the above task and the problem of *detecting copies* of directed cycles in *general directed* graphs. We prove the following:

- One can compute the number of homomorphic copies of C_{2k} and C_{2k+1} in n-vertex graphs of bounded degeneracy in time $\tilde{O}(n^{d_k})$, where the fastest *known* algorithm for detecting directed copies of C_k in general m-edge digraphs runs in time $\tilde{O}(m^{d_k})$.

- Conversely, one can transform any $O(n^{b_k})$ algorithm for computing the number of homomorphic copies of C_{2k} or of C_{2k+1} in n-vertex graphs of bounded degeneracy, into an $\tilde{O}(m^{b_k})$ time algorithm for detecting directed copies of C_k in general m-edge digraphs.

Lior Gishboliner
ETH Zurich

lior.gishboliner@math.ethz.ch

Yevgeny Levanzov, Asaf Shapira
Tel Aviv University
yevgenyl@mail.tau.ac.il, asafico@tau.ac.il

Raphael Yuster
University of Haifa
raphael.yuster@gmail.com

CP7

The Complexity of Average-Case Dynamic Subgraph Counting

Statistics of small subgraph counts such as triangles, four-cycles, and s-t paths of short lengths reveal important properties of the underlying graph. In most applications, the graphs are not only massive but also change dynamically over time. Most of these problems are hard in the worst-case dynamic setting. In this paper, we ask whether the question of small subgraph counting over dynamic graphs is hard also in the average case. We consider the simplest possible average case model where the updates follow an Erdos-Renyi graph: each update selects a pair of vertices (u, v) uniformly at random and flips the existence of the edge (u, v). We develop new lower bounds and matching algorithms in this model for counting four-cycles, counting triangles through a specified point s, or a random queried point, and st paths of length 3, 4 and 5. Our results indicate while computing st paths of length 3, and 4 are easy in the average case with $O(1)$ update time (note that they are hard in the worst case), it becomes hard when considering st paths of length 5. We introduce new techniques which allow us to get average-case hardness for these graph problems from the worst-case hardness of the Online Matrix vector problem (OMv). Our techniques rely on recent advances in fine-grained average-case complexity. Our techniques advance this literature, giving the ability to prove new lower bounds on average-case dynamic algorithms.

Andrea Lincoln
UC Berkeley
andrealincoln@berkeley.edu

Monika Henzinger
University of Vienna
mhenzinger@gmail.com

Barna Saha
University of California, Berkeley
barnas@berkeley.edu

CP7

Approximately Counting Independent Sets in Bipartite Graphs Via Graph Containers

By implementing algorithmic versions of Sapozhenko's graph container methods, we give new algorithms for approximating the number of independent sets in bipartite graphs. The first algorithm applies to d-regular, bipartite graphs satisfying a weak expansion condition: when d is constant, and the graph is a $\Omega(\log^2 d/d)$-bipartite expander, we obtain an FPTAS for the number of independent sets. Previously such a result for $d > 5$ was known only for graphs satisfying the much stronger expansion conditions of random graphs. The second algorithm applies to all d-regular, bipartite graphs, runs

in time $\exp\left(O\left(n \cdot \frac{\log^3 d}{d}\right)\right)$, and outputs a $(1 + o(1))$-approximation to the number of independent sets.

Aditya Potukuchi
Rutgers University
apotu.57@gmail.com

Matthew Jenssen
University of Birmingham
m.jenssen@bham.ac.uk

Will Perkins
University of Illinois at Chicago
math@willperkins.org

CP8
Load Balancing: The Long Road from Theory to Practice

There is a long history of approximation schemes for the problem of scheduling jobs on identical machines to minimize the makespan. Such a scheme grants a $(1 + \varepsilon)$-approximation solution for every $\varepsilon > 0$, but the running time grows exponentially in $1/\varepsilon$. For a long time, these schemes seemed like a purely theoretical concept. Even solving instances for moderate values of ε seemed completely illusional. In an effort to bridge theory and practice, we refine recent ILP techniques to develop the fastest known approximation scheme for this problem. An implementation of this algorithm reaches values of ε lower than $2/11 \approx 18.2\%$ within a reasonable timespan. This is the approximation guarantee of MULTIFIT, which, to the best of our knowledge, has the best proven guarantee of any non-scheme algorithm.

Max A. Deppert
Kiel University
made@informatik.uni-kiel.de

Sebastian Berndt
University of Lübeck
s.berndt@uni-luebeck.de

Klaus Jansen
University of Kiel
Institute for Computer Science
kj@informatik.uni-kiel.de

Lars Rohwedder
EPFL
lars.rohwedder@epfl.ch

CP8
Computing Graph Hyperbolicity using Dominating Sets

Hyperbolicity is a graph parameter related to how much a graph resembles a tree with respect to distances. Its computation is challenging as the main approaches consist in scanning all quadruples of the graph or using fast matrix multiplication as building block, both are not practical for large graphs. In this paper, we propose and evaluate an approach that uses a hierarchy of distance-k dominating sets to reduce the search space. This technique, compared to the previous best practical algorithms, enables us to compute the hyperbolicity of graphs with unprecedented size (up to a million nodes) and speeds up the computation of previously attainable graphs by up to 3 orders of mag-

nitude while reducing the memory consumption by up to more than a factor of 23.

David Coudert
Université Côte dAzur, Inria, CNRS, I3S, France
david.coudert@inria.fr

André Nusser
Max Planck Institute for Informatics
Saarland Informatics Campus
anusser@mpi-inf.mpg.de

Laurent Viennot
Inria
laurent.viennot@inria.fr

CP8
A SAT Approach to Twin-Width

Twin-width is a new graph invariant that was recently introduced by Bonnet *et al.* (FOCS'20, SODA'21, ICALP'21). Problems expressible in first-order logic, which includes many prominent NP-hard problems, are tractable on graphs of bounded twin-width if a certificate for the twin-width bound is provided as an input. Computing such a certificate, however, is an intrinsic problem, for which no nontrivial algorithm is known. In this paper, we propose the first practical approach for computing the twin-width of graphs together with the corresponding certificate. We propose efficient SAT-encodings that rely on a new characterization of twin-width based on elimination sequences. This allows us to determine the twin-width of many famous graphs with previously unknown twin-width. We utilize our encodings to identify the smallest graphs for a given twin-width bound $d \in \{1, \ldots, 4\}$.

Andre Schidler, Stefan Szeider
TU Wien
andre@schidler.at, sz@ac.tuwien.ac.at

CP8
Towards the 5/6-Density Conjecture of Pinwheel Scheduling

Pinwheel Scheduling aims to find a perpetual schedule for unit-length tasks on a single machine subject to given maximal time spans (a.k.a. frequencies) between any two consecutive executions of the same task. The density of a Pinwheel Scheduling instance is the sum of the inverses of these task frequencies; the 5/6-Conjecture (Chan and Chin, 1993) states that any Pinwheel Scheduling instance with density at most 5/6 is schedulable. We formalize the notion of Pareto surfaces for Pinwheel Scheduling and exploit novel structural insights to engineer an efficient algorithm for computing them. This allows us to (1) confirm the 5/6-Conjecture for all Pinwheel Scheduling instances with at most 12 tasks and (2) to prove that a given list of only 23 schedules solves all schedulable Pinwheel Scheduling instances with at most 5 tasks.

Benjamin M. Smith
University of Liverpool
University of Liverpool
benjaminmarksmith@outlook.com

Sebastian Wild
University of iverpool
sebastian.wild@liverpool.ac.uk

Leszek A. Gasieniec
University of Liverpool
l.a.gasieniec@liverpool.ac.uk

CP9

Sensitivity Oracles for All-Pairs Mincuts

Surender Baswana
IIT Kanpur
sbaswana@gmail.com

Abhyuday Pandey
Indian Institute of Technology, Kanpur
tbd@iitk.ac.in

CP9

Tight Bounds for Approximate Near Neighbor Searching for Time Series under the Frchet Distance

We study the following problem: Given n polygonal curves with m vertices, a radius $\delta > 0$, a parameter $k \leq m$, and a constant $c > 0$, preprocess the curves into a data structure that, given a query curve q with k vertices, either returns an input curve with Frchet distance at most $c \cdot \delta$ to q, or returns that there exists no input curve with Frchet distance at most δ to q. We give a comprehensive analysis for the case of one-dimensional polygonal curves. We obtain new upper bounds that provide different tradeoffs between approximation factor, preprocessing time, and query time. We show that for any $0 < \varepsilon \leq 1$ an approximation factor of $(1 + \varepsilon)$ can be achieved within the same time bounds as the previously best result for $(2 + \varepsilon)$. Moreover, we show that an approximation factor of $(2 + \varepsilon)$ can be obtained by using preprocessing time and space $O(nm)$ and query time in $O(\frac{1}{\varepsilon})^{k+2}$, where the previously best result used preprocessing time in $n \cdot O(\frac{m}{\varepsilon k})^k$ and query time in $O(1)^k$. We complement our upper bounds with matching conditional lower bounds based on the Orthogonal Vectors Hypothesis. Some of our lower bounds already hold for any super-constant value of k. This is achieved by proving hardness of a one-sided sparse version of the Orthogonal Vectors problem as an intermediate problem, which we believe to be of independent interest.

Karl Bringmann
Saarland University and Max Planck Institute for Informatics
Saarland Informatics Campus, Germany
bringmann@cs.uni-saarland.de

Anne Driemel
Hausdorff Center for Mathematics
University of Bonn
driemel@cs.uni-bonn.de

André Nusser
Max Planck Institute for Informatics
Saarland Informatics Campus
anusser@mpi-inf.mpg.de

Ioannis Psarros
Athena Research Center

ipsarros@di.uoa.gr

CP9

A Tail Estimate with Exponential Decay for the Randomized Incremental Construction of Search Structures

The Randomized Incremental Construction (RIC) of search DAGs for point location in planar subdivisions, nearest-neighbor search in 2D points, and extreme point search in 3D convex hulls, are well known to take $\mathcal{O}(n \log n)$ expected time for structures of $\mathcal{O}(n)$ expected size. Moreover, searching takes w.h.p. $\mathcal{O}(\log n)$ comparisons in the first and w.h.p. $\mathcal{O}(\log^2 n)$ comparisons in the latter two DAGs. However, the expected depth of the DAGs and high probability bounds for their size are unknown. Using a novel analysis technique, we show that the three DAGs have w.h.p. i) a size of $\mathcal{O}(n)$, ii) a depth of $\mathcal{O}(\log n)$, and iii) a construction time of $\mathcal{O}(n \log n)$. One application of these new and improved results are *remarkably simple* Las Vegas verifiers to obtain search DAGs with optimal worst-case bounds. This positively answers the conjectured logarithmic search cost in the DAG of Delaunay triangulations [Guibas et al.; ICALP 1990] and a conjecture on the depth of the DAG of Trapezoidal subdivisions [Hemmer et al.; ESA 2012]. It also shows that history-based RIC circumvents a lower bound on runtime tail estimates of conflict-graph RICs [Sen; STACS 2019].

Martin P. Seybold
Algorithms Group - FMI
University of Stuttgart
mpseybold@gmail.com

Joachim Gudmundsson
University of Sydney
joachim.gudmundsson@sydney.edu.au

CP9

Nearly 2-Approximate Distance Oracles in Subquadratic Time

Let $G = (V, E)$ be an unweighted undirected graph on n vertices and m edges. For a fixed pair of real values $\alpha \geq 1, \beta \geq 0$, an (α, β) distance oracle of G is a space-efficient data structure that answers, in constant time, for any pair of vertices $u, v \in V$ a distance estimate within the range of $[\mathbf{dist}(u, v), \alpha \cdot \mathbf{dist}(u, v) + \beta]$; here $\mathbf{dist}$ denotes distances in the graph G. Two main concerns in designing distance oracles are the approximation ratio (the stretch) and the construction time. A classical result was given in [Baswana, Goyaland and Sen 2005] which builds a $(2, 3)$ distance oracle with $\tilde{O}(n^{5/3})$ space in $\tilde{O}(n^2)$ time. Recently, [Akav and Roditty, 2020] broke the quadratic running time at the expense of increasing the stretch. More specifically, they obtained an algorithm that constructs a $(2 + \epsilon, 5)$ distance oracle with space $\tilde{O}(n^{11/6})$ in $O(m + n^{2-\Omega(\epsilon)})$ time for any constant $\epsilon \in (0, 1/2)$. In this paper, we show that one can beat the quadratic running time without compromising on the stretch. More specifically, our algorithm constructs a $(2, 3)$ distance oracle with $\tilde{O}(n^{5/3})$ space in $\tilde{O}(m + n^{1.987})$ time. As secondary results, we further reduce the preprocessing time to $\tilde{O}(m + n^{7/4+\epsilon})$ and $\tilde{O}(m + n^{5/3+\epsilon})$ by tolerating larger stretch.

Tianyi Zhang
IIIS, Tsinghua
tianyi-z16@tsinghua.org.cn

Shiri Chechik
Tel-Aviv University, Israel
shiri.chechik@gmail.com

CP10

A Two-Pass (Conditional) Lower Bound for Semi-Streaming Maximum Matching

We prove a lower bound on the space complexity of two-pass semi-streaming algorithms that approximate the maximum matching problem. The lower bound is parameterized by the density of Ruzsa- Szemer edi graphs: Any two-pass semi-streaming algorithm for maximum matching has approximation ratio at most $1 - O(\log(RS(n))/\log(n))$, where RS(n) denotes the maximum number of induced matchings of size T(n) in any n-vertex graph, i.e., the largest density of a Ruzsa-Szemer edi graph. Currently, it is known that $n^O(1/\log\log n) = RS(n) = n/2^O(\log *(n))$ and closing this (large) gap between upper and lower bounds has remained a notoriously difficult problem in combinatorics. Under the plausible hypothesis that $RS(n) = n^O(1)$, our lower bound is the first to rule out small-constant approximation two-pass semi-streaming algorithms for the maximum matching problem, making progress on a longstanding open question in the graph streaming literature.

Sepehr Assadi
Princeton University and Rutgers University, U.S,
sepehr.assadi@rutgers.edu

CP10

Streaming Regular Expression Membership and Pattern Matching

A regular expression is a formalism for compactly describing a set of strings, built recursively from single characters using three operators: concatenation, union, and Kleene star. Two basic problems concerning regular expressions are membership and pattern matching. By now we have a good understanding of the complexity of regular expression membership and pattern matching in the classical setting. However, only some special cases have been considered in the practically relevant streaming setting: dictionary matching and wildcard pattern matching. Both problems can be solved in the streaming model by a randomised Monte Carlo algorithm that uses $\mathcal{O}(d\log m)$ space [Golan and Porat (ESA 2017), Golan, Kopelowitz and Porat (Algorithmica 2019)], where d is either the number of strings in the dictionary (for the dictionary matching problem) or the number of wildcards (for the wildcard pattern matching). This is possibly much smaller than the length of the corresponding regular expression. In the general case, we cannot hope for a streaming algorithm with space complexity smaller than the length of R. The main contribution of this paper is that we identify the number of unions and Kleene stars, denoted by d, as the parameter that allows for an efficient streaming algorithm. We design general randomised Monte Carlo algorithms for both problems that use $\mathcal{O}(d^3 \text{polylog}\, n)$ space in the streaming setting.

Pawel Gawrychowski, Bartlomiej Dudek
University of Wroclaw
gawry@cs.uni.wroc.pl, bardek.dudek@gmail.com

Garance Gourdel
DI/ENS, PSL Research University, IRISA Inria Renne
garance.gourdel@gmail.com

Tatiana Starikovskaya
Ecole Normale Supérieure
Paris
tat.starikovskaya@gmail.com

CP10

Semi-Streaming Bipartite Matching in Fewer Passes and Optimal Space

We provide $\widetilde{O}(\epsilon^{-1})$-pass semi-streaming algorithms for computing $(1 - \epsilon)$-approximate maximum cardinality matchings in bipartite graphs. Our most efficient methods are deterministic and use optimal, $O(n)$, space, improving upon the space complexity of the previous state-of-the-art $\widetilde{O}(\epsilon^{-1})$-pass algorithm of [Ahn and Guha, Access to data and number of iterations: Dual primal algorithms for maximum matching under resource constraints, 2018]. To obtain our results we provide semi-streaming adaptations of more general continuous optimization tools. Further, we leverage these techniques to obtain improvements for streaming variants of approximate linear programming, optimal transport, exact matching, transshipment, and shortest path problems.

Yujia Jin
Stanford University
yujiajin@stanford.edu

Sepehr Assadi
Princeton University and Rutgers University, U.S,
sepehr.assadi@rutgers.edu

Arun Jambulapati, Aaron Sidford, Kevin Tian
Stanford University
jmblpati@stanford.edu, sidford@stanford.edu, kjtian@stanford.edu

CP10

Frequency Estimation with One-Sided Error

Frequency estimation is one of the most fundamental problems in streaming algorithms. Given a stream S of elements from some universe $U = \{1 \ldots n\}$, the goal is to compute, in a single pass, a short sketch of S so that for any element $i \in U$, one can estimate the number x_i of times i occurs in S based on the sketch alone. Two state of the art solutions to this problems are the Count-Min and Count-Sketch algorithms. The frequency estimator $\tilde{x}$ produced by Count-Min, using $O(1/\varepsilon \cdot \log n)$ dimensions, guarantees that $\|\tilde{x} - x\|_\infty \leq \varepsilon \|x\|_1$ with high probability, and $\tilde{x} \geq x$ holds deterministically. Also, Count-Min works under the assumption that $x \geq 0$. On the other hand, Count-Sketch, using $O(1/\varepsilon^2 \cdot \log n)$ dimensions, guarantees that $\|\tilde{x} - x\|_\infty \leq \varepsilon \|x\|_2$ with high probability. A natural question is whether it is possible to design the best of both worlds sketching method, with error guarantees depending on the ℓ_2 norm and space comparable to Count-Sketch, but (like Count-Min) also has the no-underestimation property. Our main set of results shows that the answer to the above question is negative. We show this in two incomparable computational models: linear sketching and streaming algorithms. We also study the complementary problem, where the sketch is required to not over-estimate, i.e., $\tilde{x} \leq x$ should hold always.

Piotr Indyk, Shyam Narayanan
Massachusetts Institute of Technology
indyk@mit.edu, shyamsn@mit.edu

David Woodruff
Carnegie Mellon University
dwoodruf@cs.cmu.edu

CP11
Strong Recovery of Geometric Planted Matchings

We study the problem of efficiently recovering the matching between an unlabelled collection of n points in $\mathbb{R}^d$ and a small random perturbation of those points. We consider a model where the initial points are i.i.d. standard Gaussian vectors, perturbed by adding i.i.d. Gaussian vectors with variance σ^2. In this setting, the maximum likelihood estimator (MLE) can be found in polynomial time as the solution of a linear assignment problem. We establish thresholds on σ^2 for the MLE to perfectly recover the planted matching (making no errors) and to strongly recover the planted matching (making $o(n)$ errors) both for d constant and $d = d(n)$ growing arbitrarily. Between these two thresholds, we show that the MLE makes $n^{\delta+o(1)}$ errors for an explicit $\delta \in (0,1)$. These results extend a recent line of work on recovering matchings planted in random graphs with independently-weighted edges to the geometric setting. Our proof techniques rely on careful analysis of the combinatorial structure of partial matchings in large, weakly dependent random graphs using the first and second moment methods.

Dmitriy Kunisky
Courant Institute of Mathematical Sciences
New York University
kunisky@cims.nyu.edu

Jonathan Niles-Weed
New York University
jnw@cims.nyu.edu

CP11
The Sparse Parity Matrix

The last decade witnessed several pivotal results on random inference problems where the aim is to learn a hidden ground truth from indirect randomised observations; much of this research has been guided by statistical physics intuition. Prominent examples include the stochastic block model, low-density parity check codes or compressed sensing. In all random inference problems studied so far the posterior distribution of the ground truth given the observations appears to enjoy a key property called 'strong replica symmetry'. This means that the overlap of the posterior distribution with the ground truth (basically the number of bits that can be learned correctly) concentrates on a deterministic value. Whether this is generally true has been an open question. In this paper we discover an example of an inference problem based on a very simple random matrix over $\mathbb{F}_2$ that fails to exhibit strong replica symmetry. Beyond its impact on random inference problems, the random matrix model, reminiscent of the binomial Erds-Rnyi random graph, gives rise to a natural random constraint satisfaction problem related to the intensely studied random k-XORSAT problem.

Amin Coja-Oghlan
Goethe University, Germany
amin.coja-oghlan@tu-dortmund.de

Oliver Cooley
Graz University of Technology
cooley@math.tugraz.at

Mihyun Kang
Technische Universitaet Graz
Institut fuer Optimierung und Diskrete Mathematik
kang@math.tugraz.at

Joon Lee, Jean Ravelomanana
TU Dortmund
joon.lee@tu-dortmund.de,   jean.ravelomanana@tu-dortmund.de

CP11
Near-Optimal Average-Case Approximate Trace Reconstruction from Few Traces

In the standard trace reconstruction problem, the goal is to *exactly* reconstruct an unknown source string $x \in \{0,1\}^n$ from independent "traces", which are copies of x that have been corrupted by a δ-deletion channel which independently deletes each bit of x with probability δ and concatenates the surviving bits. We study the *approximate* trace reconstruction problem, in which the goal is only to obtain a high-accuracy approximation of x rather than an exact reconstruction. We give an efficient algorithm, and a near-matching lower bound, for approximate reconstruction of a random n-bit source string x from few traces. Our main algorithmic result is a polynomial-time algorithm with the following property: for any deletion rate $0 < \delta < 1$ (which may depend on n), for almost every n-bit source string x, given any number $M \leq \Theta(1/\delta)$ of traces from $\mathrm{Del}_\delta(x)$, the algorithm constructs a hypothesis string $\widehat{x}$ that has edit distance at most $n \cdot (\delta M)^{\Omega(M)}$ from x. We also prove a near-matching information-theoretic lower bound showing that given $M \leq \Theta(1/\delta)$ traces from $\mathrm{Del}_\delta(x)$ for a random n-bit string x, the smallest possible expected edit distance that any algorithm can achieve, regardless of its running time, is $n \cdot (\delta M)^{O(M)}$.

Sandip Sinha, Xi Chen
Columbia University
sandip@cs.columbia.edu, xichen@cs.columbia.edu

Anindya De
University of Pennsylvania
anindyad@cis.upenn.edu

Chin Ho Lee
Harvard University
chlee@seas.harvard.edu

Rocco Anthony Servedio
Columbia University
rocco@cs.columbia.edu

CP11
Average-Case Subset Balancing Problems

Given a set of n input integers, the *Equal Subset Sum* problem asks us to find two distinct subsets with the same sum. In this paper we present an algorithm that runs in time $O^*(3^{0.386n})$ in the average case, significantly improving over the $O^*(3^{0.488n})$ running time of the best known worst-case algorithm and the Meet-in-the-Middle benchmark of $O^*(3^{0.5n})$. Our algorithm generalizes to a number of related problems, such as the "*Generalized Equal Subset Sum*' problem which asks us to assign a coefficient c_i from a set C to each input number x_i such that $\sum_i c_i x_i = 0$. Our algorithm for the average-case version of this problem runs in time $|C|^{(0.5-c_0/|C|)n}$ for some positive constant c_0, when-

ever $C = \{0, \pm 1, \ldots, \pm d\}$ or $\{\pm 1, \ldots, \pm d\}$ for some positive integer d (with runtime $O^*(|C|^{0.45n})$ when $|C| < 10$). Our results extend to that of finding "nearly balanced" solutions in which the target is not 0 but some not-too-large offset τ. Our approach relies on new structural results that characterize the probability that $\sum_i c_i x_i = \tau$ has a solution $c \in C^n$ when x_i's are chosen randomly; these results may be of independent interest. Our algorithm is inspired by the "representation technique" introduced by Howgrave-Graham and Joux.

Xi Chen, Yaonan Jin, Tim Randolph, Rocco A. Servedio
Columbia University
xichen@cs.columbia.edu, yj2552@columbia.edu, t.randolph@columbia.edu, rocco@cs.columbia.edu

CP12
Shared-Memory N-Level Hypergraph Partitioning

We present a shared-memory algorithm to compute high-quality solutions to the balanced k-way hypergraph partitioning problem. This problem asks for a partition of the vertex set into k disjoint blocks of bounded size that minimizes the connectivity metric (i.e., the sum of the number of different blocks connected by each hyperedge). High solution quality is achieved by parallelizing the core technique of the currently best sequential partitioner KaHyPar: the most extreme n-level version of the widely used multilevel paradigm, where only a single vertex is contracted on each level. This approach is made fast and scalable through intrusive algorithms and data structures that allow precise control of parallelism through atomic operations and fine-grained locking. We perform extensive experiments on more than 500 real-world hypergraphs with up to 140 million vertices and two billion pins (sum of hyperedge sizes). We find that our algorithm computes solutions that are on par with a comparable configuration of KaHyPar while being a factor of 9 faster using 10 threads.

Lars Gottesbüren, Tobias Heuer
Karlsruhe Institute of Technology
lars.gottesbueren@kit.edu, tobias.heuer@kit.edu

Peter Sanders
Karlsruhe Institue of Technology
Institute of Theoretical Informatics
sanders@kit.edu

Sebastian Schlag
Karlsruhe Institute of Technology
research@sebastianschlag.de

CP12
Fast Multimodal Journey Planning for Three Criteria

We study the journey planning problem for multimodal networks consisting of public transit and a non-schedule-based transfer mode (e.g., walking, bicycle, e-scooter). So far, all efficient algorithms for this problem either restrict usage of the transfer mode or Pareto-optimize only two criteria: arrival time and the number of used public transit trips. However, we show that both limitations must be lifted in order to obtain high-quality solutions. In particular, the time spent using the (unrestricted) transfer mode must be optimized as a third criterion. We present McTB, the first algorithm that optimizes three criteria efficiently by avoiding costly data structures for maintaining Pareto sets. To enable unlimited transfers, we combine it with a

three-criteria extension of the ULTRA preprocessing technique. Furthermore, since full Pareto sets become impractically large for more than two criteria, we adapt an approach by Delling et al. to restrict the Pareto set in a methodical manner. Extensive experiments on real-world data show that our algorithms are fast enough for interactive queries even on large country-sized networks. Compared to the state of the art for multicriteria multimodal journey planning, MCR, we achieve a speedup of up to 80.

Moritz Potthoff, Jonas Sauer
Karlsruhe Institute of Technology
moritz.potthoff@student.kit.edu, jonas.sauer2@kit.edu

CP12
Perturbation Analysis of Practical Algorithms for the Maximum Scatter Travelling Salesman Problem

The Maximum Scatter Traveling Salesman Problem (MSTSP) is a variant of the famous Traveling Salesman Problem (TSP) and finds its use in several real-world applications including manufacturing, imaging and laser melting processes. The objective of this NP-hard problem is to maximize the cost of the least cost edge in a tour of an input graph. While several approximation algorithms have been proposed for this problem, many of them suffer from bad worst-case complexities and present challenges in scalability and practical use. Besides, these algorithms have often been designed and evaluated with a sole focus on theoretical approximation quality, while practical applications require detailed experimental evaluations to study the stability, quality and runtime over a large and diverse set of inputs. In this work, we describe six algorithms for MSTSP inspired by prior work in the area, with improved formulations that enhance their utility in real-world scenarios. Further, we perform experimental studies motivated by smoothed analysis to comprehensively evaluate these algorithms on various performance metrics. We demonstrate that despite having bad worst-case complexities, certain algorithms perform exceedingly well in practical use cases. Our experiments reveal trade-offs among the runtime, quality and stability of different algorithms that must be considered when making a design choice depending on the objectives and constraints associated with the use of the algorithm.

