



Searchable Abstracts Document

ACM-SIAM Symposium on Discrete Algorithms (SODA23)

January 22–25, 2023

Symposium on Algorithm Engineering and Experiments (ALENEX23)

January 22–23, 2023

Symposium on Simplicity in Algorithms (SOSA23)

January 23–25, 2023

Symposium on Algorithmic Principles of Computer Systems (APOCS23)

January 25, 2023

This document was current as of January 8, 2023. Abstracts appear as submitted.



3600 Market Street, 6th Floor

Philadelphia, PA 19104-2688 U.S.

Telephone: 800-447-7426 (U.S. & Canada) +1-215-382-9800 (Worldwide)

meetings@siam.org

IP1**Games for Fair Allocation**

We consider allocation of m indivisible items to n agents, in settings without payments. For this purpose, we design allocation games in which agents have safe strategies, where a strategy is safe if any agent that uses it in the allocation game gets a bundle of items that meets, or approximately meets, some fairness benchmark, regardless of strategies used by other agents. The fairness benchmarks that we consider are mostly share-based benchmarks, such as the maximin share (MMS) and the anyprice share (APS), though we shall also touch upon envy-based benchmarks, such as envy-free up to one item (EF1). Using these games, for some classes of valuation functions and for some values of r , we establish the existence of allocations in which every agent gets a bundle that offers an r -approximation to the associated benchmark. In some cases, the value of r guaranteed by our allocation games is the best currently known, even compared to general allocation mechanisms that receive as input a full description of the true valuation functions of all agents. We shall also present some intriguing open questions in this area.

Uriel Feige

Weizmann Institute and Microsoft Research, Israel
uriel.feige@weizmann.ac.il

IP2**New Approximation Algorithms for Graph Crossing Number**

In the classical Graph Crossing Number problem, the input is an n -vertex graph G , and the goal is to draw G in the plane while minimizing the number of crossings between the images of its edges. This is a fundamental and extensively studied problem, with connections to many different areas, whose approximability status remains widely open. In all currently known approximation algorithms, the approximation factor depends polynomially on Δ – the maximum vertex degree in G . Even for the special case where maximum vertex degree is bounded by a constant, until recently, the best approximation algorithm achieved a factor- $O(\sqrt{n})$ approximation, while the best current negative results do not rule out a constant-factor approximation. In this talk we survey several recent results and techniques that led to the first subpolynomial-factor approximation algorithm for the Graph Crossing Number problem in low-degree graphs.

Julia Chuzhoy

Toyota Technological Institute at Chicago
jchulia@ttic.edu

IP3**Almost Linear Time Algorithms for All Flows: Success of the Laplacian Paradigm**

Over the last decade or so, combining methods from continuous optimization and analysis with graph theoretic insights has led to a revolution in algorithms for classic problems on graphs such as maximum flow. This has often been referred to as the Laplacian paradigm, a hat-tip to the pioneering work of Spielman and Teng for solving Laplacian linear systems in almost-linear time. In this talk, I will present some of our key ideas behind our recent work that gives almost-linear time algorithms for solving all convex flow problems on graphs, including maximum-flow, and connect them in retrospect to some of key themes devel-

oped as part of the Laplacian paradigm.

Sushant Sachdeva

University of Toronto
sachdeva@cs.toronto.edu

IP4**Explainable Clustering and Noisy Labels**

In this talk, we explore novel aspects of k -median and k -means clustering, which are central optimization problems with many applications in data analysis and machine learning. We focus on what makes a clustering explainable and on how to use side information, such as noisy labels, in the design of algorithms. We present simple algorithms that have strong theoretical guarantees for both of these settings. We complement these findings with several interesting related directions and open questions.

Ola Svensson

EPFL
ola.svensson@epfl.ch

CP1**Dynamic Algorithms for Maximum Matching Size**

We study fully dynamic algorithms for maximum matching. This is a well-studied problem, known to admit several update-time/approximation trade-offs. For instance, it is known how to maintain a $1/2$ -approximate matching in $(polylog n)$ update time or a $2/3$ -approximate matching in $O(\sqrt{n})$ update time, where n is the number of vertices. It has been a long-standing open problem to determine whether either of these bounds can be improved. In this paper, we show that when the goal is to maintain just the size of the matching (and not its edge-set), then these bounds can indeed be improved. First, we give an algorithm that takes $(polylog n)$ update-time and maintains a .501-approximation (.585-approximation if the graph is bipartite). Second, we give an algorithm that maintains a $(2/3 + \Omega(1))$ -approximation in $O(\sqrt{n})$ time for bipartite graphs. Our results build on new connections to sublinear time algorithms. In particular, a key tool for both is an algorithm of the author for estimating the size of maximal matchings in $\tilde{O}(n)$ time [Behnezhad; FOCS 2021]. Our second result also builds on the edge-degree constrained subgraph (EDCS) of Bernstein and Stein [ICALP'15, SODA'16]. In particular, while it has been known that EDCS may not include a better than $2/3$ -approximation, we give a new characterization of such tight instances which allows us to break it. We believe this characterization might be of independent interest.

Soheil Behnezhad

Stanford University
soheil.behnezhad@gmail.com

CP1**Fully Dynamic Exact Edge Connectivity in Sublinear Time**

Given a simple n -vertex, m -edge graph G undergoing edge insertions and deletions, we give two new fully dynamic algorithms for exactly maintaining the edge connectivity of G in $\tilde{O}(n)$ and $\tilde{O}(m^{1-1/16})$ update time, respectively. Prior to our work, all dynamic edge connectivity algorithms either guaranteed an approximate solution or were restricted to edge insertions only. Our results answer in the affirma-

tive an open question posed by Thorup [Combinatorica'07].

Gramoz Goranci
ETH Zurich
gramoz.goranci@gmail.com

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

Danupon Nanongkai
KTH Royal Institute of Technology, Sweden
danupon@gmail.com

Thatchaphol Saranurak
University of Michigan
thsa@umich.edu

Mikkel Thorup
University of Copenhagen
mikkel2thorup@gmail.com

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

CP1

Maintaining Expander Decompositions via Sparse Cuts

In this article, we show that the algorithm of maintaining expander decompositions in graphs undergoing edge deletions directly by removing sparse cuts repeatedly can be made efficient. Formally, for an m -edge undirected graph G , we say a cut (S, \bar{S}) is ϕ -sparse if $|E_G(S, \bar{S})| < \phi \cdot \min\{\text{vol}_G(S), \text{vol}_G(\bar{S})\}$. A ϕ -expander decomposition of G is a partition of V into sets X_1, X_2, \dots, X_k such that each cluster $G[X_i]$ contains no ϕ -sparse cut (meaning it is a ϕ -expander) with $\tilde{O}(\phi m)$ edges crossing between clusters. A natural way to compute a ϕ -expander decomposition is to decompose clusters by ϕ -sparse cuts until no such cut is contained in any cluster. We show that even in graphs undergoing edge deletions, a slight relaxation of this meta-algorithm can be implemented efficiently with amortized update time $m^{o(1)}/\phi^2$. Our approach naturally extends to maintaining directed ϕ -expander decompositions and ϕ -expander hierarchies and thus gives a unifying framework while having simpler proofs than previous state-of-the-art work. In all settings, our algorithm matches the run-times of previous algorithms up to subpolynomial factors. Moreover, our algorithm provides stronger guarantees for ϕ -expander decompositions.

Yiding Hua
ETH Zürich
yidhua@student.ethz.ch

Rasmus Kyng
ETH Zurich
kyng@inf.ethz.ch

Maximilian Probst Gutenberg, Zihang Wu
ETH Zürich

maxprobst@ethz.ch, zihangwu98@gmail.com

CP1

Dynamic Matching with Better-Than-2 Approximation in Polylogarithmic Update Time

We present dynamic algorithms with polylogarithmic update time for estimating the size of the maximum matching of a graph undergoing edge insertions and deletions with approximation ratio strictly better than 2. Specifically, we obtain a $1 + \frac{1}{\sqrt{2}} + \epsilon \approx 1.707 + \epsilon$ approximation in bipartite graphs and a $1.973 + \epsilon$ approximation in general graphs. We thus answer in the affirmative the value version of the major open question repeatedly asked in the dynamic graph algorithms literature. Our randomized algorithms' approximation and worst-case update time bounds both hold w.h.p. against adaptive adversaries. Our algorithms are based on simulating new two-pass streaming matching algorithms in the dynamic setting. Our key new idea is to invoke the recent sublinear-time matching algorithm of Behnezhad (FOCS'21) in a white-box manner to efficiently simulate the second pass of our streaming algorithms, while bypassing the well-known vertex-update barrier.

Sayan Bhattacharya, Peter Kiss
University of Warwick
jucse.sayan@gmail.com, peter.kiss.maths@gmail.com

Thatchaphol Saranurak
University of Michigan
thsa@umich.edu

David Wajc
Google Research
david.wajc@gmail.com

CP1

Dynamic Algorithms for Packing-Covering LPS via Multiplicative Weight Updates

In the dynamic linear program (LP) problem, we are given an LP undergoing updates and we need to maintain an approximately optimal solution. Recently, significant attention (e.g. [Gupta et al. STOC'17; Arar et al. ICALP'18, Wajc STOC'20]) has been devoted to the study of special cases of dynamic packing and covering LPs, such as the dynamic fractional matching and set cover problems. But until now, there is no non-trivial dynamic algorithm for general packing and covering LPs. In this paper, we settle the complexity of dynamic packing and covering LPs, up to a polylogarithmic factor in update time. More precisely, in the partially dynamic setting (where updates can either only relax or only restrict the feasible region), we give near-optimal deterministic ϵ -approximation algorithms with polylogarithmic amortized update time. Then, we show that both partially dynamic updates and amortized update time are necessary; without any of these conditions, the trivial algorithm that recomputes the solution from scratch after every update is essentially the best possible, assuming SETH. To obtain our results, we initiate a systematic study of the multiplicative weights update method in the dynamic setting. As by-products, we also obtain the first online $(1 + \epsilon)$ -competitive algorithms for both covering and packing LPs with polylogarithmic recourse, and the first streaming algorithms for covering and packing LPs with linear space and polylogarithmic passes.

Sayan Bhattacharya, Peter Kiss

University of Warwick
 jucse.sayan@gmail.com, peter.kiss.maths@gmail.com

Thatchaphol Saranurak
 University of Michigan
 thsa@umich.edu

CP1

Faster Deterministic Worst-Case Fully Dynamic All-Pairs Shortest Paths via Decremental Hop-Restricted Shortest Paths

Dynamic all-pairs shortest paths is a well-studied problem in the field of dynamic graph algorithms. More specifically, given a directed weighted graph $G = (V, E, \omega)$ on n vertices which undergoes a sequence of vertex or edge updates, the goal is to maintain distances between any pair of vertices in V . In a classical work by [Demetrescu and Italiano, 2004], the authors showed that all-pairs shortest paths can be maintained deterministically in amortized $\tilde{O}(n^2)$ time, which is nearly optimal. For worst-case update time guarantees, so far the best randomized algorithm has $\tilde{O}(n^{3-1/3})$ time [Abraham, Chechik, Krinninger, 2017], and the best deterministic algorithm needs $\tilde{O}(n^{3-2/7})$ time [Probst Gutenberg, Wulff-Nilsen, 2020]. We provide a faster deterministic worst-case update time of $\tilde{O}(n^{3-20/61})$ for fully dynamic all-pairs shortest paths. To achieve this improvement, we study a natural variant of this problem where a hop constraint is imposed on shortest paths between vertices; that is, given a parameter h , the h -hop shortest path between any pair of vertices $s, t \in V$ is a path from s to t with at most h edges whose total weight is minimized. As a result which might be of independent interest, we give a deterministic algorithm that maintains all-pairs h -hop shortest paths under vertex deletions in total update time $\tilde{O}(n^3 h + Kn^2 h^2)$, where K bounds the total number of vertex deletions.

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

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

CP2

Fast Discrepancy Minimization with Hereditary Guarantees

Efficiently computing low discrepancy colorings of various set systems, has been studied extensively since the breakthrough work by Bansal (FOCS 2010), who gave the first polynomial time algorithms for several important settings, including for general set systems, sparse set systems and for set systems with bounded hereditary discrepancy. The hereditary discrepancy of a set system, is the maximum discrepancy over all set systems obtainable by deleting a subset of the ground elements. While being polynomial time, Bansal's algorithms were not practical, with e.g. his algorithm for the hereditary setup running in time $\Omega(mn^{4.5})$ for set systems with m sets over a ground set of n elements. More efficient algorithms have since then been developed for general and sparse set systems, however, for the hereditary case, Bansal's algorithm remains state-of-the-art. In this work, we give a significantly faster algorithm with hereditary guarantees, running in $O(mn^2 \lg(2 + m/n) + n^3)$ time. Our algorithm is based on new structural insights

into set systems with bounded hereditary discrepancy. We also implement our algorithm and show experimentally that it computes colorings that are significantly better than random and finishes in a reasonable amount of time, even on set systems with thousands of sets over a ground set of thousands of elements.

Kasper Green Larsen
 Aarhus University
 larsen@cs.au.dk

CP2

Near-Linear Time Approximations for Cut Problems via Fair Cuts

We introduce the notion of *fair cuts* as an approach to leverage approximate (s, t) -mincut (equivalently (s, t) -maxflow) algorithms in undirected graphs to obtain near-linear time approximation algorithms for several cut problems. Informally, for any $\alpha \geq 1$, an α -fair (s, t) -cut is an (s, t) -cut such that there exists an (s, t) -flow that uses $1/\alpha$ fraction of the capacity of *every* edge in the cut. (So, any α -fair cut is also an α -approximate mincut, but not vice-versa.) We give an algorithm for $(1 + \epsilon)$ -fair (s, t) -cut in $\tilde{O}(m)$ -time, thereby matching the best run-time for $(1 + \epsilon)$ -approximate (s, t) -mincut [Peng, SODA '16]. We then demonstrate the power of this approach by showing that this result almost immediately leads to several applications: 1. the first nearly-linear time $(1 + \epsilon)$ -approximation algorithm that computes all-pairs maxflow values (by constructing an approximate Gomory-Hu tree). Prior to our work, such a result was not known even for the special case of Steiner mincut; 2. the first almost-linear-work subpolynomial-depth parallel algorithms for computing $(1 + \epsilon)$ -approximations for all-pairs maxflow values (again via an approximate Gomory-Hu tree) in unweighted graphs; 3. the first near-linear time expander decomposition algorithm that works even when the expansion parameter is polynomially small.

Jason Li
 Simons Institute, UC Berkeley
 jmli@alumni.cmu.edu

Danupon Nanongkai
 KTH Royal Institute of Technology, Sweden
 danupon@gmail.com

Debmalya Panigrahi
 Duke University
 debmalya@cs.duke.edu

Thatchaphol Saranurak
 University of Michigan
 thsa@umich.edu

CP2

A Tight Quasi-Polynomial Bound for Global Label Min-Cut

We study a generalization of the classic Global Min-Cut problem, called Global Label Min-Cut: the edges of the input (multi)graph are labeled, and removing all edges of the same label costs one. The problem asks to disconnect the graph at minimum cost. While the st -cut version of the problem is known to be NP-hard, the above global cut version is known to admit a quasi-polynomial randomized $n^{O(\log \text{OPT})}$ -time algorithm due to Ghaffari, Karger

and Panigrahi [SODA 2017]. They consider this as "strong evidence that this problem is in P". We show that this is actually not the case. We complete the study of the complexity of Global Label Min-Cut by showing that the quasi-polynomial running time is probably optimal: We show that the existence of a $(np)^{o(\log n / (\log \log n)^2)}$ -time algorithm would contradict the ETH, where n is the number of vertices, and p is the number of labels in the input. The key step for the lower bound is a proof that Global Label Min-Cut is W[1]-hard when parameterized by the number of uncut labels. To turn this lower bound into a quasi-polynomial-time lower bound, we revisit the framework due to Marx [Theory Comput. 2010] of proving lower bounds assuming ETH through the Subgraph Isomorphism problem parameterized by the number of edges of the pattern. Here, we provide an alternative simplified proof of the hardness of this problem that is more versatile with respect to the choice of the regimes of the parameters.

Paloma Lima

IT University of Copenhagen
palt@itu.dk

Lars Jaffke

University of Bergen, Norway
lars.jaffke@uib.no

Tomáš Masarík, Marcin Pilipczuk

University of Warsaw
masarik@mimuw.edu.pl, m.pilipczuk@mimuw.edu.pl

Uéverton Souza

Universidade Federal Fluminense
University of Warsaw
ueverton@ic.uff.br

CP2

Maximal k -Edge-Connected Subgraphs in Weighted Graphs via Local Random Contraction

The *maximal k -edge-connected subgraphs* problem is a classical graph clustering problem studied since the 70's. Surprisingly, no non-trivial technique for this problem in weighted graphs is known: a very straightforward recursive-min-cut algorithm with $\Omega(mn)$ time has remained the fastest algorithm until now. All previous progress gives a speed-up only when the graph is unweighted, and k is small enough (e.g. Henzinger et al. (ICALP'15), Chechik et al. (SODA'17), and Forster et al. (SODA'20)). We give the first algorithm that breaks through the long-standing $\tilde{O}(mn)$ -time barrier in *weighted undirected* graphs. More specifically, we show a maximal k -edge-connected subgraphs algorithm that takes only $\tilde{O}(m \cdot \min\{m^{3/4}, n^{4/5}\})$ time. As an immediate application, we can $(1+\epsilon)$ -approximate the *strength* of all edges in undirected graphs in the same running time. Our key technique is the first local cut algorithm with *exact* cut-value guarantees whose running time depends only on the output size. All previous local cut algorithms either have running time depending on the cut value of the output, which can be arbitrarily slow in weighted graphs or have approximate cut guarantees.

Chaitanya Nalam

University of Michigan, Ann Arbor, MI 48105
nalamsai@umich.edu

Thatchaphol Saranurak

University of Michigan

thsa@umich.edu

CP2

Faster and Unified Algorithms for Diameter Reducing Shortcuts and Minimum Chain Cover

For an n -vertex m -edge digraph G , a D -*shortcut* is a small set H of directed edges taken from the transitive closure of G , satisfying that the diameter of $G \cup H$ is at most D . In this paper, we present faster and unified shortcut algorithms for general digraphs. We show: - A unified and faster shortcutting algorithm which implements the [KP'22] framework in an almost *optimal* time, conditioned on the Boolean Matrix Multiplication (BMM) conjecture. - An improved algorithm for computing the Minimum Chain Cover (MCC) of DAGs. For an n -vertex m -edge DAG G of width k , the algorithm runs in $\tilde{O}(\sqrt{k} \cdot n + m^{1+o(1)})$ time. For sparse digraphs, there are faster $\tilde{O}(k^{1/3} \cdot n + n^{1+o(1)})$ -time algorithms. This improves the time bounds of $\tilde{O}(n^{3/2} + m)$ [KP, ICALP'22] and $\tilde{O}(k^2 \cdot n + m)$ [Caceres et al., SODA 2022]. - An MCC-based shortcut algorithm for DAGs with improved size and time bounds, as a function of the width k . For example, providing a linear-size \sqrt{k} -shortcut in time $\tilde{O}(\min\{\sqrt{k} \cdot m + m^{1+o(1)}, n^2\})$, improving the general graph's size and time bounds for $k = o(n^{2/3})$.

Merav Parter, Shimon Kogan

Weizmann Institute

merav.parter@weizmann.ac.il,

shimon.kogan@weizmann.ac.il

shi-

CP2

Closing the Gap Between Directed Hopsets and Shortcut Sets

For an n -vertex directed graph $G = (V, E)$, a β -shortcut set H is a set of additional edges $H \subseteq V \times V$ such that $G \cup H$ has the same transitive closure as G , and for every pair $u, v \in V$, there is a uv -path in $G \cup H$ with at most β edges. A natural generalization of shortcut sets to distances is a (β, ϵ) -hopset $H \subseteq V \times V$, where H and $G \cup H$ have the same shortest-path distances, and for every $u, v \in V$, there is a $(1+\epsilon)$ -approximate shortest path in $G \cup H$ with at most β edges. We consider the tradeoff between the optimal size of a shortcut set / hopset and the value of β . In particular we highlight a natural point on this tradeoff: what is the minimum value of β , such that for any graph G , there exists a β -shortcut set H with $O(n)$ edges? What is the minimum value of β such that there exists a (β, ϵ) -hopset with $O(n)$ edges? Until very recently the best known upper bound was a folklore construction showing $\beta = O(n^{1/2})$, but in a breakthrough result Kogan and Parter [SODA 2022] improve this to $\beta = \tilde{O}(n^{1/3})$ for shortcut sets and $\tilde{O}(n^{2/5})$ for hopsets. Our result is to close the gap between shortcut sets and hopsets introduced by the result of Kogan and Parter. That is, we show that for any graph G and any fixed ϵ there is a $(\tilde{O}(n^{1/3}), \epsilon)$ hopset with $O(n)$ edges.