Emil Biju, Sundar Raman
Indian Institute of Technology Madras
emilbiju7@gmail.com, sundarramanp2000@gmail.com

CP12
Efficient Access History for Race Detection

While there has been extensive research on race-detection algorithms for task parallel programs, most of this research has focused on optimizing a particular component — namely reachability analysis, which checks whether two instructions are logically in parallel. Little attention has been paid to the other important component, namely the access history, which stores all memory locations previous instructions have accessed. In theory, the access history component adds no asymptotic overhead; however, in practice, it is often the most expensive component of race detection since it is queried and (possibly) updated at each memory access. We optimize this component based on the observation that, typically, strands within parallel programs access contiguous blocks of memory. Therefore, instead of maintaining the access history at the granular-

ity of individual memory locations, we maintain it at the granularity of these (varying size) intervals. To enable this access history, we propose (1) compiler and runtime mechanisms that allow us to efficiently collect these intervals and (2) a tree-based access history data structure that allows us to update and query it at this interval granularity. The resulting tool can race detect fork-join code with amortized constant overhead, assuming the number of intervals is small compared to the total work of the computation. Our evaluations indicate that this technique improves the performance of race detection on several benchmarks.

Yifan Xu, Anchengcheng Zhou
Washington University in St. Louis
xuyifan@wustl.edu, ann.zhou@wustl.edu

Grace Yin
Massachusetts Institute of Technology
graceyin@mit.edu

Kunal Agrawal
Washington University in St. Louis
kunal@wustl.edu

I-Ting A. Lee
Washington University in St. Louis
Washington University in St. Louis
angelee@wustl.edu

Tao Schardl
Massachusetts Institute of Technology (CSAIL)
neboat@mit.edu

CP13

Approximation Schemes for Capacitated Vehicle Routing on Graphs of Bounded Treewidth, Bounded Doubling, Or Highway Dimension

In this paper, we present Approximation Schemes for Capacitated Vehicle Routing Problem (CVRP) on several classes of graphs. In CVRP we are given a graph $G = (V, E)$ with metric edges costs, a depot $r \in V$, and a vehicle of bounded capacity Q. The goal is to find a minimum cost collection of tours for the vehicle that returns to the depot, each visiting at most Q nodes, such that they cover all the nodes. This generalizes classic TSP and has been studied extensively. In the more general setting, each node v has a demand d_v and the total demand of each tour must be no more than Q. Either the demand of each node must be served by one tour (unsplittable) or can be served by multiple tours (splittable). The best known approximation algorithm for general graphs has ratio $\alpha + 2(1 - \epsilon)$ (for the unsplittable) and $\alpha + 1 - \epsilon$ (for the splittable) for some fixed $\epsilon > \frac{1}{3000}$, where α is the best approximation for TSP. Even for the case of trees, the best approximation ratio is $4/3$. Das and Mathieu presented an approximation scheme with time $n^{\log^{O(1/\epsilon)} n}$ for Euclidean plane $\mathbb{R}^2$. No other approximation scheme is known for any other class of metrics. In this paper, we make significant progress on this classic problem by presenting Quasi-Polynomial Time Approximation Schemes (QPTAS) for graphs of bounded treewidth, graphs of bounded highway dimensions, and graphs of bounded doubling dimensions.

Aditya Jayapraka, Mohammad Salavatipour
University of Alberta
jayaprak@ualberta.ca, mrs@ualberta.ca

CP13

A 3-Approximation Algorithm for Maximum Independent Set of Rectangles

We study the Maximum Independent Set of Rectangles (MISR) problem, where we are given a set of axis-parallel rectangles in the plane and the goal is to select a subset of non-overlapping rectangles of maximum cardinality. In a recent breakthrough, Mitchell [2021] obtained the first constant-factor approximation algorithm for MISR. His algorithm achieves an approximation ratio of 10 and it is based on a dynamic program that intuitively recursively partitions the input plane into special polygons called corner-clipped rectangles, without intersecting certain special horizontal line segments called fences. In this paper, we present a 3-approximation algorithm for MISR which is based on a similar recursive partitioning scheme. First, we use a partition into a more general class of axis-parallel polygons with constant complexity each, which allows us to provide an arguably simpler analysis and at the same time already improves the approximation ratio to 6. Then, using a more elaborate charging scheme and a recursive partitioning into general axis-parallel polygons with constant complexity, we improve our approximation ratio to 3. In particular, our partitioning uses more general fences that can be sequences of up to $O(1)$ line segments each. This and our other new ideas may be useful for future work towards a PTAS for MISR.

Madhusudhan R. Pittu
IIT Kharagpur
Carnegie Mellon University
pmsreddifeb18@gmail.com

Waldo Galvez
Universidad de OHiggins, Chile
Technical University of Munich in Germany
waldo.galvez@uoh.cl

Arindam Khan
Indian Institute of Science, Bengaluru, India
arindamkhan@iisc.ac.in

Mathieu Mari
Ecole Normale Superieure, Paris
mathieu.mari@ens.fr

Tobias Mömke
University of Augsburg, Germany
moemke@informatik.uni-augsburg.de

Andreas Wiese
Universidad de Chile
awiese@dii.uchile.cl

CP13

Approximate Hypergraph Vertex Cover and Generalized Tuza's Conjecture

A famous conjecture of Tuza states that the minimum number of edges needed to cover all the triangles in a graph is at most twice the maximum number of edge-disjoint triangles. This conjecture was couched in a broader setting by Aharoni and Zerbib who proposed a hypergraph version of this conjecture, and also studied its implied fractional versions. We establish the fractional version of the Aharoni-

Zerbib conjecture up to lower order terms. Specifically, we give a factor $t/2 + O(\sqrt{t \log t})$ approximation based on LP rounding for an algorithmic version of the hypergraph Turán problem (AHTP). The objective in AHTP is to pick the smallest collection of $(t-1)$-sized subsets of vertices of an input t-uniform hypergraph such that every hyperedge contains one of these subsets. Aharoni and Zerbib also posed whether Tuza's conjecture and its hypergraph versions could follow from non-trivial duality gaps between vertex covers and matchings on hypergraphs that exclude certain sub-hypergraphs, for instance, a "tent" structure that cannot occur in the incidence of triangles and edges. We give a strong negative answer to this question, by exhibiting tent-free hypergraphs, and indeed $\mathcal{F}$-free hypergraphs for any finite family $\mathcal{F}$ of excluded sub-hypergraphs, whose vertex covers must include almost all the vertices.

Venkatesan Guruswami, Sai Sandeep
Carnegie Mellon University
venkatg@cs.cmu.edu, spallerl@andrew.cmu.edu

CP13
Unsplittable Flow on a Path: The Game!

The unsplittable flow on a path (UFP) problem is a well-studied optimization problem, and it has applications in various settings like bandwidth allocation, caching, and scheduling. We are given a path with capacities on its edges and a set of n tasks, each of them defined via a demand, a subpath, and a profit. The goal is to select the most profitable set of tasks that together respect the edge capacities, i.e., for each edge e the total demand of the selected tasks whose subpath contains e is at most the capacity of e. Informally, a task is large if its demand is at least an ϵ-fraction of the capacity of some edge on its path, and small otherwise. If all tasks are large, a PTAS can be obtained via dynamic programming. The same approach fails for small tasks since then this number can be up to $\Omega(n)$ which would yield an exponential number of states. In this paper we introduce a novel randomized sketching technique to address this issue. We model the computation of a solution as a solitary game where tasks are presented one by one to a player, who has to decide for each task i whether to select i (hence getting its profit) or not. We can use the game to construct a feasible UFP solution which is $1 + \frac{1}{1+e} + \epsilon < 1.269$ approximate in expectation. It is potentially possible that a more sophisticated probabilistic analysis gives a PTAS for the problem.

Fabrizio Grandoni
IDSIA
fabrizio@idsia.ch

Tobias Mömke
University of Augsburg, Germany
moemke@informatik.uni-augsburg.de

Andreas Wiese
Universidad de Chile
awiese@dii.uchile.cl

CP14
On Complete Classes of Valuated Matroids

We characterize a rich class of valuated matroids, called R-minor valuated matroids that includes the indicator functions of matroids, and is closed under operations such as taking minors, duality, and induction by network. We refute the refinement of a 2003 conjecture by Frank, exhibit-ing valuated matroids that are not R-minor. The class of counterexamples is based on sparse paving matroids. Valuated matroids are inherently related to gross substitute valuations in mathematical economics. By the same token we refute the Matroid Based Valuation Conjecture by Ostrovsky and Paes Leme (Theoretical Economics 2015) asserting that every gross substitute valuation arises from weighted matroid rank functions by repeated applications of merge and endowment operations. Our result has also implications in the context of Lorentzian polynomials: it reveals the limitations of known construction operations.

Edin Husic
London School of Economics (LSE)
London School of Economics (LSE)
e.husic@lse.ac.uk

Georg Loho
University of Twente
g.loho@utwente.nl

Ben Smith
University of Manchester
Heilbronn Institute for Mathematical Research
benjamin.smith-3@manchester.ac.uk

László Végh
London School of Economics
l.vegh@lse.ac.uk

CP14
Twin-Width Vi: the Lens of Contraction Sequences

The recently introduced twin-width graph invariant is based on iterative mergers, called a contraction sequence, of two vertices until only one vertex remains. If one puts red edges between two vertices representing non-homogeneous vertex subsets, the twin-width is the minimum integer d such that a contraction sequence exists that keeps red degree at most d. In this paper, we explore new avenues along the general theme of contraction sequences by changing the condition imposed on the trigraphs and possibly slightly tweaking the notion of contractions. The main results are the following.

- We characterize the well-established graph classes such as graphs of bounded rank-width, tree-width, linear rank-width, path-width, and proper minor-closed classes by means of contraction sequences. As an application we give a transparent alternative proof of the celebrated Courcelle's theorem that MSO_2 model checking on graphs with bounded tree-width is fixed-parameter tractable in the size of the input sentence.

- We define an oriented version of twin-width, which turns out to be equivalent to twin-width. This greatly simplifies the task of showing that a class has bounded twin-width.

- We show that FO model checking (resp. $\exists$FO model checking) is fixed-parameter tractable on classes with partial contraction sequences to a class of bounded degree (resp. bounded expansion), provided such a sequence is given.

Édouard Bonnet
Univ Lyon, CNRS, ENS de Lyon, France
edouard.bonnet@ens-lyon.fr

Eun Jung Kim
CNRS, LAMSADE-Universite Paris Dauphine

eunjungkim78@gmail.com

Amadeus Reinald
Univ Lyon, CNRS, ENS de Lyon, France
amadeus.reinald@ens-lyon.fr

Stéphan Thomassé
ENS de Lyon
stephan.thomasse@ens-lyon.fr

CP14

A Sublinear Bound on the Page Number of Upward Planar Graphs

The page number of a directed acyclic graph G is the minimum k for which there is a topological ordering of G and a k-coloring of the edges such that no two edges of the same color cross, i.e., have alternating endpoints along the topological ordering. We address the long-standing open problem asking for the largest page number among all upward planar graphs. We improve the best known lower bound to 5 and present the first asymptotic improvement over the trivial $\mathcal{O}(n)$ upper bound, where n denotes the number of vertices in G. Specifically, we first prove that the page number of every upward planar graph is bounded in terms of its width, as well as its height. We then combine both approaches to show that every n-vertex upward planar graph has page number $\mathcal{O}(n^{2/3} \log^{2/3}(n))$.

Paul Jungeblut, Laura Merker, Torsten Ueckerdt
Karlsruhe Institute of Technology
paul.jungeblut@kit.edu,   laura.merker2@kit.edu,
torsten.ueckerdt@kit.edu

CP14

A Deterministic Parallel Reduction from Weighted Matroid Intersection Search to Decision

Given two matroids on the same ground set, the matroid intersection problem asks for a common base, i.e., a subset of the ground set that is a base in both the matroids. The weighted version of the problem asks for a common base with maximum weight. In the general case, when the two matroids are given via rank oracles, the question of its parallel complexity is completely open. In the case of linearly representable matroids, the problem is known to have randomized parallel (RNC) algorithms, when the given weights are polynomially bounded. Finding a deterministic parallel (NC) algorithm in this case, even for the decision question, has been a long-standing open question. We make some progress towards understanding the parallel complexity of matroid intersection by showing that the weighted matroid intersection (WMI) search problem is equivalent to its decision version, in a parallel model of computation. More precisely, we give an NC algorithm for WMI-search using oracle access to WMI-decision. This resolves an open question posed by Anari and Vazirani (ITCS 2020).

Sumanta Ghosh, Rohit Gurjar, Roshan Raj
IIT Bombay
besusumanta@gmail.com, rgurjar@cse.iitb.ac.in, roshan-raj@cse.iitb.ac.in

CP14

Perfect Matching in Random Graphs is as Hard as

Tseitin

We study the complexity of proving that a sparse random regular graph on an odd number of vertices does not have a perfect matching, and related problems involving each vertex being matched some pre-specified number of times. We show that this requires proofs of degree $\Omega(n/\log n)$ in the Polynomial Calculus (over fields of characteristic $\neq 2$) and Sum-of-Squares proof systems, and exponential size in the bounded-depth Frege proof system. This resolves a question by Razborov asking whether the Lovsz-Schrijver proof system requires n^{δ} rounds to refute these formulas for some $\delta > 0$. The results are obtained by a worst-case to average-case reduction of these formulas relying on a topological embedding theorem which may be of independent interest.

Per Austrin, Kilian Risse
KTH Royal Institute of Technology
austrin@kth.se, kilianr@kth.se

CP15

CLAP: A New Algorithm For Promise CSPs

We propose a new algorithm for Promise Constraint Satisfaction Problems (PCSPs). It is a combination of the **C**onstraint Basic **L**P relaxation and the **A**ffine I**P** relaxation (CLAP). We give a characterisation of the power of CLAP in terms of a minion homomorphism. Using this characterisation, we identify a certain weak notion of symmetry which, if satisfied by infinitely many polymorphisms of PCSPs, guarantees tractability. We demonstrate that there are PCSPs solved by CLAP that are not solved by any of the existing algorithms for PCSPs; in particular, not by the BLP+AIP algorithm of Brakensiek and Guruswami [SODA'20] and not by a reduction to tractable finite-domain CSPs.

Lorenzo Ciardo
University of Oxford
United Kingdom
lorenzo.ciardo@cs.ox.ac.uk

Stanislav Zivny
University of Oxford, United Kingdom
standa.zivny@cs.ox.ac.uk

CP15

Promise Constraint Satisfaction and Width

We study the power of the bounded-width consistency algorithm for Promise Constraint Satisfaction(PCSP). Our main technical finding is that the template of every PCSP that is solvable in bounded width satisfies a certain structural condition implying that its algebraic closure-properties include weak near unanimity polymorphisms (WNUs) of all large arities. While this parallels the standard (non-promise) CSP theory, the method of proof is quite different and applies even to the regime of sublinear width. We also show that, in contrast with the CSP world, the presence of WNUs of all large arities does not guarantee solvability in bounded width. The separating example is even solvable in the second level of the Sherali-Adams (SA) hierarchy of linear programming relaxations. This shows that, unlike for CSPs, LP can be stronger than bounded width. A direct application of these methods also show that the problem of q-coloring p-colorable graphs is not solvable in bounded or even sublinear width, for any two constants p and q

such that $3 \leq p \leq q$. Turning to algorithms, we note that Wigderson's algorithm for $O(\sqrt{n})$-coloring 3-colorable graphs with n vertices is implementable in width 4. Indeed, by generalizing the method we see that, for any $\epsilon > 0$ smaller than $1/2$, the optimal width for solving the problem of $O(n^\epsilon)$-coloring 3-colorable graphs with n vertices lies between $n^{1-3\epsilon}$ and $n^{1-2\epsilon}$.

Victor Dalmau
University Pompeu Fabra, Barcelona
vdalmau@gmail.com

Albert Atserias
Universitat Politècnica de Catalunya
atserias@cs.upc.edu

CP15

High Dimensional Expanders: Eigenstripping, Pseudorandomness, and Unique Games

Higher order random walks (HD-walks) on high dimensional expanders (HDX) have seen an incredible amount of study and application since their introduction by [Kaufman and Mass "High Dimensional Random Walks and Colorful Expansion"], yet their broader combinatorial and spectral properties remain poorly understood. We develop a combinatorial characterization of the spectral structure of HD-walks on two-sided local-spectral expanders [Dinur and Kaufman "High Dimensional Expanders Imply Agreement Expanders"], which offer a broad generalization of the well-studied Johnson and Grassmann graphs. Our characterization, which shows that the spectra of HD-walks lie tightly concentrated in a few combinatorially structured strips, leads to novel structural theorems such as a tight ℓ_2-characterization of edge-expansion, as well as to a new understanding of local-to-global graph algorithms on HDX. Towards the latter, we introduce a novel spectral complexity measure called Stripped Threshold Rank, and show how it can replace the (much larger) threshold rank as a parameter controlling the performance of algorithms on structured objects. Combined with a sum-of-squares proof for the former ℓ_2-characterization, we give a concrete application of this framework to algorithms for unique games on HD-walks, where in many cases we improve the state of the art from nearly-exponential to polynomial time.

Mitali Bafna
John A. Paulson School of Engineering and Applied Sciences
Harvard University
mitalibafna@g.harvard.edu

Max Hopkins
University of California San Diego
nmhopkin@eng.ucsd.edu

Tali Kaufman
Bar-Ilan University, Israel
kaufmant@mit.edu

Shachar Lovett
University of California San Diego
slovett@cs.ucsd.edu

CP15

Algorithmic Thresholds for Refuting Random

Polynomial Systems

Consider a system of m polynomial equations $\{p_i(x) = b_i\}_{i \leq m}$ of degree $D \geq 2$ in n-dimensional variable $x \in \mathbb{R}^n$ such that each coefficient of every p_i and b_is are chosen at random and independently from some continuous distribution. We study the basic question of determining the smallest m – the algorithmic threshold – for which efficient algorithms can find refutations (i.e. certificates of unsatisfiability) for such systems. This setting generalizes problems such as refuting random SAT instances, low-rank matrix sensing and certifying pseudo-randomness of Goldreich's candidate generators and generalizations. We show that for every $d \in \mathbb{N}$, the $(n+m)^{O(d)}$-time canonical sum-of-squares (SoS) relaxation refutes such a system with high probability whenever $m \geq O(n) \cdot (\frac{n}{d})^{D-1}$. We prove a lower bound in the restricted low-degree polynomial model of computation which suggests that this trade-off between SoS degree and the number of equations is nearly tight for all d. We also confirm the predictions of this lower bound in a limited setting by showing a lower bound on the canonical degree-4 sum-of-squares relaxation for refuting random quadratic polynomials. Together, our results provide evidence for an algorithmic threshold for the problem at $m \gtrsim \widetilde{O}(n) \cdot n^{(1-\delta)(D-1)}$ for 2^{n^δ}-time algorithms for all δ.

Jun-Ting Hsieh
Carnegie Mellon University
juntingh@cs.cmu.edu

Pravesh Kothari
CMU
praveshk@cs.cmu.edu

CP15

Combinatorial Gap Theorem and Reductions Between Promise CSPs

A value of a CSP instance is typically defined as a fraction of constraints that can be simultaneously met. We propose an alternative definition of a value of an instance and show that, for purely combinatorial reasons, a value of an unsolvable instance is bounded away from one; we call this fact a gap theorem. We show that the gap theorem implies NP-hardness of a gap version of the Layered Label Cover Problem. The same result can be derived from the PCP Theorem, but a full, self-contained proof of our reduction is quite short and the result can still provide PCP-free NP-hardness proofs for numerous problems. The simplicity of our reasoning also suggests that weaker versions of Unique-Games-type conjectures, e.g., the d-to-1 conjecture, might be accessible and serve as an intermediate step for proving these conjectures in their full strength. As the second, main application we provide a sufficient condition under which a fixed template Promise Constraint Satisfaction Problem (PCSP) reduces to another PCSP. The correctness of the reduction hinges on the gap theorem, but the reduction itself is very simple. As a consequence, we obtain that every CSP can be canonically reduced to most of the known NP-hard PCSPs, such as the approximate hypergraph coloring problem.

Marcin Kozik
Jagiellonian University
marcin.kozik@uj.edu.pl

Barto Libor
Charles University

libor.barto@gmail.com

CP16
Learning to Prune Instances of K-Median and Related Problems

In a large number of industrial applications, combinatorial optimization problems are repeatedly solved with datasets from similar distribution. In recent years, machine learning techniques have been shown to be quite effective in speeding up such computations. However, black-box end-to-end machine learning approaches suffer from poor interpretability and the requirement for a large amount of labelled data. In this paper, we demonstrate a simple and highly effective way to incorporate the insights from the algorithmic and optimization literature on these problems into a machine learning framework to speed-up the solutions of these problems. We study the k-median problem and the following closely related combinatorial optimization problems: set cover, max coverage and uncapacitated facility location. We look at the kind of quantities employed by the approximation algorithms for these problems and use these to derive useful features for training a classifier that helps quickly reduce the problem size by identifying the difficult core of the problem and pruning the remainder. The difficult core is then solved using an ILP solver. Several of the features were derived from approximation algorithms that were designed for the metric instances of the k-median and the uncapacitated facility location problem. However, remarkably, the use of these features also leads to significant speed up in the non-metric instances of these problems and even the other two related problems.

Deepak Ajwani
University College Dublin
deepak.ajwani@gmail.com

Dena Tayebi
University College Dublin, Ireland
dena.tayebi@ucdconnect.ie

Saurabh Ray
Ben-Gurion University, Israel
saurabh.ray@gmail.com

CP16
Parallel Nearest Neighbors in Low Dimensions with Batch Updates

We present a set of parallel algorithms for computing exact k-nearest neighbors in low dimensions. Many k-nearest neighbor algorithms use either a kd-tree or the Morton ordering of the point set; our algorithms combine these approaches using a data structure we call the zd-tree. We show that this combination is both theoretically efficient under common assumptions, and fast in practice. For point sets of size n with bounded expansion constant and bounded ratio, the zd-tree can be built in $O(n)$ work with $O(n^\epsilon)$ span for constant $\epsilon < 1$, and searching for the k-nearest neighbors of a point takes expected $O(k \log k)$ time. We benchmark our k-nearest neighbor algorithms against existing parallel k-nearest neighbor algorithms, showing that our implementations are generally faster than the state of the art as well as achieving 75x speedup on 144 hyperthreads. Furthermore, the zd-tree supports parallel batch-dynamic insertions and deletions; to our knowledge, it is the first k-nearest neighbor data structure to support such updates. On point sets with bounded expansion constant and bounded ratio, a batch-dynamic update of size k requires $O(k \log n/k)$ work with $O(k^\epsilon + \text{polylog}(n))$ span.

Magdalen Dobson
CMU
mrdobson@andrew.cmu.edu

Guy Blelloch
Computer Science Department
Carnegie Mellon University
guyb@cs.cmu.edu

CP16
An Efficient Branch-and-Bound Solver for Hitting Set

The hitting set problem asks for a collection of sets over a universe U to find a minimum subset of U that intersects each of the given sets. It is NP-hard and equivalent to the problem set cover. We give a branch-and-bound algorithm to solve hitting set. Though it requires exponential time in the worst case, it can solve many practical instances from different domains in reasonable time. Our algorithm outperforms a modern ILP solver, the state-of-the-art for hitting set, by at least an order of magnitude on most instances.

Thomas Bläsius
Karlsruhe Institute of Technology
thomas.blaesius@kit.edu

Tobias Friedrich
Hasso Plattner Institute
University of Potsdam
tobias.friedrich@hpi.de

David Stangl
Hasso Plattner Institute
d4v1dstangl@gmail.com

Christopher Weyand
Karlsruhe Institute of Technology
christopher.weyand@kit.edu

CP16
Approximating 1-Wasserstein Distance Between Persistence Diagrams by Graph Sparsification

Persistence diagrams (PDs) play a central role in topological data analysis. This analysis requires computing distances among such diagrams such as the 1-Wasserstein distance. Accurate computation of these PD distances for large data sets that render large diagrams may not scale appropriately with the existing methods. The main source of difficulty ensues from the size of the bipartite graph on which a matching needs to be computed for determining these PD distances. We address this problem by making several observations in order to obtain an approximation. First, taking advantage of the proximity of PD points, we condense them thereby decreasing the number of nodes in the graph for computation. The increase in point multiplicities is addressed by reducing the matching problem to a min-cost flow problem on a transshipment network. Second, we use Well Separated Pair Decomposition to sparsify the graph to a size that is linear in the number of nodes. Both node and arc sparsifications contribute to an approximation factor for which we leverage a lower bound given by the Relaxed Word Mover's distance. Third, we eliminate bottlenecks during the sparsification procedure

by introducing parallelism. Fourth, we develop an open source software called PDoptFlow based on our algorithm, exploiting parallelism by GPU and multicore. We perform extensive experiments and show that we can achieve high performance with a guarantee of low relative error, improving upon the state of the arts.

Tamal Dey
Purdue
tamaldey@purdue.edu

Simon Zhang
Purdue University
szhang31415@gmail.com

CP17
Single Sample Prophet Inequalities via Greedy-Ordered Selection

We study *single-sample prophet inequalities* (SSPIs), i.e., prophet inequalities where only a single sample from each prior distribution is available. Besides a direct, and optimal, SSPI for the basic single choice problem [Rubinstein et al., 2020], most existing SSPI results were obtained via an elegant, but inherently lossy reduction to order-oblivious secretary (OOS) policies [Azar et al., 2014]. Motivated by this discrepancy, we develop an intuitive and versatile greedy-based technique that yields SSPIs *directly* rather than through the reduction to OOSs. Our results can be seen as generalizing and unifying a number of existing results in the area of prophet and secretary problems. Our algorithms significantly improve on the competitive guarantees for a number of interesting scenarios (including general matching with edge arrivals, bipartite matching with vertex arrivals, and certain matroids), and capture new settings (such as budget additive combinatorial auctions). Complementing our algorithmic results, we also consider mechanism design variants. Finally, we analyze the power and limitations of different SSPI approaches by providing a partial converse to the reduction from SSPI to OOS given by Azar et al.

Constantine Caramanis
The University of Texas at Austin
caramanis@mail.utexas.edu

Paul Duetting
London School of Economics
p.d.duetting@lse.ac.uk

Matthew Faw
University of Texas at Austin
matthewfaw@utexas.edu

Federico Fusco, Philip Lazos
Sapienza University of Rome
federico.fusco@uniroma1.it, plazos@gmail.com

Stefano Leonardi
Sapienza Università di Roma, Italy
leonardi@diag.uniroma1.it

Orestis Papadigenopoulos
University of Texas at Austin
papadig@cs.utexas.edu

Emmanouil Pountourakis
Drexel University
ep556@drexel.edu

Rebecca Reiffenhaeuser
RWTH Aachen University
rebeccar@diag.uniroma1.it

CP17
Robust Secretary and Prophet Algorithms for Packing Integer Programs

We study the problem of solving Packing Integer Programs (PIPs) in the online setting, where columns in $[0, 1]^d$ of the constraint matrix are revealed sequentially, and the goal is to pick a subset of the columns that sum to at most B in each coordinate while maximizing the objective. Excellent results are known in the secretary setting, where the columns are adversarially chosen, but presented in a uniformly random order. However, these existing algorithms are susceptible to adversarial attacks: they try to "learn' characteristics of a good solution, but tend to over-fit to the model, and hence a small number of adversarial corruptions can cause the algorithm to fail. In this paper, we give the first robust algorithms for Packing Integer Programs, specifically in the recently proposed Byzantine Secretary framework. Our techniques are based on a two-level use of online learning, to robustly learn an approximation to the optimal value, and then to use this robust estimate to pick a good solution. These techniques are general and we use them to design robust algorithms for PIPs in the prophet model as well, specifically in the Prophet-with-Augmentations framework. We also improve known results in the Byzantine Secretary framework: we make the non-constructive results algorithmic and improve the existing bounds for single-item and matroid constraints.

C.J. Argue, Anupam Gupta
Carnegie Mellon University
cjargue@andrew.cmu.edu, anupamg@cs.cmu.edu

MARCO Molinaro
PUC Rio
marco.molinaro@gmail.com

Sahil Singla
Georgia Tech
ssingla@gatech.edu

CP17
Tight Guarantees for Multi-Unit Prophet Inequalities and Online Stochastic Knapsack

Prophet inequalities are a useful tool for designing online allocation procedures and comparing their performance to the optimal offline allocation. In the basic setting of k-unit prophet inequalities, the magical procedure of Alaei (2011) with its celebrated performance guarantee has found widespread adoption in mechanism design and other online allocation problems. Despite being commonly used for implementing online allocation, the tightness of Alaei's procedure for a given k has remained unknown. In this paper we resolve this question, characterizing the tight bound by identifying the structure of the optimal online implementation and consequently improving the best-known guarantee for all k¿1. We also consider a more general online stochastic knapsack problem where each individual allocation can consume an arbitrary fraction of the initial capacity. We introduce a new "best-fit" procedure for implementing a fractionally-feasible knapsack solution online, with a performance guarantee of 0.319, which we also show is tight. Our analysis differs from existing ones by eschewing the need to split items into "large" or "small" based on capac-

ity consumption, using instead an invariant for the overall utilization on different sample paths. All in all, our results imply tight Online Contention Resolution Schemes for k-uniform matroids and the knapsack polytope, respectively, which has further implications in mechanism design.

Jiashuo Jiang
NYU Stern School of Business
jj2398@stern.nyu.edu

Will Ma
Massachusetts Institute of Technology
willma353@gmail.com

Jiawei Zhang
NYU Stern School of Business
jiazhangstern@gmail.com

CP17
Online Weighted Matching with a Sample

We study the greedy-based online algorithm for edge-weighted matching with (one-sided) vertex arrivals in bipartite graphs, and edge arrivals in general graphs. This algorithm was first studied more than a decade ago by Korula and Pl for the bipartite case in the random-order model. While the weighted bipartite matching problem is solved in the random-order model, this is not the case in recent and exciting online models in which the online player is provided with a sample, and the arrival order is adversarial. The greedy-based algorithm is arguably the most natural and practical algorithm to be applied in these models. Despite its simplicity and appeal, and despite being studied in multiple works, the greedy-based algorithm was not fully understood in any of the studied online models, and its actual performance remained an open question for more than a decade. We provide a thorough analysis of the greedy-based algorithm in several online models. For vertex arrivals in bipartite graphs, we characterize the exact competitive-ratio of this algorithm in the random-order model, for any arrival order of the vertices subsequent to the sampling phase (adversarial and random orders in particular). We use it to derive tight analysis in the recent adversarial-order model with a sample (AOS model) for any sample size, providing the first result in this model beyond the simple secretary problem. Then, we generalize and strengthen the black box method of converting results in the random-order model to single-sample prophet inequalities, and use it to derive the state-of-the-art single-sample prophet inequality for the problem. Finally, we use our new techniques to analyze the greedy-based algorithm for edge arrivals in general graphs and derive results in all the mentioned online models. In this case as well, we improve upon the state-of-the-art single-sample prophet inequality.

Haim Kaplan
Tel-Aviv University
haimk@post.tau.ac.il

David Naori, Danny Raz
The Technion Israel Institute of Technology
dnaori@cs.technion.ac.il, danny@cs.technion.ac.il

CP18
Monotone Edge Flips to An Orientation of Maximum Edge-Connectivity à La Nash-Williams

We initiate the study of k-edge-connected orientations of undirected graphs through edge flips for $k \geq 2$. We prove that in every orientation of an undirected $2k$-edge-connected graph, there exists a sequence of edges such that flipping their directions one by one does not decrease the edge-connectivity, and the final orientation is k-edge-connected. This yields an "edge-flip based' new proof of Nash-Williams's theorem: an undirected graph G has a k-edge-connected orientation if and only if G is $2k$-edge-connected. As another consequence of the theorem, we prove that the edge-flip graph of k-edge-connected orientations of an undirected graph G is connected if G is $(2k+2)$-edge-connected. This has been known to be true only when $k = 1$.

Yusuke Kobayashi
Kyoto University
yusuke@kurims.kyoto-u.ac.jp

Takehiro Ito
Graduate School of Information Sciences
Tohoku University, Japan
takehiro@ecei.tohoku.ac.jp

Yuni Iwamasa
Kyoto University
iwamasa@i.kyoto-u.ac.jp

Naonori Kakimura
Keio University
kakimura@math.keio.ac.jp

Naoyuki Kamiyama
Kyushu University
kamiyama@imi.kyushu-u.ac.jp

Shun-Ichi Maezawa
The University of Electro-Communications
maezawa.mw@gmail.com

Yuta Nozaki
Hiroshima University
nozakiy@hiroshima-u.ac.jp

Yoshio Okamoto
The University of Electro-Communications
okamotoy@uec.ac.jp

Kenta Ozeki
Yokohama National University
ozeki-kenta-xr@ynu.ac.jp

CP18
Recognizing k-Leaf Powers in Polynomial Time, for Constant k

A graph G is a k-leaf power if there exists a tree T whose leaf set is $V(G)$, and such that $uv \in E(G)$ if and only if the distance between u and v in T is at most k. The graph classes of k-leaf powers have several applications in computational biology, but recognizing them has remained a challenging algorithmic problem for the past two decades. The best known result is that 6-leaf powers can be recognized in polynomial time. In this paper, we present an algorithm that decides whether a graph G is a k-leaf power in time $O(n^{f(k)})$ for some function f that depends only on k (but has the growth rate of a power tower function). Our techniques are based on the fact that either a k-leaf power has a corresponding tree of low maximum degree, in

which case finding it is easy, or every corresponding tree has large maximum degree. In the latter case, large degree vertices in the tree imply that G has redundant substructures which can be pruned from the graph. In addition to solving a longstanding open problem, we hope that the structural results presented in this work can lead to further results on k-leaf powers.

Manuel Lafond
University of Sherbrooke
manuel.lafond@USherbrooke.ca

CP18

New Diameter-Reducing Shortcuts and Directed Hopsets: Breaking the $O(\sqrt{n})$ Barrier

For an n-vertex digraph $G = (V, E)$, a *shortcut set* is a (small) subset of edges H taken from the transitive closure of G that, when added to G guarantees that the diameter of $G \cup H$ is small. Shortcut sets, introduced by Thorup in 1993, have a wide range of applications in algorithm design, especially in the context of parallel, distributed and dynamic computation on directed graphs. A folklore result in this context shows that every n-vertex digraph admits a shortcut set of linear size (i.e., of $O(n)$ edges) that reduces the diameter to $\widetilde{O}(\sqrt{n})$. The question of whether one can reduce the diameter to $o(\sqrt{n})$ with $\widetilde{O}(n)$ shortcut edges has been left open. We provide the first improved diameter-sparsity tradeoff for this problem, breaking the $\sqrt{n}$ diameter barrier. Specifically, we show an $O(n^\omega)$-time randomized algorithm for computing a linear shortcut set that reduces the diameter of the digraph to $\widetilde{O}(n^{1/3})$. This narrows the gap w.r.t. the current diameter lower bound of $\Omega(n^{1/6})$ by [Huang and Pettie, SWAT'18]. Moreover, we show that a diameter of $\widetilde{O}(n^{1/2})$ can in fact be achieved with a *sublinear* number of $O(n^{3/4})$ shortcut edges. We also extend our algorithms to provide improved (β, ϵ) hopsets for n-vertex weighted directed graphs.

Merav Parter, Shimon Kogan
Weizmann Institute
merav.parter@weizmann.ac.il,   shimon.kogan@weizmann.ac.il

CP19

Faster Algorithms for Bounded-Difference Min-Plus Product

Min-plus product of two $n \times n$ matrices is a fundamental problem in algorithm research. It is known to be equivalent to APSP, and in general it has no truly subcubic algorithms. In this paper, we focus on the min-plus product on a special class of matrices, δ-bounded-difference matrices, in which the difference between any two adjacent entries is bounded by $\delta = O(1)$. Our algorithm runs in randomized time $O(n^{2.779})$ by the fast rectangular matrix multiplication algorithm [Le Gall & Urrutia 18], better than $\tilde{O}(n^{2+\omega/3}) = O(n^{2.791})$ ($\omega < 2.373$ [Alman & V.V.Williams 20]). This improves the previous result of $\tilde{O}(n^{2.824})$ [Bringmann et al. 16]. When $\omega = 2$ in the ideal case, our complexity is $\tilde{O}(n^{2+2/3})$, improving Bringmann et al.'s result of $\tilde{O}(n^{2.755})$.

Tianle Xie, Shucheng Chi
Tsinghua University
xtl17@tsinghua.org.cn, chisc21@mails.tsinghua.edu.cn

Ran Duan
Tsinghua University, Beijing, China
duanran@mail.tsinghua.edu.cn

CP19

Polynomial-Time Algorithm for Maximum Independent Set in Bounded-Degree Graphs with No Long Induced Claws

For graphs G and H, we say that G is H-free if it does not contain H as an induced subgraph. Already in the early 1980s Alekseev observed that if H is connected, then the Max Weight Independent Set problem (MWIS) remains NP-hard in H-free graphs, unless H is a path or a subdivided claw, i.e., a graph obtained from the three-leaf star by subdividing each edge some number of times (possibly zero). Since then determining the complexity of MWIS in these remaining cases is one of the most important problems in algorithmic graph theory. We make an important step towards solving the problem by showing that for any subdivided claw H, MWIS is polynomial-time solvable in H-free graphs of bounded degree.

Tara Abrishami, Maria Chudnovsky, Cemil Dibek
Princeton University
taraa@princeton.edu, mchudnov@math.princeton.edu, cdibek@princeton.edu

Pawel Rzazewski
Warsaw University of Technology, Poland
p.rzazewski@mini.pw.edu.pl

CP19

Isomorphism Testing for Graphs Excluding Small Topological Subgraphs

We give an isomorphism test that runs in time $n^{\text{polylog}(h)}$ on all n-vertex graphs excluding some h-vertex graph as a topological subgraph. Previous results state that isomorphism for such graphs can be tested in time $n^{\text{polylog}(n)}$ [Babai, STOC 2016] and $n^{f(h)}$ for some function f [Grohe and Marx, SIAM J. Comp., 2015]. Our result also unifies and extends previous isomorphism tests for graphs of maximum degree d running in time $n^{\text{polylog}(d)}$ [FOCS 2018] and for graphs of Hadwiger number h running in time $n^{\text{polylog}(h)}$ [FOCS 2020].

Daniel Neuen
CISPA Helmholtz Center for Information Security
daniel.neuen@cispa.de

CP19

Algorithmic Trade-Offs for Girth Approximation in Undirected Graphs

We present several new efficient algorithms for approximating the girth, g, of weighted and unweighted n-vertex, m-edge undirected graphs. For undirected graphs with polynomially bounded, integer, non-negative edge weights, we provide an algorithm that for every integer $k \geq 1$, runs in $\emptyset(m + n^{1+1/k} \log g)$ time and returns a cycle of length at most $2kg$. For unweighted, undirected graphs we present an algorithm that for every $k \geq 1$, runs in $\emptyset(n^{1+1/k})$ time and returns a cycle of length at most $2k\lceil g/2 \rceil$, an almost k-approximation. Both algorithms provide trade-offs between the running time and the quality of the approximation. We also obtain faster algorithms for approximation

factors better than 2, and improved approximations when the girth is odd or small (e.g., 3 and 4).

Liam Roditty, Avi Kadria
Bar-Ilan University
liamr@macs.biu.ac.il, avi.kadria3@gmail.com

Aaron Sidford
Stanford University
sidford@stanford.edu

Virginia Vassilevska Williams
MIT
virgi@mit.edu

Uri Zwick
Tel Aviv University
zwick@tau.ac.il

CP20

The Parameterized Complexity of the Survivable Network Design Problem

For the Survivable Network Design Problem (SNDP) we are given an undirected graph G with edge costs, a set R of terminal vertices, and an integer demand $d_{s,t}$ for every terminal pair $s, t \in R$. The task is to compute a subgraph H of G of minimum cost, such that there are at least $d_{s,t}$ disjoint paths between s and t in H. If the paths are required to be edge-disjoint we obtain the edge-connectivity variant (EC-SNDP), while internally vertex-disjoint paths result in the vertex-connectivity variant (VC-SNDP). Another case is the element-connectivity variant (LC-SNDP), where the paths are disjoint on edges and non-terminals. We shed light on the parameterized complexity of the above problems. We consider the solution size ℓ, the sum of demands D, the number of terminals k, and the maximum demand $d_{\max}$. - We give a complete picture of the parameterized tractability of the three variants w.r.t. parameter ℓ: both EC-SNDP and LC-SNDP are FPT, while VC-SNDP is W[1]-hard. - We identify some cases of VC-SNDP that are FPT: * when $d_{\max} \leq 3$ for parameter ℓ, * on locally bounded treewidth graphs for parameter ℓ, and * on graphs of treewidth tw for parameter $tw + D$. - The Directed Steiner Tree (DST) problem is FPT parameterized by k [Dreyfus & Wagner 1971]. We show that the 2-DST problem is W[1]-hard, when parameterized by ℓ.

Andreas E. Feldmann
KAM
Charles University
feldmann.a.e@gmail.com

Anish Mukherjee
University of Warsaw
anish343@gmail.com

Erik Jan van Leeuwen
Utrecht University
e.j.vanleeuwen@uu.nl

CP20

Simpler Adjacency Labeling for Planar Graphs with B-Trees

An adjacency labeling scheme consists of an encoder and a decoder. The encoder assigns a binary string, called a label, to each vertex of a given graph G. Then, the decoder should decide, given only the labels assigned to two vertices u and v of the same graph G, whether (u, v) is an edge. While for planar graphs labels consisting of $\mathcal{O}(\log n)$ bits are not too hard to design, determining the exact constant factor in the upper bound remained a challenging open problem, with the only lower bound being $\log n$. Only very recently, Dujmovic et al. [FOCS 2020] were able to bring the upper bound down to $\log n + \mathcal{O}(\sqrt{\log n \log \log n}) = \log n + o(\log n)$. At the heart of their construction lies a graph product structure theorem that is used to translate the problem into a data-structure question. The latter is then solved by designing a tailored balanced binary search trees that allow for efficient bulk operations. We show how this can be achieved with an arguably simpler approach based on B-Trees, while the other parts of the whole framework remain relatively unchanged. This allows us to obtain a cleaner upper bound of $\log n + \mathcal{O}(\sqrt{\log n})$ bits on the length of the labels, and additionally decrease the decoding time to constant.

Pawel Gawrychowski, Wojciech Janczewski
University of Wroclaw
gawry@cs.uni.wroc.pl,   wojciech.janczewski@cs.uni.wroc.pl

CP20

A Simple Algorithm for Multiple-Source Shortest Paths in Planar Digraphs

Given an n-vertex planar embedded digraph G with non-negative edge weights and a face f of G, Klein presented a data structure with $O(n \log n)$ space and preprocessing time which can answer any query (u, v) for the shortest path distance in G from u to v or from v to u in $O(\log n)$ time, provided u is on f. This data structure is a key tool in a number of state-of-the-art algorithms and data structures for planar graphs. Klein's data structure relies on dynamic trees and the persistence technique as well as a highly non-trivial interaction between primal shortest path trees and their duals. The construction of our data structure follows a completely different and in our opinion very simple divide-and-conquer approach that solely relies on Single-Source Shortest Path computations and contractions in the primal graph. Our space and preprocessing time bound is $O(n \log |f|)$ and query time is $O(\log |f|)$ which is an improvement over Klein's data structure when f has small size.

Evangelos Kipouridis
BARC (Basic Algorithms Research Copenhagen)
University of Copenhagen
kipouridis@di.ku.dk

Debarati Das
BARC, University of Copenhagen
debaratix710@gmail.com

Maximilian Probst Gutenberg
ETH Zurich
maximilian.probst@inf.ethz.ch

Christian Wulff-Nilsen
Department of Computer Science
University of Copenhagen
koolooz@di.ku.dk

CP20

Deterministic Approximation of Random Walks via

Queries in Graphs of Unbounded Size

Consider the following computational problem: given a regular digraph $G = (V, E)$, two vertices $u, v \in V$, and a walk length $t \in \mathbb{N}$, estimate the probability that a random walk of length t from u ends at v to within $\pm\varepsilon$. A randomized algorithm can solve this problem by carrying out $O(1/\varepsilon^2)$ random walks of length t from u and outputting the fraction that end at v. In this paper, we study *deterministic* algorithms for this problem that are also restricted to carrying out walks of length t from u and seeing which ones end at v. Specifically, if G is d-regular, the algorithm is given oracle access to a function $f : [d]^t \to \{0, 1\}$ where $f(x)$ is 1 if the walk from u specified by the edge labels in x ends at v. We assume that G is *consistently labelled*, meaning that the edges of label i for each $i \in [d]$ form a permutation on V. We show that there exists a deterministic algorithm that makes poly(dt/ε) nonadaptive queries to f, regardless of the number of vertices in the graph G. Crucially, and in contrast to the randomized algorithm, our algorithm does not simply output the average value of its queries. Indeed, Hoza, Pyne, and Vadhan (ITCS 2021) showed that any deterministic algorithm of the latter form that works for graphs of unbounded size must have query complexity at least $\exp(\tilde{\Omega}(\log(t)\log(1/\varepsilon)))$.

Edward Pyne, Salil Vadhan
Harvard University
epyne@college.harvard.edu, salil_vadhan@harvard.edu

CP20

Maintaining An EDCS in General Graphs: Simpler, Density-Sensitive and with Worst-Case Time Bounds

In ICALP'15, Bernstein and Stein gave an algorithm for maintaining $(3/2 + \epsilon)$-approximate matching in bipartite graphs with worst-case update time $O_\epsilon(m^{1/4})$; O_ϵ hides the ϵ-dependence. Their main technical contribution was in presenting a new type of bounded-degree subgraph, an edge degree constrained subgraph (EDCS), which contains a matching of size no smaller than the maximum matching size of the graph up to a factor of $3/2 + \epsilon$. They show that an EDCS can be maintained with worst-case update time $O_\epsilon(m^{1/4})$, from which their main result follows. In SODA'16, Bernstein and Stein generalized their result for general graphs, achieving the same update time of $O_\epsilon(m^{1/4})$, albeit with an amortized bound. To date, the best deterministic worst-case update time bound for better-than-2 approximate matching is $O(\sqrt{m})$ [Gupta-Peng, FOCS'13]; allowing randomization (against an oblivious adversary) one can achieve a better (polynomial) update time for approximation slightly below 2 [Behnezhad et al., SODA'20]. In this work we simplify the approach of Bernstein and Stein for bipartite graphs, which allows us to generalize it for general graphs while maintaining the same bound of $O_\epsilon(m^{1/4})$ on the worst-case update time. Our approach is density-sensitive: If the arboricity of the dynamic graph is always at most α, the algorithm's worst-case update time is $O_\epsilon(\sqrt{\alpha})$.

Fabrizio Grandoni
IDSIA
fabrizio.grandoni@gmail.com

Chris Schwiegelshohn
Aarhus University
cschwiegelshohn@gmail.com

Shay Solomon, Amitai Uzrad
Tel Aviv University
solo.shay@gmail.com, amitaiuzrad@mail.tau.ac.il

CP21

Johnson Coverage Hypothesis: Inapproximability of k-Means and k-Median in ℓ_p-Metrics

We introduce a new hypothesis called the Johnson Coverage Hypothesis (JCH), which roughly asserts that the well-studied Max k-coverage problem on set systems is hard to approximate to a factor greater than $\left(1 - \frac{1}{e}\right)$, even when the membership graph of the set system is a subgraph of the Johnson graph. We then show that together with generalizations of the embedding techniques introduced by Cohen-Addad and Karthik (FOCS '19), JCH implies hardness of approximation results for k-median and k-means in ℓ_p-metrics for factors which are close to the ones obtained for general metrics. In particular, assuming JCH we show that it is hard to approximate the k-means objective: • Discrete case: To a factor of 3.94 in the ℓ_1-metric and to a factor of 1.73 in the ℓ_2-metric; this improves upon the previous factor of 1.56 and 1.17 respectively, obtained under the Unique Games Conjecture (UGC). • Continuous case: To a factor of 2.10 in the ℓ_1-metric and to a factor of 1.36 in the ℓ_2-metric; this improves upon the previous factor of 1.07 in the ℓ_2-metric obtained under UGC (and to the best of our knowledge, the continuous case of k-means in ℓ_1-metric was not previously analyzed in literature). We futher obtain similar improvements under JCH for the k-median objective.

Karthik C. S.
NEW YORK UNIVERSITY
NEW YORK UNIVERSITY
karthik0112358@gmail.com

Vincen Cohen-Addad
Google Research
vcohenad@gmail.com

Euiwoong Lee
University of Michigan
euiwoong@umich.edu

CP21

An Improved Local Search Algorithm for K-Median

We present a new local-search algorithm for the k-median clustering problem. We show that local optima for this algorithm give a $(2.836 + \epsilon)$-approximation; our result improves upon the $(3+\epsilon)$-approximate local-search algorithm of Arya et al. [STOC'01]. Moreover, a computer-aided analysis of a natural extension suggests that this approach may lead to an improvement over the best-known approximation guarantee for the problem. The new ingredient in our algorithm is the use of a potential function based on both the closest and second-closest facilities to each client. Specifically, the potential is the sum over all clients, of the distance of the client to its closest facility, plus (a small constant times) the truncated distance to its second-closest facility. We move from one solution to another only if the latter can be obtained by swapping a constant number of facilities, and has a smaller potential than the former. This refined potential allows us to avoid the bad local optima given by Arya et al. for the local-search algorithm based

only on the cost of the solution.

David Saulpic
LIP6, Sorbonne Université
david.saulpic@lip6.fr

Vincent Cohen-Addad
Sorbonne Universités and CNRS
vcohenad@gmail.com

Anupam Gupta
Carnegie Mellon University
anupamg@cs.cmu.edu

Lunjia Hu
Stanford University
lunjia@stanford.edu

Hoon Oh
Carnegie Mellon University
hooh@andrew.cmu.edu

CP21

Densest Subgraph: Supermodularity, Iterative Peeling, and Flow

In the densest subgraph problem (DSG), the input is a graph $G = (V, E)$, and the goal is to find a subset $S \subseteq V$ that maximizes the ratio $|E(S)|/|S|$ where $E(S)$ is the set of edges with both endpoints in S. We study fast algorithms and structural aspects of DSG via the lens of supermodularity. For DSG we describe a simple flow-based algorithm that outputs a $(1-\epsilon)$-approximation in deterministic $\tilde{O}(m/\epsilon)$ time where m is the number of edges. This is the first algorithm to have a near-linear dependence on m and $1/\epsilon$, and improves previous methods based on an LP relaxation. It generalizes to hypergraphs, and also yields a faster algorithm for directed DSG. Boob et al. [Boob, D., Gao, Y., Peng, R., Sawlani, S., Tsourakakis, C.,Wang, D., and Wang, J. "Flowless: Extracting densest subgraphs without flow computations." Proceedings of The Web Conference 2020. 2020.]developed an iterative peeling algorithm for DSG and made a conjecture about its convergence to optimality.We affirmatively answer their conjecture, and in fact prove that a natural generalization of their algorithm converges to a $(1 - \epsilon)$-approximation for any supermodular function f. We also extend some previous results on generalizations of DSG to the supermodular setting.

Chandra Chekuri
University of Illinois at Urbana-Champaign
chekuri@illinois.edu

Kent Quanrud
Purdue University
krq@purdue.edu

Manuel R. Torres
University of Illinois at Urbana-Champaign
manuelt2@illinois.edu

CP21

Massively Parallel and Dynamic Algorithms for Minimum Size Clustering

Clustering of data in metric spaces is a fundamental problem. In many scenarios where we are concerned with the privacy implications of clustering users, clusters are re-quired to have minimum-size constraint. Our work is motivated by real-world applications where a min size clustering algorithm needs to handle very large amount of data and the data may also change over time. Thus efficient parallel or dynamic algorithms are desired. We study the r-gather problem, a natural formulation of min-size clustering. Its goal is to partition n points into clusters such that each cluster has size at least r, and the maximum radius of the clusters is minimized. We propose algorithms both in the Massively Parallel Computation (MPC) model and in the dynamic setting. Our MPC algorithm handles input points from the Euclidean space $\mathbb{R}^d$. It computes an $O(1)$-approximation for r-gather in $O(\log^\varepsilon n)$ rounds using total space $O(n^{1+\gamma} \cdot d)$ for arbitrarily small constants $\varepsilon, \gamma > 0$. In addition our algorithm is fully scalable, i.e., there is no lower bound on the memory per machine. Our dynamic algorithm maintains an $O(1)$-approximate r-gather solution under insertions/deletions of points in a metric space with doubling dimension d. The update time is $r \cdot 2^{O(d)} \cdot \log^{O(1)} \Lambda$ and the query time is $2^{O(d)} \cdot \log^{O(1)} \Lambda$, where Λ is the ratio between the largest and the smallest distance.

Alessandro Epasto, Mohammad Mahdian
Google Research
aepasto@google.com, mahdian@google.com

Vahab Mirrokni
Google Inc.
mirrokni@google.com

Peilin Zhong
Google Research
peilinz@google.com

CP22

The Complexity of Testing All Properties of Planar Graphs, and the Role of Isomorphism

Consider property testing on bounded degree graphs with $\varepsilon > 0$ as the proximity parameter. Newman-Sohler (SICOMP 2013) assert that all properties of planar graphs are testable with query complexity only depending on ε. Recent advances have proven that all additive and monotone properties of planar graphs can be tested in $\text{poly}(\varepsilon^{-1})$ queries. Some properties outside this class, such as Hamiltonicity, also have a similar complexity for planar graphs. Motivated by this, we ask: can all properties of planar graphs can be tested in $\text{poly}(\varepsilon^{-1})$ queries? We discover that any property of bounded degree planar graphs can be tested in $\exp(O(\varepsilon^{-2}))$ queries. Moreover, the natural property of testing isomorphism to a fixed graph requires $\exp(\Omega(\varepsilon^{-2}))$ queries; so isomorphism to an explicit fixed graph is the hardest property of planar graphs up to polynomial dependencies. The upper bound is a straightforward, more careful adaptation of the Newman-Sohler analysis. The main technical contribution is the lower bound construction, achieved by a special family of planar graphs that are all mutually far from each other. Our techniques imply analogous results for bounded treewidth graphs. We prove that all properties of bounded treewidth graphs can be tested in $\exp(O(\varepsilon^{-1} \log \varepsilon^{-1}))$ queries, and testing isomorphism to a fixed forest requires $\exp(\Omega(\varepsilon^{-1}))$ queries.

Sabyasachi Basu, Comandur Seshadhri
University of California, Santa Cruz
sbasu3@ucsc.edu, sesh@ucsc.edu

Akash Kumar

EPFL
akash.kumar@epfl.ch

CP22

Distribution-Free Testing for Halfspaces (Almost) Requires PAC Learning

It is well known that halfspaces over $\mathbb{R}^n$ and $\{0,1\}^n$ are PAC-learnable with $\Theta(n)$ samples. Recently Blais et al. [STOC 2021] showed that even the easier task of distribution-free sample-based testing requires $\Omega(n/\log n)$ samples for halfspaces. In this work we study the distribution-free testing of halfspaces with queries, for which we show that the complexity remains to be $\tilde{\Omega}(n)$. Indeed we prove the following stronger tradeoff result: any distribution-free testing algorithm for halfspaces over $\{0,1\}^n$ that receives k samples must make $\exp(\tilde{\Omega}(\sqrt{n/k}))$ queries on the input function, when $n^{.99} \leq k \leq O(n/\log^3 n)$. For $\mathbb{R}^n$ we show that any algorithm that makes a finite number of queries must draw $\Omega(n/\log n)$ many samples.