Aaron Bernstein

Rutgers University

bernstei@gmail.com

Nicole Wein

MIT

nicole.s.wein@gmail.com

CP3

Interdependent Public Projects

In the interdependent values (IDV) model introduced by Milgrom and Weber [1982], agents have private signals that capture their information about different social alternatives, and the valuation of every agent is a function of all agent signals. While interdependence has been mainly studied for auctions, it is very relevant for a large variety of social choice settings, including the canonical and practically important setting of public projects. Welfare guarantees for IDV were achieved mainly through two conditions known as single-crossing and submodularity over signals (SOS). In either case, the existing theory falls short of solving the public projects setting. Our contribution is twofold: (i) We give a useful characterization of truthfulness for IDV public projects, parallel to the known characterization for independent private values, and identify the domain frontier for which this characterization applies; (ii) Using this, we provide possibility and impossibility results for welfare approximation in public projects with SOS valuations. Our main impossibility result is that, in contrast to auctions, no universally truthful mechanism performs better for public projects with SOS than choosing a project at random. Our main positive result applies to excludable public projects with SOS, for which we establish a constant factor approximation similar to auctions. Our results suggest that exclusion may be a key tool for achieving welfare guarantees in the IDV model.

Avi Cohen

StarkWare Industries Ltd.
chn.avi@gmail.com

Michal Feldman
Tel Aviv University
Microsoft Research Israel
mfeldman@tauex.tau.ac.il

Divyarthi Mohan
Tel Aviv University
divyarthim@tau.ac.il

Inbal Talgam-Cohen
Technion
italgam@cs.technion.ac.il

CP3

The Price of Stability for First Price Auction

This paper establishes the Price of Stability (PoS) for First Price Auctions, for all equilibrium concepts that have been studied in the literature: Bayesian Nash Equilibrium \subsetneq Bayesian Correlated Equilibrium \subsetneq Bayesian Coarse Correlated Equilibrium. Bayesian Nash Equilibrium: For independent valuations, the tight PoS is $1 - 1/e^2 \approx 0.8647$, matching the counterpart Price of Anarchy (PoA) bound [JL22]. For correlated valuations, the tight PoS is $1 - 1/e \approx 0.6321$, matching the counterpart PoA bound [ST13, S14]. This result indicates that, in the worst cases, efficiency degradation depends not on different selections among Bayesian Nash Equilibria. Bayesian (Coarse) Correlated Equilibrium: For independent or correlated valuations, the tight PoS is always $1 = 100\%$, i.e., no efficiency degradation. This result indicates that First Price Auctions can be fully efficient when we allow the more general equilibrium

concepts.

Pinyan Lu

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

Yaonan Jin

Columbia University
yj2552@columbia.edu

CP3

Fair Allocation of a Multiset of Indivisible Items

We study the problem of fairly allocating a multiset M of m indivisible items among n agents with additive valuations. Specifically, we introduce a parameter t for the number of distinct types of items and study fair allocations of multisets that contain only items of these t types, under two standard notions of fairness: 1. Envy-freeness (EF): For arbitrary n, t , we show that a complete EF allocation exists when at least one agent has a unique valuation and the number of items of each type exceeds a particular finite threshold. We give explicit upper and lower bounds on this threshold in some special cases. 2. Envy-freeness up to any good (EFX): For arbitrary n, m , and for $t \leq 2$, we show that a complete EFX allocation always exists. We give two different proofs of this result. One proof is constructive and runs in polynomial time; the other is geometrically inspired.

Pranay Gorantla

Princeton University
gorantla@princeton.edu

Kunal Marwaha

Berkeley Center for Quantum Information and Computation
kmarw@uchicago.edu

Santhoshini Velusamy

Harvard University
svelusamy@g.harvard.edu

CP3

Pricing Query Complexity of Revenue Maximization

The common way to optimize auction and pricing systems is to set aside a small fraction of traffic to run experiments. This leads to the question: how can we learn the most with the smallest amount of data? For truthful auctions, this is the sample complexity problem. For posted price auctions, we no longer have access to samples. Instead, the algorithm is allowed to choose a price p_t ; then for a fresh sample $v_t \sim \mathcal{D}$ we learn the sign $s_t = \text{sign}(p_t - v_t) \in \{-1, +1\}$. How many pricing queries are needed to estimate a given parameter of the underlying distribution? We give tight upper and lower bounds on the number of pricing queries required to find an approximately revenue optimal reserve price for general, regular and MHR distributions. Interestingly, for regular distributions, the pricing query and sample complexities match. But for general and MHR distributions, we show a strict separation between them. All known results on sample complexity for revenue optimization follow from a variant of using the optimal reserve price of the empirical distribution. In the pricing query complexity setting, we show that learning the entire distribution within an error of ϵ in Levy distance requires strictly more pricing queries than to estimate the reserve. Instead, our

algorithm uses a new property we identify called relative flatness to quickly zoom into the right region of the distribution to get the optimal results.

Balasubramanian Sivan, Renato Paes Leme, Yifeng Teng, Pratik Worah
Google Research
balusivan@google.com, renatoppl@google.com, yifengt@google.com, pworah@google.com

CP3

Optimal Pricing Schemes for an Impatient Buyer

A patient seller aims to sell a good to an impatient buyer (i.e., one who discounts utility over time). The buyer will remain in the market for a period of time T , and her private value is drawn from a publicly known distribution. What is the revenue-optimal pricing-curve (sequence of (price, time) pairs) for the seller? Is randomization of help here? Is the revenue-optimal pricing curve computable in polynomial time? We answer these questions in this paper. We give an efficient algorithm for computing the revenue-optimal pricing curve. We show that pricing curves, that post a price at each point of time and let the buyer pick her utility maximizing time to buy, are revenue-optimal among a much broader class of sequential lottery mechanisms: namely, mechanisms that allow the seller to post a menu of lotteries at each point of time cannot get any higher revenue than pricing curves. We also show that the even broader class of mechanisms that allow the menu of lotteries to be adaptively set, can earn strictly higher revenue than that of pricing curves, and the revenue gap can be as big as the support size of the buyer's value distribution.

Balasubramanian Sivan, Yuan Deng
Google Research
balusivan@google.com, dengyuan@google.com

Jieming Mao
Google Research New York
maojm@google.com

Kangning Wang
University of California, Berkeley
kangning@berkeley.edu

CP3

Competitive Information Design for Pandoras Box

We study a natural competitive-information-design variant for the Pandoras Box problem [Wei-79], where each box is associated with a strategic information sender who can design what information about the box's prize value to be revealed to the agent when she inspects the box. This variant with strategic boxes is motivated by a wide range of real-world economic applications for Pandora's box. The main contributions of this article are two-fold: (1) we study informational properties of Pandora's Box by analyzing how a box's partial information revelation affects the agent's optimal decisions; and (2) we fully characterize the pure symmetric equilibrium for the boxes' competitive information revelation, and reveal various insights regarding information competition and the resultant agent utility at equilibrium.

Bolin Ding
Alibaba Group
bolin.ding@alibaba-inc.com

Yiding Feng
Northwestern University
Department of Electrical Engineering & Computer Science
yidingfeng2021@u.northwestern.edu

Chien-Ju Ho
Washington University in St. Louis
chienju.ho@wustl.edu

Wei Tang
Columbia University
wt2359@columbia.edu

Haifeng Xu
The University of Chicago
haifengxu@uchicago.edu

CP4

Filtration-Domination in Bifiltered Graphs

Bifiltered graphs are a versatile tool for modelling relations between data points across multiple grades of a two-dimensional scale. They are especially popular in topological data analysis, where the homological properties of the induced clique complexes are studied. To reduce the large size of these clique complexes, we identify filtration-dominated edges of the graph, whose removal preserves the relevant topological properties. We give two algorithms to detect filtration-dominated edges in a bifiltered graph and analyze their complexity. These two algorithms work directly on the bifiltered graph, without first extracting the clique complexes, which are generally much bigger. We present extensive experimental evaluation which shows that in most cases, more than 90% of the edges can be removed. In turn, we demonstrate that this often leads to a substantial speedup, and reduction in the memory usage, of the computational pipeline of multiparameter topological data analysis.

ngel Javier Alonso, Michael Kerber
University of Technology, Graz, Austria
alonsohernandez@tugraz.at, kerber@tugraz.at

Siddharth Pritam
Shiv Nadar University
siddharth.pritam@snu.edu.in

CP4

Lossy Reduction Rules for the Directed Feedback Vertex Set Problem

Given a directed graph $G(V, A)$, the DIRECTED FEEDBACK VERTEX SET (DFVS) problem asks for the smallest sized subset of V whose removal makes G acyclic. The problem is NP-complete and efficient constant-factor approximation algorithms are ruled out under UGC. Attempting to get an exact DFVS in practice usually involves the application of reduction rules that decrease the instance size without compromising the optimal solution. If the reduced graph gets sufficiently small, the respective instance can then be solved to optimality e.g. by a branching algorithm. However, one might need to resort to heuristics in the end in case the reduced instance is still huge. In this paper, we propose novel reduction rules for DFVS with a special focus on lossy rules. Here, the idea is that an optimal solution on the reduced graph combined with the information gained in the reduction process provides an α -approximation for

the original instance. We present several rules that ensure small α , and discuss how to combine and engineer them. We also propose a taxonomy to study general types of lossy rules. In an extensive experimental analysis, we evaluate the impact of exact and lossy rules on the running time, the size of the reduced instance, and the solution quality. It turns out that the lossy rules are indeed very effective and that it is often possible to solve instances by using reduction rules only.

Timon Behr

University of Konstanz
timon.behr@uni-konstanz.de

Sabine Storandt

Universität Konstanz
sabine.storandt@uni-konstanz.de

CP4

A Closer Cut: Computing Near-Optimal Lawn Mowing Tours

For a given polygonal region P , the Lawn Mowing Problem (LMP) asks for a shortest tour T that gets within Euclidean distance 1 of every point in P ; this is equivalent to computing a shortest tour for a unit-disk cutter C that covers all of P . As a geometric optimization problem of natural practical and theoretical importance, the LMP generalizes and combines several notoriously difficult problems, including minimum covering by disks, the Traveling Salesman Problem with neighborhoods (TSPN), and the Art Gallery Problem (AGP). In this paper, we conduct the first study of the Lawn Mowing Problem with a focus on practical computation of near-optimal solutions. We provide new theoretical insights: Optimal solutions are polygonal paths with a bounded number of vertices, allowing a restriction to straight-line solutions; on the other hand, there can be relatively simple instances for which optimal solutions require a large class of irrational coordinates. On the practical side, we present a primal-dual approach with provable convergence properties based on solving a special case of the TSPN restricted to witness sets. In each iteration, this establishes both a valid solution and a valid lower bound, and thereby a bound on the remaining optimality gap. As we demonstrate in an extensive computational study, this allows us to achieve provably optimal and near-optimal solutions for a large spectrum of benchmark instances with up to 2000 vertices.

Sándor P. Fekete

TU Braunschweig
Department of Computer Science
s.fekete@tu-bs.de

Dominik M. Krupke, Michael Perk, Christian Rieck

TU Braunschweig
krupke@ibr.cs.tu-bs.de, perk@ibr.cs.tu-bs.de,
c.rieck@tu-bs.de

Christian Scheffer

Bochum University of Applied Sciences
christian.scheffer@hs-bochum.de

CP4

A Dynamic Maxsat-Based Approach to Directed Feedback Vertex Sets

We propose a new approach to the Directed Feedback Vertex Set Problem (DFVSP), where the input is a directed

graph and the solution is a minimum set of vertices whose removal makes the graph acyclic. Our approach, implemented in the solver DAGger, is based on two novel contributions: Firstly, we add a wide range of data reductions that are partially inspired by reductions for the similar vertex cover problem. For this, we give a theoretical basis for lifting reductions from vertex cover to DFVSP but also incorporate novel ideas into strictly more general and new DFVSP reductions. Secondly, we propose dynamically encoding DFVSP in propositional logic using cycle propagation for improved performance. Cycle propagation builds on the idea that already a limited number of the constraints in a propositional encoding is usually sufficient for finding an optimal solution. Our algorithm, therefore, starts with a small number of constraints and cycle propagation adds additional constraints when necessary. We propose an efficient integration of cycle propagation into the workflow of MaxSAT solvers, further improving the performance of our algorithm. Our extensive experimental evaluation shows that DAGger significantly outperforms the state-of-the-art solvers and our data reductions alone directly solve many of the instances.

Andre Schidler, Rafael Kiesel

TU Wien
andre@schidler.at, rafael.kiesel@tuwien.ac.at

CP4

A Uniform Sampling Procedure for Abstract Triangulations of Surfaces

We present a procedure to sample uniformly from the set of combinatorial isomorphism types of balanced triangulations of surfaces - also known as graph-encoded surfaces. For a given number of triangles n , the sample is a weighted set of graph-encoded surfaces with $2n$ triangles. The sampling procedure relies on connections between graph-encoded surfaces and permutations, and basic properties of the symmetric group. We implement our method and present a number of experimental findings based on the analysis of 138 million runs of our sampling procedure, producing graph-encoded surfaces with up to 280 triangles. Namely, we determine that, for n fixed, the empirical mean genus of our sample is very close to the maximum possible genus minus an explicit smaller order term. Moreover, we present experimental evidence that the associated genus distribution more and more concentrates on a vanishing portion of all possible genera as n tends to infinity. Finally, we observe from our data that the mean number of non-trivial symmetries of a uniformly chosen graph encoding of a surface decays to zero at a rate super-exponential in n .

Rajan Shankar

The University of Sydney
rsha5949@uni.sydney.edu.au

Jonathan Spreer

The University of Queensland
jonathan.spreer@sydney.edu.au

CP5

Stronger 3SUM-Indexing Lower Bounds

The 3SUM-Indexing problem was introduced as a data structure version of the 3SUM problem, with the goal of proving strong conditional lower bounds for static data structures via reductions. Ideally, the conjectured hardness of 3SUM-Indexing should be replaced by an uncondi-

tional lower bound. Unfortunately, we are far from proving this, with the strongest current lower bound being a logarithmic query time lower bound by Golovnev et al. from STOC'20. Moreover, their lower bound holds only for non-adaptive data structures and they explicitly asked for a lower bound for adaptive data structures. Our main contribution is precisely such a lower bound against adaptive data structures. As a secondary result, we also strengthen the non-adaptive lower bound of Golovnev et al. and prove strong lower bounds for 2-bit-probe non-adaptive 3SUM-Indexing data structures via a completely new approach that we find interesting in its own right.

Eldon Chung

Center for Quantum Technologies
echung.math@gmail.com

Kasper Green Larsen
Aarhus University
larsen@cs.au.dk

CP5

Tiny Pointers

This paper introduces a new data-structural object that we call the tiny pointer. We develop a comprehensive theory of tiny pointers, giving optimal constructions, and showing that in many applications, traditional $\log n$ -bit pointers can be replaced with $o(\log n)$ -bit tiny pointers at the cost of only a constant-factor time overhead and a small failure probability. As applications, we show how to obtain new results on five classic data-structure problems: succinct value retrieval, succinct search trees, stable dictionaries, dictionaries with variable-size values, and internal-memory stashes.

Michael A. Bender
Stony Brook University
bender@cs.stonybrook.edu

Alex Conway
VMware Research Group
aconway@vmware.com

Martin Farach-Colton
Rutgers University
martin@farach-colton.com

William Kuszmaul
Massachusetts Institute of Technology
kuszmaul@mit.edu

Guido Tagliavini
Rutgers University
guido.tag@rutgers.edu

CP5

Tight Bounds for Monotone Minimal Perfect Hashing

The monotone minimal perfect hash function (MMPHF) problem is the following indexing problem. Given a set $S = \{s_1, \dots, s_n\}$ of n distinct keys from a universe U of size u , create a data structure DS that answers the following query:

$$\text{RankOp}(q) = \begin{cases} \text{rank of } q \text{ in } S & q \in S \\ \text{arbitrary answer} & \text{otherwise.} \end{cases}$$

Solutions to the MMPHF problem are in widespread use in both theory and practice. The best upper bound known for the problem encodes DS in $O(n \log \log \log u)$ bits and performs queries in $O(\log u)$ time. It has been an open problem to either improve the space upper bound or to show that this somewhat odd looking bound is tight. In this paper, we show the latter: any data structure (deterministic or randomized) for monotone minimal perfect hashing of any collection of n elements from a universe of size u requires $\Omega(n \cdot \log \log \log u)$ expected bits to answer every query correctly. We achieve our lower bound by defining a graph \mathbf{G} where the nodes are the possible $\binom{u}{n}$ inputs and where two nodes are adjacent if they cannot share the same DS. The size of DS is then lower bounded by the log of the chromatic number of \mathbf{G} . Finally, we show that the fractional chromatic number (and hence the chromatic number) of \mathbf{G} is lower bounded by $2^{\Omega(n \log \log \log u)}$.

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

Martin Farach-Colton
Rutgers University
martin@farach-colton.com

William Kuszmaul
Massachusetts Institute of Technology
kuszmaul@mit.edu

CP5

A Nearly-Tight Analysis of Multipass Pairing Heaps

The pairing heap, introduced by Fredman et al., is a self-adjusting heap data structure that is both simple and efficient. A variant introduced in the same paper is the multipass pairing heap. Standard pairing heaps do just two linking passes during delete-min, a pairing pass and an assembly pass. In contrast, multipass pairing heaps do repeated pairing passes, in which nodes are linked in adjacent pairs, until only a minimum-key node remains. We obtain the following amortized time bounds for operations on n -item multipass pairing heaps: $O(\log n)$ for delete-min and delete; $O(\log \log n \log \log \log n)$ for decrease-key; and $O(1)$ for all other heap operations, including insert and meld. This is the first analysis giving an $O(\log n)$ bound for delete-min. Our analysis is tight for all operations except possibly decrease-key, for which Fredman and separately Iacono and Zkan proved an $\Omega(\log \log n)$ lower bound.

Corwin Sinnamon
University of Waterloo, Waterloo, Ontario
sinncore@gmail.com

Robert Tarjan
Princeton University
and Intertrust Technologies
ret@cs.princeton.edu

CP5

A Tight Analysis of Slim Heaps and Smooth Heaps

The smooth heap and the closely related slim heap are recently invented self-adjusting implementations of the heap (priority queue) data structure. They are simple to describe and seem to be efficient in practice. For both heaps,

we give a tight analysis of the amortized time per operation: $O(\log n)$ for delete-min and delete; $O(\log \log n)$ for decrease-key; and $O(1)$ for make-heap, find-min, insert, and meld, where n is the current number of items in the heap. These bounds are tight not only for smooth and slim heaps, but for any heap in Iacono and Ozkan's pure heap model, intended to capture all "self-adjusting" heaps. Slim and smooth heaps are the first known data structures to match Iacono and zkan's lower bounds and to satisfy the constraints of their model.

Corwin Sinnamon

University of Waterloo, Waterloo, Ontario
sinncore@gmail.com

Robert Tarjan
Princeton University
and Intertrust Technologies
ret@cs.princeton.edu

CP5

Improved Pattern-Avoidance Bounds for Greedy BSTs via Matrix Decomposition

Greedy is an online BST defined in the geometric view ([Lucas, 1988; Munro, 2000; Demaine, Harmon, Iacono, Kane, Patrascu, SODA 2009]). Along with Splay trees (Sleator, Tarjan 1985), Greedy is considered the most promising candidate for being dynamically optimal. However, despite having received a lot of attention in the past decades, the question has remained elusive even for highly restricted input. In this paper, we prove new bounds on the cost of Greedy in the "pattern avoidance" regime. Our results include:

- The (preorder) traversal for Greedy holds up to a factor of $O(2^{\alpha(n)})$. This is the best-known bound obtained by any online BSTs.
- We settle the postorder traversal for Greedy.
- The deque conjecture for Greedy holds up to a factor of $O(\alpha(n))$.
- The split conjecture holds for Greedy up to a factor of $O(2^{\alpha(n)})$.

Key to all these results is to partition (based on the input structures) the execution log of Greedy into several simpler-to-analyze subsets for which forbidden submatrix bounds can be leveraged. Finally, we show the applicability of this technique to handle a class of increasingly complex pattern-avoiding input sequences, called k -increasing sequences. As a bonus, we discover a new class of permutation matrices whose extremal bounds are polynomially bounded. This gives partial progress on an open question by Jacob Fox (2013).