XI Chen, Shyamal Patel
Columbia University
csxichen@gmail.com, shyamalpatelb@gmail.com

CP22

Testing Matrix Product States

Matrix product states (MPS) are a class of quantum states that arise in the study of many-body quantum physics. A quantum state $|\psi_{1,\ldots,n}\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_n}$ comprised of n qudits is said to be an MPS of bond dimension r if the reduced density matrix $\psi_{1,\ldots,k}$ has rank r for each $k \in \{1,\ldots,n\}$. Here, we study the problem of testing whether an unknown state $|\psi\rangle$ is an MPS in the property testing model. In this model, one is given m identical copies of $|\psi\rangle$, and the goal is to determine if $|\psi\rangle$ is an MPS of bond dimension r or if $|\psi\rangle$ is far from all such states. For the case of $r = 1$, we study the product test, a simple two-copy test previously analyzed by Harrow and Montanaro (2010) and a key ingredient in their proof that $\mathsf{QMA}(2) = \mathsf{QMA}(k)$ for $k \geq 2$. We give a new and simpler analysis of the product test that achieves an optimal bound for a wide range of parameters, answering open problems in Harrow and Montanaro (2010) and Montanaro and de Wolf (2016). For the case of $r \geq 2$, we give an algorithm for testing whether $|\psi\rangle$ is an MPS of bond dimension r using $m = O(nr^2)$ copies, and we show that $\Omega(n^{1/2})$ copies are necessary for this task. This lower bound shows that a dependence on the number of qudits n is necessary, in sharp contrast to the case of product states where a constant number of copies suffices.

Mehdi Soleimanifar
MIT
mehdis@mit.edu

John Wright
UT Austin
wright@cs.utexas.edu

CP22

Truly Low-Space Element Distinctness and Subset Sum via Pseudorandom Hash Functions

We consider low-space algorithms for the classic Element Distinctness problem: given an array of n input integers with $O(\log n)$ bit-length, decide whether or not all elements are pairwise distinct. Beame, Clifford, and Machmouchi [FOCS 2013] gave an $\tilde{O}(n^{1.5})$-time randomized algorithm for Element Distinctness using only $O(\log n)$ bits of working space. However, their algorithm assumes a random oracle (in particular, read-only random access to polynomially many random bits), and it was asked as an open question whether this assumption can be removed. In this paper, we positively answer this question by giving an $\tilde{O}(n^{1.5})$-time randomized algorithm using $O(\log^3 n \log\log n)$ bits of space, with one-way access to random bits. As a corollary, we also obtain a poly(n)-space $O^*(2^{0.86n})$-time randomized algorithm for the Subset Sum problem, removing the random oracles required in the algorithm of Bansal, Garg, Nederlof, and Vyas [STOC 2017]. The main technique underlying our results is a pseudorandom hash family based on iterative restrictions, which can fool the cycle-finding procedure in the algorithms of Beame et al. and Bansal et al.

Lijie Chen
MIT
lijieche@mit.edu

Ce Jin
EECS
MIT
cejin@mit.edu

Ryan Williams
Massachusetts Institute of Technology
rrw@mit.edu

Hongxun Wu
Institute for Interdisciplinary Information Sciences
Tsinghua University
wuhx18@mails.tsinghua.edu.cn

CP23

Stochastic Vertex Cover with Few Queries

We study the minimum vertex cover problem in the following stochastic setting. Let G be an arbitrary given graph, $p \in (0,1]$ a parameter of the problem, and let G_p be a random subgraph that includes each edge of G independently with probability p. We are unaware of the realization G_p, but can learn if an edge e exists in G_p by querying it. The goal is to find an approximate minimum vertex cover (MVC) of G_p by querying few edges of G. This stochastic setting has been studied extensively for various problems such as MST, matroids, matchings, etc. To our knowledge, however, no non-trivial bound was known for MVC prior to our work. In this work, we present a: $(2+\epsilon)$-approx for general graphs which queries $O(\frac{1}{\epsilon^3 p})$ edges per vertex, and a 1.367-approx for bipartite graphs which queries $(1/p)^{O(1)}$ edges per vertex. Additionally, we show that at the expense of a triple-exponential dependence on p^{-1} in the number of queries, the approximation ratio can be improved down to $(1+\epsilon)$ for bipartite graphs. Our techniques also lead to improved bounds for bipartite stochastic matching. We obtain a 0.731-approximation with nearly-linear in $1/p$ per-vertex queries. This is the first result to break the prevalent $(2/3 \sim 0.66)$-approximation barrier in the poly$(1/p)$ query regime, improving algorithms of [Behnezhad et al., SODA'19] and [Assadi and Bernstein, SOSA'19].

Soheil Behnezhad
Stanford University
soheil.behnezhad@gmail.com

Avrim Blum
TTIC
avrim@ttic.edu

Mahsa Derakhshan
Princeton University
mah.derakhshan@gmail.com

CP23

Deterministic Algorithms for the Lovász Local Lemma: Simpler, More General, and More Parallel

The Lovász Local Lemma (LLL) is a keystone principle in probability theory, guaranteeing the existence of configurations which avoid a collection $\mathcal{B}$ of "bad' events which are mostly independent and have low probability. In its simplest "symmetric' form, it asserts that whenever a bad-event has probability p and affects at most d bad-events, and $epd < 1$, then a configuration avoiding all $\mathcal{B}$ exists. A seminal algorithm of Moser & Tardos (2010) (which we call the MT algorithm) gives nearly-automatic randomized algorithms for most constructions based on the LLL. However, deterministic algorithms have lagged behind. We address three specific shortcomings of the prior deterministic algorithms. First, our algorithm applies to the LLL criterion of Shearer (1985); this is more powerful than alternate LLL criteria and also removes a number of nuisance parameters and leads to cleaner and more legible bounds. Second, we provide parallel algorithms with much greater flexibility in the functional form of the bad-events. Third, we provide a derandomized version of the MT-distribution, that is, the distribution of the variables at the termination of the MT algorithm. We show applications to non-repetitive vertex coloring, independent transversals, strong coloring, and other problems. These give deterministic algorithms which essentially match the best previous randomized sequential and parallel algorithms.

David G. Harris
University of Maryland
davidgharris29@gmail.com

CP23

Multi-Token Markov Game with Switching Costs

We study a general Markov game with metric switching costs: in each round, the player adaptively chooses one of several Markov chains to advance with the objective of minimizing the expected cost for at least k chains to reach their target states. If the player decides to play a different chain, an additional switching cost is incurred. The special case in which there is no switching cost was solved optimally by Dumitriu, Tetali, and WinklerDTW03 by a variant of the celebrated Gittins Index for the classical multi-armed bandit (MAB) problem with Markovian rewards Git74, Git79. However, for Markovian multi-armed bandit with nontrivial switching cost, even if the switching cost is a constant, the classic paper by Banks and Sundaram BS94 showed that no index strategy can be optimal. In this paper, we complement their result and show there is a simple index strategy that achieves a constant approximation factor if the switching cost is constant and $k = 1$. To the best of our knowledge, this index strategy is the first strategy that achieves a constant approximation factor for a general Markovian MAB variant with switching costs. For the general metric, we propose a more involved constant-factor approximation algorithm, via a nontrivial reduction to the stochastic k-TSP problem, in which a Markov chain is ap-

proximated by a random variable. Our analysis makes extensive use of various interesting properties of the Gittins index.

Daogao Liu
University of Washington
liudaogao@gmail.com

Jian Li
Tsinghua University
lijian83@mail.tsinghua.edu.cn

CP23

Balanced Allocations: Caching and Packing, Twinning and Thinning

We consider the sequential allocation of m balls (jobs) into n bins (servers) by allowing each ball to choose from some bins sampled uniformly at random. The goal is to maintain a small gap between the maximum load and the average load. In this paper, we present a general framework that allows us to analyze various allocation processes that slightly prefer allocating into underloaded, as opposed to overloaded bins. Our analysis covers several natural instances of processes, including: • The *Caching process* (a.k.a. memory protocol) as studied by Mitzenmacher, Prabhakar and Shah (2002). • The *Packing process*: At each round we only take one bin sample. If the load is below some threshold (e.g., the average load), then we place as many balls until the threshold is reached; otherwise, we place only one ball. • The *Twinning process*: At each round, we only take one bin sample. If the load is below some threshold, then we place two balls; otherwise, we place only one ball. • The *Thinning process* as recently studied by Feldheim and Gurel-Gurevich (2021). As we demonstrate, using an interplay between several potential functions our general framework implies for all these processes a gap of $\mathcal{O}(\log n)$ for any number of balls $m \geq n$.

Dimitrios Los, Thomas Sauerwald
University of Cambridge
dimitrios.los@cl.cam.ac.uk,
thomas.sauerwald@cl.cam.ac.uk

John Sylvester
University of Glasgow
john.sylvester@glasgow.ac.uk

CP24

Understanding Nesterov's Acceleration via Proximal Point Method

The proximal point method (PPM) is a fundamental method in optimization that is often used as a building block for designing optimization algorithms. In this work, we use the PPM method to provide conceptually simple derivations along with convergence analyses of different versions of Nesterov's accelerated gradient method (AGM). The key observation is that AGM is a simple approximation of PPM, which results in an elementary derivation of the update equations and stepsizes of AGM. This view also leads to a transparent and conceptually simple analysis of AGM's convergence by using the analysis of PPM. The derivations also naturally extend to the strongly convex case. Ultimately, the results presented in this paper are of both didactic and conceptual value; they unify and explain existing variants of AGM while motivating other

accelerated methods for practically relevant settings.

Kwangjun Ahn
MIT
kjahn@mit.edu

Suvrit Sra
Massachusetts Institute of Technology
suvrit@mit.edu

CP24

Frequency-Domain Representation of First-Order Methods: A Simple and Robust Framework of Analysis

Motivated by recent applications in min-max optimization, we employ tools from nonlinear control theory in order to analyze a class of "historical' gradient-based methods, for which the next step lies in the span of the previously observed gradients within a time horizon. Specifically, we leverage techniques developed by Hu and Lessard (2017) to build a frequency-domain framework which reduces the analysis of such methods to numerically-solvable algebraic tasks, establishing linear convergence under a class of strongly monotone and co-coercive operators. On the applications' side, we focus on the Optimistic Gradient Descent (OGD) method, which augments the standard Gradient Descent with an additional past-gradient in the optimization step. The proposed framework leads to a simple and sharp analysis of OGD—and generalizations thereof—under a broad regime of parameters. Notably, this characterization directly extends under adversarial noise in the observed value of the gradient. Moreover, our frequency-domain framework provides an exact quantitative comparison between simultaneous and alternating updates of OGD. An interesting byproduct is that OGD—and variants thereof—is an instance of PID control, arguably one of the most influential algorithms of the last century; this observation sheds more light to the stabilizing properties of "optimism'.

Ioannis Anagnostides
Carnegie Mellon University
ioannis.anagnostides@gmail.com

Ioannis Panageas
Georgia Institute of Technology
panageasj@gmail.com

CP24
Topological Art in Simple Galleries

Let P be a simple polygon, then the art gallery problem is looking for a minimum set of points (guards) that can see every point in P. We say two points $a, b \in P$ can see each other if the line segment $seg(a, b)$ is contained in P. We denote by $V(P)$ the family of all minimum guard placements. The Hausdorff distance makes $V(P)$ a metric space and thus a topological space. We show homotopy-universality, that is for every semi-algebraic set S there is a polygon P such that $V(P)$ is homotopy equivalent to S. Furthermore, for various concrete topological spaces T, we describe instances I of the art gallery problem such that $V(I)$ is homeomorphic to T.

Daniel Bertschinger, Nicolas El Maalouly
Department of Computer Science
ETH Zürich
daniel.bertschinger@inf.ethz.ch,

nicolas.elmaalouly@inf.ethz.ch

Tillmann Miltzow
Department of Information and Computing Sciences
Utrecht University
t.miltzow@googlemail.com

Patrick Schnider
Department of Mathematical Sciences
University of Copenhagen
ps@math.ku.dk

Simon Weber
Department of Computer Science
ETH Zürich
simon.weber@inf.ethz.ch

CP24

Finding Relevant Points for Nearest-Neighbor Classification

In nearest-neighbor classification problems, a set of d-dimensional training points are given, each with a known classification, and are used to infer unknown classifications of other points by using the same classification as the nearest training point. A training point is *relevant* if its omission from the training set would change the outcome of some of these inferences. We provide a simple algorithm for thinning a training set down to its subset of relevant points, using as subroutines algorithms for finding the minimum spanning tree of a set of points and for finding the extreme points (convex hull vertices) of a set of points. The time bounds for our algorithm, in any constant dimension $d \geq 3$, improve on a previous algorithm for the same problem by Clarkson (FOCS 1994).

David Eppstein
University of California, Irvine
Department of Computer Science
eppstein@uci.edu

CP24

A Simple Algorithm for Computing the Zone of a Line in An Arrangement of Lines

Let L be a set of n lines in the plane. The zone $Z(\ell)$ of a line ℓ in the arrangement $\mathcal{A}(L)$ of L is the set of faces of $\mathcal{A}(L)$ whose closure intersects ℓ. It is known that the combinatorial size of $Z(\ell)$ is $O(n)$. Given L and ℓ, computing $Z(\ell)$ is a fundamental problem. Linear-time algorithms exist for computing $Z(\ell)$ if $\mathcal{A}(L)$ has already been built, but building $\mathcal{A}(L)$ takes $O(n^2)$ time. On the other hand, $O(n \log n)$-time algorithms are also known for computing $Z(\ell)$ without relying on $\mathcal{A}(L)$, but these algorithms are relatively complicated. In this paper, we present a simple algorithm that can compute $Z(\ell)$ in $O(n \log n)$ time. More specifically, once the sorted list of the intersections between ℓ and the lines of L is known, the algorithm runs in $O(n)$ time. A big advantage of our algorithm, which mainly involves a Graham's scan style procedure, is its simplicity.

Haitao Wang
Department of Computer Science
Utah State University

haitao.wang@usu.edu

CP25
Splay Trees on Trees

Search trees on trees (STTs) are a far-reaching generalization of binary search trees (BSTs), allowing the efficient exploration of tree-structured domains. Trees on trees have been extensively studied under various guises in computer science and discrete mathematics. Recently Bose, Cardinal, Iacono, Koumoutsos, and Langerman (SODA 2020) considered adaptive STTs and observed that, apart from notable exceptions, the machinery developed for BSTs in the past decades does not readily transfer to STTs. In particular, they asked whether the optimal STT can be efficiently computed or approximated (by analogy to Knuths algorithm for optimal BSTs), and whether natural self-adjusting BSTs such as Splay trees (Sleator, Tarjan, 1983) can be extended to this more general setting. We answer both questions affirmatively. First, we show that a $(1 + 1/t)$-approximation of an optimal size-n STT for a given search distribution can be computed in time $O(n^{2t+1})$ for all integers $t \geq 1$. Second, we identify a broad family of STTs with linear rotation-distance, allowing the generalization of Splay trees to the STT setting. We show that our generalized Splay satisfies a static optimality theorem, asymptotically matching the cost of the optimal STT in an online fashion, i.e. without knowledge of the search distribution. Our results suggest an extension of the dynamic optimality conjecture for Splay trees to the broader setting of trees on trees.

Benjamin A. Berendsohn, László Kozma
Freie Universität Berlin
beab@zedat.fu-berlin.de, lkozma@zedat.fu-berlin.de

CP25
Average Sensitivity of Dynamic Programming

When processing data with uncertainty, it is desirable that the output of the algorithm is stable against small perturbations in the input. Varma and Yoshida [SODA'21] recently formalized this idea and proposed the notion of average sensitivity of algorithms, which is roughly speaking, the average Hamming distance between solutions for the original input and that obtained by deleting one element from the input, where the average is taken over the deleted element. In this work, we consider average sensitivity of algorithms for problems that can be solved by dynamic programming. We first present a $(1 - \delta)$-approximation algorithm for finding a maximum weight chain (MWC) in a transitive directed acyclic graph with average sensitivity $O(\delta^{-1} \log^3 n)$, where n is the number of vertices in the graph. We then show algorithms with small average sensitivity for various dynamic programming problems by reducing them to the MWC problem while preserving average sensitivity, including the longest increasing subsequence problem, the interval scheduling problem, the longest common subsequence problem, the longest palindromic subsequence problem, the knapsack problem with integral weight, and the RNA folding problem. For the RNA folding problem, our reduction is highly nontrivial because a naive reduction generates an exponentially large graph, which only provides a trivial average sensitivity bound.

Soh Kumabe
The University of Tokyo
sohkuma0213@gmail.com

Yuichi Yoshida
National Institute of Informatics
yyoshida@nii.ac.jp

CP25
Selectable Heaps and Optimal Lazy Search Trees

Input your abstract, including TeX commands, here. The abstract should be no longer than 1500 characters, including spaces. Only input the abstract text. Don't include title or author information here.

Bryce Sandlund, Lingyi Zhang
University of Waterloo
bcsandlund@gmail.com, l367zhang@uwaterloo.ca

CP25
Simulating a Stack using Queues

It is well known that a queue can be simulated by two stacks using a constant number of stack operations per queue operation. In this paper we consider the forgotten converse problem of simulating a stack using several queues. We consider several variants of this problem. For the *offline* variant, we obtain a tight $\Theta\big(\big(\frac{1}{k} + \frac{k}{\log n}\big)n^{1+1/k} + n\log_k n\big)$ upper and lower bounds for the worst-case number of queue operations needed to simulate a sequence of n stack operations using k queues. For the *online* variant, when the number of queues k is constant, and n is the maximum number of items in the stack at any given time, we obtain tight $\Theta(n^{1/k})$ upper and lower bounds on the worst-case and amortized number of queue operations needed to simulate one stack operation. When k is allowed to grow with n, we prove an upper bound of $O\big(n^{1/k} + \log_k n\big)$ and a lower bound of $\Omega\big(\big(\frac{1}{k} + \frac{k}{\log n}\big)n^{1/k} + \log_k n\big)$ on the amortized number of queue operations per stack operation. We also prove an upper bound of $O\big(kn^{1/k}\big)$ and a lower bound of $\Omega\big(n^{1/k} + \log_k n\big)$ on the worst-case number of queue operations per stack operation. We also analyze the specific but interesting sequence of n pushes followed by n pops.

Robert Tarjan
Princeton University
Microsoft Research
ret@cs.princeton.edu

Haim Kaplan
Tel-Aviv University
haimk@post.tau.ac.il

Or Zamir, Uri Zwick
Tel Aviv University
orzamir90@gmail.com, zwick@tau.ac.il

CP26
Deleting, Eliminating and Decomposing to Hereditary Classes are all FPT-Equivalent

Vertex-deletion problems have been at the heart of parameterized complexity throughout its history. Here, the aim is to determine the minimum size (denoted by mod_H) of a modulator to a graph class H, i.e., a set of vertices whose deletion results in a graph in H. Recent years have seen the development of a research programme where the complexity of modulators is measured in ways other than size. For instance, for a graph class H, the graph parameters elimination distance to H (denoted by ed_H and H-treewidth

(denoted by tw_H) aim to minimize the treedepth and treewidth, respectively, of the torso of the graph induced on a modulator to the graph class H. Here, the torso of a vertex set S in a graph G is the graph with vertex set S and an edge between two vertices $u, v \in S$ if there is a path between u and v in G whose internal vertices all lie outside S. In this paper, we show that from the perspective of (non-uniform) fixed-parameter tractability (FPT), the three parameters described above give equally powerful parameterizations for every hereditary graph class H that satisfies mild additional conditions. In fact, we show that for every hereditary graph class H satisfying mild additional conditions, with the exception of ed_H parameterized by tw_H, for every pair of these parameters, computing one parameterized by itself or any of the others is FPT-equivalent to the standard vertex-deletion (to H) problem.

Ramanujan M. Sridharan
TU Wien
R.Maadapuzhi-Sridharan@warwick.ac.uk

Akanksha Agrawal
Indian Institute of Technology Madras
akanksha.agrawal.2029@gmail.com

Lawqueen Kanesh
National University of Singapore
lawqueen@comp.nus.edu.sg

Daniel Lokshtanov
UCSB
daniello@ucsb.edu

Fahad Panolan
IIT Hyderabad, India
fahad@cse.iith.ac.in

Meirav Zehavi
BGU
meiravze@bgu.ac.il

Saket Saurabh
IMSc +UiB
saket@imsc.res.in

CP26

Subexponential Parameterized Algorithms for Cut and Cycle Hitting Problems on H-Minor-Free Graphs

We design the first subexponential-time parameterized algorithms for several cut and cycle-hitting problems on H-minor free graphs. In particular, we obtain algorithms for Edge Bipartization and Odd Cycle Transversal, Edge/Vertex Multiway Cut , and Edge/Vertex Multicut, and Group Feedback Edge/Vertex Set with running time exponential in $\sqrt{k}\log^{O(1)} k$, where k is the solution-size parameter. We obtain our results by giving a new decomposition theorem on graphs of bounded genus, or more generally, an h-almost-embeddable graph for an arbitrary but fixed constant h. Our new decomposition theorem generalizes known Contraction Decomposition Theorem. In particular we show the following. Let G be a graph of bounded genus, or more generally, an h-almost-embeddable graph for an arbitrary but fixed constant h. Then for every $p \in \mathbb{N}$, there exist disjoint sets $Z_1, \ldots, Z_p \subseteq V(G)$ such that for every $i \in \{1, \ldots, p\}$ and every $Z' \subseteq Z_i$, the treewidth of $G/(Z_i \backslash Z')$ is upper bounded by $O(p + |Z'|)$, where the con-

stant hidden in $O(\cdot)$ depends on h. Here $G/(Z_i \backslash Z')$ denotes the graph obtained from G by contracting every edge with both endpoints in $Z_i \backslash Z'$. When $Z' = \emptyset$, this corresponds to classical Contraction Decomposition Theorem.

Sayan Bandyapadhyay, William Lochet
University of Bergen
sayan.bandyapadhyay@gmail.com,
william.lochet@gmail.com

Daniel Lokshtanov
UCSB
daniello@ucsb.edu

Saket Saurabh
IMSc +UiB
saket@imsc.res.in

Jie Xue
New York University Shanghai
jiexue@nyu.edu

CP26

A Framework for Parameterized Subexponential Algorithms for Generalized Cycle Hitting Problems on Planar Graphs

Subexponential parameterized algorithms are known for a wide range of natural problems on planar graphs, but the techniques are usually highly problem specific. The goal of this paper is to introduce a framework for obtaining $n^{O(\sqrt{k})}$ time algorithms for a family of graph modification problems that includes problems that can be seen as generalized cycle hitting problems. Our starting point is the Node Unique Label Cover problem (that is, given a CSP instance where each constraint is a permutation of values on two variables, the task is to delete k variables to make the instance satisfiable). We introduce a variant of the problem where k vertices have to be deleted such that every 2-connected component of the remaining instance is satisfiable. Then we extend the problem with cardinality constraints that restrict the number of times a certain value can be used (globally or within a 2-connected component of the solution). We show that there is an $n^{O(\sqrt{k})}$ time algorithm on planar graphs for any problem that can be formulated this way, which includes a large number of well-studied problems, for example, Odd Cycle Transversal, Subset Feedback Vertex Set, Group Feedback Vertex Set, Vertex Multiway Cut, and Component Order Connectivity. We use or adapt known kernelization results to obtain $2^{O(\sqrt{k} \cdot \text{polylog}(k))} n^{O(1)}$ time (randomized) algorithms for Vertex Multiway Cut, Group Feedback Vertex Set, and Subset Feedback Vertex Set.

Dániel Marx
CISPA Helmholtz Center for Information Security
marx@cispa.de

Pranabendu Misra
Chennai Mathematical Institute
pranabendu@cmi.ac.in

Daniel Neuen, Prafullkumar Tale
CISPA Helmholtz Center for Information Security

daniel.neuen@cispa.de, prafullkumar.tale@cispa.de

CP26

Planar Multiway Cut with Terminals on Few Faces

We consider the Edge Multiway Cut problem on planar graphs. It is known that this can be solved in $n^{O(\sqrt{t})}$ time [Klein, Marx, ICALP 2012] and not in $n^{o(\sqrt{t})}$ time under the Exponential Time Hypothesis [Marx, ICALP 2012], where t is the number of terminals. A generalization of this parameter is the number k of faces of the planar graph that jointly cover all terminals. For the related Steiner Tree problem, an $n^{O(\sqrt{k})}$ time algorithm was recently shown [Kisfaludi-Bak et. al., SODA 2019]. By a completely different approach, we prove in this paper that Edge Multiway Cut can be solved in $n^{O(\sqrt{k})}$ time as well. Our approach employs several major concepts on planar graphs, including homotopy and sphere-cut decomposition. We also mix a global treewidth dynamic program with a Dreyfus-Wagner style dynamic program to deal with large numbers of terminals locally.

Sukanya Pandey, Erik Jan Van Leeuwen
Utrecht University
s.pandey1@uu.nl, e.jvanleeuwen@uu.nl

CP26

Subexponential Parameterized Algorithms on Disk Graphs

One of the most celebrated results in Parameterized Complexity is the Bidimensionality theory of Demaine et al., which has yielded, over the past two decades, numerous subexponential-time fixed-parameter tractable (FPT) algorithms for various problems on planar (and H-minor-free) graphs. At the heart of this theory is the proof of sublinear bounds in terms of solution size on the treewidth of a given graph. Inspired by this theory, in recent years, significant efforts have been devoted to design subexponential-time FPT algorithms for problems on geometric graph classes that utilize new treewidth bounds, in particular (but not only) for unit disk graphs. In this paper, we aim to attain such results on disk graphs, a broad class of graphs that generalizes both the classes of planar graphs and unit disk graphs, and thereby unify the aforementioned research frontiers for planar and unit disk graphs. Our main contribution is an approach to design subexponential-time FPT algorithms for problems on disk graphs, which we apply to several well-studied graph problems. At the heart of our approach lie two new combinatorial theorems concerning the treewidth of disk graphs having a realization of bounded ply (or maximum clique size) that are of independent interest. Among our applications are the first subexponential-time FPT algorithms for several problems on disk graphs, including Triangle Hitting, Feedback Vertex Set and Odd Cycle Transversal.

Daniel Lokshtanov
UCSB
daniello@ucsb.edu

Fahad Panolan
IIT Hyderabad, India
fahad@cse.iith.ac.in

Saket Saurabh
IMSc +UiB
saket@imsc.res.in

Jie Xue
New York University Shanghai
jiexue@nyu.edu

Meirav Zehavi
BGU
meiravze@bgu.ac.il

CP27

Deterministic Enumeration of All Minimum K-Cut-Sets in Hypergraphs for Fixed K

We consider the problem of deterministically enumerating all minimum k-cut-sets in a given hypergraph for any fixed k. The input is a hypergraph G = (V, E) with nonnegative hyperedge costs. A subset F of hyperedges is a k-cut-set if the number of connected components in G - F is at least k and it is a minimum k-cut-set if it has the least cost among all k-cut-sets. For fixed k, we call the problem of finding a minimum k-cut-set Hypergraph-k-Cut and the problem of enumerating all minimum k-cut-sets Enum-Hypergraph-k-Cut. While the graph versions of these problems are well-known to be solvable in polynomial time, it is only recently that polynomial-time algorithms for Hypergraph-k-Cut were developed. The randomized polynomial-time algorithm for Hypergraph-k-Cut that was designed in 2018 (Chandrasekaran, Xu, and Yu, SODA 2018) showed that the number of minimum k-cut-sets in a hypergraph is $O(n^{2k-2})$, where n is the number of vertices in the input hypergraph, and that they can all be enumerated in randomized polynomial time. A deterministic polynomial-time algorithm for Hypergraph-k-Cut was subsequently designed in 2020 (Chandrasekaran and Chekuri, FOCS 2020), but it is not guaranteed to enumerate all minimum k-cut-sets. In this work, we give the first deterministic polynomial-time algorithm to solve Enum-Hypergraph-k-Cut. Our algorithms are based on new structural results that allow for efficient recovery of all minimum k-cut-sets by solving minimum (S,T)-terminal cuts.

Calvin Beideman
University of Illinois at Urbana-Champaign
calvinb2@illinois.edu

Karthekeyan Chandrasekaran
Department of Industrial and Enterprise Systems
Engineering
University of Illinois, Urbana-Champaign
karthe@illinois.edu

Weihang Wang
University of Illinois at Urbana-Champaign
weihang3@illinois.edu

CP27

Sampling Colorings and Independent Sets of Random Regular Bipartite Graphs in the Non-Uniqueness Region

We give an FPRAS for counting q-colorings for even $q = O\left(\frac{\Delta}{\log \Delta}\right)$ on almost every Δ-regular bipartite graph. This improves significantly upon the previous best bound of $q = O\left(\frac{\sqrt{\Delta}}{(\log \Delta)^2}\right)$ by Jenssen, Keevash, and Perkins (SODA'19). Analogously, for the hard-core model on independent sets weighted by $\lambda > 0$, we present an FPRAS for estimating the partition function when $\lambda = \Omega\left(\frac{\log \Delta}{\Delta}\right)$, which improves upon previous results by an $\Omega(\log \Delta)$ factor. Our results for the colorings and hard-core models follow from a general re-

sult that applies to arbitrary spin systems. Our main contribution is to show how to elevate probabilistic/analytic bounds on the marginal probabilities for the typical structure of phases on random bipartite regular graphs into efficient algorithms, using the polymer method. We further show evidence that our results for colorings and independent sets are within a constant factor of best possible using current polymer-method approaches.

Zongchen Chen
Georgia Institute of Technology
zongchen@mit.edu

Andreas Galanis
University of Oxford
andreas.galanis@cs.ox.ac.uk

Daniel Stefankovic
Department of Computer Science
University of Rochester
stefanko@cs.rochester.edu

Eric Vigoda
University of California Santa Barbara
vigoda@ucsb.edu

CP27
Enumerating k-SAT Functions

How many k-SAT functions on n boolean variables are there? What does a typical such function look like? Bollobs, Brightwell, and Leader conjectured that, for each fixed $k \geq 2$, the number of k-SAT functions on n variables is $(1+o(1))2^{\binom{n}{k}+n}$, or equivalently: a $1-o(1)$ fraction of all k-SAT functions are unate, i.e., monotone after negating some variables. They proved a weaker version of the conjecture for $k = 2$. The conjecture was confirmed for $k = 2$ by Allen and $k = 3$ by Ilinca and Kahn. We show that the problem of enumerating k-SAT functions is equivalent to a Turn density problem for partially directed hypergraphs. Our proof uses the hypergraph container method. Furthermore, we confirm the Bollobs-Brightwell-Leader conjecture for $k = 4$ by solving the corresponding Turn density problem.

Dingding Dong
Harvard
ddong@math.harvard.edu

Nitya Mani
MIT
nmani@mit.edu

Yufei Zhao
Massachusetts Institute of Technology
yufeiz@mit.edu

CP27
Efficient Generation of Elimination Trees and Hamilton Paths on Graph Associahedra

An elimination tree for a connected graph G is a rooted tree on the vertices of G obtained by choosing a root x and recursing on the connected components of $G - x$ to produce the subtrees of x. Elimination trees appear in many guises in computer science and discrete mathematics, and they encode many interesting combinatorial objects, such as bitstrings, permutations and binary trees. We apply the recent Hartung-Hoang-Mtze-Williams combinatorial generation framework (SODA 2020) to elimination trees, and prove that all elimination trees for a chordal graph G can be generated by tree rotations using a simple greedy algorithm. This yields a short proof for the existence of Hamilton paths on graph associahedra of chordal graphs. Graph associahedra are a general class of high-dimensional polytopes introduced by Carr, Devadoss, and Postnikov, whose vertices correspond to elimination trees and whose edges correspond to tree rotations. As special cases of our results, we recover several classical Gray codes for bitstrings, permutations and binary trees, and we obtain a new Gray code for partial permutations. Our algorithm for generating all elimination trees for a chordal graph G can be implemented in time $\mathcal{O}(m + n)$ per generated elimination tree, where m and n are the number of edges and vertices of G, respectively. Furthermore, if G is a tree, we improve this to a loopless algorithm running in time $\mathcal{O}(1)$ per generated elimination tree.

Arturo Merino
TU Berlin
merino@math.tu-berlin.de

Jean Cardinal
ULB, Brussels
jcardin@ulb.ac.be

Torsten Mütze
Department of Computer Science, University of Warwick
torsten.mutze@warwick.ac.uk

CP27
A Lower Bound for the n-Queens Problem

The n-queens puzzle is to place n mutually non-attacking queens on an $n \times n$ chessboard. We present a simple two stage randomized algorithm to construct such configurations. In the first stage, a random greedy algorithm constructs an approximate *toroidal* n-queens configuration. In this well-known variant the diagonals wrap around the board from left to right and from top to bottom. We show that with high probability this algorithm succeeds in placing $(1 - o(1))n$ queens on the board. In the second stage, the method of absorbers is used to obtain a complete solution to the non-toroidal problem. By counting the number of choices available at each step of the random greedy algorithm we conclude that there are more than $\left((1 - o(1)) \, ne^{-3}\right)^n$ solutions to the n-queens problem. This proves a conjecture of Rivin, Vardi, and Zimmerman in a strong form. Recently, using different methods, Bowtell and Keevash proved the same lower bound for the toroidal problem, giving an independent proof of the result.

Zur Luria
Azrieli College of Engineering
zluria@gmail.com

Michael Simkin
Harvard University
msimkin@cmsa.fas.harvard.edu

CP28
Adversarially Robust Streaming Via DenseSparse Trade-Offs

Input your abstract, including TeX commands, here. The abstract should be no longer than 1500 characters, including spaces. Only input the abstract text. Don't include

title or author information here.

Krzysztof Onak
Boston University
krzysztof@onak.pl

Omri Ben-Eliezer
Tel Aviv University
omrib@mail.tau.ac.il

Talya Eden
MIT and Boston University
talyaa01@gmail.com

CP28
Nearly Tight Bounds for Discrete Search under Outlier Noise

Binary search is one of the most fundamental search routines, exploiting the hidden structure of the search space. In particular, it cuts down exponentially on the complexity of the search assuming that the search space is monotone. This paper is prompted by a basic question – how does the query complexity of the search problem change if the data has corruption? In particular, we study the powerful *outlier noise model* and assuming a bound on the fraction of such corruptions, establish nearly matching upper and lower bounds for the following problems: (i) binary search on an ordered set of size $[n]$; (ii) search on the posets $\{0,1\}^d$; and (iii) search on the posets $[n]^d$. In all three cases, we use randomization to create robust versions of classical algorithms for these problems that handle corrupted data with relatively small performance penalties, specified as a function of the amount of corruption K. We complement these algorithmic results with almost matching lower bounds that show that no randomized algorithm can solve these problems with a smaller performance hit on the query complexity as a function of K.

Huan Li, Anindya De, Sanjeev Khanna, Hesam Nikpey
University of Pennsylvania
huanli@seas.upenn.edu, anindyad@cis.upenn.edu, sanjeev@cis.upenn.edu, hesam@cis.upenn.edu

CP28
Uniformity Testing in the Shuffle Model: Simpler, Better, Faster

Uniformity testing, or testing whether independent observations are uniformly distributed, is the prototypical question in distribution testing. Over the past years, a line of work has been focusing on uniformity testing under privacy constraints on the data, and obtained private and data-efficient algorithms under various privacy models such as central differential privacy (DP), local privacy (LDP), pan-privacy and, very recently, the shuffle model of differential privacy. In this work, we considerably simplify the analysis of the known uniformity testing algorithm in the shuffle model, and, using a recent result on privacy amplification via shuffling, provide an alternative algorithm attaining the same guarantees with an elementary and streamlined argument.

Hongyi Lyu
The University of Melbourne
hongyi.lyu@student.unimelb.edu.au

Clément Canonne
The University of Sydney
clement.canonne@sydney.edu.au

CP28
Tight Bounds for Differentially Private Anonymized Histograms

In this note, we consider the problem of differentially privately (DP) computing an anonymized histogram, which is defined as the multiset of counts of the input dataset (without bucket labels). In the low-privacy regime $\epsilon \geq 1$, we give an ϵ-DP algorithm with an expected ℓ_1-error bound of $O(\sqrt{n}/e^\epsilon)$. In the high-privacy regime $\epsilon < 1$, we give an $\Omega(\sqrt{n \log(1/\epsilon)/\epsilon})$ lower bound on the expected ℓ_1 error. In both cases, our bounds asymptotically match the previously known lower/upper bounds due to [Suresh, NeurIPS 2019].

Pasin Manurangsi
Google Research, U.S.
pasin@google.com

CP28
Parameterized Convexity Testing

In this work, we develop new insights into the fundamental problem of convexity testing of real-valued functions over the domain $[n]$. Specifically, we present a nonadaptive algorithm that, given inputs $\varepsilon \in (0,1), s \in \mathbb{N}$, and oracle access to a function, ε-tests convexity in $O(\log(s)/\varepsilon)$, where s is an upper bound on the number of distinct discrete derivatives of the function. We also show that this bound is tight. Since $s \leq n$, our query complexity bound is at least as good as that of the optimal convexity tester (Ben Eliezer; ITCS 2019) with complexity $O(\frac{\log \varepsilon n}{\varepsilon})$; our bound is strictly better when $s = o(n)$. The main contribution of our work is to appropriately parameterize the complexity of convexity testing to circumvent the worst-case lower bound (Belovs et al.; SODA 2020) of $\Omega(\frac{\log(\varepsilon n)}{\varepsilon})$ expressed in terms of the input size and obtain a more efficient algorithm.

Nithin Varma
Chennai Mathematical Institute
nithvarma@gmail.com

Abhiruk Lahiri
Ariel University, Israel
abhiruk@ariel.ac.il

Ilan I. Newman
Haifa University
ilan@cs.haifa.ac.il

CP29
Polynomial Time Algorithms to Find an Approximate Competitive Equilibrium for Chores

Competitive equilibrium with equal income (CEEI) is considered one of the best mechanisms to allocate a set of items among agents fairly and efficiently. In this paper, we study the computation of CEEI when items are chores that are disliked (negatively valued) by agents, under 1-homogeneous and concave utility functions which includes linear functions as a subcase. It is well-known that, even with linear utilities, the set of CEEI may be non-convex and disconnected, and the problem is PPAD-hard in the more general exchange model. In contrast to these negative results, we design FPTAS: A polynomial-time algo-

rithm to compute ?-approximate CEEI where the running-time depends polynomially on 1/?. Our algorithm relies on the recent characterization due to Bogomolnaia et al. (2017) of the CEEI set as exactly the KKT points of a non-convex minimization problem that have all coordinates non-zero. Due to this non-zero constraint, naive gradient-based methods fail to find the desired local minima as they are attracted towards zero. We develop an exterior-point method that alternates between guessing non-zero KKT points and maximizing the objective along supporting hyperplanes at these points. We show that this procedure must converge quickly to an approximate KKT point which then can be mapped to an approximate CEEI; this exterior point method may be of independent interest.

Shant Boodaghians
University of Illinois at Urbana-Champaign
boodagh2@illinois.edu

Bhaskar R. Chaudhury
Max Planck Institute for Informatics
braycha@illinois.edu

Ruta Mehta
University of Illinois at Urbana-Champaign
rutameht@ilinois.edu

CP29
Computational Hardness of the Hylland-Zeckhauser Scheme

We study the complexity of the classic Hylland-Zeckhauser scheme [HZ'79] for one-sided matching markets. We show that the problem of finding an ?-approximate equilibrium in the HZ scheme is PPAD-hard, and this holds even when ? is polynomially small and when each agent has no more than four distinct utility values. Our hardness result, when combined with the PPAD membership result of [VY'21], resolves the approximation complexity of the HZ scheme. We also show that the problem of approximating the optimal social welfare (the weight of the matching) achievable by HZ equilibria within a certain constant factor is NP-hard.

Binghui Peng
Tsinghua University
bp2601@columbia.edu

Xi Chen, Thomas Chen, Mihalis Yannakakis
Columbia University
xc2198@columbia.edu,   tc3039@columbia.edu,   mihalis@cs.columbia.edu

CP29
Approximating Equilibrium under Constrained Piecewise Linear Concave Utilities with Applications to Matching Markets

We study the equilibrium computation problem in the Fisher market model with constrained piecewise linear concave (PLC) utilities. This general class captures many well-studied special cases, including markets with PLC utilities, markets with satiation, and matching markets. For the special case of PLC utilities, although the problem is PPAD-hard, Devanur and Kannan (FOCS 2008) gave a polynomial-time algorithm when the number of items is constant. Our main result is a fixed parameter approximation scheme for computing an approximate equilibrium, where the parameters are the number of agents and the ap-

proximation accuracy. This provides an answer to an open question by Devanur and Kannan for PLC utilities, and gives a simpler and faster algorithm for matching markets as the one by Alaei, Jalaly, and Tardos (EC 2017). The main technical idea is to work with the stronger concept of thrifty equilibria, and approximating the input utility functions by 'robust' utilities that have favorable marginal properties. With some restrictions, the results also extend to the Arrow–Debreu exchange market model.

Yixin Tao
London School of Economics
Y.Tao16@lse.ac.uk

Jugal Garg
University of Illinois at Urbana-Champaign
jugal@illinois.edu

László Végh
London School of Economics
l.vegh@lse.ac.uk

CP29
Approximate Core for Committee Selection via Multilinear Extension and Market Clearing

Motivated by civic problems such as participatory budgeting and multiwinner elections, we consider the problem of public good allocation: Given a set of indivisible projects (or candidates) of different sizes, and voters with different monotone utility functions over subsets of these candidates, the goal is to choose a budget-constrained subset of these candidates (or a committee) that provides fair utility to the voters. The notion of fairness we adopt is core stability from cooperative game theory: No subset of voters should be able to choose another blocking committee of proportionally smaller size that provides strictly larger utility to all voters that deviate. It is well-known that an exact core need not exist even when utility functions of the voters are additive across candidates. We therefore relax the problem to allow approximation. Our main result is that an α-core, for $\alpha < 67.37$, always exists when utilities of the voters are arbitrary monotone submodular functions, and this can be computed in polynomial time. This result improves to $\alpha < 9.27$ for additive utilities, albeit without the polynomial time guarantee. Our results are a significant improvement over prior work that only shows logarithmic approximations for the case of additive utilities. We complement our results with a lower bound of $\alpha > 1.015$ for submodular utilities, and a lower bound of any function in the number of voters and candidates for general monotone utilities.

Kamesh Munagala
Duke University, USA
kamesh@cs.duke.edu

Yiheng Shen, Kangning Wang, Zhiyi Wang
Duke University
yiheng.shen@duke.edu,   knwang@cs.duke.edu,   zhiyi.wang@duke.edu

CP30
Approximating The Arboricity in Sublinear Time

We consider the problem of approximating the arboricity of a graph $G = (V, E)$, which we denote by $\mathsf{arb}(G)$, in sublinear time, where the arboricity of a graph is the minimal number of forests required to cover its edges. An algo-

rithm for this problem may perform degree and neighbor queries, and is allowed a small error probability. We design an algorithm that outputs an estimate $\hat{\alpha}$, such that with probability $1 - 1/\text{poly}(n)$, $\text{arb}(G)/c \log^2 n \leq \hat{\alpha} \leq \text{arb}(G)$, where $n = |V|$ and c is a constant. The expected query complexity and running time of the algorithm are $O(n/\text{arb}(G)) \cdot \text{poly}(\log n)$, and this upper bound also holds with high probability. This bound is optimal for such an approximation up to a $\text{poly}(\log n)$ factor.

Talya Eden
MIT and Boston University
talyaa01@gmail.com

Dana Ron
Tel Aviv University
danar@eng.tau.ac.il

Saleet Mossel
MIT
saleet@mit.edu

CP30
Approximating Sumset Size

Given a subset A of the n-dimensional Boolean hypercube $\mathbb{F}_2^n$, the sumset $A+A$ is the set $\{a+a' : a, a' \in A\}$ where addition is in $\mathbb{F}_2^n$. Sumsets play an important role in additive combinatorics, where they feature in many central results of the field. The main result of this paper is a sublinear-time algorithm for the problem of sumset size estimation. In more detail, our algorithm is given oracle access to (the indicator function of) an arbitrary $A \subseteq \mathbb{F}_2^n$ and an accuracy parameter $\varepsilon > 0$, and with high probability it outputs a value $0 \leq v \leq 1$ that is $\pm\varepsilon$-close to $\text{Vol}(A' + A')$ for some perturbation $A' \subseteq A$ of A satisfying $\text{Vol}(A \setminus A') \leq \varepsilon$. It is easy to see that without the relaxation of dealing with A' rather than A, any algorithm for estimating $\text{Vol}(A + A)$ to any nontrivial accuracy must make $2^{\Omega(n)}$ queries. In contrast, we give an algorithm whose query complexity depends only on ε and is completely independent of the ambient dimension n.

Shivam Nadimpalli
Columbia University
sn2855@columbia.edu

Anindya De
University of Pennsylvania
de.anindya@gmail.com

Rocco Servedio
Columbia University
ras2105@columbia.edu

CP30
Better Sum Estimation via Weighted Sampling

Given a large set U where each item $a \in U$ has weight $w(a)$, we want to estimate the total weight $W = \sum_{a \in U} w(a)$ to within factor of $1 \pm \varepsilon$ with some constant probability $> 1/2$. Since $n = |U|$ is large, we want to do this without looking at the entire set U. In the traditional setting in which we are allowed to sample elements from U uniformly, sampling $\Omega(n)$ items is necessary to provide any non-trivial guarantee on the estimate. Therefore, we investigate this problem in different settings: in the proportional setting we can sample items with probabilities proportional to their weights, and in the hybrid setting we can sample both proportionally and uniformly. Sum estimation in the proportional and hybrid setting has been considered before by Motwani, Panigrahy, and Xu [ICALP, 2007]. In their paper, they give both upper and lower bounds in terms of n. Their bounds are near-matching in terms of n, but not in terms of ε. In this paper, we improve both their upper and lower bounds. Our bounds are matching up to constant factors in both settings, in terms of both n and ε. No lower bounds with dependency on ε were known previously. In the proportional setting, we improve their $\tilde{O}(\sqrt{n}/\varepsilon^{7/2})$ algorithm to $O(\sqrt{n}/\varepsilon)$. In the hybrid setting, we improve $\tilde{O}(\sqrt[3]{n}/\varepsilon^{9/2})$ to $O(\sqrt[3]{n}/\varepsilon^{4/3})$.

Lorenzo Beretta, <u>Jakub Tetek</u>
Basic Algorithms Research Copenhagen
University of Copenhagen
beretta@di.ku.dk, j.tetek@gmail.com

CP30
Improved Algorithms for Low Rank Approximation from Sparsity

We overcome two major bottlenecks in the study of low rank approximation by assuming the low rank factors themselves are sparse. Specifically, (1) For low rank approximation with spectral norm error, we show how to improve the best known $\text{nnz}(A)k/\sqrt{\varepsilon}$ running time to $\text{nnz}(A)/\sqrt{\varepsilon}$ running time plus low order terms depending on the sparsity of the low rank factors, and (2) For streaming algorithms for Frobenius norm error, we show how to bypass the known $\Omega(nk/\varepsilon)$ memory lower bound and obtain an $sk(\log n)/\text{poly}(\varepsilon)$ memory bound, where s is the number of non-zeros of each low rank factor. Although this algorithm runs in exponential time, as it must under standard complexity theoretic assumptions, we also present polynomial time algorithms using $\text{poly}(s, k, \log n, \varepsilon^{-1})$ memory that output rank k approximations supported on an $O(sk/\varepsilon) \times O(sk/\varepsilon)$ submatrix. Both the prior $\text{nnz}(A)k/\sqrt{\varepsilon}$ running time and the nk/ε memory for these problems were long-standing barriers; our results give a natural way of overcoming them assuming sparsity of the low rank factors.

David Woodruff, <u>Taisuke Yasuda</u>
Carnegie Mellon University
dwoodruf@cs.cmu.edu, taisukey@cs.cmu.edu

CP31
Distributed Zero-Knowledge Proofs Over Networks

Zero-knowledge proofs are one of the most influential concepts in theoretical computer science. In the seminal definition due to Goldwasser, Micali and Rackoff, a computationally-bounded verifier interacts with a powerful but untrusted prover, with the goal of becoming convinced that the input is in some language. In addition to the usual requirements of completeness and soundness, we protect the prover's knowledge: anything that the verifier can deduce after interacting with the prover, it could have deduced by itself. We define and study the notion of *distributed zero knowledge proofs*. In our setting, a *network* of verifiers interacts with an untrusted prover to decide some *distributed* language. The verifiers have local views of the network, and each only knows its neighbors. The prover, on the other hand, is assumed to know the entire network graph, as well as all inputs. As in the computational centralized setting, the protocol we design should protect

this knowledge. We construct communication-efficient distributed zero-knowledge proofs for two central problems: the 3-coloring problem, one of the poster children of computational zero-knowledge, and for the spanning-tree verification problem. We also give a general scheme for converting proof labeling-schemes to distributed zero-knowledge protocols. Our protocols combine ideas from computational complexity, distributed computing, and cryptography.

Rotem Oshman, <u>Aviv Bick</u>
Tel Aviv University
roshman@tau.ac.il, avivbick@mail.tau.ac.il

Gillat Kol
IAS
gillat.kol@gmail.com

CP31

Optimal Oblivious Parallel Ram

An oblivious RAM (ORAM), introduced by Goldreich and Ostrovsky (STOC '87 and J. ACM '96), is a technique for hiding RAM's access pattern. That is, for every input the distribution of the observed locations accessed by the machine is essentially independent of the machine's secret inputs. Recent progress culminated in a work of Asharov et al. (EUROCRYPT '20), obtaining an ORAM with (amortized) logarithmic overhead in total work, which is known to be optimal. Oblivious *Parallel* RAM (OPRAM) is a natural extension of ORAM to the (more realistic) parallel setting where several processors make concurrent accesses to a shared memory. It is known that any OPRAM must incur logarithmic work overhead (in the balls and bins model). Despite the significant recent advances for constructing ORAM, there is still a significant gap for OPRAM: all existing OPRAM schemes incur a *poly*-logarithmic overhead either in total work or in depth. Our main result closes the aforementioned gap and provides an *optimal* OPRAM. Specifically, assuming one-way functions, we show that any Parallel RAM with memory capacity N can be obliviously simulated in space $O(N)$, incurring only $O(\log N)$ blowup in (amortized) total work as well as in depth. Our transformation supports all PRAMs in the CRCW (concurrent read, concurrent write) mode and the resulting simulation is in the CRCW mode as well.

Gilad Asharov
Bar-Ilan University
gilad.asharov@biu.ac.il

Ilan Komargodski
Hebrew University of Jerusalem
NTT Research
ilank@cs.huji.ac.il

<u>Wei-Kai Lin</u>
Cornell University
wklin@cs.cornell.edu

Enoch Peserico
Università degli Studi di Padova
enoch@dei.unipd.it

Elaine Shi
Carnegie Mellon University

runting@cs.cmu.edu

CP31

Limits of Preprocessing for Single-Server PIR

We present lower bounds for the static cryptographic data structure problem of private information retrieval (PIR). PIR considers the setting where a server holds a database of n entries and a client wishes to privately retrieve the i-th entry without revealing the index i to the server. In our work, we focus on PIR with preprocessing where an r-bit hint may be computed in a preprocessing stage and stored by the server to be used to perform private queries in expected t time. As our main result, we prove that for any single-server, computationally secure PIR with preprocessing, it must be that $tr = \Omega(n \log n)$ when $r = \Omega(\log n)$. If $r = O(\log n)$, then we show that $t = \Omega(n)$. Our lower bound holds even when the scheme errs with probability $1/n^2$ and the adversarys distinguishing advantage is $1/n$. Our work improves upon the $tr = \Omega(n)$ lower bound of Beimel, Ishai and Malkin [JoC04]. For information-theoretic security, we present a stronger lower bound of $t + r = \Omega(n)$ and show a matching construction. Both our lower bounds apply for public-key doubly-efficient PIRs of Boyle, Ishai, Pass and Wootters [TCC17]. Additionally, our lower bound for information-theoretic security also applies for offline-online PIRs as defined by Corrigan-Gibbs and Kogan [Eurocrypt20] where the hint is private and only viewed by the client.

<u>Kevin Yeo</u>
Google
kwlyeo@google.com

Giuseppe Persiano
University of Salerno
giuper@gmail.com

CP31

Universally-Optimal Distributed Shortest Paths and Transshipment via Graph-Based L1-Oblivious Routing

We provide universally-optimal distributed graph algorithms for $(1 + \varepsilon)$-approximate shortest path problems including shortest-path-tree and transshipment. The universal optimality of our algorithms guarantees that, on any n-node network G, our algorithm completes in $T \cdot n^{o(1)}$ rounds whenever a T-round algorithm exists for G. This includes $D \cdot n^{o(1)}$-round algorithms for any planar or excluded-minor network. Our algorithms never require more than $(\sqrt{n} + D) \cdot n^{o(1)}$ rounds, resulting in the first sub-linear-round distributed algorithm for transshipment. The key technical contribution leading to these results is the first efficient $n^{o(1)}$-competitive linear ℓ_1-oblivious routing operator that does not require the use of ℓ_1-embeddings. Our construction is simple, solely based on low-diameter decompositions, and—in contrast to all known constructions—directly produces an oblivious flow instead of just an approximation of the optimal flow cost. This also has the benefit of simplifying the interaction with Sherman's multiplicative weight framework [SODA'17] in the distributed setting and its subsequent rounding procedures.

<u>Goran Zuzic</u>
ETH Zurich
zuza777@gmail.com

Gramoz Goranci
University of Glasgow
gramoz.goranci@glasgow.ac.uk

Mingquan Ye
University of Illinois at Chicago
mye9@uic.edu

Bernhard Haeupler
Carnegie Mellon U.
haeupler@cs.cmu.edu

Xiaorui Sun
University of Illinois at Chicago
xiaorui@uic.edu

CP32

A Simple Approximation Algorithm for Vector Scheduling and Applications to Stochastic Min-Norm Load Balancing

We consider the Vector Scheduling problem on identical machines: we have m machines, and a set J of n jobs, where each job j has a processing-time vector $p_j \in \mathbb{R}^d_{\geq 0}$. The goal is to find an assignment $\sigma : J \to [m]$ of jobs to machines so as to minimize the makespan $\max_{i \in [m]} \max_{r \in [d]} (\sum_{j:\sigma(j)=i} p_{j,r})$. A natural lower bound on the optimal makespan is lb := $\max\{\max_{j \in J, r \in [d]} p_{j,r}, \max_r(\sum_j p_{j,r}/m)\}$. Our main result is a very simple O(log d)-approximation algorithm for vector scheduling with respect to the lower bound lb. As an application, we show that the above guarantee leads to an O(log log m)-approximation for Stochastic Minimum-Norm Load Balancing (StochNormLB). In StochNormLB, we have m identical machines, a set J of n independent stochastic jobs whose processing times are nonnegative random variables, and a monotone, symmetric norm $f : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$. The goal is to find an assignment $\sigma : J \to [m]$ that minimizes the expected f-norm of the induced machine-load vector, where the load on a machine is the (random) total processing time assigned to it. Our O(log log m)-approximation guarantee is in fact much stronger: we obtain an assignment that is simultaneously an O(log log m)-approximation for StochNormLB with all monotone, symmetric norms.