Parinya Chalermsook
Aalto University, Espoo, Finland
parinya.chalermsook@aalto.fi

Manoj Gupta
Indian Institute of Technology Gandhinagar
India
gmanoj@iitgn.ac.in

Wanchote Jiamjitrak, Nidia Acosta
Aalto University
wanchote.jiamjitrak@aalto.fi, nidia.obscuraacosta@aalto.fi

Akash Pareek

IIT Gandhinagar
pareek.akash@iitgn.ac.in

Sorrachai Yingchareonthawornchai
Aalto University
sorrachai.cp@gmail.com

CP6

Elliptic Curve Fast Fourier Transform (ecfft) Part I: Low-Degree Extension in Time $O(n \log n)$ over All Finite Fields

Given disjoint sets $S, S' \subseteq \mathbb{F}_q$ of size n and a function $f : S \rightarrow \mathbb{F}_q$, where \mathbb{F}_q is a finite field, the low-degree extension (LDE) of f to S' is the function $f' : S' \rightarrow \mathbb{F}_q$ obtained by restricting the interpolating polynomial of f to S' . LDE computation is a fundamental primitive of modern algebraic coding theory and cryptography. The best asymptotic running time for LDE with parameter n is $O(n \log n)$ arithmetic operations over \mathbb{F}_q , when q and the sets S, S' are special. This running time is achieved via the Fast Fourier Transform (FFT), and requires the field to contain a multiplicative or additive subgroup of smooth order $\geq n$. Most finite fields have no such subgroup, which raises the question of computing the LDE in time $O(n \log n)$ over general finite fields, for some disjoint pair of sets S, S' of size n . The main result of this paper is a positive answer to this question, presenting $O(n \log n)$ -time LDE over all fields, as long as $q = \Omega(n^2)$. This result is achieved by introducing a new FFT-like transform, the Elliptic Curve Fast Fourier Transform (ECFFT), which gives fast algorithms for polynomial operations over all large finite fields, including the above LDE. The key idea is to replace the group of roots of unity with a set of points $L \subset \mathbb{F}_q$ related to a well-chosen elliptic curve group over \mathbb{F}_q .

Eli Ben-Sasson, Dan Carmon
StarkWare Industries
eli@starkware.co, dancar@starkware.co

Swastik Kopparty
University of Toronto
swastik@cs.toronto.edu

David Levit
StarkWare Industries
david@starkware.co

CP6

Short Synchronizing Words for Random Automata

We prove that a uniformly random automaton with n states on a 2-letter alphabet has a synchronizing word of length $O(\sqrt{n} \log n)$ with high probability (whp). That is to say, whp there exists a word ω of such length, and a state v_0 , such that ω sends all states to v_0 . This confirms a conjecture of Kisielewicz, Kowalski, Szykula (2013) based on numerical simulations, up to a log factor – the previous best partial result towards the conjecture was the quasilinear bound $O(n \log^3 n)$ due to Nicaud (2019). Moreover, the synchronizing word ω we obtain has small entropy, in the sense that it can be encoded with only $O(\log(n))$ bits whp. Our proof introduces the concept of w -trees, for a word w , that is, automata in which the w -transitions induce a (loop-rooted) tree. We prove a strong structure result that says that, whp, a random automaton on n states is a w -tree for some word w of length at most $(1 + \epsilon) \log_2(n)$, for any $\epsilon > 0$. The existence of the (random) word w is proved

by the probabilistic method. This structure result is key to proving that a short synchronizing word exists.

Guillaume Chapuy
CNRS, IRIF, Université Paris Diderot
France
guillaume.chapuy@irif.fr

Guillem Perarnau
Universitat Politècnica de Catalunya
guillem.perarnau@upc.edu

CP6

Hierarchies Of Minion Tests For Pcsp Through Tensors

We provide a unified framework to study hierarchies of relaxations for Constraint Satisfaction Problems and their Promise variant. The idea is to split the description of a hierarchy into an algebraic part, depending on a minion capturing the base level of the hierarchy, and a geometric part – which we call tensorisation – inspired by multilinear algebra. We show that the hierarchies of minion tests obtained in this way are general enough to capture the (combinatorial) bounded width and also the Sherali-Adams LP, Sum-of-Squares SDP, and affine IP hierarchies. We exploit the geometry of the tensor spaces arising from our construction to prove general properties of such hierarchies. We identify certain classes of minions, which we call linear and conic, whose corresponding hierarchies have particularly fine features. Finally, in order to analyse the Sum-of-Squares SDP hierarchy we also characterise the solvability of the standard SDP relaxation through a new minion.

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

Stanislav ivn
University of Oxford
standa.zivny@cs.ox.ac.uk

CP6

On Complex Roots of the Independence Polynomial

It is known from the work of Shearer and of Scott and Sokal that the independence polynomial $Z_G(\lambda)$ of a graph G of maximum degree at most $d + 1$ does not vanish provided that $|\lambda| \leq \frac{d^d}{(d+1)^{d+1}}$. Significant extensions of this result have recently been given in the case when λ is in the right half-plane (i.e., when $\Re \lambda \geq 0$) by Peters and Regts, and by Bencs and Csikvári. In this paper, our motivation is to further extend these results to find new zero free regions not only in the right half plane, but also in the left half-plane, that is, when $\Re \lambda \leq 0$. We give new geometric criterions for establishing zero-free regions as well as for carrying out semi-rigorous numerical explorations. We then provide two examples of the (rigorous) use of these criterions, by establishing two new zero-free regions in the left-half plane. We also extend the results of Bencs and Csikvári for the right half-plane using our framework. By a direct application of the interpolation method of Barvinok, combined with extensions due to Patel and Regts, our results also imply deterministic polynomial time approximation algorithms for the independence polynomial of bounded degree graphs in

the new zero-free regions.

Piyush Srivastava
Tata Institute of Fundamental Research
piyushs@caltech.edu

Ferenc Bencs
University of Amsterdam
ferenc.bencs@gmail.com

Péter Csikvári
Alfréd Rényi Institute of Mathematics
peter.csikvari@gmail.com

Jan Vondrák
Stanford University
jvondrak@stanford.edu

CP6

Approximate Trace Reconstruction from a Single Trace

The well-known trace reconstruction problem is the problem of inferring an unknown source string $x \in \{0, 1\}^n$ from independent "traces", i.e. 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 consider the extreme data-limited regime in which only a single trace is provided to the reconstruction algorithm. In this setting exact reconstruction is of course impossible, and the question is to what accuracy the source string x can be approximately reconstructed. We give a detailed study of this question, providing algorithms and lower bounds for the high, intermediate, and low deletion rate regimes in both the worst-case and average-case models. In several cases the lower bounds we establish are matched by computationally efficient algorithms that we provide. We highlight our results for the high deletion rate regime: roughly speaking, they show that - having access to a single trace is already quite useful for worst-case trace reconstruction: an efficient algorithm can perform much more accurate reconstruction, given one trace that is even only a few bits long, than it could given no traces at all. But in contrast, - in the average-case setting, having access to a single trace is provably not very useful: no algorithm can achieve significantly higher accuracy given one trace that is $o(n)$ bits long than it could with no traces.

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

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

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

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

CP6

Sharp Threshold Sequence and Universality for

Ising Perceptron Models

We study a family of Ising perceptron models with $\{0, 1\}$ -valued activation functions. This includes the classical half-space models, as well as some of the symmetric models considered in recent works. For each of these models we show that the free energy is self-averaging, there is a sharp threshold sequence, and the free energy is universal with respect to the disorder. A prior work of C. Xu (2019) used very different methods to show a sharp threshold sequence in the half-space Ising perceptron with Bernoulli disorder. Recent works of Perkins–Xu (2021) and Abbe–Li–Sly (2021) determined the sharp threshold and limiting free energy in a symmetric perceptron model. The results of this paper apply in more general settings, and are based on new “add one constraint” estimates extending Talagrand’s estimates for the half-space model (1999, 2011).

Nike Sun

Massachusetts Institute of Technology, U.S.
nsun@mit.edu

Shuta Nakajima
Meiji University
njima@meiji.ac.jp

CP7

Low Degree Testing over the Reals

We study the problem of testing whether a function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a polynomial of degree at most d in the *distribution-free* testing model. Here, the distance between functions is measured with respect to an unknown distribution \mathcal{D} over \mathbb{R}^n from which we can draw samples. In contrast to previous work, we do not assume that \mathcal{D} has finite support. We design a tester that given query access to f , and sample access to \mathcal{D} , makes $\text{poly}(d/\epsilon)$ many queries to f , accepts with probability 1 if f is a polynomial of degree d , and rejects with probability at least $2/3$ if every degree- d polynomial P disagrees with f on a set of mass at least ϵ with respect to \mathcal{D} . Our result also holds under mild assumptions when we receive only a polynomial number of bits of precision for each query to f , or when f can only be queried on rational points representable using a logarithmic number of bits. Along the way, we prove a new stability theorem for multivariate polynomials that may be of independent interest.

Vipul Arora, Arnab Bhattacharyya

Department of Computer Science, School of Computing,
National University of Singapore
vipul@comp.nus.edu.sg, arnab@comp.nus.edu.sg

Noah Fleming
Department of Computer Science,
Memorial University, Canada
nffleming@mun.ca

Esty Kelman
Hebrew University, and
Reichman University, Israel
esther.kelman@post.idc.ac.il

Yuichi Yoshida
National Institute of Informatics

yyoshida@nii.ac.jp

CP7

Single-Pass Streaming Algorithms for Correlation Clustering

We study correlation clustering in the streaming setting. This problem has been studied extensively and numerous algorithms have been developed, most requiring multiple passes over the stream. For the important case of single-pass algorithms, recent work of Assadi and Wang (2021) obtains a c -approximation using $\tilde{O}(n)$ space where $c > 10^5$ is a constant and n is the number of vertices to be clustered. We present a single-pass algorithm that obtains a 5-approximation using $O(n)$ space. The algorithm itself is extremely simple and has implications beyond the streaming setting (such as for dynamic and local computation algorithms). The approximation analysis, on the other hand, is delicate and in fact tight.

Soheil Behnezhad

Stanford University
soheil.behnezhad@gmail.com

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

Weiyun Ma, Li-Yang Tan
Stanford University
wyma@cs.stanford.edu, liyang@cs.stanford.edu

CP7

The ℓ_p -Subspace Sketch Problem in Small Dimensions with Applications to Support Vector Machines

In the ℓ_p -subspace sketch problem, we are given an $n \times d$ matrix A with $n > d$, and asked to build a small memory data structure $Q(A, \epsilon)$ so that, for any query vector $x \in \mathbb{R}^d$, we can output a number in $(1 \pm \epsilon)\|Ax\|_p^p$ given only $Q(A, \epsilon)$. This problem is known to require $\tilde{\Omega}(d\epsilon^{-2})$ bits of memory for $d = \Omega(\log(1/\epsilon))$. However, for $d = o(\log(1/\epsilon))$, no data structure lower bounds were known. We resolve the memory required to solve the ℓ_p -subspace sketch problem for any constant d and integer p , showing that it is $\Omega(\epsilon^{-2(d-1)/(d+2p)})$ bits and $\tilde{O}(\epsilon^{-2(d-1)/(d+2p)})$ words. This shows that one can beat the $\Omega(\epsilon^{-2})$ lower bound, which holds for $d = \Omega(\log(1/\epsilon))$, for any constant d . We also show how to implement the upper bound in a single pass stream, with an additional multiplicative $\text{poly}(\log \log n)$ factor and an additive $\text{poly}(\log n)$ cost in the memory. Our bounds can be applied to point queries for SVMs with additive error, yielding an optimal bound of $\tilde{\Theta}(\epsilon^{-2d/(d+3)})$ for every constant d . This is a near-quadratic improvement over the $\Omega(\epsilon^{-(d+1)/(d+3)})$ lower bound of (Andoni et al. 2020). Our techniques rely on a novel connection to low dimensional techniques from geometric functional analysis.

Yi Li
Nanyang Technological University
yili@ntu.edu.sg

Honghao Lin, David Woodruff
Carnegie Mellon University

honghaol@andrew.cmu.edu, dwoodruf@andrew.cmu.edu

CP7

Learning Hierarchical Cluster Structure of Graphs in Sublinear Time

Learning graph cluster structure using few queries is a classical question in property testing, with the fundamental special case, namely expansion testing, considered in the seminal work of Goldreich and Ron[STOC'96]. The most recent results in this line of work design clustering oracles for (k, ϵ) -clusterable graphs, which are graphs that can be partitioned into k induced expanders with outer conductance bounded by $\epsilon \ll 1$. These oracles, given a graph whose vertex set can be partitioned into a disjoint union of k clusters (i.e., good expanders) with outer conductances bounded by $\epsilon \ll 1$, provide query access to an $O(\epsilon \log k)$ -approximation to this ground truth clustering in time $\approx 2^{\text{poly}(k/\epsilon)} n^{1/2+O(\epsilon)}$ per query. Motivated by the rising interest in learning hierarchical structures in large networks, in this paper we introduce (k, γ) -hierarchically clusterable graphs, a natural hierarchical analog of classical (k, ϵ) -clusterable graphs; intuitively, these are graphs that exhibit pronounced hierarchical structure. We give a hierarchical clustering oracle for this model, i.e. a small space data structure that provides query access to a good hierarchical clustering at cost $\approx \text{poly}(k) \cdot n^{1/2+O(\gamma)}$ per query; notable, the dependence on k is polynomial, in contrast to best known flat clustering oracles.

Michael Kapralov, Akash Kumar
EPFL
michael.kapralov@epfl.ch, akash.mnnit@gmail.com

Silvio Lattanzi
Google Zurich
silviol@google.com

Aida Mousavifar
Google
aida.mousavifar@gmail.com

CP7

Towards Multi-Pass Streaming Lower Bounds for Optimal Approximation of Max-Cut

We consider the Max-Cut problem, asking how much space is needed by a streaming algorithm in order to estimate the value of the maximum cut in a graph. This problem has been extensively studied over the last decade, and we now have an optimal lower bound for one-pass streaming algorithms, showing that they require linear space to guarantee a better-than-2 approximation. At the core of this result is a lower bound for the cycle-finding problem, showing that it is hard for a one-pass streaming algorithm to find a cycle in a union of matchings. The end-goal of our research is to prove a similar lower bound for multi-pass streaming algorithms that guarantee a better-than-2 approximation for Max-Cut, a highly challenging open problem. In this paper, we take a significant step in this direction, showing that even $o(\log n)$ -pass streaming algorithms need $n^{\Omega(1)}$ space to solve the cycle-finding problem. Our proof is quite involved, dividing the cycles in the graph into “short” and “long” cycles, and using tailor-made lower bound techniques to handle each case.

Lijie Chen
MIT

lijieche@mit.edu

Gillat Kol
IAS
gillat.kol@gmail.com

Dmitry Paramonov
Princeton University
dp20@princeton.edu

Raghuvansh Saxena
Microsoft Research
raghuvansh.saxena@gmail.com

Zhao Song
Adobe Research
zsong@adobe.com

Huacheng Yu
Princeton University
yuhch123@gmail.com

CP7

Streaming Algorithms for the Missing Item Finding Problem

Many problems on data streams have been studied at two extremes of difficulty: either allowing randomized algorithms, in the static setting (where they should err with bounded probability on the worst case stream); or when only deterministic and infallible algorithms are required. Some recent works have considered the adversarial setting, in which a randomized streaming algorithm must succeed even on data streams provided by an adaptive adversary that can see the intermediate outputs of the algorithm.? In order to better understand the differences between these models, we study a streaming task called “Missing Item Finding”. In this problem, for $r < n$, one is given a data stream a_1, \dots, a_r of elements in $[n]$, (possibly with repetitions), and must output some $x \in [n]$ which does not equal any of the a_i . We prove a set of almost tight bounds for the space (excluding random bits) needed for algorithms in the adversarial setting, for deterministic algorithms, and for algorithms robust against “white-box” adversaries that can see the internal state of the algorithm, but not predict its future random decisions. We also investigate a “random start” model of streaming algorithms where all random bits used are included in the space cost. Here we find a conditional lower bound on the space usage, which depends on the space that would be needed for a pseudo-deterministic algorithm to solve the problem.

Manuel Stoeckl
Dartmouth College
manuel.stoeckl.gr@dartmouth.edu

CP8

Parallel and I/O-Efficient Algorithms for Non-Linear Preferential Attachment

Preferential attachment lies at the heart of many network models aiming to replicate features of real world networks. To simulate the attachment process, conduct statistical tests, or obtain input data for benchmarks, efficient algorithms are required that are capable of generating large graphs according to these models. Existing graph generators are optimized for the most simple model, where new nodes that arrive in the network are connected to earlier

nodes with a probability $P(h) \propto d$ that depends linearly on the degree d of the earlier node h . Yet, some networks are better explained by a more general attachment probability $P(h) \propto f(d)$ for some function $f: \mathbb{N} \rightarrow \mathbb{R}$. Here, the polynomial case $f(d) = d^\alpha$ where $\alpha \in \mathbb{R}_{>0}$ is of particular interest. In this paper, we present efficient algorithms that generate graphs according to the more general models. We first design a simple yet optimal sequential algorithm for the polynomial model. We then parallelize the algorithm by identifying batches of independent samples and obtain a near-optimal speedup when adding many nodes. In addition, we present an I/O-efficient algorithm that can even be used for the fully general model. To showcase the efficiency and scalability of our algorithms, we conduct an experimental study and compare their performance to existing solutions.

Daniel Allendorf, Ulrich Meyer
Goethe University Frankfurt
dallendorf@ae.cs.uni-frankfurt.de,
umeyer@ae.cs.uni-frankfurt.de

Manuel Penschuck
Goethe University Frankfurt, Germany
mpenschuck@ae.cs.uni-frankfurt.de

Hung Tran
Goethe University Frankfurt
htran@ae.cs.uni-frankfurt.de

CP8

Local Motif Clustering via (Hyper)Graph Partitioning

A widely-used operation on graphs is local clustering, i.e., extracting a well-characterized community around a seed node without the need to process the whole graph. Recently local motif clustering has been proposed: it looks for a local cluster based on the distribution of motifs. Since this local clustering perspective is relatively new, most approaches proposed for it are extensions of statistical and numerical methods previously used for edge-based local clustering, while the available combinatorial approaches are still few and relatively simple. In this work, we build a hypergraph and a graph model which both represent the motif-distribution around the seed node. We solve these models using sophisticated combinatorial algorithms designed for (hyper)graph partitioning. In extensive experiments with the triangle motif, we observe that our algorithm computes communities with a motif conductance value being one third on average in comparison against the communities computed by the state-of-the-art tool MAPPR while being 6.3 times faster on average.

Adil Chhabra, Marcelo Fonseca Faraj
Heidelberg University
adil.chhabra@stud.uni-heidelberg.de,
marcelofaraj@informatik.uni-heidelberg.de

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

CP8

Applying Skeletons to Speed Up the Arc-Flags Routing Algorithm

The Single-Source Shortest Path problem is classically solved by applying Dijkstra's algorithm. However, the

plain version of this algorithm is far too slow for real-world applications such as routing in large road networks. To amend this, many speed-up techniques have been developed that build on the idea of computing auxiliary data in a preprocessing phase, that is used to speed up the queries. One well-known example is the Arc-Flags algorithm that is based on the idea of precomputing edge flags to make the search more goal-directed. To explain the strong practical performance of such speed-up techniques, several graph parameters have been introduced. The skeleton dimension is one such parameter that has already been used to derive runtime bounds for some speed-up techniques. Moreover, it was experimentally shown to be low in real-world road networks. We introduce a method to incorporate skeletons, the underlying structure behind the skeleton dimension, to improve routing speed-up techniques even further. As a proof of concept, we develop new algorithms called SKARF and SKARF+ that combine skeletons with Arc-Flags, and demonstrate via extensive experiments on large real-world road networks that SKARF+ yields a significant reduction of the search space and the query time of about 30% to 40% over Arc-Flags. We also prove theoretical bounds on the query time of SKARF, which is the first time an Arc-Flags variant has been analyzed in terms of skeleton dimension.

Ivan Khomutovskiy
Student at ETH Zürich
ivankhomutovskiy@gmx.de

Rebekka Dunker, Jessica Dierking, Julian Egbert,
Christian Helms, Finn Schoellkopf, Katrin Casel, Philipp
Fischbeck

Hasso Plattner Institute
bm.rebekka@googlemail.com,
jessica.dierking@student.hpi.de,
julian.egbert@student.hpi.de,
christian.helms@student.hpi.de,
finn.schoellkopf@student.hpi.de, katrin.casel@hpi.de,
philipp.fischbeck@hpi.de

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

Davis Issac, Simon Krogmann, Pascal Lenzner
Hasso Plattner Institute
davis.issac@hpi.de, simon.krogmann@hpi.de, pas-
cal.lenzner@hpi.de

CP8

Tailored Vertex Ordering for Faster Triangle Listing in Large Graphs

Listing triangles is a fundamental graph problem with many applications, and large graphs require fast algorithms. Vertex ordering allows the orientation of edges from lower to higher vertex indices, and state-of-the-art triangle listing algorithms use this to accelerate their execution and to bound their time complexity. Yet, only basic orderings have been tested. In this paper, we show that studying the precise cost of algorithms instead of their bounded complexity leads to faster solutions. We introduce cost functions that link ordering properties with the running time of a given algorithm. We prove that their minimization is NP-hard and propose heuristics to obtain new orderings with different trade-offs between cost reduction and ordering time. Using datasets with up to two billion edges, we show that our heuristics accelerate the

listing of triangles by an average of 38% when the ordering is already given as an input, and 16% when the ordering time is included.

Fabrice Lécuyer

Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
fabrice.lecuyer@lip6.fr

Louis Jachiet

LTCI, Télécom Paris, Institut polytechnique de Paris, France
louis.jachiet@telecom-paris.fr

Clémence Magnien, Lionel Tabourier

Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
clemence.magnien@lip6.fr, lionel.tabourier@lip6.fr

CP8

Efficiently Computing Directed Minimum Spanning Trees

Computing a directed minimum spanning tree, called arborescence, is a fundamental algorithmic problem, although not as common as its undirected counterpart. In 1967, Edmonds discussed an elegant solution. It was refined to run in $O(\min(n^2, m \log n))$ by Tarjan which is optimal for very dense and very sparse graphs. Gabow et al. gave a version of Edmonds' algorithm that runs in $O(n \log n + m)$, thus asymptotically beating the Tarjan variant in the regime between sparse and dense. Despite the attention the problem received theoretically, there exists, to the best of our knowledge, no empirical evaluation of either of these algorithms. In fact, the version by Gabow et al. has never been implemented and, aside from coding competitions, all readily available Tarjan implementations run in $O(n^2)$. In this paper, we provide the first implementation of the version by Gabow et al. as well as five variants of Tarjan's version with different underlying data structures. We evaluate these algorithms and existing solvers on a large set of real-world and random graphs.

Christopher Weyand

Karlsruhe Institute of Technology
christopher.weyand@kit.edu

Maximilian Böther

ETH Zurich
maximilian.boether@student.hpi.de

Otto Kißig

Hasso Plattner Institute
otto.kissig@student.hpi.de

CP9

Modified Iterative Quantum Amplitude Estimation is Asymptotically Optimal

In this work, we provide the first QFT-free algorithm for Quantum Amplitude Estimation (QAE) that is asymptotically optimal while maintaining the leading numerical performance. QAE algorithms appear as a subroutine in many applications for quantum computers. The optimal query complexity achievable by a quantum algorithm for QAE is $O(\frac{1}{\epsilon} \log \frac{1}{\epsilon})$ queries, providing a speedup of a factor of $1/\epsilon$ over any other classical algorithm for the same problem. The original algorithm for QAE utilizes the quantum Fourier transform (QFT) which is expected to be a challenge for near-term quantum hardware. To solve this

problem, there has been interest in designing a QAE algorithm that avoids using QFT. Recently, the iterative QAE algorithm (IQAE) [?] was introduced with a near-optimal $O(\frac{1}{\epsilon} \log(\frac{1}{\epsilon} \log \frac{1}{\epsilon}))$ query complexity and small constant factors. In this work, we combine ideas from the preceding line of work to introduce a QFT-free QAE algorithm that maintains the asymptotically optimal $O(\frac{1}{\epsilon} \log \frac{1}{\epsilon})$ query complexity while retaining small constant factors. We supplement our analysis with numerical experiments comparing our performance with IQAE where we find that our modifications retain the high performance, and in some cases even improve the numerical results.

Sandy Irani, Shion Fukuzawa, Christopher Ho, Jasen Zion

University of California, Irvine
irani@ics.uci.edu, fukuzaws@uci.edu, christh9@uci.edu, jzion@uci.edu

CP9

Lz77 via Prefix-Free Parsing

Aaron Hong

University of Florida
yu.hong@ufl.edu

CP10

Improved Bi-Point Rounding Algorithms and a Golden Barrier for k-Median

The current best approximation algorithms for k -median rely on first obtaining a structured fractional solution known as a bi-point solution, and then rounding it to an integer solution. We improve this second step by unifying and refining previous approaches. We describe a hierarchy of increasingly-complex partitioning schemes for the facilities, along with corresponding sets of algorithms and factor-revealing non-linear programs. We prove that the third layer of this hierarchy is a 2.613-approximation, improving upon the current best ratio of 2.675, while no layer can be proved better than 2.588 under the proposed analysis. On the negative side, we give a family of bi-point solutions which cannot be approximated better than the square root of the golden ratio, even if allowed to open $k+o(k)$ facilities. This gives a barrier to current approaches for obtaining an approximation better than $2\sqrt{\phi} \approx 2.544$. Altogether we reduce the approximation gap of bi-point solutions by two thirds.

Kishen N. Gowda

University of Maryland
kishen19@cs.umd.edu

Thomas Pensyl

Unaffiliated
tommy.pensyl@gmail.com

Aravind Srinivasan

University of Maryland
srin@cs.umd.edu

Khoa Trinh

Google
khoatrinh@google.com

CP10

On the Integrality Gap of MFN Relaxation for the

Capacitated Facility Location Problem

The Multicommodity Flow Network (MFN) relaxation, developed in [An, Singh, Svensson, FOCS 2014], is the only polynomial-time solvable relaxation that is known to provide a bounded integrality gap for the classic capacitated facility location (CFL) problem. The best upper-bound known for the integrality gap of this strong LP relaxation, however, is in the order of 288. In this paper, we show that the MFN relaxation has an integrality gap at most $(10 + \sqrt{67})/2 \approx 9.0927$ for the CFL problem. This significantly narrows down the range of the integrality gap to one digit. Our ingredient is an iterative rounding algorithm for this sophisticated LP relaxation.

Mong-Jen Kao

National Yang-Ming Chiao-Tung University
mjkao@nycu.edu.tw

CP10

Passing the Limits of Pure Local Search for Weighted k -Set Packing

We study the weighted k -Set Packing problem, which is defined as follows: Given a collection \mathcal{S} of sets, each of cardinality at most k , together with a positive weight function $w : \mathcal{S} \rightarrow \mathbb{Q}_{>0}$, the task is to compute a sub-collection $A \subseteq \mathcal{S}$ of maximum total weight such that the sets in A are pairwise disjoint. For $k \leq 2$, the weighted k -Set Packing problem can be solved in polynomial time, but for $k \geq 3$, already the unweighted variant where all weights equal 1 is NP -hard. The state-of-the-art algorithms for both the unweighted and the weighted k -Set Packing problem rely on local search. In the unweighted setting, the best known approximation guarantee is $\frac{k+1}{3} + \epsilon$, which is a result by Fürer and Yu. For general weights, Berman's $\frac{k+1}{2} + \epsilon$ -approximation algorithm, has remained unchallenged for twenty years. Only recently, Neuwöhner managed to obtain approximation guarantees of $\frac{k+\epsilon_k}{2}$ with $\lim_{k \rightarrow \infty} \epsilon_k = 0$. She further established a lower bound of $\frac{k}{2}$ for algorithms considering local improvements of logarithmically bounded size. In this paper, we show how to beat the threshold of $\frac{k}{2}$ for the weighted k -Set Packing problem by $\Omega(k)$. We achieve this by combining local search with the application of a black box algorithm for the unweighted k -Set Packing problem to carefully chosen sub-instances.

Meike Neuwöhner

University of Bonn
neuwohner@or.uni-bonn.de

CP10

Breaching the 2 LMP Approximation Barrier for Facility Location with Applications to k -Median

The Uncapacitated Facility Location (UFL) problem is one of the most fundamental clustering problems: Given a set of clients C and a set of facilities F in a metric space $(C \cup F, \text{dist})$ with facility costs $\text{open} : F \rightarrow \mathbb{R}^+$, the goal is to find a set of facilities $S \subseteq F$ to minimize the opening cost $\text{open}(S)$ and the connection cost $d(S) := \sum_{p \in C} \min_{c \in S} \text{dist}(p, c)$. An algorithm for UFL is called a Lagrangian Multiplier Preserving (LMP) α -approximation if it outputs a solution $S \subseteq F$ satisfying $\text{open}(S) + d(S) \leq \text{open}(S^*) + \alpha d(S^*)$ for any $S^* \subseteq F$. The best-known LMP approximation ratio for UFL is 2 by the JMS algorithm of Jain, Mahdian, and Saberi [STOC'02, J.ACM'03] based on the Dual-Fitting technique. The lack of progress on improving the upper bound on α_{LMP} in

the last two decades raised the natural question whether $\alpha_{LMP} = 2$. We answer this question negatively by presenting a (slightly) improved LMP approximation algorithm for UFL. This is achieved by combining the Dual-Fitting technique with Local Search, another popular technique to address clustering problems. Our result directly implies a (slightly) improved approximation for the related k -Median problem. Embedding our LMP algorithm into the state of the art approximation algorithm frameworks gives a 2.67059-approximation.

Vincent Cohen-Addad

Google Research
cohenaddad@google.com

Fabrizio Grandoni

IDSIA
fabrizio@idsia.ch

Euiwoong Lee

University of Michigan
euiwoong@umich.edu

Chris Schwiegelshohn

Aarhus University
cschwiegelshohn@gmail.com

CP10

A Nearly Tight Analysis of Greedy K -Means++

The famous k -means++ algorithm of Arthur and Vassilvitskii [SODA 2007] is the most popular way of solving the k -means problem in practice. The algorithm is very simple: it samples the first center uniformly at random and each of the following $k - 1$ centers is then always sampled proportional to its squared distance to the closest center so far. Afterward, Lloyd's iterative algorithm is run. The k -means++ algorithm is known to return a $\Theta(\log k)$ approximate solution in expectation. In their seminal work, Arthur and Vassilvitskii [SODA 2007] asked about the guarantees for its following greedy variant: in every step, we sample ℓ candidate centers instead of one and then pick the one that minimizes the new cost. This is also how k -means++ is implemented in e.g. the popular Scikit-learn library [Pedregosa et al.; JMLR 2011]. We present nearly matching lower and upper bounds for the greedy k -means++: We prove that it is an $O(\ell^3 \log^3 k)$ -approximation algorithm. On the other hand, we prove a lower bound of $\Omega(\ell^3 \log^3 k / \log^2(\ell \log k))$. Previously, only an $\Omega(\ell \log k)$ lower bound was known [Bhattacharya, Eube, Rglin, Schmidt; ESA 2020] and there was no known upper bound.

Christoph Grunau, Ahmet Ozudogru, Václav Rozhon

ETH
christoph.grunau@inf.ethz.ch, oahmet@student.ethz.ch, vaclavrozhon@gmail.com

Jakub Tetek

Basic Algorithms Research Copenhagen
University of Copenhagen
j.tetek@gmail.com

CP10

An Improved Approximation for Maximum Weighted k -Set Packing

We consider the weighted k -set packing problem, in which

we are given a collection of weighted sets, each with at most k elements and must return a collection of pairwise disjoint sets with maximum total weight. For $k = 3$, this problem generalizes the classical 3-dimensional matching problem listed as one of the Karp's original 21 NP-complete problems. We give an algorithm attaining an approximation factor of 1.786 for 3-set packing, improving on the recent best result of $2 - \frac{1}{63,700,992}$ due to Neuwöhner. Our algorithm is based on the local search procedure of Berman that attempts to improve the sum of squared weights rather than the problem's objective. When using exchanges of size at most k , this algorithm attains an approximation factor of $\frac{k+1}{2}$. Using exchanges of size $k^2(k-1) + k$, we provide a relatively simple analysis to obtain an approximation factor of 1.811 when $k = 3$. We then show that the tools we develop can be adapted to larger exchanges of size $2k^2(k-1) + k$ to give an approximation factor of 1.786. Although our primary focus is on the case $k = 3$, our approach in fact gives slightly stronger improvements on the factor $\frac{k+1}{2}$ for all $k > 3$. As in previous works, our guarantees hold also for the more general problem of finding a maximum weight independent set in a $(k+1)$ -claw free graph.

Theophile Thiery, Justin Ward
Queen Mary University of London
t.f.thiery@qmul.ac.uk, justin.ward@qmul.ac.uk

CP11

Testing and Learning Quantum Juntas Nearly Optimally

We consider the problem of testing and learning quantum k -juntas: n -qubit unitary matrices which act non-trivially on just k of the n qubits and as the identity on the rest. As our main algorithmic results, we give 1. A $\tilde{O}(\sqrt{k})$ -query quantum algorithm that can distinguish quantum k -juntas from unitary matrices that are “far” from every quantum k -junta; and 2. A $O(4^k)$ -query algorithm to learn quantum k -juntas. We complement our upper bounds for testing and learning quantum k -juntas with near-matching lower bounds of $\Omega(\sqrt{k})$ and $\Omega(4^k/k)$, respectively. Our techniques are Fourier-analytic and make use of a notion of “influence” of qubits on unitaries.

Thomas Chen, Shivam Nadimpalli, Henry Yuen
Columbia University
tc3039@columbia.edu, sn2855@columbia.edu, hyuen@cs.columbia.edu

CP11

A Sublinear-Time Quantum Algorithm for Approximating Partition Functions

We present a novel quantum algorithm for estimating Gibbs partition functions in sublinear time with respect to the logarithm of the size of the state space. This is the first speed-up of this type to be obtained over the seminal nearly-linear time algorithm of tefankovic, Vempala and Vigoda [JACM, 2009]. Our result also preserves the quadratic speed-up in precision and spectral gap achieved in previous work by exploiting the properties of quantum Markov chains. As an application, we obtain new polynomial improvements over the best-known algorithms for computing the partition function of the Ising model, and counting the number of k -colorings, matchings or independent sets of a graph. Our approach relies on developing new variants of the quantum phase and amplitude estimation algorithms that return nearly unbiased estimates with

low variance and without destroying their initial quantum state. We extend these subroutines into a nearly unbiased quantum mean estimator that reduces the variance quadratically faster than the classical empirical mean. No such estimator was known to exist prior to our work. These properties, which are of general interest, lead to better convergence guarantees within the paradigm of simulated annealing for computing partition functions.

Arjan Cornelissen
University of Amsterdam
arjan_rekenmachine@hotmail.com

Yassine Hamoudi
Simons Institute for the Theory of Computing
ys.hamoudi@gmail.com

CP11

Unique Games Hardness of Quantum Max-Cut, and a Conjectured Vector-Valued Borell's Inequality

The Gaussian noise stability of a function $f : \mathbb{R}^n \rightarrow \{-1, 1\}$ is the expected value of $f(\mathbf{x}) \cdot f(\mathbf{y})$ over ρ -correlated Gaussian random variables \mathbf{x} and \mathbf{y} . Borell's inequality states that for $\rho \in [-1, 0]$, this is minimized by the $f(\mathbf{x}) = \text{sign}(x_1)$. We conjecture that a generalization of this result, the *vector-valued Borell's inequality*, holds for functions $f : \mathbb{R}^n \rightarrow S^{k-1}$ which output k -dimensional unit vectors. The conjecture states that the expected value of $\langle f(\mathbf{x}), f(\mathbf{y}) \rangle$ is minimized by $f(\mathbf{x}) = x_{\leq k} / \|x_{\leq k}\|$, where $x_{\leq k} = (x_1, \dots, x_k)$. We give a proof when $n = k$. The conjecture implies several hardness of approximation results for a special case of the local Hamiltonian problem known as Quantum Max-Cut, a quantum analogue of the classical Max-Cut problem. Assuming the conjecture, we show: (1) The integrality gap of the basic SDP is 0.498, matching an existing rounding algorithm. This shows that the basic SDP does not achieve the optimal approximation ratio. (2) It is Unique Games-hard (UG-hard) to compute a $(0.956 + \epsilon)$ -approximation to the value of the best product state, matching an existing approximation algorithm. (3) It is UG-hard to compute a $(0.956 + \epsilon)$ -approximation to the value of the best (possibly entangled) state.

Yeongwoo Hwang
Harvard University
yeongwoohwang@g.harvard.edu

Joe Neeman
The University of Texas at Austin
jneeman@math.utexas.edu

Ojas Parekh, Kevin Thompson
Sandia National Laboratories
odparek@sandia.gov, kevthom@sandia.gov

John Wright
The University of California, Berkeley
jswright@berkeley.edu

CP11

Mean Estimation When You Have the Source Code; Or, Quantum Monte Carlo Methods

Suppose \mathbf{y} is a real random variable, and one is given access to “the code” that generates it (for example, a randomized or quantum circuit whose output is \mathbf{y}). We give a quantum

procedure that runs the code $O(n)$ times and returns an estimate $\hat{\mu}$ for $\mu = \mathbb{E}[\mathbf{y}]$ that with high probability satisfies $|\hat{\mu} - \mu| \leq \sigma/n$, where $\sigma = \text{stddev}[\mathbf{y}]$. This dependence on n is optimal for quantum algorithms. One may compare with classical algorithms, which can only achieve the quadratically worse $|\hat{\mu} - \mu| \leq \sigma/\sqrt{n}$. Our method improves upon previous works, which either made additional assumptions about \mathbf{y} , and/or assumed the algorithm knew an a priori bound on σ , and/or used additional logarithmic factors beyond $O(n)$. The central subroutine for our result is essentially Grover's algorithm but with complex phases. ally Grover's algorithm but with complex phases.

Robin Kothari
Google Research
robin@robinkothari.com

Ryan O'Donnell
Carnegie Mellon University, USA
odonnell@cs.cmu.edu

CP11

Quantum Tomography Using State-Preparation Unitaries

We describe algorithms to obtain an approximate classical description of a d -dimensional quantum state when given access to a unitary (and its inverse) that prepares it. For pure states we characterize the query complexity for ℓ_q -norm error up to logarithmic factors. As a special case, we show that it takes $\tilde{\Theta}(d/\varepsilon)$ applications of the unitaries to obtain an ε - ℓ_2 -approximation of the state. For mixed states we consider a similar model, where the unitary prepares a purification of the state. We characterize the query complexity for obtaining Schatten q -norm estimates of a rank- r mixed state, up to polylogarithmic factors. In particular, we show that a trace-norm ($q = 1$) estimate can be obtained with $\tilde{\Theta}(dr/\varepsilon)$ queries. This improves (assuming our stronger input model) the ε -dependence over the algorithm of Haah et al. (IEEE Trans. Inf. Theory, 63.9, 2017) that uses a joint measurement on $\tilde{O}(dr/\varepsilon^2)$ copies of the state. To our knowledge, the most sample-efficient results for pure-state tomography come from setting the rank to 1 in generic mixed-state tomography algorithms, which can be computationally demanding. We describe sample-optimal algorithms for pure states that are easy and fast to implement.

Joran van Apeldoorn, Arjan Cornelissen
University of Amsterdam
work@bitofbytes.com, arjan_rekenmachine@hotmail.com

András Gilyén
Alfréd Rényi Institute of Mathematics
gilyen@renyi.hu

Giacomo Nannicini
University of Southern California
g.nannicini@usc.edu

CP11

Efficient Decoding Up to a Constant Fraction of the Code Length for Asymptotically Good Quantum Codes

We introduce and analyse an efficient decoder for quantum Tanner codes that can correct adversarial errors of linear weight. Previous decoders for quantum low-density parity-

check codes could only handle adversarial errors of weight $O(\sqrt{n \log n})$. We also work on the link between quantum Tanner codes and the Lifted Product codes of Panteleev and Kalachev, and show that our decoder can be adapted to the latter. The decoding algorithm alternates between sequential and parallel procedures and converges in linear time.

Anthony Leverrier
INRIA
anthony.leverrier@inria.fr

Gilles Zemor
Ecole Nationale Supérieure des Telecommunications (ENST)
zemor@math.u-bordeaux.fr

CP12

On the Number of Incidences When Avoiding an Induced Biclique in Geometric Settings

Given a set of points P and a set of regions \mathcal{O} , an incidence is a pair $(p, o) \in P \times \mathcal{O}$ such that $p \in o$. We obtain a number of new results on a classical question in combinatorial geometry: What is the maximum number of incidences (under certain restrictive conditions)? We prove a bound of $O(kn(\log n / \log \log n)^{d-1})$ on the number of incidences between n points and n axis-parallel boxes in \mathbb{R}^d , if no k boxes contain k common points, that is, if the incidence graph between the points and the boxes does not contain $K_{k,k}$ as a subgraph. This new bound improves over previous work, by Basit, Chernikov, Starchenko, Tao, and Tran (2021), by more than a factor of $\log^d n$ for $d > 2$. Furthermore, it matches a lower bound implied by the work of Chazelle (1990), for $k = 2$, thus settling the question for points and boxes. We also study several other variants of the problem. For halfspaces, using shallow cuttings, we get a linear bound in two and three dimensions. We also present linear (or near linear) bounds for shapes with low union complexity, such as pseudodisks and fat triangles.

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

Sariel Har-Peled
Department of Computer Science
University of Illinois, Urbana Champaign
sariel@illinois.edu

CP12

Gap-ETH-Tight Approximation Schemes for Red-Green-Blue Separation and Bicolored Noncrossing Euclidean Travelling Salesman Tours

In this paper, we study problems of connecting classes of points via noncrossing structures. Given a set of colored terminal points, we want to find a graph for each color that connects all terminals of its color with the restriction that no two graphs cross each other. We consider these problems both on the Euclidean plane and in planar graphs. On the algorithmic side, we give a Gap-ETH-tight EPTAS for the bicolored noncrossing travelling salesman tours problem as well as for the red-blue-green separation problem (in which we want to separate terminals of three colors with two noncrossing polygons of minimum length), both on the Euclidean plane. This improves the work of Arora and Chang (ICALP 2003) who gave a slower PTAS for

the simpler red-blue separation problem. For the case of unweighted plane graphs, we also show a PTAS for the bicolored noncrossing travelling salesman tours problem. All these results are based on our new patching procedure that might be of independent interest. On the negative side, we show that the problem of connecting terminal pairs with noncrossing paths is NP-hard on the Euclidean plane, and that the problem of finding two noncrossing spanning trees is NP-hard in plane graphs.