Sharat Ibrahimpur, Chaitanya Swamy
University of Waterloo
sharat.ibrahimpur@uwaterloo.ca, cswamy@uwaterloo.ca

CP32

Towards Non-Uniform K-Center with Constant Types of Radii

In the Non-Uniform k-Center problem we need to cover a finite metric space using k balls of different radii that can be scaled uniformly. The goal is to minimize the scaling factor. If the number of different radii is unbounded, the problem does not admit a constant-factor approximation algorithm but it has been conjectured that such an algorithm exists if the number of radii is constant. Yet, this is known only for the case of two radii. Our first contribution is a simple black box reduction which shows that if one can handle the variant of $t-1$ radii with outliers, then one can also handle t radii. Together with an algorithm by Chakrabarty and Negahbani for two radii with outliers, this immediately implies a constant-factor approximation algorithm for three radii; thus making further progress on

the conjecture. Furthermore, using algorithms for the k-center with outliers problem, that is the one radii with outliers case, we also get a simple algorithm for two radii. The algorithm by Chakrabarty and Negahbani uses a top-down approach, starting with the larger radius and then proceeding to the smaller one. Our reduction, on the other hand, looks only at the smallest radius and eliminates it, which suggests that a bottom-up approach is promising. In this spirit, we devise a modification of the Chakrabarty and Negahbani algorithm which runs in a bottom-up fashion, and in this way we recover their result with the advantage of having a simpler analysis.

Xinrui Jia
School of Computer and Communication Sciences,
EPFL, Switzerland
xinrui.jia@epfl.ch

Lars Rohwedder
EPFL
lars.rohwedder@epfl.ch

Kshiteej Sheth
School of Computer and Communication Sciences,
EPFL, Switzerland
kshiteej.sheth@epfl.ch

Ola Svensson
EPFL
ola.svensson@epfl.ch

CP32

The Short-Side Advantage in Random Matching Markets

A breakthrough of Ashlagi, Kanoria, and Leshno [AKL17] found that imbalance in the number of agents on either side of a random matching market has a profound effect on the market's expected characteristics. Specifically, across all stable matchings, the "long side" (i.e. the side with a greater number of agents) receives significantly worse matches in expectation than the short side. Intuitively, this occurs because an agent on the long side is essentially unneeded to create a stable matching – a matching could form almost as easily without them. Thus, an agent on the long side has very little market power, and must settle for a match which is not much better than a random assignment. We provide a new and simpler proof for a result of [AKL17] which formalizes this intuition.

Linda Cai, Clayton Thomas
Princeton University
tcai@princeton.edu, claytont@princeton.edu

CP32

An Explicit Vector Algorithm for High-Girth Max-Cut

We give an approximation algorithm for MaxCut and provide guarantees on the average fraction of edges cut on d-regular graphs of girth $\geq 2k$. For every $d \geq 3$ and $k \geq 4$, our approximation guarantees are better than those of all other classical and quantum algorithms known to the authors. Our algorithm constructs an explicit vector solution to the standard semidefinite relaxation of MaxCut and applies hyperplane rounding. It may be viewed as a simplification of the previously best known technique, which approximates Gaussian wave processes on the infinite d-

regular tree.

Jessica Thompson
University of Maryland
Sandia National Laboratories
jktho@umd.edu

Ojas Parekh
Sandia National Laboratories
odparek@sandia.gov

Kunal Marwaha
Berkeley Center for Quantum Information and
Computation
marwahaha@berkeley.edu

CP32
Oblivious Online Contention Resolution Schemes

Contention resolution schemes (CRSs) are powerful tools for obtaining "ex post feasible' solutions from candidates that are drawn from "ex ante feasible' distributions. Online contention resolution schemes (OCRSs), the online version, have found myriad applications in Bayesian and stochastic problems, such as prophet inequalities and stochastic probing. When the ex ante distribution is unknown, it was unknown whether good CRSs/OCRSs exist with no sample (in which case the scheme is oblivious) or few samples from the distribution. In this work, we give a simple $\frac{1}{e}$-selectable oblivious single item OCRS by mixing two simple schemes evenly, and show, via a Ramsey theory argument, that it is optimal. On the negative side, we show that no CRS or OCRS with $O(1)$ samples can be $\Omega(1)$-balanced/selectable (i.e., preserve every active candidate with a constant probability) for graphic or transversal matroids.

Qianfan Zhang
IIIS, Tsinghua University
zqf18@mails.tsinghua.edu.cn

Hu Fu
ITCS, Shanghai University of Finance and Economics
fuhu@mail.shufe.edu.cn

Pinyan Lu
Shanghai University of Finance and Economics
lu.pinyan@mail.shufe.edu.cn

Zhihao Tang
ITCS, Shanghai University of Finance and Economics
tang.zhihao@mail.shufe.edu.cn

Abner Turkieltaub
University of British Columbia
abner7@cs.ubc.ca

Hongxun Wu
Institute for Interdisciplinary Information Sciences
Tsinghua University
wuhx18@mails.tsinghua.edu.cn

Jinzhao Wu
CFCS, Computer Science Dept., Peking University
jinzhao.wu@pku.edu.cn

CP33
Near-Optimal Explainable k-Means for All Dimen-

sions

Many clustering algorithms are guided by certain cost functions such as the widely-used k-means cost. These algorithms divide data points into clusters with often complicated boundaries, creating difficulties in explaining the clustering decision. In a recent work, Dasgupta, Frost, Moshkovitz, and Rashtchian (ICML 2020) introduced explainable clustering, where the cluster boundaries are axis-parallel hyperplanes and the clustering is obtained by applying a decision tree to the data. The central question here is: how much does the explainability constraint increase the value of the cost function? Given d-dimensional data points, we show an efficient algorithm that finds an explainable clustering whose k-means cost is at most $k^{1-2/d} \operatorname{poly}(d \log k)$ times the minimum cost achievable by a clustering without the explainability constraint, assuming $k, d \geq 2$. Taking the minimum of this bound and the $k \operatorname{polylog}(k)$ bound in independent work by Makarychev-Shan (ICML 2021), Gamlath-Jia-Polak-Svensson (2021), or Esfandiari-Mirrokni-Narayanan (2021), we get an improved bound of $k^{1-2/d} \operatorname{polylog}(k)$, which we show is optimal for every choice of $k, d \geq 2$ up to a poly-logarithmic factor in k. For $d = 2$ in particular, we show an $O(\log k \log \log k)$ bound, improving near-exponentially over the previous best bound of $O(k \log k)$ by Laber and Murtinho (ICML 2021).

Moses Charikar
Stanford University, California
moses@cs.stanford.edu

Lunjia Hu
Stanford University
lunjia@stanford.edu

CP33
How Many Clusters ? - An Algorithmic Answer

Many algorithms for clustering high dimensional data assume that k, the number of clusters, is given. However, there has been little work on provably inferring k from the data. This paper gives polynomial time algorithms for finding k from the data assuming it satisfies certain natural deterministic conditions. Informally, we assume that there is a Ground Truth (GT) clustering of the data such that: (i) Each cluster has a certain minimum size, (ii) the inter-mean separation of any two distinct clusters in the GT is large enough, and (iii) we define a novel "no large sub-cluster'(NLSC) property that characterizes the notion of a cluster by stipulating that there be no subsets of low "directional variance'. NLSC is indeed satisfied by large class of distributions. The first main contribution is an algorithm for finding k where m, the minimum GT cluster size, is known. This algorithm uses a novel rounding procedure which finds subsets of size m with low Directional Variance by rounding a SDP relaxation. It is shown that k is precisely the number of such sets whose means are well-separated. The harder problem of finding k when m not given is addressed by running the previous algorithm for each value of m to find candidate values of k and the corresponding k-clustering. The second major contribution of this paper is a test which certifies the correct candidate thereby yielding a polynomial time algorithm which finds k.

Amit Kumar
IIT Delhi
amitk@cse.iitd.ac.in

Ravindran Kannan
Indian. Institute of Science
Microsoft Research India
kannan@microsoft.com

Chiranjib Bhattacharyya
Indian Institute of Science, Bangalore
India - 560012
chiru@iisc.ac.in

CP33
Almost Tight Approximation Algorithms for Explainable Clustering

Recently, due to an increasing interest for transparency in artificial intelligence, several methods of explainable machine learning have been developed with the simultaneous goal of accuracy and interpretability by humans. In this paper, we study a recent framework of explainable clustering first suggested by Dasgupta et al. Specifically, we focus on the k-means and k-median problems and provide nearly tight upper and lower bounds. First, we provide an $O(\log k \log \log k)$-approximation algorithm for explainable k-median, improving on the best known algorithm of $O(k)$ (Dasgupta et al.) and nearly matching the known $\Omega(\log k)$ lower bound (Dasgupta et al.). In addition, in low-dimensional spaces $d \ll \log k$, we show that our algorithm also provides an $O(d \log^2 d)$-approximate solution for explainable k-median. This improves over the best known bound of $O(d \log k)$ for low dimensions (Laber and Murtinho), and is a constant for constant dimensional spaces. To complement this, we show a nearly matching $\Omega(d)$ lower bound. Next, we study the k-means problem in this context and provide an $O(k \log k)$-approximation algorithm for explainable k-means, improving over the $O(k^2)$ bound of Dasgupta et al. and the $O(dk \log k)$ bound of Laber and Murtinho. To complement this we provide an almost tight $\Omega(k)$ lower bound, improving over the $\Omega(\log k)$ lower bound of Dasgupta et al.

Hossein Esfandiari
Google Research
esfandiari@google.com

Vahab Mirrokni
Google Inc.
mirrokni@google.com

Shyam Narayanan
Massachusetts Institute of Technology
shyamsn@mit.edu

CP33
Approximating Fair Clustering with Cascaded Norm Objectives

We introduce the (p, q)-Fair Clustering problem. In this problem, we are given a set of points P and a collection of different weight functions W. We would like to find a clustering which minimizes the ℓ_q-norm of the vector over W of the ℓ_p-norms of the weighted distances of points in P from the centers. This generalizes various clustering problems, including Socially Fair k-Median and k-Means, and is closely connected to other problems such as Densest k-Subgraph and Min k-Union. We utilize convex programming techniques to approximate the (p, q)-Fair Clustering problem for different values of p and q. When $p \geq q$, we get an $O(k^{(p-q)/(2pq)})$, which nearly matches a $k^{\Omega((p-q)/(pq))}$

lower bound based on conjectured hardness of Min k-Union and other problems. When $q \geq p$, we get an approximation which is independent of the size of the input for bounded p, q, and also matches the recent $O((\log n/(\log \log n))^{1/p})$-approximation for (p, ∞)-Fair Clustering by Makarychev and Vakilian (COLT 2021).

Eden Chlamtac
Ben-Gurion University
chlamtac@cs.bgu.ac.il

Yury Makarychev
Toyota Technological Institute at Chicago
yury@ttic.edu

Ali Vakilian
MIT
CSAIL
vakilian@ttic.edu

CP34
On The Complexity Of Binary Polynomial Optimization Over Acyclic Hypergraphs

In this work we advance the understanding of the fundamental limits of computation for Binary Polynomial Optimization (BPO), which is the problem of maximizing a given polynomial function over all binary points. In our main result we provide a novel class of BPO that can be solved efficiently both from a theoretical and computational perspective. In fact, we give a strongly polynomial-time algorithm for instances whose corresponding hypergraph is β-acyclic. We note that the β-acyclicity assumption is natural in several applications including relational database schemes and the lifted multicut problem on trees. Due to the novelty of our proving technique, we obtain an algorithm which is interesting also from a practical viewpoint. This is because our algorithm is very simple to implement and the running time is a polynomial of very low degree in the number of nodes and edges of the hypergraph. Our result completely settles the computational complexity of BPO over acyclic hypergraphs, since the problem is NP-hard on α-acyclic instances. Our algorithm can also be applied to any general BPO problem that contains β-cycles. For these problems, the algorithm returns a smaller instance together with a rule to extend any optimal solution of the smaller instance to an optimal solution of the original instance.

Alberto Del Pia
University of Wisconsin-Madison
delpia@wisc.edu

Silvia Di Gregorio
TU Dresden
silvia.di_gregorio@tu-dresden.de

CP34
Congruency-Constrained TU Problems Beyond the Bimodular Case

A long-standing open question in Integer Programming is whether integer programs with constraint matrices with bounded subdeterminants are efficiently solvable. An important special case thereof are congruency-constrained integer programs $\min\{c^\top x : Tx \leq b, \gamma^\top x \equiv r \pmod{m}, x \in \mathbb{Z}^n\}$ with a totally unimodular constraint matrix T. Such problems have been shown to be polynomial-time solvable

for $m = 2$, which led to an efficient algorithm for integer programs with bimodular constraint matrices, i.e., full-rank matrices whose $n \times n$ subdeterminants are bounded by two in absolute value. Whereas these advances heavily relied on existing results on well-known combinatorial problems with parity constraints, new approaches are needed beyond the bimodular case, i.e., for $m > 2$. We make first progress in this direction through several new techniques. In particular, we show how to efficiently decide feasibility of congruency-constrained integer programs with a totally unimodular constraint matrix for $m = 3$. Furthermore, for general m, our techniques also allow for identifying flat directions of infeasible problems, and deducing bounds on the proximity between solutions of the problem and its relaxation.

Martin Nägele, Rico Zenklusen, Richard Santiago
ETH Zurich
martinn@ethz.ch, ricoz@ethz.ch, rtorres@ethz.ch

CP34
Computing Lewis Weights to High Precision

We present an algorithm for computing approximate ℓ_p Lewis weights to high precision. Given a full-rank $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m \geq n$ and a scalar $p > 2$, our algorithm computes ϵ-approximate ℓ_p Lewis weights of $\mathbf{A}$ in $\widetilde{O}_p(\log(1/\epsilon))$ iterations; the cost of each iteration is linear in the input size plus the cost of computing the leverage scores of $\mathbf{DA}$ for diagonal $\mathbf{D} \in \mathbb{R}^{m \times m}$. Prior to our work, such a computational complexity was known only for $p \in (0, 4)$ [**?**], and combined with this result, our work yields the first polylogarithmic-depth polynomial-work algorithm for the problem of computing ℓ_p Lewis weights to high precision for *all* constant $p > 0$. An important consequence of this result is also the first polylogarithmic-depth polynomial-work algorithm for computing a nearly optimal self-concordant barrier for a polytope.

Maryam Fazel, Yin Tat Lee, Swati Padmanabhan
University of Washington
mfazel@uw.edu, yintat@uw.edu, pswati@uw.edu

Aaron Sidford
Stanford University
sidford@stanford.edu

CP34
On Finding Exact Solutions of Linear Programs in the Oracle Model

We consider linear programming in the oracle model: $\min c^\top x$ s.t. $x \in P$, where the polyhedron $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ is given by a separation oracle that returns violated inequalities from the system $Ax \leq b$. We present an algorithm that finds exact primal and dual solutions using $O(n^2 \log(n/\delta))$ oracle calls and $O(n^4 \log(n/\delta) + n^5 \log \log(1/\delta))$ arithmetic operations, where δ is a geometric condition number associated with the system (A, b). These bounds do not depend on the cost vector c. The algorithm works in a black box manner, requiring a subroutine for approximate primal and dual solutions; the above running times are achieved when using the cutting plane method of Jiang, Lee, Song, and Wong (STOC 2020) for this subroutine. Whereas approximate solvers may return primal solutions only, we develop a general framework for extracting dual certificates based on the work of Burrell and Todd (Math. Oper. Res. 1985). Our algorithm works in the real model of computation, and extends results

by Grötschel, Lovász, and Schrijver (Prog. Comb. Opt. 1984), and by Frank and Tardos (Combinatorica 1987) on solving LPs in the bit-complexity model. We show that under a natural assumption, simultaneous Diophantine approximation in these results can be avoided.

Laszlo A. Vegh
London School of Economics
l.vegh@lse.ac.uk

Daniel Dadush
Centrum Wiskunde & Informatica
dadush@cwi.nl

Giacomo Zambelli
London School of Economics
g.zambelli@lse.ac.uk

CP35
Pattern Matching on Grammar-Compressed Strings in Linear Time

In the compressed pattern matching problem we are given a compressed representation of the text, with n being the length of the compressed representation and N being the length of the text, and an uncompressed pattern of length m. The most challenging (and yet relevant when working with highly repetitive data, say biological information) scenario is when the chosen compression method is capable of describing a string of exponential length (in the size of its representation). An elegant formalism for such a compression method is that of straight-line programs, which are simply context-free grammars describing exactly one string. While it has been known that compressed pattern matching problem can be solved in $O(m + n \log N)$ time for this compression method, designing a linear-time algorithm remained open. We resolve this open question by presenting an $O(n + m)$ time algorithm that, given a context-free grammar of size n that produces a single string t and a pattern p of length m, decides whether p occurs in t as a substring. To this end, we devise improved solutions for the weighted ancestor problem and the substring concatenation problem.

Pawel Gawrychowski
University of Wroclaw
gawry@cs.uni.wroc.pl

Moses Ganardi
Max Planck Institute for Software Systems
ganardi@mpi-sws.org

CP35
Near-Optimal Quantum Algorithms for String Problems

We study quantum algorithms for several fundamental string problems, including Longest Common Substring, Lexicographically Minimal String Rotation, and Longest Square Substring. These problems have been widely studied in the stringology literature since the 1970s, and are known to be solvable by near-linear time classical algorithms. In this work, we give quantum algorithms for these problems with near-optimal query complexities and time complexities. Specifically, we show that: LCS can be solved by a quantum algorithm in $\tilde{O}(n^{2/3})$ time, improving upon the recent $\tilde{O}(n^{5/6})$-time algorithm by Le Gall and Seddighin (2020). Our algorithm uses the MNRS quantum

walk framework, together with a careful combination of string synchronizing sets (Kempa and Kociumaka, 2019) and generalized difference covers. LMSR can be solved by a quantum algorithm in $n^{1/2+o(1)}$ time, improving upon the recent $\tilde{O}(n^{3/4})$-time algorithm by Wang and Ying (2020). We design our algorithm by first giving a new classical divide-and-conquer algorithm in near-linear time based on exclusion rules, and then speeding it up quadratically using nested Grover search and quantum minimum finding. LSS can be solved by a quantum algorithm in $\tilde{O}(\sqrt{n})$ time. Our algorithm is an adaptation of the algorithm by Le Gall and Seddighin for the Longest Palindromic Substring problem, but uses additional techniques to overcome the difficulty that binary search no longer applies.

Shyan Akmal, <u>Ce Jin</u>
EECS
MIT
naysh@mit.edu, cejin@mit.edu

CP35

An Upper Bound and Linear-Space Queries on the LZ-End Parsing

Lempel-Ziv (LZ77) compression is the most commonly used lossless compression algorithm. The basic idea is to greedily break the input string into blocks (called "phrases"), every time forming as a phrase the longest prefix of the unprocessed part that has an earlier occurrence. In 2010, Kreft and Navarro introduced a variant of LZ77 called LZ-End, that additionally requires the previous occurrence of each phrase to end at the boundary of an already existing phrase. Due to its excellent practical performance, they conjectured that it achieves a compression that can be provably upper-bounded in terms of the LZ77 size. Despite the recent progress in understanding such relation for other compression algorithms, no such result is known for LZ-End. We prove that for any string of length n, the number z_e of phrases in the LZ-End parsing satisfies $z_e = \mathcal{O}(z \log^2 n)$, where z is the number of phrases in the LZ77 parsing. This puts LZ-End among the strongest dictionary compressors and solves a decade-old open problem of Kreft and Navarro. Using our techniques we also derive bounds for other variants of LZ-End and with respect to other compression measures. Our second contribution is a data structure that implements random access queries to the text in $\mathcal{O}(z_e)$ space and $\mathcal{O}(\text{polylog}\, n)$ time. All previous data structures either incur a logarithmic penalty in the space or have slow queries.

<u>Dominik Kempa</u>
Stony Brook University
kempa@cs.stonybrook.edu

Barna Saha
University of California, Berkeley
barnas@berkeley.edu

CP35

How Compression and Approximation Affect Efficiency in String Distance Measures

In this work, we study the time complexity of sequence similarity problems for highly compressible data. We consider grammar compression, which unifies many relevant compression schemes. For two strings of total length N and total compressed size n, it is known that the edit distance (ED) and longest common subsequence (LCS) can

be computed exactly in time $\tilde{O}(nN)$, as opposed to $O(N^2)$ for the uncompressed setting. When $k > 2$ strings are concerned, the fastest known exact algorithms for median-ED and LCS take $O(N^k)$ time, whereas the one for center-ED takes $O(N^{2k})$ time. This naturally raises the question if compression can help to reduce the running times significantly for $k > 2$. Unfortunately, we show new lower bounds that rule out any improvement beyond $\Omega(N^{k-1}n)$ time for any of these problems assuming the Strong Exponential Time Hypothesis. We thus ask if allowing approximation can help, and we show that it can be surprisingly effective for compressed data. We develop an $\tilde{O}(N^{k/2}n^{k/2})$-time FPTAS for median-ED and an $\tilde{O}(N^{k/2+o(k)}n^{k/2})$-time FPTAS for center-ED. In comparison, no $O(N^{k-\Omega(1)})$-time PTAS is known for these problems in the uncompressed setting. For $k = 2$, we get an $\tilde{O}(N^{2/3}n^{4/3})$-time FPTAS for both ED and LCS. In contrast, there is no $O(N^{2-\Omega(1)})$-time algorithm that approximates LCS up to an $N^{o(1)}$ factor.

Arun Ganesh, <u>Tomasz Kociumaka</u>, Andrea Lincoln
UC Berkeley
arunganesh@berkeley.edu, kociumaka@mimuw.edu.pl,
andrealincoln@berkeley.edu

Barna Saha
University of California, Berkeley
barnas@berkeley.edu

CP36

A Simple Deterministic Algorithm for Systems of Quadratic Polynomials over $\mathbb{F}_2$

This article discusses a simple deterministic algorithm for solving quadratic Boolean systems which is essentially a special case of more sophisticated methods. The main idea fits in a single sentence: guess enough variables so that the remaining quadratic equations can be solved by linearization (i.e. by considering each remaining monomial as an independent variable and solving the resulting linear system) and restart until the solution is found. Under strong heuristic assumptions, this finds all the solutions of m quadratic polynomials in n variables with $\tilde{}$

$O\left(2^{n-\sqrt{2m}}\right)$ operations. Although the best known algorithms require exponentially less time, the present technique has the advantage of being simpler to describe and easy to implement. In strong contrast with the state-of-the-art, it is also quite efficient in practice.

<u>Charles Bouillaguet</u>
Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
charles.bouillaguet@lip6.fr

Claire Delaplace, Monika Trimoska
MIS Laboratory, Université de Picardie Jules Verne,
Amiens,
claire.delaplace@u-picardie.fr,   monika.trimoska@u-picardie.fr

CP36

Parameterizing the Permanent: Hardness for Fixed Excluded Minors

In the 1960s, statistical physicists discovered a fascinating algorithm for counting perfect matchings in planar graphs.

Valiant later showed that the same problem is #P-hard for general graphs. Since then, the algorithm for planar graphs was extended to bounded-genus graphs, to graphs excluding $K_{3,3}$ or K_5 as a minor, and more generally, to any graph class excluding a fixed minor H that can be drawn in the plane with a single crossing. This stirred up hopes that counting perfect matchings might be polynomial-time solvable for graph classes excluding any fixed minor H. Alas, in this paper, we show #P-hardness for K_8-minor-free graphs by a simple and self-contained argument.

Radu Curticapean
IT University of Copenhagen
radu.curticapean@gmail.com

Mingji Xia
Institute of Software, Chinese Academy of Sciences.
University of Chinese Academy of Sciences
mingji@ios.ac.cn

CP36

A Tighter Relation Between Hereditary Discrepancy and Determinant Lower Bound

In seminal work, Lovász, Spencer, and Vesztergombi [European J. Combin., 1986] proved a lower bound for the hereditary discrepancy of a matrix $A \in \mathbb{R}^{m \times n}$ in terms of the maximum $|\det(B)|^{1/k}$ over all $k \times k$ submatrices B of A. We show algorithmically that this determinant lower bound can be off by at most a factor of $O(\sqrt{\log(m) \cdot \log(n)})$, improving over the previous bound of $O(\log(mn) \cdot \sqrt{\log(n)})$ given by Matoušek [Proc. of the AMS, 2013]. Our result immediately implies $\mathsf{herdisc}(\mathcal{F}_1 \cup \mathcal{F}_2) \leq O(\sqrt{\log(m) \cdot \log(n)} \cdot \max(\mathsf{herdisc}(\mathcal{F}_1), \mathsf{herdisc}(\mathcal{F}_2))$, for any two set systems $\mathcal{F}_1, \mathcal{F}_2$ over $[n]$ satisfying $|\mathcal{F}_1 \cup \mathcal{F}_2| = m$. Our bounds are tight up to constants when $m = O(\mathsf{poly}(n))$ due to a construction of Pálvölgyi [Discrete Comput. Geom., 2010] or the counterexample to Beck's three permutation conjecture by Newman, Neiman and Nikolov [FOCS, 2012].

Haotian Jiang, Victor Reis
University of Washington
jhtdavid96@gmail.com, voreis@cs.washington.edu

CP36

Faster Exponential Algorithm for Permutation Pattern Matching

The Permutation Pattern Matching problem asks, given two permutations σ on n elements and π, whether σ admits a subsequence with the same relative order as π (or, in the counting version, how many such subsequences are there). This natural problem was shown by Bose, Buss and Lubiw [IPL 1998] to be NP-complete, and hence we should seek exact exponential time solutions. The asymptotically fastest such solution up to date, by Berendsohn, Kozma and Marx [IPEC 2019], works in $\mathcal{O}(1.6181^n)$ time. We design a simple and faster $\mathcal{O}(1.415^n)$ time algorithm for both the detection and the counting version. We also prove that this is optimal among a certain natural class of algorithms.

Pawel Gawrychowski
University of Wroclaw
gawry@cs.uni.wroc.pl

Mateusz Rzepecki
University ofWroclaw
mateuszrzepecki99@gmail.com

CP36

The Quantum Union Bound Made Easy

We give a short proof of Gao's Quantum Union Bound and Gentle Sequential Measurement theorems.

Ryan O'Donnell, Ramgopal Venkateswaran
Carnegie Mellon University, USA
odonnell@cs.cmu.edu, ramvenkat98@gmail.com

CP37

Private Interdependent Valuations

We consider the single-item interdependent value setting, where a single item is auctioned to n buyers, and each buyer has a private signal s_i describing a piece of information about the item. Additionally, each bidder i has a valuation function $vi(s_1, \ldots, s_n)$ mapping the private signals of all buyers into their value for the item. This setting captures scenarios where the items information is asymmetric or dispersed among agents. It is generally assumed that each bidders valuation function v_i is public knowledge, but in many situations, the seller may not know the bidders valuation functions. In this paper, we design mechanisms that guarantee approximately-optimal social welfare while satisfying ex-post incentive compatibility and individually rationality for the case where the valuation functions are private to the bidders. When the valuations are private, we show that no finite bound on the social welfare can be achieved by any deterministic mechanism even under single-crossing. Moreover, no randomized mechanism can guarantee better than n-approximation. We thus consider valuation functions that are submodular over signals (SOS), introduced by Eden et al. [EC19]. Our main result is an $O(log^2(n))$-approximation randomized mechanism for buyers with private signals and valuations under the SOS condition. We also give a tight $T(k)$-approximation mechanism for the case each agent's valuation depends on at most k other signals.

Alon Eden
Tel Aviv University
alonarden@gmail.com

CP37

Fixed-Price Approximations in Bilateral Trade

We consider the bilateral trade problem, in which two agents trade a single indivisible item. It is known that the only dominant-strategy truthful mechanism is the fixed-price mechanism: given commonly known distributions of the buyer's value B and the seller's value S, a price p is offered to both agents and trade occurs if $S \leq p \leq$

B. The objective is to maximize either expected welfare $\mathbb{E}[S + (B - S)\mathbf{1}_{S \leq p \leq B}]$ or expected gains from trade $\mathbb{E}[(B - S)\mathbf{1}_{S \leq p \leq B}]$. We improve the approximation ratios for several welfare maximization variants of this problem. When the agents' distributions are identical, we show that the optimal approximation ratio for welfare is $\frac{2+\sqrt{2}}{4}$. With just one prior sample from the common distribution, we show that a 3/4-approximation to welfare is achievable. When agents' distributions are not required to be identical, we show that a previously best-known $(1 - 1/e)$-approximation can be strictly improved, but $1 - 1/e$ is optimal if only the seller's distribution is known.