François Dross
University of Bordeaux, France
francois.dross@u-bordeaux.fr

Krzysztof Fleszar
University of Warsaw
krzysztof.fleszar@mimuw.edu.pl

Karol Węgrzycki
Saarland University and MPI, Germany
wegrzycki@cs.uni-saarland.de

Anna Zych-Pawlewicz
University of Warsaw
anka@mimuw.edu.pl

CP12

Curve Simplification and Clustering under Fréchet Distance

We present new approximation results on curve simplification and clustering under Fréchet distance. Let T be a set of polygonal curves in R^d of m vertices each. Let l be any integer from $[m]$. We study a generalized curve simplification problem: given error bounds $\delta_i > 0$ for $i \in [n]$, find a curve σ of at most l vertices such that $d_F(\sigma, \tau_i) \leq \delta_i$ for $i \in [n]$. We present an algorithm that returns a null output or a curve σ of at most l vertices such that $d_F(\sigma, \tau_i) \leq \delta_i + \epsilon \delta_{\max}$ for $i \in [n]$, where $\delta_{\max} = \max_{i \in [n]} \delta_i$. If the output is null, we can assert that there is no satisfactory σ . The running time is $\tilde{O}(n^{O(l)} m^{O(l^2)} (dl/\epsilon)^{O(dl)})$. This algorithm yields the first polynomial-time bicriteria approximation scheme to simplify a curve τ to another curve σ , where the vertices of σ can be anywhere in R^d . By combining our technique with some previous results in the literature, we obtain an approximation algorithm for (k, l) -median clustering. Given T , it computes a set Σ of k curves, each of l vertices, such that the clustering cost is within a factor $1 + \epsilon$ of the optimum with probability at least $1 - \mu$ for any given $\mu, \epsilon \in (0, 1)$. The running time is $\tilde{O}(nm^{O(kl^2)} \mu^{-O(kl)} (dkl/\epsilon)^{O((dkl/\epsilon) \log(1/\mu))})$.

Siu-Wing Cheng
The Hong Kong University of Science and Technology
scheng@cse.ust.hk

Haoqiang Huang
The Hong Kong University of Science and Technology
The Hong Kong University of Science and Technology
hhuangbm@connect.ust.hk

CP12

Map Matching Queries on Realistic Input Graphs under the Fréchet Distance

Map matching is a common preprocessing step for analysing vehicle trajectories. In the theory community, the most popular approach for map matching is to com-

pute a path on the road network that is the most spatially similar to the trajectory, where spatial similarity is measured using the Fréchet distance. A shortcoming of existing map matching algorithms under the Fréchet distance is that every time a trajectory is matched, the entire road network needs to be reprocessed from scratch. An open problem is whether one can preprocess the road network into a data structure, so that map matching queries can be answered in sublinear time. In this paper, we investigate map matching queries under the Fréchet distance. We provide a negative result for geometric planar graphs. We show that, unless SETH fails, there is no data structure that can be constructed in polynomial time that answers map matching queries in $O((pq)^{1-\delta})$ query time for any $\delta > 0$, where p and q are the complexities of the geometric planar graph and the query trajectory, respectively. We provide a positive result for realistic input graphs, which we regard as the main result of this paper. We show that for c -packed graphs, one can construct a data structure of $\tilde{O}(cp)$ size that can answer $(1 + \epsilon)$ -approximate map matching queries in $\tilde{O}(c^4 q \log^4 p)$ time, where $\tilde{O}(\cdot)$ hides lower-order factors and dependence of ϵ .

Sampson Wong, Joachim Gudmundsson, Martin P. Seybold

University of Sydney
sampson.wong123@gmail.com,
joachim.gudmundsson@gmail.com,
bold@gmail.com

mpsey-

CP12

Range Searching Revisited: How to Shave Logs in Multi-Level Data Structures

We revisit the classic problem of simplex range searching and related problems in computational geometry. We present a collection of new results which improve previous bounds by multiple logarithmic factors that were caused by the use of multi-level data structures. Highlights include: • For a set of n points in a constant dimension d , we give data structures with $O(n^d)$ (or better) space that can answer simplex range counting queries in optimal $O(\log n)$ time and simplex range reporting queries in optimal $O(\log n + k)$ time, where k denotes the output size. For semigroup range searching, we obtain $O(\log n)$ query time with $O(n^d \text{polylog } n)$ space. Previous data structures with similar space bounds by Matoušek from nearly three decades ago had $O(\log^{d+1} n)$ or $O(\log^{d+1} n + k)$ query time. • For a set of n simplices in a constant dimension d , we give data structures with $O(n)$ space that can answer stabbing counting queries (counting simplices containing a query point) in $O(n^{1-1/d})$ time, and stabbing reporting queries in $O(n^{1-1/d} + k)$ time. Previous data structures had extra $\log^d n$ factors in space and query time. • For a set of n line segments in 2D, we give a data structure with $O(n)$ space that can answer ray shooting queries in $O(\sqrt{n})$ time. This improves Wang's recent data structure [SoCG'20] by $O(n \log n)$ space and $O(\sqrt{n} \log n)$ 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

ssingla@gatech.edu

CP12**Halving by a Thousand Cuts Or Punctures**

For point sets P_1, \dots, P_k , a set of lines L is halving if any face of the arrangement $\mathcal{A}(L)$ contains at most $|P_i|/2$ points of P_i , for all i . We study the problem of computing a halving set of lines of minimal size. Surprisingly, we show a polynomial time algorithm that outputs a halving set of size $O(o^{3/2})$, where o is the size of the optimal solution. Our solution relies on solving a new variant of the weak ε -net problem for corridors, which we believe to be of independent interest. We also study other variants of this problem, including an alternative setting, where one needs to introduce a set of guards (i.e., points), such that no convex set avoiding the guards contains more than half the points of each point set.

Sariel Har-Peled
Department of Computer Science
University of Illinois, Urbana Champaign
sariel@illinois.edu

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

CP13**Online and Bandit Algorithms Beyond ℓ_p Norms**

Vector norms play a fundamental role in computer science and optimization, so there is an ongoing effort to generalize existing algorithms to settings beyond ℓ_∞ and ℓ_p norms. We show that many online and bandit applications for general norms admit good algorithms as long as the norm can be approximated by a function that is gradient-stable, a notion that we introduce. Roughly it says that the gradient of the function should not drastically decrease (multiplicatively) in any component as we increase the input vector. We prove that several families of norms, including all monotone symmetric norms, admit a gradient-stable approximation, giving us the first online and bandit algorithms for these norm families. In particular, our notion of gradient-stability gives $O(\log^2(\text{dimension}))$ -competitive algorithms for the symmetric norm generalizations of Online Generalized Load Balancing and Bandits with Knapsacks. Our techniques extend to applications beyond symmetric norms as well, e.g., to Online Vector Scheduling and to Online Generalized Assignment with Convex Costs. Some key properties underlying our applications that are implied by gradient-stable approximations are a smooth game inequality and an approximate converse to Jensen's inequality.

Thomas Kesselheim
University of Bonn
thomas.kesselheim@uni-bonn.de

Marco Molinaro
Microsoft Research (Redmond) and Computer Science
Department
mmolinaro@microsoft.com

Sahil Singla
Georgia Tech

CP13**Player-Optimal Stable Regret for Bandit Learning in Matching Markets**

Matching market has been studied for a long time in the literature due to its wide range of applications. Finding a stable matching is a common equilibrium objective in this problem. Since market participants are usually uncertain of their preferences, a rich line of recent works study the online setting where one-side participants (players) learn their unknown preferences from iterative interactions with the other side (arms). Most previous works in this line are only able to derive theoretical guarantees for player-pessimal stable regret, which is defined compared with the players least-preferred stable matching. However, under the pessimal stable matching, players only obtain the least reward among all stable matchings. To maximize their profits, player-optimal stable matching would be the most desirable. Though basu21beyond successfully bring an upper bound for player-optimal stable regret, their result can be exponentially large if players preference gap is small. Whether a polynomial guarantee for this regret exists is a significant but still open problem. In this work, we provide a new algorithm and show the optimal stable regret can be upper bounded by $O(K \log T / \Delta^2)$ where K is the number of arms, T is the horizon and Δ is the players minimum preference gap. This result significantly improves previous works and also matches the previously derived lower bound when the preferences of participants satisfy some special conditions.

Fang Kong, Shuai Li
Shanghai Jiao Tong University
fangkong@sjtu.edu.cn, shuaili8@sjtu.edu.cn

CP13**Almost Tight Bounds for Online Facility Location in the Random-Order Model**

We study the online facility location problem with uniform facility costs in the random-order model. Meyerson's algorithm [FOCS'01] is arguably the most natural and simple online algorithm for the problem with several advantages and appealing properties. Its analysis in the random-order model is one of the cornerstones of random-order analysis beyond the secretary problem. Meyerson's algorithm was shown to be (asymptotically) optimal in the standard worst-case adversarial-order model and 8-competitive in the random order model. While this bound in the random-order model is the long-standing state-of-the-art, it is not known to be tight, and the true competitive-ratio of Meyerson's algorithm remained an open question for more than two decades. We resolve this question and prove tight bounds on the competitive-ratio of Meyerson's algorithm in the random-order model, showing that it is exactly 4-competitive. Following our tight analysis, we introduce a generic parameterized version of Meyerson's algorithm that retains all the advantages of the original version. We show that the best algorithm in this family is exactly 3-competitive. On the other hand, we show that no online algorithm for this problem can achieve a competitive-ratio better than 2. Finally, we prove that the algorithms in this family are robust to partial adversarial arrival orders.

David Naori
The Technion Israel Institute of Technology
dnaori@cs.technion.ac.il

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

Danny Raz
The Technion Israel Institute of Technology
danny@cs.technion.ac.il

CP13

Online Prediction in Sub-Linear Space

We provide the first sub-linear space and sub-linear regret algorithm for online learning with expert advice (against an oblivious adversary), addressing an open question raised recently by Srinivas, Woodruff, Xu and Zhou (STOC 2022). We also demonstrate a separation between oblivious and (strong) adaptive adversaries by proving a linear memory lower bound of any sub-linear regret algorithm against an adaptive adversary. Our algorithm is based on a novel pool selection procedure that bypasses the traditional wisdom of leader selection for online learning, and a generic reduction that transforms any weakly sub-linear regret $o(T)$ algorithm to $T^{1-\alpha}$ regret algorithm, which may be of independent interest. Our lower bound utilizes the connection of no-regret learning and equilibrium computation in zero-sum games, leading to a proof of a strong lower bound against an adaptive

Binghui Peng
Columbia University
bp2601@columbia.edu

Fred Zhang
Harvard University
z0@berkeley.edu

CP13

The Power of Clairvoyance for Multi-Level Aggregation and Set Cover with Delay

Most online problems with delay require clairvoyance, the future delay of a request is known upon its arrival, to achieve polylogarithmic competitiveness. An exception is Set Cover with Delay: Azar et al. (ESA 2020) gave a non-clairvoyant randomized algorithm with polylogarithmic competitive ratio. However, no non-trivial algorithms are known for other non-clairvoyant online problems with delay and it is also unclear if non-clairvoyance requires randomization. In this work, we make progress towards understanding the power of clairvoyance for online problems with delay by providing deterministic non-clairvoyant algorithms for Multi-Level Aggregation and Set Cover with Delay. Our main contribution is a deterministic $O(\sqrt{n} + D)$ -competitive algorithm for Multi-Level Aggregation, where D is the depth of the aggregation tree. For the special case of Joint Replenishment ($D = 1$), we give an $\Omega(\sqrt{n})$ lower bound against non-clairvoyant randomized algorithms. Thus, we get a tight characterization of the competitive ratio for non-clairvoyant Joint Replenishment. Finally, we show that clairvoyance is not required at all for Set Cover with Delay by derandomizing the algorithm of Azar et al. losing at most a constant factor in the competitiveness. Together with the above bounds, this also implies that randomization does not help in the non-clairvoyant setting.

Ngoc Mai Le
University of Sydney
maile191205@gmail.com

William Umboh
The University of Sydney
Australia
william.umbogh@sydney.edu.au

Ningyuan Xie
University of Sydney
nxie0271@uni.sydney.edu.au

CP13

Online Min-Max Paging

Motivated by fairness requirements in communication networks, we introduce a natural variant of the online paging problem, called *min-max* paging, where the objective is to minimize the maximum number of faults on any page. While the classical paging problem, whose objective is to minimize the total number of faults, admits k -competitive deterministic and $O(\log k)$ -competitive randomized algorithms, we show that min-max paging does not admit a $c(k)$ -competitive algorithm for any function c . Specifically, we prove that the randomized competitive ratio of min-max paging is $\Omega(\log(n))$ and its deterministic competitive ratio is $\Omega(k \log(n) / \log(k))$, where n is the total number of pages ever requested. We design a fractional algorithm for paging with a more general objective – minimize the value of an n -variate differentiable convex function applied to the vector of the number of faults on each page. This gives an $O(\log(n) \log(k))$ -competitive fractional algorithm for min-max paging. We show how to round such a fractional algorithm with at most a k factor loss in the competitive ratio, resulting in a deterministic $O(k \log(n) \log(k))$ -competitive algorithm for min-max paging. This matches our lower bound modulo a $\text{poly}(\log(k))$ factor. We also give a randomized rounding algorithm that results in a $O(\log^2 n \log k)$ -competitive algorithm.

Ashish Chiplunkar
Department of Computer Science and Engineering
Indian Institute of Technology Delhi
ashishc@iitd.ac.in

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

Sagar Sudhir Kale, Maximilian Vötsch
Faculty of Computer Science
University of Vienna
sagar.kale@univie.ac.at, maximilian.voetsch@univie.ac.at

CP14

Small Shadows of Lattice Polytopes

The diameter of the graph of a d -dimensional lattice polytope $P \subseteq [0, k]^n$ is known to be at most dk due to work by Kleinschmidt and Onn. However, it is an open question whether the monotone diameter, the shortest guaranteed length of a monotone path, of a d -dimensional lattice polytope $P = \{\mathbf{x} : A\mathbf{x} \leq \mathbf{b}\} \subseteq [0, k]^n$ is bounded by a polynomial in d and k . This question is of particular interest in linear optimization, since paths traced by the Simplex method must be monotone. We introduce partial results in this direction including a monotone diameter bound of $3d$ for $k = 2$, a monotone diameter bound of $(d - 1)m + 1$ for d -dimensional $(\ell + 1)$ -level polytopes, a pivot rule such that the Simplex method is guaranteed to take at most

$dnk\|A\|_\infty$ non-degenerate steps to solve a linear program on P , and a bound of dk for lengths of paths from certain fixed starting points. Finally, we present a constructive approach to a diameter bound of $(3/2)dk$ and describe how to translate this final bound into an algorithm that solves a linear program by tracing such a path.

Alexander E. Black
UC Davis
aebblack@ucdavis.edu

CP14

Integrality Gaps for Random Integer Programs via Discrepancy

We prove new bounds on the additive gap between the value of a random integer program $\max c^T x$, $Ax \leq b$, $x \in \{0, 1\}^n$ with m constraints and that of its linear programming relaxation for a wide range of distributions on (A, b, c) . We are motivated by the work of Dey, Dubey, and Molinaro (SODA '21), who gave a framework for relating the size of Branch-and-Bound (B&B) trees to additive integrality gaps. Dyer and Frieze (MOR '89) and Borst et al. (IPCO '21), respectively, showed that for certain random IPs, where the entries of A, c are independently distributed according to either the uniform distribution on $[0, 1]$ or the standard normal distribution, the integrality gap is bounded by $O_m(\log^2 n/n)$ with probability at least $1 - 1/n - e^{-\Omega_m(1)}$. We generalize these results to the case where the entries of A are uniformly distributed on an integer interval, and where the columns of A are distributed according to an isotropic logconcave distribution. Using the connection to Branch-and-Bound, our results imply that for these IPs BB trees have size $n^{\text{poly}(m)}$ whp. Our main technical contribution is a new linear discrepancy theorem for random matrices. Our theorem gives general conditions under which a target vector is equal to or very close to a $\{0, 1\}$ combination of the columns of a random matrix A . The proof uses a Fourier analytic approach, building on work of Hoberg and Rothvoss (SODA '19) and Franks and Saks (RSA '20).

Sander Borst
Centrum Wiskunde & Informatica (CWI)
sander.borst@cwi.nl

Daniel Dadush
Centrum Wiskunde & Informatica
dadush@cwi.nl

Dan Mikulincer
MIT
danmiku@mit.edu

CP14

Discrepancy Minimization via Regularization

We introduce a new algorithmic framework for discrepancy minimization based on *regularization*. We demonstrate how varying the regularizer allows to us to re-interpret several breakthrough works in algorithmic discrepancy, ranging from Spencer's theorem [Spencer'1985, Bansal'2010] to Banaszczyk's bounds [Banaszczyk'1998, Bansal-Dadush-Garg'2016]. We also show that our techniques imply Beck-Fiala and Komlós conjectures for a new regime of *pseudo-random* instances.

Lucas Pesenti
Bocconi University

lucas.pesenti@phd.unibocconi.it

Adrian Vladu
CNRS & IRIF, Université Paris Cité
adrian.vladu@irif.fr

CP14

Shrunk Subspaces via Operator Sinkhorn Iteration

A recent breakthrough in Edmonds' problem showed that the noncommutative rank can be computed in deterministic polynomial time, and various algorithms for it were devised. However, only quite complicated algorithms are known for finding a so-called shrunk subspace, which acts as a dual certificate for the value of the noncommutative rank. In particular, the operator Sinkhorn algorithm, perhaps the simplest algorithm to compute the noncommutative rank with operator scaling, does not find a shrunk subspace. Finding a shrunk subspace plays a key role in applications, such as separation in the Brascamp-Lieb polytope, one-parameter subgroups in the null-cone membership problem, and primal-dual algorithms for matroid intersection and fractional matroid matching. In this paper, we provide a simple Sinkhorn-style algorithm to find the smallest shrunk subspace over the complex field in deterministic polynomial time. To this end, we introduce a generalization of the operator scaling problem, where the spectra of the marginals must be majorized by specified vectors. Then we design an efficient Sinkhorn-style algorithm for the generalized operator scaling problem. Applying this to the shrunk subspace problem, we show that a sufficiently long run of the algorithm also finds an approximate shrunk subspace close to the minimum exact shrunk subspace. Finally, we show that the approximate shrunk subspace can be rounded if it is sufficiently close.

Cole Franks, Tasuku Soma
MIT
franks@mit.edu, tasuku@mit.edu

Michel Goemans
Massachusetts Institute of Technology
goemans@math.mit.edu

CP14

A Polynomial Time Algorithm for Finding a Minimum 4-Partition of a Submodular Function

In this paper, we study the minimum k -partition problem of submodular functions, i.e., given a finite set V and a submodular function $f : 2^V \rightarrow \mathbb{R}$, computing a k -partition $\{V_1, \dots, V_k\}$ of V with minimum $\sum_{i=1}^k f(V_i)$. The problem is a natural generalization of the minimum k -cut problem in graphs and hypergraphs. It is known that the problem is NP-hard for general k , and solvable in polynomial time for $k \leq 3$. In this paper, we construct the first polynomial-time algorithm for the minimum 4-partition problem.

Chao Xu
Yahoo! Research NYC
thechaoxu@gmail.com

Tsuyoshi Hirayama
Toshiba Digital Solutions Corporation
tsuyoshi1.hirayama@toshiba.co.jp

Yuhao Liu
University of Electronic Science and Technology of China

yuhaoliu126@gmail.com

Kazuhisa Makino
Kyoto University
makino@kurims.kyoto-u.ac.jp

Ke Shi
University of Electronic Science and Technology of China
self.ke.shi@gmail.com

CP14

The Exact Bipartite Matching Polytope Has Exponential Extension Complexity

Given a graph with edges colored red or blue and an integer k , the exact perfect matching problem asks if there exists a perfect matching with exactly k red edges. There exists a randomized polylogarithmic-time parallel algorithm to solve this problem, dating back to the eighties, but no deterministic polynomial-time algorithm is known, even for bipartite graphs. In this paper we show that there is no sub-exponential sized linear program that can describe the convex hull of exact matchings in bipartite graphs. In fact, we prove something stronger, that there is no sub-exponential sized linear program to describe the convex hull of perfect matchings with an odd number of red edges.

Weiqliang Yuan
EPFL
weiqliang.yuan@epfl.ch

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

Ola Svensson
EPFL
ola.svensson@epfl.ch

CP15

Finding Triangles and Other Small Subgraphs in Geometric Intersection Graphs

We consider problems related to finding short cycles, small cliques, small independent sets, and small subgraphs in geometric intersection graphs. We obtain a plethora of new results. For example: For the intersection graph of n line segments in the plane, we give algorithms to find a 3-cycle in $O(n^{1.408})$ time, a size-3 independent set in $O(n^{1.652})$ time, a 4-clique in near- $O(n^{24/13})$ time, and a k -clique (or any k -vertex induced subgraph) in $O(n^{0.565k+O(1)})$ time for any constant k ; we can also compute the girth in near- $O(n^{3/2})$ time. For the intersection graph of n axis-aligned boxes in a constant dimension d , we give algorithms to find a 3-cycle in $O(n^{1.408})$ time for any d , a 4-clique (or any 4-vertex induced subgraph) in $O(n^{1.715})$ time for any d , a size-4 independent set in near- $O(n^{3/2})$ time for any d , and a k -clique (or any k -vertex induced subgraph) in $O(n^{0.429k+O(1)})$ time for any d and any constant k . For the intersection graph of n fat objects in any constant dimension d , we give an algorithm to find any k -vertex (non-induced) subgraph in $O(n \log n)$ time for any constant k , generalizing a result by Kaplan et al. (2019) for 3-cycles in 2D disk graphs. We also prove a near- $\Omega(n^{4/3})$ conditional lower bound for finding a size-4 independent set for boxes.

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

CP15

Online Sorting and Translational Packing of Convex Polygons

We investigate several online packing problems in which convex polygons arrive one by one and have to be placed irrevocably into a container, while the aim is to minimize the used space. Among other variants, we consider strip packing and bin packing, where the container is the infinite horizontal strip $[0, \infty) \times [0, 1]$ or a collection of 1×1 bins, respectively. If polygons may be rotated, there exist $O(1)$ -competitive online algorithms for all problems at hand [Baker and Schwarz, SIAM J. Comput., 1983]. Likewise, if the polygons may not be rotated but only translated, then using a result from [Alt, de Berg and Knauer, JoCG, 2017] we can derive $O(1)$ -approximation algorithms for all problems at hand. Thus, it is natural to conjecture that the online version of these problems, in which only translations are allowed, also admits a $O(1)$ -competitive algorithm. We disprove this conjecture by showing a superconstant lower bound on the competitive ratio for several online packing problems. We prove lower bounds on the competitive ratio of translation-only online packing problems by reducing from a purpose-built novel and natural combinatorial problem that we call *online sorting*. We prove a superconstant lower bound on the competitive ratio of online sorting. We believe that this technique is of independent interest since it uncovers a deep connection between inherently geometrical and purely combinatorial problems.

Linda Kleist
TU Braunschweig
kleist@ibr.cs.tu-bs.de

Anders Aamand
MIT
aamand@mit.edu

Mikkel Abrahamsen
University of Copenhagen
miab@di.ku.dk

Lorenzo Beretta
Basic Algorithms Research Copenhagen
University of Copenhagen
beretta@di.ku.dk

CP15

Approximate Distance Oracles for Planar Graphs with Subpolynomial Error Dependency

Thorup [FOCS'01, JACM'04] and Klein [SODA'01] independently showed that there exists a $(1 + \epsilon)$ -approximate distance oracle for planar graphs with $O(n(\log n)\epsilon^{-1})$ space and $O(\epsilon^{-1})$ query time. While the dependency on n is nearly linear, the space-query product of their oracles depend quadratically on $1/\epsilon$. Many follow-up results either improved the space or the query time of the oracles while having the same, sometimes worst, dependency on $1/\epsilon$. Kawarabayashi, Sommer, and Thorup [SODA'13] were the first to improve the dependency on $1/\epsilon$ from quadratic to nearly linear (at the cost of $\log^*(n)$ factors). It is plausible

to conjecture that the linear dependency on $1/\epsilon$ is optimal: for many known distance-related problems in planar graphs, it was proved that the dependency on $1/\epsilon$ is at least linear. In this work, we disprove this conjecture by reducing the dependency of the space-query product on $1/\epsilon$ from linear all the way down to subpolynomial $(1/\epsilon)^{o(1)}$. More precisely, we construct an oracle with $O(n \log(n)(\epsilon^{-o(1)} + \log^* n))$ space and $\log^{2+o(1)}(1/\epsilon)$ query time. Our construction is the culmination of several different ideas developed over the past two decades.

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

CP15

4D Range Reporting in the Pointer Machine Model in Almost-Optimal Time

In the orthogonal range reporting problem we must preprocess a set P of multi-dimensional points, so that for any axis-parallel query rectangle q all points from $q \cap P$ can be reported efficiently. In this paper we study the query complexity of multi-dimensional orthogonal range reporting in the pointer machine model. We present a data structure that answers four-dimensional orthogonal range reporting queries in almost-optimal time $O(\log n \log \log n + k)$ and uses $O(n \log^4 n)$ space, where n is the number of points in P and k is the number of points in $q \cap P$. This is the first data structure with nearly-linear space usage that achieves almost-optimal query time in 4d. This result can be immediately generalized to $d \geq 4$ dimensions: we show that there is a data structure supporting d -dimensional range reporting queries in time $O(\log^{d-3} n \log \log n + k)$ for any constant $d \geq 4$.

Yakov Nekrich
Michigan Technological University
yakov.nekrich@googlemail.com

Saladi Rahul
Indian Institute of Science
saladi@iisc.ac.in

CP15

A Subquadratic n^ϵ -Approximation for the Continuous Frchet Distance

The Frchet distance is a commonly used similarity measure between curves. It is known how to compute the continuous Frchet distance between two polylines with m and n vertices in \mathbb{R}^d in $O(mn(\log \log n)^2)$ time; doing so in strongly subquadratic time is a longstanding open problem. Recent conditional lower bounds suggest that it is unlikely that a strongly subquadratic algorithm exists. Moreover, it is unlikely that we can approximate the Frchet distance to within a factor 3 in strongly subquadratic time, even if $d = 1$. The best current results establish a tradeoff between approximation quality and running time. Specifically, Colombe and Fox (SoCG, 2021) give an $O(\alpha)$ -approximate algorithm that runs in $O((n^3/\alpha^2) \log n)$ time for any $\alpha \in [\sqrt{n}, n]$, assuming $m \leq n$. In this paper, we improve this result with an $O(\alpha)$ -approximate algorithm that runs in $O((n + mn/\alpha) \log^3 n)$ time for any $\alpha \in [1, n]$, assuming $m \leq n$ and constant dimension d .

Thijs Van Der Horst
TU Eindhoven
Utrecht University

t.w.j.vanderhorst@uu.nl

Marc van Kreveld
Utrecht University
m.j.vankreveld@uu.nl

Tim Ophelders
Utrecht University
TU Eindhoven
t.a.e.ophelders@uu.nl

Bettina Speckmann
Dept. of Mathematics and Computer Science
TU Eindhoven
b.speckmann@tue.nl

CP15

Economical Convex Coverings and Applications

Coverings of convex bodies have emerged as a central component in the design of efficient solutions to approximation problems involving convex bodies. Intuitively, given a convex body K and $\epsilon > 0$, a *covering* is a collection of convex bodies whose union covers K such that a constant factor expansion of each body lies within an ϵ expansion of K . Coverings have been employed in many applications, such as approximations for diameter, width, and ϵ -kernels of point sets, approximate nearest neighbor searching, polytope approximations with low combinatorial complexity, and approximations to the Closest Vector Problem (CVP). It is known how to construct coverings of size $n^{O(n)}/\epsilon^{(n-1)/2}$ for general convex bodies in \mathbb{R}^n . In special cases, such as when the convex body is the ℓ_p unit ball, this bound has been improved to $2^{O(n)}/\epsilon^{(n-1)/2}$. This raises the question of whether such a bound generally holds. In this paper we answer the question in the affirmative. We demonstrate the power and versatility of our coverings by applying them to the problem of approximating a convex body by a polytope, where the error is measured through the Banach-Mazur metric.

Guilherme da Fonseca
Aix-Marseille Universite
guilherme.fonseca@lis-lab.fr

Sunil Arya
The Hong Kong University of Science and Technology
arya@cse.ust.hk

David M. Mount
University of Maryland
mount@umd.edu

CP16

Pachash: Packed and Compressed Hash Tables

We introduce PaCHash, a hash table that stores its objects contiguously in an array without intervening space, even if the objects have variable size. In particular, each object can be compressed using standard compression techniques. A small search data structure allows locating the objects in constant expected time. PaCHash is most naturally described as a static external hash table where it needs a constant number of bits of internal memory per block of external memory. Here, in some sense, PaCHash beats a lower bound on the space consumption of k -perfect hashing. An implementation for fast SSDs needs about 5 bits of internal memory per block of external memory, requires

only one disk access (of variable length) per search operation, and has small internal search overhead compared to the disk access cost. Our experiments show that it has lower space consumption than all previous approaches even when considering objects of identical size.

Florian Kurpicz, Hans-Peter Lehmann
Karlsruhe Institute of Technology
kurpicz@kit.edu, hans-peter.lehmann@kit.edu

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

CP16

SicHash Small Irregular Cuckoo Tables for Perfect Hashing

A Perfect Hash Function (PHF) is a hash function that has no collisions on a given input set. PHFs can be used for space efficient storage of data in an array, or for determining a compact representative of each object in the set. In this paper, we present the PHF construction algorithm SicHash Small Irregular Cuckoo Tables for Perfect Hashing. At its core, SicHash uses a known technique: it places objects in a cuckoo hash table and then stores the final hash function choice of each object in a retrieval data structure. We combine the idea with irregular cuckoo hashing, where different objects can have a different number of hash functions. Additionally, we use many small tables that we overload beyond their asymptotic maximum load factor. The most space efficient competitors often use brute force methods to determine the PHFs. SicHash provides a more direct construction algorithm that only rarely needs to re-compute parts. Our implementation improves the state of the art in terms of space usage versus construction time for a wide range of configurations. For some configurations, SicHash is up to 4.3 times faster than the next best competitor. At the same time, it provides very fast queries.

Hans-Peter Lehmann
Karlsruhe Institute of Technology
hans-peter.lehmann@kit.edu

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

Stefan Walzer
Cologne University
walzer@cs.uni-koeln.de

CP16

Multiway Powersort

Powersort (Munro & Wild, ESA2018) has recently replaced Timsort's suboptimal merge policy in the CPython reference implementation of Python, as well as in PyPy and further libraries. We present a stable mergesort variant, *Multiway Powersort*, that exploits existing runs and finds nearly-optimal merging orders for k -way merges with negligible overhead. As observed with Multiway Quicksort (Kushagra et al., ALENEX 2014; Aumiller & Dietzfelbinger, TALG 2016; Wild, PhD thesis 2016) and the inclusion of Dual-Pivot Quicksort in the Java runtime library, memory transfers increasingly determine the cost of inter-

nal sorting. We demonstrate that our 4-way Powersort implementation can achieve substantial speedups over standard (2-way) Powersort and other stable sorting methods without compromising the optimally run-adaptive performance of Powersort.

William Cawley Gelling
University of Liverpool
william.cawleygelling@gmail.com

Markus Nebel
University of Bielefeld
nebel@tf.unibi.de

Benjamin Smith
University of Liverpool
b.m.smith@liverpool.ac.uk

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

CP16

Lower Bounds for Sorting 16, 17, and 18 Elements

It is a long-standing open question to determine the minimum number of comparisons $S(n)$ that suffice to sort an array of n elements. Indeed, before this work, $S(n)$ has been known only for $n \leq 22$ with the exception of $n = 16, 17$, and 18 . In this work, we fill that gap by proving that sorting $n = 16, 17$, and 18 elements requires $46, 50$, and 54 comparisons respectively. This fully determines $S(n)$ for these values and disproves a conjecture by Knuth that $S(16) = 45$. Moreover, we show that for sorting 28 elements at least 99 comparisons are needed. We obtain our result via an exhaustive computer search which extends previous work by Wells (1965) and Pecarski (2002, 2004, 2007, 2012). Our progress is both based on advances in hardware and on novel algorithmic ideas such as applying a bidirectional search to this problem.

Armin Weiß, Florian Stober
Universität Stuttgart
armin.weiss@fmi.uni-stuttgart.de, florian.stober@fmi.uni-stuttgart.de

CP16

Optimizing Search Layouts in Packed Memory Arrays

This paper introduces Search-optimized Packed Memory Arrays (SPMAs), a collection of data structures based on Packed Memory Arrays (PMAs) that address suboptimal search via cache-optimized search layouts. Traditionally, PMAs and B-trees have tradeoffs between searches/inserts and scans: B-trees were faster for searches and inserts, while PMAs were faster for scans. Our empirical evaluation shows that SPMAs overcome this tradeoff for unsorted input distributions: on average, SPMAs are faster than B+-trees (a variant of B-trees optimized for scans) on all major operations. We generated datasets and search/insert workloads from the Yahoo! Cloud Serving Benchmark (YCSB) and found that SPMAs are about $2\times$ faster than B+-trees regardless of the ratio of searches to inserts. On uniform random inputs, SPMAs are on average between $1.3\times$ – $2.3\times$ faster than B+-trees on all operations. Finally, we vary the amount of sortedness in the inputs to stress the worst-case insert distribution in the PMA. We find that the worst-case B+-tree insertion throughput is about $1.5\times$ faster

than the worst-case PMA insertion throughput. However, the worst-case input for the PMA is sorted and highly unlikely to appear naturally in practice. The SPMA's maintain higher insertion throughput than the B+-tree when the input is up to 25% sorted.

Brian Wheatman, Randal Burns
Johns Hopkins University
bwheatman@gmail.com, randal@cs.jhu.edu

Aydin Buluc, Helen Xu
Lawrence Berkeley National Laboratory
abuluc@lbl.gov, hjxu@lbl.gov

CP17

Beating $(1 - 1/e)$ -Approximation for Weighted Stochastic Matching

In the stochastic weighted matching problem, the goal is to find a large-weight matching of a graph when we are uncertain about the existence of its edges. In particular, each edge e has a known weight w_e but is realized independently with some probability p_e . The algorithm may query an edge to see whether it is realized. We consider the well-studied query commit version of the problem, in which any queried edge that happens to be realized must be included in the solution. Gamlath, Kale, and Svensson [SODA'19] showed that when the input graph is bipartite, the problem admits a $(1 - 1/e)$ -approximation. In this paper, we give an algorithm that for an absolute constant $\delta > 0.0014$ obtains a $(1 - 1/e + \delta)$ -approximation, therefore breaking this prevalent bound.

Mahsa Derakhshan
Princeton University
mah.derakhshan@gmail.com

Alireza Farhadi
CMU
mahsaderakhshan

CP17

On (Random-Order) Online Contention Resolution Schemes for the Matching Polytope of (Bipartite) Graphs

We present new results for online contention resolution schemes for the matching polytope of graphs, in the random-order (RCRS) and adversarial (OCRS) arrival models. Our results include improved selectability guarantees, as well as new impossibility results. By well-known reductions to the prophet (secretary) matching problem, a c -selectable OCRS (RCRS) implies a c -competitive algorithm for adversarial (random order) edge arrivals. For the adversarial arrival model, we present a new analysis of the OCRS of Ezra et al. We show that this scheme is 0.344-selectable for general graphs and 0.349-selectable for bipartite graphs, improving on the previous 0.337 selectability result for this algorithm. We then show that no OCRS can achieve a selectability greater than 0.4 for general graphs, and 0.433 for bipartite graphs. For random-order arrivals, we present two attenuation-based schemes which use new attenuation functions. Our first RCRS is 0.474-selectable for general graphs, and our second is 0.476-selectable for bipartite graphs. These results improve upon the recent 0.45 (and 0.456) selectability results for general graphs (respectively, bipartite graphs) due to Pollner et al. On general graphs, our 0.474-selectable RCRS provides the best-known positive result even for offline contention resolution,

and also for the correlation gap. We conclude by proving a fundamental upper bound of 0.5 on the selectability of RCRS, using bipartite graphs.

Calum Macrury
University of Toronto
cmacrury@cs.toronto.edu

Will Ma
Columbia University
wm2428@gsb.columbia.edu

Nathaniel Grammel
University of Maryland
ngrammel@umd.edu

CP17

Minimizing Completion Times for Stochastic Jobs via Batched Free Times

We study the classic problem of minimizing the expected total completion time of jobs on m identical machines in the setting where the sizes of the jobs are stochastic. Specifically, the size of each job is a random variable whose distribution is known to the algorithm, but whose realization is revealed only after the job is scheduled. While minimizing the total completion time is easy in the deterministic setting, the stochastic problem has long been notorious: all known algorithms have approximation ratios that either depend on the variances, or depend linearly on the number of machines. We give an $\tilde{O}(\sqrt{m})$ -approximation for stochastic jobs which have Bernoulli processing times. This is the first approximation for this problem that is both independent of the variance in the job sizes, and is sublinear in the number of machines m . Our algorithm is based on a novel reduction from minimizing the total completion time to a natural makespan-like objective, which we call the weighted free time. We hope this free time objective will be useful in further improvements to this problem, as well as other stochastic scheduling problems.

Rudy Zhou, Anupam Gupta, Benjamin Moseley
Carnegie Mellon University
rbz@andrew.cmu.edu, anupamg@cs.cmu.edu, moseleyb@andrew.cmu.edu

CP17

Secretary Problems: The Power of a Single Sample

In this paper, we investigate two variants of the secretary problem. In these variants, we are presented with a sequence of numbers X_i that come from distributions \mathcal{D}_i , and that arrive in either random or adversarial order. We do not know what the distributions are, but we have access to a single sample Y_i from each distribution \mathcal{D}_i . After observing each number, we have to make an irrevocable decision about whether we would like to accept it or not with the goal of maximizing the probability of selecting the largest number. The random order version of this problem was first studied by Correa et al. [SODA 2020] who managed to construct an algorithm that achieves a probability of 0.4529. In this paper, we improve this probability to 0.5009, almost matching an upper bound of ≈ 0.5024 which we show follows from earlier work. We also show that there is an algorithm which achieves the probability of ≈ 0.5024 asymptotically if no particular distribution is especially likely to yield the largest number. For the adversarial order version of the problem, we show that we can select the maximum number with a probability of $1/4$, and

that this is best possible. Our work demonstrates that unlike in the case of the expected value objective studied by Rubinfeld et al. [ITCS 2020], knowledge of a single sample is not enough to recover the factor of success guaranteed by full knowledge of the distribution.

Pranav Nuti, Jan Vondrák
Stanford University
pranavn@stanford.edu, jvondrak@stanford.edu

CP17

Superpolynomial Lower Bounds for Decision Tree Learning and Testing

We establish new hardness results for decision tree optimization problems, adding to work that dates back to Hyafil and Rivest in 1976. We prove, under the randomized exponential time hypothesis, superpolynomial runtime lower bounds for two basic problems: given an explicit representation of a function f and a generator for a distribution \mathcal{D} , *construct* a small decision tree approximator for f under \mathcal{D} , and *decide* if there is a small decision tree approximator for f under \mathcal{D} . Our results imply new lower bounds for distribution-free PAC learning and testing of decision trees, settings in which the algorithm only has restricted access to f and \mathcal{D} . Specifically, we show: n -variable size- s decision trees cannot be properly PAC learned in time $n^{\tilde{O}(\log \log s)}$, and depth- d decision trees cannot be tested in time $\exp(d^{O(1)})$. For learning, the previous best lower bound only ruled out $\text{poly}(n)$ -time algorithms (Alekhnovich, Braverman, Feldman, Klivans, and Pitassi, 2009). For testing, recent work gives similar though incomparable lower bounds when f is random and \mathcal{D} is nonexplicit (Blais, Ferreira Pinto Jr., and Harms, 2021). Assuming a plausible conjecture on the hardness of Set-Cover, we improve our decision tree learning lower bound to $n^{\Omega(\log s)}$, matching Ehrenfeucht and Haussler's 1989 upper bound of $n^{O(\log s)}$.

Carmen Strassle, Caleb Koch, Li-Yang Tan
Stanford University
strassle@stanford.edu, ckoch@stanford.edu, lytan@stanford.edu

CP17

Lossless Online Rounding for Online Bipartite Matching (Despite Its Impossibility)

For numerous online bipartite matching problems, such as edge-weighted matching and matching under two-sided vertex arrivals, the state-of-the-art fractional algorithms outperform their randomized integral counterparts. This gap is surprising, given that the bipartite fractional matching polytope is integral, and so lossless rounding is possible. This gap was explained by Devanur et al. (SODA'13), who showed that online lossless rounding is impossible. Despite the above, we initiate the study of lossless online rounding for online bipartite matching problems. Our key observation is that while lossless online rounding is impossible in general, randomized algorithms induce fractional algorithms of the same competitive ratio which by definition are losslessly roundable online. This motivates the addition of constraints that decrease the "online integrality gap", thus allowing for lossless online rounding. We characterize a set of non-convex constraints which allow for such lossless online rounding, with better competitive ratios than deterministic algorithms. As applications of our lossless online rounding approach, we obtain two results of independent interest: (i) a doubly-exponential improvement,

and a sharp threshold for the amount of randomness (or advice) needed to outperform deterministic online bipartite matching algorithms, and (ii) an optimal semi-OCS, matching a recent result of Gao et al. (FOCS'21) answering a question of Fahrback et al. (FOCS'20).