Francisco Pernice, Zi Yang Kang, Jan Vondrak
Stanford University
fpernice@stanford.edu,   zykang@stanford.edu,   jvon-drak@stanford.edu

CP37

Metric Distortion Bounds for Randomized Social Choice

Consider the following social choice problem. Suppose we have a set of n voters and m candidates that lie in a metric space. The goal is to design a mechanism to choose a candidate whose average distance to the voters is as small as possible. However, the mechanism doesn't get direct access to the metric space. Instead, it gets each voter's ordinal ranking of the candidates by distance. Given only this partial information, what is the smallest worst-case approximation ratio (known as the distortion) that a mechanism can guarantee? A simple example shows that no deterministic mechanism can guarantee distortion better than 3, and no randomized mechanism can guarantee distortion better than 2. It has been conjectured that both of these lower bounds are optimal, and recently, Gkatzelis, Halpern, and Shah proved this conjecture for deterministic mechanisms. We disprove the conjecture for randomized mechanisms for $m \geq 3$ by constructing elections for which no randomized mechanism can guarantee distortion better than 2.0261 for $m = 3$, 2.0496 for $m = 4$, up to 2.1126 as $m \to \infty$. We obtain our lower bounds by identifying a class of simple metrics that appear to capture much of the hardness of the problem, and we show that any randomized mechanism must have high distortion on one of these metrics. We conjecture that these bounds give the optimal distortion for every m, and provide a proof for $m = 3$, thereby resolving that case.

Prasanna Ramakrishnan
Stanford University
pras1712@stanford.edu

Moses Charikar
Stanford University, California
moses@cs.stanford.edu

CP37

Deterministic Budget-Feasible Clock Auctions

We revisit the well-studied problem of budget-feasible procurement, where a buyer with a strict budget constraint seeks to acquire services from a group of strategic providers (the sellers). During the last decade, several strategyproof budget-feasible procurement auctions have been proposed, aiming to maximize the value of the buyer, while eliciting each seller's true cost for providing their service. These solutions predominantly take the form of randomized sealed-bid auctions: they ask the sellers to report their private costs and then use randomization to determine which subset of services will be procured and how much each of the chosen providers will be paid, ensuring that the total payment does not exceed the buyer's budget. Our main result in this paper is a novel method for designing budget-feasible auctions, leading to solutions that outperform the previously proposed auctions in multiple ways. First, our solutions take the form of descending clock auctions, and thus satisfy a list of very appealing properties, such as obvious strategyproofness, group strategyproofness, transparency, and unconditional winner privacy; this makes these auctions much more likely to be used in practice. Second, in contrast to previous results that heavily depend on randomization, our auctions are deterministic. As a result, we provide an affirmative answer to one of the main open questions in this literature, asking whether a deterministic strategyproof auction can achieve a constant approximation when the buyer's valuation function is submodular over the set of services. In addition to this, we also provide the first deterministic budget-feasible auction that matches the approximation bound of the best-known randomized auction for the class of subadditive valuations. Finally, using our method, we improve the best-known approximation factor for monotone submodular valuations, which has been the focus of most of the prior work.

Eric Balkanski, Pranav Garimidi
Columbia University
eb3224@columbia.edu, tbd@columbia.edu

Vasilis Gkatzelis
Drexel University, Computer Science department
gkatz@drexel.edu

Daniel Schoepflin, Xizhi Tan
Drexel University
schoep@drexel.edu, xizhi@drexel.edu

CP38

Deterministic and Las Vegas Algorithms for Sparse Nonnegative Convolution

Computing the convolution $A \star B$ of two length-n integer vectors A, B is a core problem in several disciplines. For many of its applications it typically suffices to compute convolutions of nonnegative vectors. This problem can be classically solved in time $O(n \log n)$ using the Fast Fourier Transform. However, in many applications the involved vectors are sparse and hence one could hope for output-sensitive algorithms. This question was raised by Muthukrishnan and solved by Cole and Hariharan (STOC '02) by a randomized algorithm running in near-linear time in the (unknown) output-size t. Chan and Lewenstein (STOC '15) presented a deterministic algorithm with a $2^{O(\sqrt{\log t \cdot \log \log n})}$ overhead in running time and an additional assumption which was later removed by Bringmann and Nakos (ICALP '21). In this paper we present the

first deterministic near-linear-time algorithm for computing sparse nonnegative convolutions. This immediately gives improved deterministic algorithms for the state-of-the-art of output-sensitive Subset Sum, block-mass pattern matching, N-fold Boolean convolution, and others, matching up to log-factors the fastest known randomized algorithms for these problems. Additionally, we provide two fast Las Vegas algorithms for computing sparse nonnegative convolutions: A simple $O(t \log^2 t)$-time algorithm (as an accessible alternative to Cole and Hariharan's algorithm) and improved $O(t \log t \cdot \log \log t)$-time algorithm.

Karl Bringmann
Saarland University and Max Planck Institute for Informatics
Saarland Informatics Campus, Saarbrücken, Germany
kbringma@mpi-inf.mpg.de

Nick Fischer
Saarland University and Max Planck Institute for Informatics
Saarland Informatics Campus
nfischer@mpi-inf.mpg.de

Vasileios Nakos
Saarland University and Max Planck Institute for Informatics
Saarland Informatics Campus, Germany
vnakos@mpi-inf.mpg.de

CP38

Near-Optimal Algorithms for Linear Algebra in the Current Matrix Multiplication Time

In the numerical linear algebra community, it was suggested that to obtain nearly optimal bounds for various problems such as rank computation, finding a maximal linearly independent subset of columns (a basis), regression, or low-rank approximation, a natural way would be to resolve the main open question of Nelson and Nguyen (FOCS, 2013). This question is regarding the logarithmic factors in the sketching dimension of existing oblivious subspace embeddings that achieve constant-factor approximation. We show how to bypass this question using a refined sketching technique, and obtain optimal or nearly optimal bounds for these problems. A key technique we use is an explicit mapping of Indyk based on uncertainty principles and extractors, which after first applying known oblivious subspace embeddings, allows us to quickly spread out the mass of the vector so that sampling is now effective. We thereby avoid a logarithmic factor in the sketching dimension that is standard in bounds proven using the matrix Chernoff inequality. For the fundamental problems of rank computation and finding a basis, our algorithms improve Cheung, Kwok, and Lau (JACM, 2013), and are optimal to within a constant factor and a poly(loglog(n))-factor, respectively. Further, for constant-factor regression and low-rank approximation we give the first optimal algorithms, for the current matrix multiplication exponent.

Nadiia Chepurko
Massachusetts Institute of Technology
nadiia@mit.edu

Kenneth Clarkson
IBM
klclarks@us.ibm.com

Praneeth Kacham
Carnegie Mellon University
Pittsburgh, PA, USA
pkacham@andrew.cmu.edu

David Woodruff
Carnegie Mellon University
dwoodruf@andrew.cmu.edu

CP38

Simulating Random Walks in Random Streams

The random order graph streaming model has received significant attention recently, with problems such as matching size estimation, component counting, and the evaluation of bounded degree constant query testable properties shown to admit surprisingly space efficient algorithms. The main result of this paper is a space efficient single pass random order streaming algorithm for simulating nearly independent random walks that start at uniformly random vertices. We show that the distribution of k-step walks from b vertices chosen uniformly at random can be approximated up to error ε per walk using $(1/\varepsilon)^{O(k)} 2^{O(k^2)} \cdot b$ words of space with a single pass over a randomly ordered stream of edges, solving an open problem of Peng and Sohler [SODA '18]. Applications of our result include the estimation of the average return probability of the k-step walk (the trace of the kth power of the random walk matrix) as well as the estimation of PageRank. We complement our algorithm with a strong impossibility result for directed graphs.

John M. Kallaugher
Sandia National Laboratories
jmkall@sandia.gov

Michael Kapralov
EPFL
michael.kapralov@epfl.ch

Eric Price
The Unviersity of Texas at Austin
ecprice@cs.utexas.edu

CP38

Improved Sliding Window Algorithms for Clustering and Coverage via Bucketing-Based Sketches

Streaming computation plays an important role in large-scale data analysis. The sliding window model is a model of streaming computation which also captures the recency of the data. In this model, data arrives one item at a time, but only the latest W data items are considered for a particular problem. The goal is to output a good solution at the end of the stream by maintaining a small summary during the stream. In this work, we propose a new algorithmic framework for designing efficient sliding window algorithms via *bucketing-based sketches*. Based on this new framework, we develop space-efficient sliding window algorithms for k-cover, k-clustering and diversity maximization problems.

For each of the above problems, our algorithm achieves $(1 \pm \varepsilon)$-approximation. Compared with the previous work, it improves both the approximation ratio and the space.

Alessandro Epasto, Mohammad Mahdian
Google Research
aepasto@google.com, mahdian@google.com

Vahab Mirrokni
Google Inc.
mirrokni@google.com

Peilin Zhong
Google Research
peilinz@google.com

CP39

Optimal Angle Bounds for Steiner Triangulations of Polygons

For any simple polygon P we compute the optimal upper and lower angle bounds for triangulating P with Steiner points, and show that these bounds can be attained (except in one special case). The sharp angle bounds for an N-gon are computable in time $O(N)$, even though the number of triangles needed to attain these bounds has no bound in terms of N alone. In general, the sharp upper and lower bounds cannot both be attained by a single triangulation, although this does happen in some cases. For example, we show that any polygon with minimal interior angle θ has a triangulation with all angles in the interval $I = [\theta, 90° - \min(36°, \theta)/2]$, and for $\theta \leq 36°$ both bounds are best possible. Surprisingly, we prove the optimal angle bounds for polygonal triangulations are the same as for triangular dissections. The proof of this verifies, in a stronger form, a 1984 conjecture of Gerver.

Christopher Bishop
Stony Brook University
bishop@math.stonybrook.edu

CP39

Polygon Placement Revisited: (Degree of Freedom + 1)-Sum Hardness and an Improvement via Offline Dynamic Rectangle Union

We revisit the classical problem of determining the largest copy of a simple polygon P that can be placed into a simple polygon Q. Despite significant effort, known algorithms require high polynomial running times. (Barequet and Har-Peled, 2001) give a lower bound of $n^{2-o(1)}$ under the 3SUM conjecture when P and Q are (convex) polygons with $\Theta(n)$ vertices each. This leaves open whether we can establish (1) hardness beyond quadratic time and (2) any superlinear bound for constant-sized P or Q. In this paper, we affirmatively answer these questions under the kSUM conjecture, proving natural hardness results that increase with each degree of freedom (scaling, x-translation, y-translation, rotation): (1) Finding the largest copy of P that can be x-translated into Q requires time $n^{2-o(1)}$ under the 3SUM conjecture. (2) Finding the largest copy of P that can be arbitrarily translated into Q requires time $n^{2-o(1)}$ under the 4SUM conjecture. (3) The above lower bounds are al-

most tight when one of the polygons is of constant size: we obtain an $\tilde{O}((pq)^{2.5})$-time algorithm for orthogonal polygons P, Q with p and q vertices, respectively. (4) Finding the largest copy of P that can be arbitrarily rotated and translated into Q requires time $n^{3-o(1)}$ under the 5SUM conjecture. We are not aware of any other such natural (degree of freedom +1)-SUM hardness for a geometric optimization problem.

Marvin Künnemann
Institute for Theoretical Studies
ETH Zürich
marvin.kuennemann@eth-its.ethz.ch

André Nusser
Max Planck Institute for Informatics
Saarland Informatics Campus
anusser@mpi-inf.mpg.de

CP39

Preprocessing Imprecise Points for the Pareto Front

The preprocessing model for uncertain data models the geometric imprecision of day-to-day algorithmic input. In the preprocessing model we are given a set of regions R which model the uncertainty associated with an unknown set of points P. In this model there are two phases: a preprocessing phase, in which we have access only to R, followed by a reconstruction phase, in which we have access to points in P at a certain retrieval cost C per point. We study the following algorithmic question: how fast can we compute the maximal points in P in the preprocessing model? We show that if R is a set of pairwise-disjoint axis-aligned rectangles, then we can preprocess R to reconstruct the staircase of maximal points of P efficiently. To refine our algorithmic analysis, we introduce a new notion of algorithmic optimality which relates to the entropy of the uncertainty regions. Our proposed uncertainty-region optimality falls on the spectrum between worst-case optimality and instance optimality. We prove that instance optimality is unobtainable for a wide class of problems in the preprocessing model. Our results are worst-case optimal in the preprocessing phase; in the reconstruction phase, our results are uncertainty-region optimal with respect to real RAM instructions, and instance optimal with respect to point retrievals.

Ivor Van Der Hoog
Utrecht University
vanderhoog@gmail.com

Irina Kostitsyna
Technical University Eindhoven
i.kostitsyna@tue.nl

Maarten Löffler
Utrecht University
m.loffler@uu.nl

Bettina Speckmann
Dept. of Mathematics and Computer Science
TU Eindhoven

b.speckmann@tue.nl

CP39
Constructing Many Faces in Arrangements of Lines and Segments

We present new algorithms for computing many faces in arrangements of lines and segments. Given a set S of n lines (resp., segments) and a set P of m points in the plane, the problem is to compute the faces of the arrangements of S that contain at least one point of P. These problems have been extensively studied before and our new algorithms improve the previous work. For the line case, we give a deterministic algorithm of $O(m^{2/3}n^{2/3}\log^{2/3}(n/\sqrt{m}) + (m+n)\log n)$ time. For the segment case, we present a deterministic algorithm of $O(n^{2/3}m^{2/3}\log n + \tau(n\alpha^2(n) + n\log m + m)\log n)$ time, where $\tau = \min\{\log m, \log(n/\sqrt{m})\}$ and $\alpha(n)$ is the inverse Ackermann function. We also give a randomized algorithm of $O(m^{2/3}K^{1/3}\log n + \tau(n\alpha(n) + n\log m + m)\log n \log K)$ expected time, where K is the number of intersections of all segments of S. In addition, we consider the query version of the problem, that is, preprocess S to compute the face of the arrangement of S that contains any query point. We present new results that improve the previous work for both the line and the segment cases.

Haitao Wang
Department of Computer Science
Utah State University
haitao.wang@usu.edu

CP40
Routed Be-Trees and Splinterdb: An Optimal External-Memory Dictionary Data Structure and Implementation

Rob Johnson
VMWare Research
robj@vmware.com

CP40
Compact Data Structures for Running the Internet

Jennifer Rexford
Princeton University
jrex@cs.princeton.edu

CP40
Improved Bounds for Scheduling Flows under Endpoint Capacity Constraints

We study flow scheduling with node capacity constraints. We are given capacitated nodes and an online sequence of jobs, each with a release time and a demand to be routed between two nodes. A schedule specifies which jobs are routed in each step, guaranteeing that the total demand on a node in any step is at most its capacity. A key metric in this scenario is response time: the time between a job's release and its completion. Prior work shows no unaugmented algorithm is competitive for average response time, and that a constant factor competitive ratio is achievable with augmentation exceeding 2 (Dinitz-Moseley Infocom 2020). For maximum response time, the best known result is a 2-competitive algorithm with augmentation 4 (Jahanjou et al SPAA 2020). We improve these bounds under various response time objectives. We show that, without resource augmentation, the best competitive ratio for maximum response time is $\Omega(n)$, where n is the number of nodes. Our Proportional Allocation algorithm uses $(1+\varepsilon)$ resource augmentation to achieve a $(1/\varepsilon)$-competitive ratio in the setting with general demands and capacities, and splittable jobs. Our Batch Decomposition algorithm is 2-competitive (resp., optimal) for maximum response time using resource augmentation 2 (resp., 4) in the setting with unit demands and capacities, and unsplittable jobs. We also derive bounds for the simultaneous approximation of average and maximum response time metrics.

David Stalfa, Searidang Pa
Northeastern University
stalfa.d@northeastern.edu, pa.s@northeastern.edu

Rajmohan Rajaraman
Northeastern University
USA
rraj@ccs.neu.edu

CP40
Semi-Streaming Dynamic Connectivity: To Infinity and Beyond

David Tench
Rutgers
dtench@pm.me

CP40
Unbiased Delay Measurement in the Data Plane

Network administrators are interested in measuring the distribution of delays (the time between a request and its response), by directly running succinct algorithms within high-speed network devices. Unfortunately, the considerable gap between the small available memory and the huge volume of arriving traffic makes it challenging to design an algorithm that accurately measures delays. Existing algorithms exhibit bias against samples with higher delays. We present *fridges*, a novel data structure that corrects for the *survivorship bias* due to hash collisions, producing *unbiased* estimates of the delay distribution. The key idea is to consider a sample that was lucky enough to survive many insertions into the data structure as a representative for other similar samples that did not survive. We also show how to combine results from multiple fridges, each optimized for a different range of delays, for further accuracy gains. Simulation experiments show our design outperforms prior work using naive hash-indexed arrays, achieving 2x-4x memory saving. We implement a prototype P4 program running on the Intel Tofino programmable switch, using only moderate hardware resources.

Yufei Zheng, Xiaoqi Chen, Mark Braverman, Jennifer Rexford
Princeton University

yufei@cs.princeton.edu,   xiaoqic@cs.princeton.edu,
mbraverm@cs.princeton.edu, jrex@cs.princeton.edu

CP41

An Improved Analysis of Greedy for Online Steiner Forest

This paper considers the classic Online Steiner Forest problem where one is given a (weighted) graph G and an arbitrary set of k terminal pairs $\{\{s_1, t_1\}, \ldots, \{s_k, t_k\}\}$ that are required to be connected. The goal is to maintain a minimum-weight sub-graph that satisfies all the connectivity requirements as the pairs are revealed one by one. It has been known for a long time that no algorithm (even randomized) can be better than $\Omega(\log(k))$-competitive for this problem. Interestingly, a simple greedy algorithm is already very efficient for this problem. This algorithm can be informally described as follows: *Upon arrival of a new pair $\{s_i, t_i\}$, connect s_i and t_i with the shortest path in the current metric, contract the metric along the chosen path and wait for the next pair.* Although simple and intuitive, greedy proved itself challenging to analyze and its competitive ratio is a long-standing open problem in the area of online algorithms. The last progress on this problem is due to an elegant analysis by Awerbuch, Azar, and Bartal [SODA 1996], who showed that greedy is $O(\log^2(k))$-competitive, and conjectured that is should be $\Theta(\log(k))$-competitive. In this paper, we show that greedy is $O(\log(k) \log \log(k))$-competitive on a wide class of instances, which contains all the lower bounds that appeared in the literature so far.

Etienne Bamas, Marina Drygala, Andreas Maggiori
EPFL
etienne.bamas@epfl.ch,   marina.drygala@epfl.ch,   andreas.maggiori@epfl.ch

CP41

Augmenting Edge Connectivity via Isolating Cuts

We give an algorithm for augmenting the edge connectivity of an undirected graph by using the isolating cuts framework (Li and Panigrahi, FOCS '20). Our algorithm uses poly-logarithmic calls to any max-flow algorithm, which yields a running time of $\tilde{O}(m + n^{3/2})$ and improves on the previous best time of $\tilde{O}(n^2)$ (Benczur and Karger, SODA '98) for this problem. We also obtain an identical improvement in the running time of the closely related edge splitting off problem in undirected graphs.

Ruoxu Cen
Duke University
ruoxu.cen@duke.edu

Jason M. Li
Carnegie Mellon University
jmli@cs.cmu.edu

Debmalya Panigrahi
Duke University

debmalya@cs.duke.edu

CP41

Polynomial Integrality Gap of Flow LP for Directed Steiner Tree

In the Directed Steiner Tree (DST) problem, we are given a directed graph $G = (V, E)$ on n vertices with edge-costs $c \in \mathbb{R}_{\geq 0}^E$, a root vertex r, and a set K of k terminals. The goal is to find a minimum-cost subgraph of G that contains a path from r to every terminal $t \in K$. DST has been a notorious problem for decades as there is a large gap between the best-known polynomial-time approximation ratio of $O(k^\epsilon)$ for any constant $\epsilon > 0$, and the best quasi-polynomial-time approximation ratio of $O\left(\frac{\log^2 k}{\log \log k}\right)$. Towards understanding this gap, we study the integrality gap of the standard flow LP relaxation for the problem. We show that the LP has an integrality gap polynomial in n. Previously, the integrality gap LP is only known to be $\Omega\left(\frac{\log^2 n}{\log \log n}\right)$ [Halperin et al., SODA'03 & SIAM J. Comput.] and $\Omega(\sqrt{k})$ [Zosin-Khuller, SODA'02] in some instance with $\sqrt{k} = O\left(\frac{\log n}{\log \log n}\right)$. Our result gives the first known lower bound on the integrality gap of this standard LP that is polynomial in n, the number of vertices. Consequently, we rule out the possibility of developing a poly-logarithmic approximation algorithm for the problem based on the flow LP relaxation.

Shi Li
University at Buffalo
shil@buffalo.edu

Bundit Laekhanukit
Shanghai University of Finance and Economics
lbundit+sufe@gmail.com

CP41

Local Search for Weighted Tree Augmentation and Steiner Tree

We present a technique that allows for improving on some relative greedy procedures by well-chosen (non-oblivious) local search algorithms. Relative greedy procedures are a particular type of greedy algorithm that start with a simple, though weak, solution, and iteratively replace parts of this starting solution by stronger components. Some well-known applications of relative greedy algorithms include approximation algorithms for Steiner Tree and, more recently, for connectivity augmentation problems. The main application of our technique leads to a $(1.5 + \epsilon)$-approximation for Weighted Tree Augmentation, improving on a recent relative greedy based method with approximation factor $1 + \ln 2 + \epsilon \approx 1.69$. Furthermore, we show how our local search technique can be applied to Steiner Tree, leading to an alternative way to obtain the currently best known approximation factor of $\ln 4 + \epsilon$. Contrary to prior methods, our approach is purely combinatorial without the need to solve an LP. Nevertheless, the solution value can still be bounded in terms of the well-known hypergraphic LP, leading to an alternative, and arguably simpler, tech-

nique to bound its integrality gap by $\ln 4$.

Vera Traub, Rico Zenklusen
ETH Zurich
vera.traub@ifor.math.ethz.ch, ricoz@ethz.ch

CP42

Partially Optimal Edge Fault-Tolerant Spanners

Recent work has established that, for every positive integer k, every n-node graph has a $(2k-1)$-spanner with $O(f^{1-1/k}n^{1+1/k})$ edges that is resilient to f edge or vertex faults. For vertex faults, this bound is tight. However, the case of edge faults is not as well understood: the best known lower bound for general k is $\Omega(f^{\frac{1}{2}-\frac{1}{2k}}n^{1+1/k} + fn)$. Our main result is to nearly close this gap with an improved upper bound, thus separating the cases of edge and vertex faults. For odd k, our new upper bound is $O_k(f^{\frac{1}{2}-\frac{1}{2k}}n^{1+1/k} + fn)$, which is tight up to hidden $\text{poly}(k)$ factors. For even k, our new upper bound is $O_k(f^{1/2}n^{1+1/k} + fn)$, which leaves a gap of $\text{poly}(k)\ f^{1/(2k)}$. Our proof is an analysis of the fault-tolerant greedy algorithm, which requires exponential time, but we also show that there is a polynomial-time algorithm which creates edge fault tolerant spanners that are larger only by factors of k.

Greg Bodwin
University of Michigan
bodwin@umich.edu

Michael Dinitz
Johns Hopkins University
mdinitz@cs.jhu.edu

Caleb Robelle
MIT
robelle@mit.edu

CP42

Near-Optimal Spanners for General Graphs in (Nearly) Linear Time

Let $G = (V, E, w)$ be a weighted undirected graph on $|V| = n$ vertices and $|E| = m$ edges, $k \geq 1$ be an integer, and $\epsilon < 1$ be any parameter. We present the following results on fast constructions of spanners with near-optimal sparsity and lightness (normalized notions of size and weight), culminating a long line of work in this area. 1. One can construct $(2k-1)(1+\epsilon)$-spanners for G with near-optimal sparsity $O(n^{1/k} \cdot \log(1/\epsilon)/\epsilon))$: In time $O(m \log(1/\epsilon)/\epsilon))$ in the WORD-RAM model and in time $O(m\alpha(m,n) \cdot \log(1/\epsilon)/\epsilon) + SORT(m))$ in the pointer-machine model; $SORT(m)$ is the time needed to sort m integers. 2. One can construct $(2k-1)(1+\epsilon)$-spanners for G with near-optimal bound $O(n^{1/k} \cdot poly(1/\epsilon))$ on both sparsity and lightness: In time $O(m\alpha(m,n) \cdot poly(1/\epsilon))$ in the WORD-RAM model and in time $O(m\alpha(m,n) \cdot poly(1/\epsilon) + SORT(m))$ in the pointer-machine model. The previous fastest constructions of $(2k-1)(1+\epsilon)$-spanners with near-optimal sparsity incur runtime $O(\min\{m(n^{1+1/k}) + n\log n, k \cdot n^{2+1/k}\})$, regardless of lightness. The greedy spanner for stretch $2k-1$ has sparsity $O(n^{1/k})$, with no ϵ-dependence, but its runtime is

$O(m(n^{1+1/k}+n\log n))$. The previous best lightness bound of any $(2k-1)$-spanner was poor regardless of sparsity and runtime.

Hung Le
University of Massachusetts Amherst
hungle@cs.umass.edu

Shay Solomon
Tel Aviv University
solo.shay@gmail.com

CP42

Greedy Spanners in Euclidean Spaces Admit Sublinear Separators

The greedy spanner in a low dimensional Euclidean space is a fundamental geometric construction that has been extensively studied over three decades. Recently, Eppstein and Khodabandeh showed that the greedy spanner in $\mathbb{R}^2$ admits a sublinear separator. Their technique is inherently planar and is not extensible to higher dimensions. They left showing the existence of a small separator for the greedy spanner in $\mathbb{R}^d$ for any constant $d \geq 3$ as an open problem. In this paper, we resolve the problem of Eppstein and Khodabandeh by showing that any subgraph of k vertices of the greedy spanner in $\mathbb{R}^d$ has a separator of size $O(k^{1-1/d})$. We introduce a new technique that gives a simple criterion for any geometric graph to have a sublinear separator that we dub τ-lanky: a geometric graph is τ-lanky if any ball of radius r cuts at most τ edges of length at least r in the graph. We show that any τ-lanky geometric graph of n vertices in $\mathbb{R}^d$ has a separator of size $O(\tau n^{1-1/d})$. We then derive our main result by showing that the greedy spanner is $O(1)$-lanky. Our technique naturally extends to doubling metrics. We use the τ-lanky criterion to show that there exists a $(1 + \epsilon)$-spanner for doubling metrics of dimension d with a constant maximum degree and a separator of size $O(n^{1-\frac{1}{d}})$; this result resolves an open problem posed by Abam and Har-Peled a decade ago.

Hung Le, Cuong Than
University of Massachusetts Amherst
hungle@cs.umass.edu, cthan@umass.edu

CP42

Better Lower Bounds for Shortcut Sets and Additive Spanners via An Improved Alternation Product

We obtain improved lower bounds for additive spanners, additive emulators, and diameter-reducing shortcut sets. Spanners and emulators are sparse graphs that approximately preserve the distances of a given graph. A shortcut set is a set of edges that when added to a directed graph, decreases its diameter. The previous best known lower bounds for these three structures are given by Huang and Pettie [HP18]. For $O(n)$-sized spanners, we improve the lower bound on the additive stretch from $\Omega(n^{1/11})$ to $\Omega(n^{2/21})$. For $O(n)$-sized emulators, we improve the lower bound on the additive stretch from $\Omega(n^{1/18})$ to $\Omega(n^{2/29})$. For $O(m)$-sized shortcut sets, we improve the lower bound

on the graph diameter from $\Omega(n^{1/11})$ to $\Omega(n^{1/8})$. Our key technical contribution, which is the basis of all of our bounds, is an improvement of a graph product known as an alternation product.