Niv Buchbinder
Statistics and Operations Research Dept.
Tel Aviv University, Israel
niv.buchbinder@gmail.com

Joseph (Seffi) Naor
Technion
naor@cs.technion.ac.il

David Wajc
Google Research
david.wajc@gmail.com

CP18

A Distanced Matching Game, Decremental ApSP in Expanders, and Faster Deterministic Algorithms for Graph Cut Problems

We consider expander graphs and their use in fast deterministic algorithms for graph cut problems, and in algorithms for dynamic distance-based graph problems, such as APSP. A powerful toolkit of expander-based algorithmic techniques, including the Cut-Matching game, expander pruning, and algorithms for decremental APSP in expanders has led to many new results in these areas. Unfortunately, the use of expanders in these settings has several drawbacks. For example, the best current algorithm for decremental APSP in expanders only achieves a $(\log n)^{O(1/\epsilon^2)}$ -approximation with $n^{1+O(\epsilon)}$ total update time. All currently known algorithms for the Cut-Matching game are either randomized, or provide rather weak expansion guarantees. This, in turn, leads to somewhat weak algorithmic guarantees for several central cut problems: the best current almost linear time deterministic algorithms for Sparsest Cut only achieve $(\log n)^{\omega(1)}$ -approximation. Lastly, when relying on expanders in distance-based problems, such as dynamic APSP, it seems inevitable that an $\Omega(\log n)$ factor in the approximation must be lost. We propose to replace expanders in these settings with well-connected graphs, and provide an algorithmic toolkit for such graphs that mirrors the above mentioned tools for expanders. We demonstrate the power of these new tools by obtaining better results for several of the problems mentioned above.

Julia Chuzhoy
Toyota Technological Institute at Chicago
cjulia@ttic.edu

CP18

Shortest Cycles With Monotone Submodular Costs

We introduce the following submodular generalization of the Shortest Cycle problem. For a nonnegative monotone submodular cost function f defined on the edges (or the vertices) of an undirected graph G , we seek for a cycle C in G of minimum cost $OPT = f(C)$. We give an algorithm that given an n -vertex graph G , parameter $\epsilon > 0$, and the function f represented by an oracle, in time $n^{O(\log 1/\epsilon)}$ finds a cycle C in G with $f(C) \leq (1 + \epsilon) \cdot OPT$. This is in sharp contrast with the non-approximability of the closely related Monotone Submodular Shortest (s, t) -Path

problem, which requires exponentially many queries to the oracle for finding an $n^{2/3-\varepsilon}$ -approximation [Goel et al., FOCS 2009]. We complement our algorithm with a matching lower bound. We show that for every $\varepsilon > 0$, obtaining a $(1+\varepsilon)$ -approximation requires at least $n^{\Omega(\log 1/\varepsilon)}$ queries to the oracle. When the function f is integer-valued, our algorithm yields that a cycle of cost OPT can be found in time $n^{O(\log OPT)}$. In particular, for $OPT = n^{O(1)}$ this gives a quasipolynomial-time algorithm computing a cycle of minimum submodular cost. Interestingly, while a quasipolynomial-time algorithm often serves as a good indication that a polynomial time complexity could be achieved, we show a lower bound that $n^{O(\log n)}$ queries are required even when $OPT = O(n)$.

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

Petr Golovach
University of Bergen, Norway
petr.golovach@uib.no

Tuukka Korhonen
University of Bergen
tuukka.korhonen@uib.no

Daniel Lokshtanov
UCSB
daniello@ucsb.edu

Giannos Stamoulis
University of Montpellier and the National Center
for Scientific Research
giannos.stamoulis@lirmm.fr

CP18

Computing Square Colorings on Bounded-Treewidth and Planar Graphs

A square coloring of a graph G is a coloring of the square G^2 of G , that is, a coloring of the vertices of G such that any two vertices that are at distance at most 2 in G receive different colors. We investigate the complexity of finding a square coloring with a given number of q colors. We show that the problem is polynomial-time solvable on graphs of bounded treewidth by presenting an algorithm with running time $n^{2^{t+4}+O(1)}$ for graphs of treewidth at most t . The somewhat unusual exponent 2^t in the running time is essentially optimal: we show that for any $\epsilon > 0$, there is no algorithm with running time $f(t)n^{(2-\epsilon)^t}$ unless the Exponential-Time Hypothesis (ETH) fails. We also show that the square coloring problem is NP-hard on planar graphs for any fixed number $q \geq 4$ of colors. Our main algorithmic result is showing that the problem (when the number of colors q is part of the input) can be solved in subexponential time $2^{O(n^{2/3} \log n)}$ on planar graphs. The result follows from the combination of two algorithms. If the number q of colors is small ($\leq n^{1/3}$), then we can exploit a treewidth bound on the square of the graph to solve the problem in time $2^{O(\sqrt{qn} \log n)}$. If the number of colors is large ($\geq n^{1/3}$), then an algorithm based on protrusion decompositions and building on our result for the bounded-treewidth case solves the problem in time $2^{O(n \log n/q)}$.

Akanksha Agrawal

Indian Institute of Technology Madras
akanksha@cse.iitm.ac.in

Dániel Marx, Daniel Neuen
CISPA Helmholtz Center for Information Security
marx@cispa.de, dneuen@sfu.ca

Jasper Slusallek
Saarland University
s8jaslus@stud.uni-saarland.de

CP18

Packing Cycles in Planar and Bounded-Genus Graphs

We devise constant-factor approximation algorithms for finding as many disjoint cycles as possible from a certain family of cycles in a given planar or bounded-genus graph. Here disjoint can mean vertex-disjoint or edge-disjoint, and the graph can be undirected or directed. The family of cycles must be uncrossable and allow for an oracle access that finds a weight-minimal cycle in that family for given non-negative edge weights or (in planar graphs) the union of all remaining cycles in that family after deleting a given subset of edges. Our setting generalizes many problems that were studied separately in the past. For example, three families that satisfy the above properties are (i) all cycles in a directed or undirected graph, (ii) all odd cycles in an undirected graph, and (iii) all cycles in an undirected graph that contain precisely one demand edge, where the demand edges form a subset of the edge set. The latter family (iii) corresponds to the classical disjoint paths problem in fully planar and bounded-genus instances. While constant-factor approximation algorithms were known for edge-disjoint paths in such instances, we improve the constant in the planar case and obtain the first such algorithms for vertex-disjoint paths. We also obtain approximate min-max theorems of the Erdős-Pósa type. For example, the minimum feedback vertex set in a planar digraph is at most 12 times the maximum number of vertex-disjoint cycles.

Niklas Schlömborg
University of Bonn
schlomborg@or.uni-bonn.de

Hanjo Thiele
Greenplan GmbH, Bonn
hanjothiele@freenet.de

Jens Vygen
University of Bonn
vygen@or.uni-bonn.de

CP18

Excluding Single-Crossing Matching Minors in Bipartite Graphs

By a result of Valiant, computing the permanent of $(0,1)$ -matrices is, in general, #P-hard. In 1913 Pólya asked for which $(0,1)$ -matrices A it is possible to change some signs such that the permanent of A equals the determinant of the resulting matrix. In 1975, Little showed these matrices to be exactly the biadjacency matrices of bipartite graphs excluding $K_{3,3}$ as a matching minor. This was turned into a polynomial time algorithm by McCuaig, Robertson, Seymour, and Thomas in 1999. Recently it was shown that the exclusion of any planar bipartite graph as a matching minor yields a class of bipartite graphs on which the per-

manent of the corresponding $(0,1)$ -matrices can be computed efficiently. In this paper we unify the two results above into a single, more general result in the style of the celebrated structure theorem for single-crossing minor-free graphs. We identify a class of bipartite graphs strictly generalising planar bipartite graphs and $K_{3,3}$ which includes infinitely many non-Pfaffian graphs. The exclusion of any member of this class as a matching minor yields a structure that allows for the efficient evaluation of the permanent. Moreover, we show that the evaluation of the permanent remains $\#P$ -hard on bipartite graphs which exclude $K_{5,5}$ as a matching minor. This establishes a first computational lower bound for the problem of counting perfect matchings on matching minor closed classes.

Sebastian Wiederrecht

Institute for Basic Science, DIMAG
sebastian.wiederrecht@gmail.com

Archontia Giannopoulou

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

Dimitrios Thilikos

CNRS, LIRMM, Montpellier
sedthilk@thilikos.info

CP18

A Framework for Approximation Schemes on Disk Graphs

In this paper, we initiate a systematic study of approximation schemes for fundamental optimization problems on *disk graphs*, a common generalization of both planar graphs and unit-disk graphs. Our main contribution is a general framework for designing efficient polynomial-time approximation schemes (EPTASes) for vertex-deletion problems on disk graphs, which results in EPTASes for many fundamental problems including Vertex Cover, Feedback Vertex Set, Small Cycle Hitting (in particular, Triangle Hitting), P_k -Vertex Deletion for $k \in \{3, 4, 5\}$, Path Deletion, Pathwidth 1-Deletion, Component Order Connectivity, Bounded Degree Deletion, Pseudoforest Deletion, Finite-Type Component Deletion, etc. All EPTASes obtained using our framework are *robust* in the sense that they do not require a realization of the input disk graph (in fact, we allow the input to be any graph, and our algorithms either output a correct approximation solution for the problem or conclude that the input graph is not a disk graph). To the best of our knowledge, prior to this work, the only problems known to admit PTASes or EPTASes on disk graphs are Maximum Clique, Independent Set, Dominating set, and Vertex Cover, among which the existing PTAS [Erlebach et al., SICOMP'05] and EPTAS [Leeuwen, SWAT'06] for Vertex Cover require a realization of the input disk graph (while ours does not).

Jie Xue

New York University Shanghai
jiexue@nyu.edu

Daniel Lokshtanov

UCSB
daniello@ucsb.edu

Fahad Panolan

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

Saket Saurabh

IMSc +UiB
saket@imsc.res.in

Meirav Zehavi

BGU
meiravze@bgu.ac.il

CP19

Fast Algorithms for Solving the Hamilton Cycle Problem with High Probability

We study the Hamilton cycle problem with input a random graph $G \sim G(n, p)$ in two different settings. In the first one, G is given to us in the form of randomly ordered adjacency lists while in the second one, we are given the adjacency matrix of G . In each of the two settings we derive a deterministic algorithm that w.h.p. either finds a Hamilton cycle of G or returns a certificate that such a cycle does not exist for $p = p(n) \geq 0$. The running times of our algorithms are $O(n)$ and $O(\frac{n}{p})$ respectively, each being best possible in its own setting.

Michael Anastos

Institute of Science and Technology Austria
michael.anastos@ist.ac.at

CP19

Approximate Graph Colouring And Crystals

We show that approximate graph colouring is not solved by any level of the affine integer programming (AIP) hierarchy. To establish the result, we translate the problem of exhibiting a graph fooling a level of the AIP hierarchy into the problem of constructing a highly symmetric crystal tensor. In order to prove the existence of crystals in arbitrary dimension, we provide a combinatorial characterisation for realisable systems of tensors; i.e., sets of low-dimensional tensors that can be realised as the projections of a single high-dimensional tensor.

Lorenzo Ciardo

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

Stanislav ivn

University of Oxford
standa.zivny@cs.ox.ac.uk

CP19

Small Subgraphs with Large Average Degree

In this paper we study the fundamental problem of finding small dense subgraphs in a given graph. For a real number $s > 2$, we prove that every graph on n vertices with average degree at least d contains a subgraph of average degree at least s on at most $nd^{-\frac{s-2}{s-1}} (\log d)^{O_s(1)}$ vertices. This is optimal up to the polylogarithmic factor, and resolves a conjecture of Feige and Wagner.

Oliver Janzer, Benny Sudakov

ETH Zurich
janzerol94@gmail.com, benjamin.sudakov@math.ethz.ch

István Tomon

Umea University

istvan.tomon@umu.se

CP19

Improved Girth Approximation in Weighted Undirected Graphs

Let $G = (V, E, \ell)$ be a n -nodes m -edges weighted undirected graph, where $\ell : E \rightarrow (0, \infty)$ is a real *length* function defined on its edges. Let g be the length of the shortest cycle in G . We present an algorithm that in $O(kn^{1+1/k} \log n + m(k + \log n))$ expected running time finds a cycle of length at most $\frac{4k}{3}g$, for every integer $k \geq 1$. This improves upon the previous best algorithm that in $O((n^{1+1/k} \log n + m) \log(nM))$ time, where $\ell : E \rightarrow [1, M]$ is an integral length function, finds a cycle of length at most $2kg$ [?]. For $k = 1$ our algorithm also improves the result of Roditty and Tov [?].

Avi Katria

Bar-Ilan University
avi.katria3@gmail.com

Liam Roditty

Bar Ilan University
liam.roditty@biu.ac.il

Aaron Sidford

Stanford University
sidford@stanford.edu

Virginia Vassilevska Williams

MIT
virg@mit.edu

Uri Zwick

Tel Aviv University
zwick@tau.ac.il

CP19

A Simple and Sharper Proof of the Hypergraph Moore Bound

The hypergraph Moore bound is an elegant statement that characterizes the extremal trade-off between the girth - the number of hyperedges in the smallest cycle or even cover (a subhypergraph with all degrees even) and size - the number of hyperedges in a hypergraph. For graphs (i.e., 2-uniform hypergraphs), a bound tight up to the leading constant was proven in a classical work of Alon, Hoory and Linial (AHL). For hypergraphs of uniformity $k > 2$, an appropriate generalization was conjectured by Feige. The conjecture was settled up to an additional $\log^{4k+1} n$ factor in the size in a recent work of Guruswami, Kothari and Manohar (GKM). Their argument relies on a connection between the existence of short even covers and the spectrum of a certain randomly signed Kikuchi matrix. Their analysis, especially for the case of odd k , is significantly complicated. In this work, we present a substantially simpler and shorter proof of the hypergraph Moore bound, which loses only a single logarithmic factor for all $k > 2$ -uniform hypergraphs. As in GKM, our ideas naturally extend to yield a simpler proof of the full trade-off for strongly refuting smoothed instances of constraint satisfaction problems with similarly improved parameters.

Jun-Ting Hsieh, Pravesh K. Kothari

Carnegie Mellon University
juntingh@cs.cmu.edu, praveshk@cs.cmu.edu

Sidhanth Mohanty

University of California Berkeley
sidhanthm@berkeley.edu

CP19

Weisfeiler-Leman and Graph Spectra

Two simple undirected graphs are cospectral if their respective adjacency matrices have the same multiset of eigenvalues. Cospectrality yields an equivalence relation on the family of graphs which is provably weaker than isomorphism. In this paper, we study cospectrality in relation to another well-studied relaxation of isomorphism, namely k -dimensional Weisfeiler-Leman (k -WL) indistinguishability. Cospectrality with respect to standard graph matrices such as the adjacency or the Laplacian matrix yields a strictly finer equivalence relation than 2-WL indistinguishability. We show that individualising one vertex plus running 1-WL already subsumes cospectrality with respect to all such graph matrices. Building on this result, we resolve an open problem of Fürer (2010) on spectral invariants and strengthen a result of Godsil (1981) on commute distances. Looking beyond 2-WL, we devise a hierarchy of graph matrices generalising the adjacency matrix such that k -WL indistinguishability after a fixed number of iterations can be captured as a spectral condition on these matrices. Precisely, we provide a spectral characterisation of k -WL indistinguishability after d iterations, for $k, d \in \mathbb{N}$. Our results can be viewed as characterisations of homomorphism indistinguishability over certain graph classes in terms of matrix equations.

Tim Seppelt

RWTH Aachen University
seppelt@cs.rwth-aachen.de

Gaurav Rattan

RWTH Aachen University
rattan@informatik.rwth-aachen.de

CP20

Approximate Minimum Cuts and Their Enumeration

We show that every α -approximate minimum cut in a connected graph is the unique minimum (S, T) -terminal cut for some subsets S and T of vertices each of size at most $\lfloor 2\alpha \rfloor + 1$. This leads to an alternative proof that the number of α -approximate minimum cuts in a n -vertex connected graph is $n^{O(\alpha)}$ and they can all be enumerated in deterministic polynomial time for constant α .

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

CP20

A Simple Framework for Finding Balanced Sparse Cuts via APSP

We present a very simple and intuitive algorithm to find balanced sparse cuts in a graph via shortest-paths. Our algorithm combines a new multiplicative-weights framework for solving unit-weight multi-commodity flows with standard ball growing arguments. Using Dijkstra's algorithm for computing the shortest paths afresh every time gives a very simple algorithm that runs in time $\tilde{O}(m^2/\phi)$ and finds an $\tilde{O}(\phi)$ -sparse balanced cut, when the given graph has a ϕ -sparse balanced cut. Combining our algorithm with known deterministic data-structures for answering approximate All Pairs Shortest Paths (APSP) queries under increasing edge weights (decremental setting), we obtain a simple deterministic algorithm that finds $m^{o(1)}\phi$ -sparse balanced cuts in $m^{1+o(1)}/\phi$ time. Our deterministic almost-linear time algorithm matches the state-of-the-art in randomized and deterministic settings up to subpolynomial factors, while being significantly simpler to understand and analyze, especially compared to the only almost-linear time deterministic algorithm, a recent breakthrough by Chuzhoy-Gao-Li-Nanongkai-Peng-Saranurak (FOCS 2020).

Li Chen

Georgia Institute of Technology
lichen@gatech.edu

Rasmus Kyng, Maximilian Probst Gutenberg
ETH Zurich
kyng@inf.ethz.ch, maxprobst@inf.ethz.ch

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

CP20

A Simple Algorithm for Submodular Minimum Linear Ordering

We consider the Minimum Linear Ordering Problem (MLOP): given a ground set N (of size n) and a set function $f : 2^N \rightarrow \mathbb{R}_+$ the goal is to find a permutation π over N that minimizes $\sum_{i=1}^{n-1} f(N_{\pi,i})$, where $N_{\pi,i}$ is the collection of the first i elements from N with respect to the permutation π . MLOP has been studied for various types of set functions, capturing multiple application, e.g., Minimum Linear Arrangement, Minimum Sum Vertex Cover, and Minimum Sum Set Cover. Surprisingly, for the general case of a symmetric submodular f an approximation of only $O(n)$ is known and it is achieved by any permutation π . In this paper we present a simple $O(\sqrt{n}/\ln n)$ approximation for MLOP for the case f is symmetric and submodular.

Dor Katzelnick, Roy Schwartz

Technion - Israel Institute of Technology
DOR767@gmail.com, schwartz@cs.technion.ac.il.

CP20

A Local Search-Based Approach for Set Covering

In the Set Cover problem, we are given a set system with each set having a weight, and we want to find a collec-

tion of sets that cover the universe, whilst having low total weight. There are several approaches known (based on greedy approaches, relax-and-round, and dual-fitting) that achieve a $H_k \approx \ln k + O(1)$ approximation for this problem, where the size of each set is bounded by k . Moreover, getting a $\ln k - O(\ln \ln k)$ approximation is hard. Where does the truth lie? Can we close the gap between the upper and lower bounds? An improvement would be particularly interesting for small values of k , which are often used in reductions between Set Cover and other combinatorial optimization problems. We consider a non-oblivious local-search approach: to the best of our knowledge this gives the first H_k -approximation for Set Cover using an approach based on local-search. Our proof fits in one page, and gives a integrality gap result as well. Refining our approach by considering larger moves and an optimized potential function gives an $(H_k - \Omega(\log^2 k)/k)$ -approximation, improving on the previous bound of $(H_k - \Omega(1/k^8))$ (R. Hassin and A. Levin, SICOMP '05) based on a modified greedy algorithm.

Jason Li

Simons Institute, UC Berkeley
jmli@alummi.cmu.edu

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

Euiwoong Lee

University of Michigan
euiwoong@umich.edu

CP20

Minimum Cost Adaptive Submodular Cover

We consider the problem of minimum cost cover of adaptive-submodular functions, and provide a $4(\ln Q + 1)$ -approximation algorithm, where Q is the goal value. This bound is nearly the best possible as the problem does not admit any approximation ratio better than $\ln Q$ (unless $P = NP$). Our result is the first $O(\ln Q)$ -approximation algorithm for this problem. Previously, $O(\ln Q)$ -approximation algorithms were only known assuming either independent items or unit-cost items. Furthermore, our result easily extends to the setting where one wants to simultaneously cover multiple adaptive-submodular functions: we obtain the first approximation algorithm for this generalization.