Kevin Lu
Citadel Securities
lu.kevin2@gmail.com

Virginia Vassilevska Williams, Nicole Wein, <u>Zixuan Xu</u>
MIT
virgi@mit.edu, nwein@mit.edu, zixuanxu@mit.edu

CP43

Algorithms using Local Graph Features to Predict Epidemics

We study a simple model of epidemics where an infected node transmits the infection to its neighbors independently with probability p. The size of an outbreak in this model is closely related to that of the giant connected component in 'edge percolation', studied for a large class of networks including configuration model and preferential attachment. Even though these models capture the role of super-spreaders in the spread of an epidemic, they only consider graphs that are locally tree like i.e. have a few short cycles. Some generalizations of the configuration model were suggested to capture local communities, known as household models [Ball et al., 2009], or hierarchical configuration model [Hofstad et al., 2015]. Here, we ask a different question: what information is needed for general networks to predict the size of an outbreak? Is it possible to make predictions by accessing the distribution of small subgraphs (or motifs)? We answer the question in the affirmative for large-set expanders with Benjamini-Schramm limits. In particular, we show that there is an algorithm which gives a $(1-\epsilon)$ approximation of the probability and the final size of an outbreak by accessing a constant-size neighborhood of a constant number of nodes chosen uniformly at random. We also present corollaries of the theorem for the preferential attachment model, and study generalizations with household (or motif) structure. The latter was only known for the configuration model.

Yeganeh S. Alimohammadi
Stanford University
yeganeh@stanford.edu

Christian Borgs
UC Berkeley
borgs@berkeley.edu

Amin Saberi
Management Science and Engineering
Stanford University
saberi@stanford.edu

CP43

Spectral Recovery of Binary Censored Block Models

Community detection is the problem of identifying community structure in graphs. Often the graph is modeled as a sample from the Stochastic Block Model, where the probability of creating an edge depends on the communities of vertices. In this paper, we consider a model of censored community detection with two communities, where most of the data is missing as the status of only a small fraction of the edges is revealed. In this model, vertices in the same community are connected with probability p while vertices in opposite communities are connected with probability q. The connectivity status of a given pair of vertices $\{u, v\}$ is revealed with probability α, independently, where $\alpha = t \log(n)/n$. We establish the information-theoretic threshold $t_c(p, q)$, such that no algorithm succeeds in recovering the communities exactly when $t < t_c(p, q)$. We show that when $t > t_c(p, q)$, a simple spectral algorithm based on a weighted, signed adjacency matrix succeeds in recovering the communities exactly. While spectral algorithms are shown to have near-optimal performance in the symmetric case, we show that they may fail in the asymmetric case where the connection probabilities inside the two communities are allowed to be different. In particular, we show the existence of a parameter regime where a simple two-phase algorithm succeeds but any algorithm based on the top two eigenvectors of the weighted, signed adjacency matrix fails.

<u>Souvik Dhara</u>
MIT
sdhara@mit.edu

Julia Gaudio
Northwestern University
julia.gaudio@northwestern.edu

Elchanan Mossel
MIT
elmos@mit.edu

Colin Sandon
Massachusetts Institute of Technology
csandon@mit.edu

CP43

Fast Consensus via the Unconstrained Undecided State Dynamics

We consider the *plurality consensus* problem for n agents. Initially, each agent has one of k opinions. Agents choose random interaction partners and revise their state according to a fixed transition function, depending on their own state and the state of the interaction partners. The goal is to reach a configuration in which all agents agree on the same opinion. If there is initially a sufficiently large bias towards some opinions one of them should prevail. In this paper we consider a synchronized variant of the undecided state dynamics where the agents use so-called phase clocks. The phase clocks divide the time in overlapping phases. Each phase consists of a decision and a boosting part. In the decision part, any agent that encounters an agent with a different opinion becomes undecided. In the boosting part, undecided agents adopt the first opinion they encounter. We consider this dynamics both in the sequential *population model* and the parallel *gossip model*. In the population model agents interact in randomly cho-

sen pairs, one pair per time step. The runtime is measured in *parallel time* (number of interactions divided by n). We show that our protocol reaches consensus (w.h.p.) in $O(\log^2 n)$ parallel time, providing the first polylogarithmic result for $k > 2$ (w.h.p.) in this model. If there is an initial bias of $\Omega(\sqrt{n \log n})$, then (w.h.p.) that opinion wins. [Abstract truncated.]

Dominik Kaaser
Universität Hamburg
dominik.kaaser@uni-hamburg.de

Gregor Bankhamer
University of Salzburg
gbank@cs.sbg.ac.at

Petra Berenbrink, Felix Biermeier
Universität Hamburg
petra.berenbrink@uni-hamburg.de,
felix.biermeier@uni-hamburg.de

Robert Elsässer
University of Salzburg
elsa@cs.sbg.ac.at

Hamed Hosseinpour
Universität Hamburg
hamed.hosseinpour@uni-hamburg.de

Peter Kling
University of Hamburg
peter.kling@uni-hamburg.de

CP43

Co-Evolution of Opinion and Social Tie Dynamics Towards Structural Balance

In this paper, we propose co-evolution models for both dynamics of opinions (people's view on a particular topic) and dynamics of social appraisals (the approval or disapproval towards each other). We propose a co-evolution model, where each vertex i in the network has a current opinion vector v_i and each edge (i, j) has a weight w_{ij} that models the relationship between i, j. The system evolves as opinions and edge weights are updated over time. We are interested in characterizing the long-time behavior of the dynamic model – i.e., whether edge weights evolve to have stable signs (positive or negative) and structural balance (the multiplication of weights on any triangle is non-negative). Our main theoretical result solves the above dynamic system with time-evolving opinions $V(t) = [v_1(t), \cdots, v_n(t)]$ and social tie weights $W(t) = [w_{ij}(t)]_{n \times n}$. For a generic initial opinion vector $V(0)$ and weight matrix $W(0)$, one of the two phenomena must occur at the limit. The first one is that both sign stability and structural balance. The second one is that all the opinions converge to 0. We also performed extensive simulations to examine how different initial conditions affect the network evolution. Of particular interest is that our dynamic model can be used to faithfully detect community structures.

Haotian Wang, Feng Luo, Jie Gao
Rutgers University
hw487@cs.rutgers.edu,   fluo@math.rutgers.edu,
jg1555@rutgers.edu

CP44

Algorithms for Non-Volatile Ram

Guy Blelloch
Computer Science Department
Carnegie Mellon University
guyb@cs.cmu.edu

CP44

Analysis of Work-Stealing and Parallel Cache Complexity

Parallelism has become extremely popular over the past decade, and there have been a lot of new parallel algorithms and software. The randomized work-stealing (RWS) scheduler is crucial in this ecosystem. In this paper, we study two important topics related to the randomized work-stealing scheduler. Our first contribution is a simplified, classroom-ready version of analysis for the RWS scheduler. The theoretical efficiency of the RWS scheduler has been analyzed for a variety of settings, but most of them are quite complicated. In this paper, we show a new analysis, which we believe is easy to understand, and can be especially useful in education. We avoid using the potential function in the analysis, and we assume an asynchronous setting, which is more realistic for today's parallel machines. Our second and main contribution is new parallel cache bounds for algorithms using the RWS scheduler. Although the sequential I/O model has been well-studied over the past decades, so far very few results have extended it to the parallel setting. The parallel cache bounds of many existing algorithms are affected by a polynomial of the span, which causes a significant overhead for high-span algorithms. Our new analysis decouples the span from the analysis of the parallel cache complexity, which allows us to show new parallel cache bounds for a list of classic algorithms. Our results are only a polylogarithmic factor off the lower bounds, and significantly improve previous results.

Yan Gu
University of California, Riverside
yan.gu@ucr.edu

Zachary Napier, Yihan Sun
UC Riverside
znapi001@ucr.edu, yihans@cs.ucr.edu

CP44

Locality-of-Reference Optimality of Cache-Oblivious Algorithms

The program performance on modern hardware is characterized by *locality of reference*, that is, it is faster to access data that is close in address space to data that has been accessed recently than data in a random location. This is due to many architectural features including caches, prefetching, virtual address translation and the physical properties of a hard disk drive; attempting to model all the components that constitute the performance of a modern machine is impossible, especially for general algorithm

design purposes. What if one could prove an algorithm is asymptotically optimal on all systems that reward locality of reference, no matter how it manifests itself within reasonable limits? We show that this is possible, and that excluding some pathological cases, cache-oblivious algorithms that are asymptotically optimal in the ideal-cache model are asymptotically optimal in any reasonable setting that rewards locality of reference. This is surprising as the cache-oblivious framework envisions a particular architectural model involving blocked memory transfer into a multi-level hierarchy of caches of varying sizes, and was not designed to directly model locality-of-reference correlated performance.

John Iacono
Univerité libre de Bruxelles
johniacono@gmail.com

Peyman Afshani
Aarhus University
peyman@cs.au.dk¿

Varunkumar Jayapaul
Université libre de Bruxelles
varunkumarj@gmail.com

Ben Karsin
NVIDIA
bkarsin@gmail.com

Nodari Sitchinava
University of Hawaii at Manoa
nodari@hawaii.edu

CP44
Using External Memory to Improve Cyber-Security Stream Monitoring

Cynthia Phillips
Sandia National Laboratories
caphill@sandia.gov

CP44
Survivability Via Path Disjointness for HyperX

Network survivability has been recognized as an issue of a major importance in terms of security, stability and prosperity. This paper studies fundamental properties of HyperX, an emerging topology for connecting supercomputing and datacenter networks. We focus on the establishment of paths with guaranteed survivability, allowing path availability even upon a restricted number of node failures. We first examine the availability of disjoint paths connecting a pair of input nodes. Disjoint paths guarantee path existence even upon a bounded number of link failures. We explore the inherent tradeoff between allowing slightly longer paths and the ability to extend available sets of mutually disjoint paths. Second, we study the availability of paths in a HyperX topology that already observed link failures. Such failures can increase the length of available paths or even eliminate connectivity between network parts. Last, we provide an evaluation to illustrate the analytical path

availability along with the potential impact of failures.

Ori Rottenstreich
Technion
or@technion.ac.il

CP45
New Trade-Offs for Fully Dynamic Matching via Hierarchical EDCS

We study the maximum matching problem in fully dynamic graphs: a graph is undergoing both edge insertions and deletions, and the goal is to efficiently maintain a large matching after each edge update. This problem has received considerable attention in recent years. The known algorithms naturally exhibit a trade-off between the quality of the matching maintained (i.e., the approximation ratio) and the time needed per update. While several interesting results have been obtained, the optimal behavior of this trade-off remains largely unclear. Our main contribution is a new approach to designing fully dynamic approximate matching algorithms that in a unified manner not only (essentially) recovers all previously known trade-offs that were achieved via very different techniques, but reveals some new ones as well. Our approach is based on a new and of possible independent interest generalization of edge-degree constrained subgraphs (EDCS) that we call hierarchical EDCS (HEDCS).

Soheil Behnezhad
Stanford University
soheil.behnezhad@gmail.com

Sanjeev Khanna
University of Pennsylvania
soheil@cs.umd.edu

CP45
A Near-Optimal Offline Algorithm for Dynamic All-Pairs Shortest Paths in Planar Digraphs

In the planar, dynamic All-Pairs Shortest Paths (APSP) problem, a planar, weighted digraph G undergoes a sequence of edge weight updates and the goal is to maintain a data structure on G, that can quickly answer distance queries between any two vertices $x, y \in V(G)$. The currently best algorithms Fakcharoenphol and Rao [FOCS'01], Klein [SODA'05] for this problem require $\tilde{O}(n^{2/3})$ worst-case update and query time, while conditional lower bounds Abboud and Dahlgaard [FOCS'16] show that either update or query time $\tilde{\Omega}(\sqrt{n})$ is needed. In this article, we present the first algorithm with near-optimal $\tilde{O}(\sqrt{n})$ worst-case update and query time for the offline setting, where the update sequence is given initially. This result is obtained by giving the first offline dynamic algorithm for maintaining dense distance graphs (DDGs) faster than recomputing from scratch after each update. Further, we also present an online algorithm for the incremental APSP problem with $\tilde{O}(\sqrt{n})$ worst-case update/ query time. This allows us to reduce the online dynamic APSP problem to the online decremental APSP problem, which constitutes partial progress even for the online version of this notorious

problem.

Debarati Das
BARC, University of Copenhagen
debaratix710@gmail.com

Maximilian Probst Gutenberg
ETH Zurich
maximilian.probst@outlook.com

Christian Wulff-Nilsen
Department of Computer Science
University of Copenhagen
koolooz@di.ku.dk

CP45

Dynamic Geometric Set Cover, Revisited

In the dynamic geometric set cover problem, points and ranges may be inserted and deleted, and our goal is to efficiently maintain a set cover solution. We give a plethora of new dynamic geometric set cover data structures in 1D and 2D, which significantly improve and extend the previous results: * The first data structure for $(1+\varepsilon)$-approximate dynamic interval set cover with $O(\log^3 n/\varepsilon)$ amortized update time, improving $O(n^\delta/\varepsilon)$ of Agarwal et al. [SoCG'20]. * A data structure for $O(1)$-approximate dynamic unit-square set cover with $2^{O(\sqrt{\log n})}$ amortized update time, substantially improving $O(n^{1/2+\delta})$ of Agarwal et al. [SoCG'20]. * A data structure for $O(1)$-approximate dynamic square set cover with $O(n^{1/2+\delta})$ randomized amortized update time, improving $O(n^{2/3+\delta})$ of Chan and He [SoCG'21]. * A data structure for $O(1)$-approximate dynamic 2D half-plane set cover with $O(n^{17/23+\delta})$ randomized amortized update time. The previous solution for halfplanes by Chan and He [SoCG'21] only reports the solution size. * The first sublinear results for weighted dynamic geometric set cover. Specifically, we give a data structure for $(3+o(1))$-approximate dynamic weighted interval set cover with $2^{O(\sqrt{\log n \log \log n})}$ amortized update time and a data structure for $O(1)$-approximate dynamic weighted unit-square set cover with $O(n^\delta)$ amortized update time.

Timothy M. Chan
University of Illinois at Urbana-Champaign
tmc@illinois.edu

Qizheng He
Department of Computer Science
University of Illinois at Urbana-Champaign
qizheng6@illinois.edu

Subhash Suri
University of California, Santa Barbara
suri@cs.ucsb.edu

Jie Xue
New York University Shanghai
jiexue@nyu.edu

CP45

Incremental SSSP for Sparse Digraphs Beyond the

Hopset Barrier

Given a directed, weighted graph $G = (V, E)$ undergoing edge insertions, the incremental single-source shortest paths (SSSP) problem asks for the maintenance of approximate distances from a dedicated source s while optimizing the total time required to process the insertion sequence of m edges. For sparse graphs, [Chechik and Zhang, SODA'21] recently presented a deterministic $\tilde{O}(m^{5/3})$ algorithm for this problem, and an adaptive randomized version with run-time $\tilde{O}(m\sqrt{n} + m^{7/5})$. This algorithm is remarkable for two reasons: 1) in very spare graphs it reaches the directed hopset barrier of $\tilde{\Omega}(n^{3/2})$ that applied to all previous approaches for partially-dynamic SSSP and 2) it does not resort to a directed hopset technique itself. We introduce propagation synchronization, a new technique for controlling the error build-up on paths throughout batches of insertions. This leads us to a significant improvement of the approach in [Chechik and Zhang, SODA'21] yielding a deterministic $\tilde{O}(m^{3/2})$ algorithm for the problem. By a very careful combination of our new technique with the sampling approach from [Chechik and Zhang, SODA'21], we further obtain an adaptive randomized algorithm with total update time $\tilde{O}(m^{4/3})$. This is the first partially-dynamic SSSP algorithm in sparse graphs to bypass the notorious directed hopset barrier.

Rasmus Kyng, Simon Meierhans, Maximilian Probst Gutenberg
ETH Zurich
kyng@inf.ethz.ch, mesimon@inf.ethz.ch, maximilian.probst@inf.ethz.ch

CP46

An Improved Algorithm for The k-Dyck Edit Distance Problem

A Dyck sequence is a sequence of opening and closing parentheses (of various types) that is balanced. The Dyck edit distance of a given sequence of parentheses S is the smallest number of edit operations (insertions, deletions, and substitutions) needed to transform S into a Dyck sequence. We consider the threshold Dyck edit distance problem, where the input is a sequence of parentheses S and a positive integer k, and the goal is to compute the Dyck edit distance of S only if the distance is at most k, and otherwise report that the distance is larger than k. Backurs and Onak [PODS'16] showed that the threshold Dyck edit distance problem can be solved in $O(n + k^{16})$ time. In this work, we design new algorithms for the threshold Dyck edit distance problem which costs $O(n + k^{4.782036})$ time with high probability or $O(n + k^{4.853059})$ deterministically. Our algorithms combine several new structural properties of the Dyck edit distance problem, a refined algorithm for fast min-plus matrix product, and a careful modification of ideas used in Valiant's parsing algorithm.

Dvir Fried, Shay Golan
Bar Ilan University
friedvir1@gmail.com, golansh1@cs.biu.ac.il

Tomasz Kociumaka
UC Berkeley
kociumaka@mimuw.edu.pl

Tsvi Kopelowitz
Bar Ilan University, Israel
kopelot@gmail.com

Ely Porat
Bar Ilan University
porately@cs.biu.ac.il

Tatiana Starikovskaya
Ecole Normale Supérieure
Paris
tat.starikovskaya@gmail.com

CP46

On the Fine-Grained Complexity of the Unbounded SubsetSum and the Frobenius Problem

Consider positive integral solutions $x \in \mathbb{Z}_{\geq 0}^{n+1}$ to the equation $a_0 x_0 + \ldots + a_n x_n = t$. In the so called unbounded subset sum problem, the objective is to decide whether such a solution exists, whereas in the Frobenius problem, the objective is to compute the largest t such that there is no such solution. In this paper we study the algorithmic complexity of the unbounded subset sum, the Frobenius problem and a generalization of both problems. More precisely, we study pseudo-polynomial time algorithms with a running time that depends on the smallest number a_0 or respectively the largest number a_n. For the parameter a_0, we show that all considered problems are subquadratically equivalent to $(min, +)$-convolution, a fundamental algorithmic problem from the area of fine-grained complexity. By this equivalence, we obtain hardness results for the considered problems (based on the assumption that an algorithm with a subquadratic running time for $(min, +)$-convolution does not exist) as well as algorithms with improved running time. The proof for the equivalence makes use of structural properties of solutions, a technique that was developed in the area of integer programming. In case of the complexity of the problems parameterized by a_n, we present improved algorithms. For example we give a quasi linear time algorithm for the Frobenius problem as well as a hardness result based on the strong exponential time hypothesis.

Kim-Manuel Klein
Kiel University
kmk@informatik.uni-kiel.de

CP46

Optimal Sorting Circuits for Short Keys

A long-standing open question in the algorithms and complexity literature is whether there exist sorting circuits of size $o(n \log n)$. A recent work by Asharov et al. (SODA'21) showed that if the elements to be sorted have short keys whose length $k = o(\log n)$, then there exist $O(n) \cdot \min(k, \log n)$-sized sorting circuits for k-bit keys, ignoring $poly \log^*$ factors. Interestingly, the recent works by Farhadi et al. (STOC'19) and Asharov et al. also showed that the above result is essentially optimal for every key length k, assuming that the famous Li-Li network coding conjecture holds. Note also that proving any *unconditional* super-linear circuit lower bound for a wide class of prob-

lems is beyond the reach of current techniques. Unfortunately, the approach taken by previous works to achieve optimality in size relies on sacrificing the depth. Asharov et al. phrase it as an open question how to achieve optimality both in size and depth. In this paper, we close this gap and construct a sorting circuit of size $O(n) \cdot \min(k, \log n)$ (ignoring $poly \log^*$ terms) and depth $O(\log n)$. To achieve this, our approach departs significantly from the prior works. Our result can be viewed as a generalization of the landmark result by Ajtai, Komlós, and Szemerédi (STOC'83), simultaneously in terms of size and depth. Specifically, for $k = o(\log n)$, we achieve asymptotical improvements in size over the AKS sorting circuit, while preserving optimality in depth.

Wei-Kai Lin
Cornell University
wklin@cs.cornell.edu

Elaine Shi
Carnegie Mellon University
runting@cs.cmu.edu

CP46

Friendly Cut Sparsifiers and Faster Gomory-Hu Trees

We devise new cut sparsifiers that are related to the classical sparsification of Nagamochi and Ibaraki [Algorithmica, 1992], which is an algorithm that, given an unweighted graph G on n nodes and a parameter k, computes a subgraph with $O(nk)$ edges that preserves all cuts of value up to k. We put forward the notion of a friendly cut sparsifier, which is a minor of G that preserves all friendly cuts of value up to k, where a cut in G is called friendly if every node has more edges connecting it to its own side of the cut than to the other side. We present an algorithm that, given a simple graph G, computes in almost-linear time a friendly cut sparsifier with $\tilde{O}(n\sqrt{k})$ edges. Plugging these sparsifiers into the recent $n^{2+o(1)}$-time algorithms for constructing a Gomory-Hu tree of simple graphs, along with a relatively simple procedure for handling the unfriendly minimum cuts, we improve the running time for moderately dense graphs (e.g., with $m = n^{1.75}$ edges). In particular, assuming a linear-time Max-Flow algorithm, the new state-of-the-art for Gomory-Hu tree is the minimum between our $(m + n^{1.75})^{1+o(1)}$ and the known $mn^{1/2+o(1)}$. Finally, assuming an $\tilde{O}(n)$-edge sparsifier that preserves all friendly minimum st-cuts can be computed efficiently, we improve the bound to $\tilde{O}(m + n^{1.5})$ which is the best possible without breaking the cubic barrier for constructing Gomory-Hu trees in non-simple graphs.

Ohad Trabelsi
University of Michigan
ohadt@umich.edu

Robert Krauthgamer
Weizmann Institute of Science
robert.krauthgamer@weizmann.ac.il

Amir Abboud
Weizmann Institute of Science
Israel

amir.abboud@weizmann.ac.il

CP47

On Mixing of Markov Chains: Coupling, Spectral Independence, and Entropy Factorization

For general spin systems, we prove that a contractive coupling for an arbitrary local Markov chain implies optimal bounds on the mixing time and the modified log-Sobolev constant for a large class of Markov chains including the Glauber dynamics, arbitrary heat-bath block dynamics, and the Swendsen-Wang dynamics. This reveals a novel connection between probabilistic techniques for bounding the convergence to stationarity and analytic tools for analyzing the decay of relative entropy. As a corollary of our general results, we obtain $O(n \log n)$ mixing time and $\Omega(1/n)$ modified log-Sobolev constant of the Glauber dynamics for sampling random q-colorings of an n-vertex graph with constant maximum degree Δ when $q > (11/6 - \varepsilon_0)\Delta$ for some fixed $\varepsilon_0 > 0$. We also obtain $O(\log n)$ mixing time and $\Omega(1)$ modified log-Sobolev constant of the Swendsen-Wang dynamics for the ferromagnetic Ising model on an n-vertex graph of constant maximum degree when the parameters of the system lie in the tree uniqueness region. At the heart of our results are new techniques for establishing spectral independence of the spin system and block factorization of the relative entropy.

Antonio Blanca
Pennsylvania State University
ablanca@cse.psu.edu

Pietro Caputo
University of Rome Tre
pietro.caputo@uniroma3.it

Zongchen Chen
Georgia Institute of Technology
zongchen@mit.edu

Daniel Parisi
University of Rome Tre
daniel.parisi@uniroma3.it

Daniel Stefankovic
Department of Computer Science
University of Rochester
stefanko@cs.rochester.edu

Eric Vigoda
University of California Santa Barbara
vigoda@ucsb.edu

CP47

Cut Sparsification of the Clique Beyond the Ramanujan Bound: A Separation of Cut Versus Spectral Sparsification

We prove that a random d-regular graph, with high probability, is a cut sparsifier of the clique with approximation error at most $\left(2\sqrt{\frac{2}{\pi}} + o_{n,d}(1)\right)/\sqrt{d}$, where $2\sqrt{\frac{2}{\pi}} = 1.595\ldots$

and $o_{n,d}(1)$ denotes an error term that depends on n and d and goes to zero if we first take the limit $n \to \infty$ and then the limit $d \to \infty$. This is established by analyzing linear-size cuts using techniques of Jagannath and Sen [jagannath2017unbalanced] derived from ideas in statistical physics, and analyzing small cuts via martingale inequalities. We also prove new lower bounds on spectral sparsification of the clique. If G is a spectral sparsifier of the clique and G has average degree d, we prove that the approximation error is at least the "Ramanujan bound' $(2 - o_{n,d}(1))/\sqrt{d}$, which is met by d-regular Ramanujan graphs, provided that either the weighted adjacency matrix of G is a (multiple of) a doubly stochastic matrix, or that G satisfies a certain high "odd pseudo-girth' property. The first case can be seen as an "Alon-Boppana theorem for symmetric doubly stochastic matrices,' showing that a symmetric doubly stochastic matrix with dn non-zero entries has a non-trivial eigenvalue of magnitude at least $(2 - o_{n,d}(1))/\sqrt{d}$; the second case generalizes a lower bound of Srivastava and Trevisan [**?**], which requires a large girth assumption. Together, these results imply a separation between spectral sparsification and cut sparsification. If G is a random $\log n$-regular graph on n vertices (this is to ensure that G, and consequently any d-regular subgraph, has high pseudogirth), we show that, with high probability, G admits a (weighted subgraph) cut sparsifier of average degree d and approximation error at most $(1.595\ldots + o_{n,d}(1))/\sqrt{d}$, while every (weighted subgraph) spectral sparsifier of G having average degree d has approximation error at least $(2 - o_{n,d}(1))/\sqrt{d}$.

Luca Trevisan
Bocconi University
l.trevisan@unibocconi.it

Antares Chen
University of Chicago
antares21chen@gmail.com

Jonathan Shi
Bocconi University
jonathan.shi@unibocconi.it

CP47

Scalar and Matrix Chernoff Bounds from ℓ_∞-Independence

We present new scalar and matrix Chernoff-style concentration bounds for a broad class of probability distributions over the binary hypercube $\{0,1\}^n$. Motivated by recent tools developed for the study of mixing times of Markov chains on discrete distributions, we say that a distribution is ℓ_∞-independent when the infinity norm of its influence matrix $\mathcal{I}$ is bounded by a constant. We show that any distribution which is ℓ_∞-infinity independent satisfies a matrix Chernoff bound that matches the matrix Chernoff bound for independent random variables due to Tropp. Our matrix Chernoff bound is a broad generalization and strengthening of the matrix Chernoff bound of Kyng and Song (FOCS'18). Using our bound, we can conclude as a corollary that a union of $O(\log |V|)$ random spanning trees gives a spectral graph sparsifier of a graph with $|V|$ vertices with high probability, matching results for independent edge sampling, and matching lower bounds from

Kyng and Song.

Tail Kaufman
Bar-Ilan University
kaufmant@mit.edu

Rasmus Kyng, Federico Soldà
ETH Zurich
kyng@inf.ethz.ch, federico.solda@inf.ethz.ch

CP47

Compact Redistricting Plans Have Many Spanning Trees

In the design and analysis of political redistricting maps, it is often useful to be able to sample from the space of all partitions of the graph of census blocks into connected subgraphs of equal population. There are influential Markov chain Monte Carlo methods for doing so that are based on sampling and splitting random spanning trees. Empirical evidence suggests that the distributions such algorithms sample from place higher weight on more "compact" redistricting plans, which is a practically useful and desirable property. In this paper, we confirm these observations analytically, establishing an inverse exponential relationship between the total length of the boundaries separating districts and the probability that such a map will be sampled. This result provides theoretical underpinnings for algorithms that are already making a significant real-world impact.

Ariel Procaccia, Jamie Tucker-Foltz
Harvard University
arielpro@gmail.com, jtuckerfoltz@gmail.com