Viswanath Nagarajan, Yubing Cui

University of Michigan
viswa@umich.edu, yubingc@umich.edu

CP20

Revisiting Garg's 2-Approximation Algorithm for the k -MST Problem in Graphs

This paper revisits the 2-approximation algorithm for k -MST presented by Garg [STOC '05] in light of a recent paper of Paul et al. [Math. Oper. Res. '20]. In the k -MST problem, the goal is to return a tree spanning k vertices of minimum total edge cost. Paul et al. [Math. Oper. Res. '20] extend Garg's primal-dual subroutine to improve the approximation ratios for the budgeted prize-collecting traveling salesman and minimum spanning tree problems. We follow their algorithm and analysis to provide a cleaner version of Garg's result. Additionally, we introduce the novel concept of a kernel which allows an

easier visualization of the stages of the algorithm and a clearer understanding of the pruning phase. Other notable updates include presenting a linear programming formulation of the k -MST problem, including pseudocode, replacing the coloring scheme used by Garg with the simpler concept of neutral sets, and providing an explicit potential function.

Emmett Breen, Renee Mirka, Zichen Wang, David P. Williamson
Cornell University
ejb284@cornell.edu, rem379@cornell.edu, zw336@cornell.edu, davidpwilliamson@cornell.edu

CP21

Steiner Connectivity Augmentation and Splitting-off in Poly-Logarithmic Maximum Flows

We give an almost-linear time algorithm for the Steiner connectivity augmentation problem: given an undirected graph, find a smallest (or minimum weight) set of edges whose addition makes a given set of terminals τ -connected (for any given $\tau > 0$). The running time of our algorithm is dominated by polylogarithmic calls to any maximum flow subroutine; using the recent almost-linear time maximum flow algorithm (Chen et al., FOCS 2022), we get an almost-linear running time for our algorithm as well. This is tight up to the polylogarithmic factor even for just two terminals. Prior to our work, an almost-linear (in fact, near-linear) running time was known only for the special case of global connectivity augmentation, i.e., when all vertices are terminals (Cen et al., STOC 2022). We also extend our algorithm to the closely related Steiner splitting-off problem, where the edges incident on a vertex have to be split-off while maintaining the (Steiner) connectivity of a given set of terminals. Prior to our work, a nearly-linear time algorithm was known only for the special case of global connectivity (Cen et al., STOC 2022). The only known generalization beyond global connectivity was to preserve all pairwise connectivities using a much slower algorithm that makes n calls to an all-pairs maximum flow (or Gomory-Hu tree) subroutine (Lau and Yung, SICOMP 2013), as against $\text{polylog}(n)$ calls to a (single-pair) maximum flow subroutine in this work.

Ruoxu Cen, William He
Duke University
ruoxu.cen@duke.edu, william.he@duke.edu

Jason Li
Simons Institute, UC Berkeley
jmli@alumni.cmu.edu

Debmalya Panigrahi
Duke University
debmalya@cs.duke.edu

CP21

Timeliness Through Telephones: Approximating Information Freshness in Vector Clock Models

We consider an information dissemination problem where the root node in an undirected graph constantly updates its information. The goal is to keep every other node in the graph as freshly informed about the root as possible. Our synchronous information spreading model uses telephone calls at each time step, in which any node can communicate with at most one neighbor, thus forming a matching over which information is transmitted at each step. We in-

troduce two problems in minimizing two natural objectives (Maximum and Average) of the latency of the root's information at all nodes in the network. After deriving a simple reduction from the maximum rooted latency problem to the well-studied minimum broadcast time problem, we focus on the average rooted latency version. We introduce a natural problem of finding a finite schedule that minimizes the average broadcast time from a root. We show that these two problems are within a logarithmic factor of each other. Then, we derive a log-squared approximation for the average broadcast time problem via rounding a time-indexed linear programming relaxation, resulting in a log-cubed approximation for the average latency problem. Surprisingly, we show that using the average broadcast time for average rooted latency introduces a necessary logarithmic factor overhead even in trees. We overcome this hurdle and give a 40-approximation for trees by exploiting its structure and constructing a locally-periodic schedule.

Da Qi Chen
University of Virginia
wny7gj@virginia.edu

Lin An, Aidin Niaparast, R Ravi, Oleksandr Rudenko
Carnegie Mellon University
linan@andrew.cmu.edu, aniapara@andrew.cmu.edu, ravi@andrew.cmu.edu, orudenko@andrew.cmu.edu

CP21

Improved Approximation for Two-Edge-Connectivity

The basic goal of survivable network design is to construct low-cost networks which preserve a sufficient level of connectivity despite the failure or removal of a few nodes or edges. One of the most basic problems in this area is the 2-Edge-Connected Spanning Subgraph problem (2-ECSS): given an undirected graph G , find a 2-edge-connected spanning subgraph H of G with the minimum number of edges (in particular, H remains connected after the removal of one arbitrary edge). 2-ECSS is NP-hard and the best-known (polynomial-time) approximation factor for this problem is $4/3$. Interestingly, this factor was achieved with drastically different techniques by [Hunkenschroder, Vempala and Vetta '00, '19] and [Sebo and Vygen, '14]. In this paper we present an improved $\frac{118}{89} + \epsilon < 1.326$ approximation for 2-ECSS. The key ingredient in our approach (which might also be helpful in future work) is a reduction to a special type of structured graphs: our reduction preserves approximation factors up to $6/5$. While reducing to 2-vertex-connected graphs is trivial (and heavily used in prior work), our structured graphs are "almost" 3-vertex-connected: more precisely, given any 2-vertex-cut $\{u, v\}$ of a structured graph $G = (V, E)$, $G[V \setminus \{u, v\}]$ has exactly 2 connected components, one of which contains exactly one node of degree 2 in G .

Afrouz Jabal Ameli
Eindhoven University of Technology
a.jabal.ameli@tue.nl

Mohit Garg
Eindhoven University of Technology
University of Bremen
garg@uni-bremen.de

Fabrizio Grandoni
IDSIA

fabrizio@idsia.ch

CP21**Exact Flow Sparsification Requires Unbounded Size**

Given a large edge-capacitated network G and a subset of k vertices called terminals, an (exact) flow sparsifier is a small network G' that preserves (exactly) all multi-commodity flows that can be routed between the terminals. Flow sparsifiers were introduced by Leighton and Moitra [STOC 2010], and have been studied and used in many algorithmic contexts. A fundamental question that remained open for over a decade, asks whether every k -terminal network admits an exact flow sparsifier whose size is bounded by some function $f(k)$ (regardless of the size of G or its capacities). We resolve this question in the negative by proving that there exist 6-terminal networks G whose flow sparsifiers G' must have arbitrarily large size. This unboundedness is perhaps surprising, since the analogous sparsification that preserves all terminal cuts (called exact cut sparsifier or mimicking network) admits sparsifiers of size $f_0(k) \leq 2^{2^k}$ [Hagerup, Katajainen, Nishimura, and Ragde, JCSS 1998]. We prove our results by analyzing the set of all feasible demands in the network, known as the demand polytope. We identify an invariant of this polytope, essentially the slope of certain facets, that can be made arbitrarily large even for $k = 6$, and implies an explicit lower bound on the size of the network.

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

Ron Mosenzon
Toyota Technological Institute of Science
ron.mosenzon@gmail.com

CP21**Faster Computation of 3-Edge-Connected Components in Digraphs**

We present an $\tilde{O}(m^{3/2})$ time randomized (Monte Carlo) algorithm for computing the 3-edge-connected components of a digraph with m edges and n vertices. This constitutes the first improvement since the algorithm of Nagamochi Watanabe from 1993, which runs in $O(mn)$ time. Thus, our algorithm is the first that overcomes the run-time of $O(n)$ computations of 3-bounded max-flows. Our algorithm involves a combination of known and new techniques together with new structural insights on the interactions between directed min-cuts. One novel aspect that we introduce is an efficient graph operation for replacing a set of vertices S that is disconnected from $V \setminus S$ by an edge-cut of size 2 (2-out set), with a gadget of small size that preserves the pairwise connectivity among the vertices of $V \setminus S$. Another main ingredient of our approach is an extension of the framework for computing the vertex-connectivity (or edge-connectivity) in a digraph [Nanongkai et al., STOC'19]. This extension allows us to efficiently identify either all small 2-out sets of vertices, or identify enough 2-out sets whose total internal volume is a constant fraction of the edges of the graph. We believe that our techniques may be of independent interest. Finally, we augment our algorithm with a data structure that can report in constant time the edges of some edge-cut of size at most 2 that disconnects any two query vertices u, v , or report in constant time that no such edge-cut exists.

Loukas Georgiadis
University of Ioannina, Greece
loukas@uoi.gr

Evangelos Kipouridis
BARC (Basic Algorithms Research Copenhagen)
University of Copenhagen
kipouridis@cs.uni-saarland.de

Charis Papadopoulos
University of Ioannina, Greece
charis@uoi.gr

Nikos Parotsidis
Google Research
nikosp@google.com

CP21**Approximation Algorithms for Steiner Tree Augmentation Problems**

In the Steiner Tree Augmentation Problem (STAP), we are given a graph $G = (V, E)$, a set of terminals $R \subseteq V$, and a Steiner tree T spanning R . The edges $L := E \setminus E(T)$ are called links and have non-negative costs. The goal is to augment T by adding a minimum cost set of links, so that there are 2 edge-disjoint paths between each pair of vertices in R . This problem is a special case of the Survivable Network Design Problem which can be approximated to within a factor of 2 using iterative rounding [Jain 2001]. We give the first polynomial time algorithm for STAP with approximation ratio better than 2. In particular we achieve an approximation ratio of $(1.5 + \epsilon)$. To do this, we employ the Local Search approach of Traub and Zenklusen for the Tree Augmentation Problem and generalize their main decomposition theorem from links (of size two) to hyper-links. We also consider the Node-Weighted Steiner Tree Augmentation Problem (NW-STAP) in which the non-terminal nodes have non-negative costs. We seek a cheapest subset $S \subseteq V \setminus R$ so that $G[R \cup S]$ is 2-edge-connected. Using a result of Nutov, there exists an $O(\log |R|)$ -approximation for this problem. We provide a $O(\log^2(|R|))$ -approximation algorithm using a greedy algorithm leveraging the spider decomposition of optimal solutions.

R Ravi
Carnegie Mellon University
ravi@andrew.cmu.edu

Weizhong Zhang, Michael Zlatin
Carnegie Mellon University
Tepper School of Business
weizhong@andrew.cmu.edu, mzlatin@andrew.cmu.edu

CP22**Improved Distributed Network Decomposition, Hitting Sets, and Spanners, via Derandomization**

This paper presents significantly improved deterministic algorithms for some of the key problems in the area of distributed graph algorithms, including network decomposition, hitting sets, and spanners. As the main ingredient in these results, we develop novel randomized distributed algorithms that we can analyze using only pairwise independence, and we can thus derandomize efficiently. As our most prominent end-result, we obtain a deterministic

construction for $O(\log n)$ -color $O(\log n \cdot \log \log \log n)$ -strong diameter network decomposition in $\tilde{O}(\log^3 n)$ rounds. This is the first construction that achieves almost $\log n$ in both parameters, and it improves on a recent line of exciting progress on deterministic distributed network decompositions [Rozhon, Ghaffari STOC'20; Ghaffari, Grunau, Rozhon SODA'21; Chang, Ghaffari PODC'21; Elkin, Haeupler, Rozhon, Grunau FOCS'22].

Saeed Ilchi
ETH Zurich
saeed.ilchi@inf.ethz.ch

Mohsen Ghaffari
MIT
ghaffari@mit.edu

Christoph Grunau
ETH Zurich
Switzerland
cgrunau@inf.ethz.ch

Bernhard Haeupler
ETH Zurich
bernhard.haeupler@inf.ethz.ch

Václav Rozhon
Carnegie Mellon University & ETH Zurich
rozhonv@inf.ethz.ch

CP22

Optimal Fully Dynamic K-Center Clustering for Adaptive and Oblivious Adversaries

In this paper, we resolve the complexity of fully dynamic k -center clustering against both adaptive and oblivious adversaries. Against oblivious adversaries, we present the first algorithm for fully dynamic k -center in an arbitrary metric space that maintains an optimal $(2 + \epsilon)$ -approximation in $O(k \cdot \text{polylog}(n, \Delta))$ amortized update time. Here, n is an upper bound on the number of active points at any time, and Δ is the aspect ratio of the metric space. Previously, the best known amortized update time was $O(k^2 \cdot \text{polylog}(n, \Delta))$, and is due to Chan, Gourqin, and Sozio (2018). Moreover, we demonstrate that our runtime is optimal up to $\text{polylog}(n, \Delta)$ factors. For adaptive adversaries, we give the first deterministic algorithm for fully dynamic k -center which achieves a $O\left(\min\left\{\frac{\log(n/k)}{\log \log n}, k\right\}\right)$ approximation in $O(k \cdot \text{polylog}(n, \Delta))$ amortized update time. We demonstrate that any algorithm which achieves a $O\left(\min\left\{\frac{\log n}{k \log f(k, 2n)}, k\right\}\right)$ -approximation against adaptive adversaries requires $f(k, n)$ update time, for any arbitrary function f . Thus, in the regime where $k = O(\sqrt{\frac{\log n}{\log \log n}})$, we close the complexity of the problem up to $\text{polylog}(n, \Delta)$ factors in the update time.

MohammadHossein Bateni
Google Research
bateni@google.com

Hossein Esfandiari
esfandiari@google.com
bateni@google.com

Hendrik Fichtenberger
Google Research

fichtenberger@google.com

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

Rajesh Jayaram, Vahab Mirrokni
Google Research
rkjayaram@google.com, mirrokni@gmail.com

Andreas Wiese
Technical University of Munich
andreas.wiese@tum.de

CP22

Optimal Deterministic Massively Parallel Connectivity on Forests

We show fast deterministic algorithms for fundamental problems on forests in the challenging low-space regime of the well-known Massive Parallel Computation (MPC) model. A recent breakthrough result by Coy and Czumaj [STOC'22] shows that, in this setting, it is possible to deterministically identify connected components on graphs in $O(\log D + \log \log n)$ rounds, where D is the diameter of the graph and n the number of nodes. The authors left open a major question: is it possible to get rid of the additive $\log \log n$ factor and deterministically identify connected components in a runtime that is completely independent of n ? In this paper we answer the above question in the affirmative in the case of forests. We show that, on forests where each tree has diameter at most D , we can solve the connected components problem in just $O(\log D)$ deterministic rounds. Our algorithm is component-stable, and the total memory required is $O(n + m)$ words, where m is the number of edges in the input graph, which is optimal as it is only enough to store the input graph. We complement our upper bound results by showing that $\Omega(\log D)$ time is necessary even for component-unstable algorithms, conditioned on the widely believed 1 vs. 2 cycles conjecture. By slightly modifying our connected components algorithm, we get a novel deterministic forest-rooting algorithm with the same runtime and memory bounds.

Alkida Balliu
Gran Sasso Science Institute
alkida.balliu@gssi.it

Rustam Latypov
Aalto University
rustam.latypov@aalto.fi

Yannic Maus
TU Graz
yannic.maus@ist.tugraz.at

Dennis Olivetti
Gran Sasso Science Institute
dennis.olivetti@gssi.it

Jara Uitto
Aalto University
jara.uitto@aalto.fi

CP22

A New Approach to Estimating Effective Resistances and Counting Spanning Trees in Expander

Graphs

We demonstrate that for expander graphs, for all $\epsilon > 0$, there exists a data structure of size $\tilde{O}(n\epsilon^{-1})$ which can be used to return $(1 + \epsilon)$ -approximations to effective resistances in $\tilde{O}(1)$ time per query. Our construction is based on two key ideas: 1) ϵ^{-1} -sparse, ϵ -additive approximations to DL^+1_u for all u , can be used to recover $(1 + \epsilon)$ -approximations to the effective resistances, 2) In expander graphs, only $\tilde{O}(\epsilon^{-1})$ coordinates of DL^+1_u are larger than ϵ . We give an efficient construction for such a data structure in $\tilde{O}(m + n\epsilon^{-2})$ time via random walks. This results in an algorithm for computing $(1 + \epsilon)$ -approximate effective resistances for s vertex pairs in expanders that runs in $\tilde{O}(m + n\epsilon^{-2} + s)$ time. We employ the above algorithm to compute a $(1 + \delta)$ -approximation to the number of spanning trees in an expander graph in $\tilde{O}(m + n^{1.5}\delta^{-1})$ time. This improves on the previously best known result of $m^{1+o(1)} + n^{1.875+o(1)}\delta^{-1.75}$ time, and matches the best known size of determinant sparsifiers.

Lawrence Li, Sushant Sachdeva
University of Toronto
lawrenceliu@gmail.com, sachdeva@cs.toronto.edu

CP22

Fast Distributed Brooks Theorem

We give a randomized Δ -coloring algorithm in the LOCAL model that runs in $\text{poly log log } n$ rounds, where n is the number of nodes of the input graph and Δ is its maximum degree. This means that randomized Δ -coloring is a rare distributed coloring problem with an upper and lower bound in the same ballpark, $\text{poly log log } n$, given the known $\Omega(\log_\Delta \log n)$ lower bound [Brandt et al., STOC '16]. Our main technical contribution is a constant time reduction to a constant number of $(deg + 1)$ -list coloring instances, for $\Delta = \Omega(\log^4 n)$, resulting in a $\text{poly log log } n$ -round CONGEST algorithm for such graphs. This reduction is of independent interest for other settings, including providing a new proof of Brooks theorem for high degree graphs, and leading to a constant-round Congested Clique algorithm in such graphs. When $\Delta = \Omega(\log^{22} n)$, our algorithm even runs in $O(\log^* n)$ rounds, showing that the base in the $\Omega(\log_\Delta \log n)$ lower bound is unavoidable. Previously, the best LOCAL algorithm for all considered settings used a logarithmic number of rounds. Our result is the first CONGEST algorithm for Δ -coloring non-constant degree graphs.

Yannic Maus
TU Graz
yannic.maus@ist.tugraz.at

Magnús Halldórsson
Reykjavik University
magnusmh@gmail.com

Manuela Fischer
ETH Zurich
manuela.fischer@inf.ethz.ch

CP22

Distributed Maximal Matching and Maximal Independent Set on Hypergraphs

We investigate the distributed complexity of maximal matching and maximal independent set (MIS) in hyper-

graphs in the LOCAL model. A maximal matching of a hypergraph $H = (V_H, E_H)$ is a maximal disjoint set $M \subseteq E_H$ of hyperedges and an MIS $S \subseteq V_H$ is a maximal set of nodes such that no hyperedge is fully contained in S . Both problems can be solved by a simple sequential greedy algorithm, which can be implemented naively in $O(\Delta r + \log^* n)$ rounds, where Δ is the maximum degree, r is the rank, and n is the number of nodes of the hypergraph. We show that for maximal matching, this naive algorithm is optimal in the following sense. Any deterministic algorithm for solving the problem requires $\Omega(\min\{\Delta r, \log_{\Delta r} n\})$ rounds, and any randomized one requires $\Omega(\min\{\Delta r, \log_{\Delta r} \log n\})$ rounds. Hence, for any algorithm with a complexity of the form $O(f(\Delta, r) + g(n))$, we have $f(\Delta, r) \in \Omega(\Delta r)$ if $g(n)$ is not too large. For the MIS problem on hypergraphs, we show that for $\Delta \ll r$, there are significant improvements over the naive $O(\Delta r + \log^* n)$ -round algorithm.

Dennis Olivetti, Alkida Balliu
Gran Sasso Science Institute
dennis.olivetti@gssi.it, alkida.balliu@gssi.it

Sebastian Brandt
CISPA Helmholtz Center for Information Security
brandt@cispa.de

Fabian Kuhn
University of Freiburg, Germany
kuhn@cs.uni-freiburg.de

CP23

Weak Bisimulation Finiteness of Pushdown Systems With Deterministic ϵ -Transitions Is 2-ExpTime-Complete

We consider the problem of deciding whether a given pushdown system all of whose ϵ -transitions are deterministic is weakly bisimulation finite, that is, whether it is weakly bisimulation equivalent to a finite system. We prove that this problem is 2-ExpTime-complete. This consists of three elements: First, we prove that the smallest finite system that is weakly bisimulation equivalent to a fixed pushdown system, if exists, has size at most doubly exponential in the description size of the pushdown system. Second, we propose a fast algorithm deciding whether a given pushdown system is weakly bisimulation equivalent to a finite system of a given size. Third, we prove 2-ExpTime-hardness of the problem. The problem was known to be decidable, but the previous algorithm had Ackermannian complexity (6-ExpSpace in the easier case of pushdown systems without ϵ -transitions); concerning lower bounds, only ExpTime-hardness was known.

Stefan Göller
University of Kassel
stefan.goeller@uni-kassel.de

Paweł Parys
University of Warsaw
parys@mimuw.edu.pl

CP23

Query Complexity of Inversion Minimization on Trees

We consider the following computational problem: Given a rooted tree and a ranking of its leaves, what is the minimum number of inversions of the leaves that can be attained by ordering the tree? We study the complexity of

the problem in the comparison-query model, used for problems like sorting and selection. For many tree shapes, we establish lower bounds close to the lower bound of $\log_2(n!)$ for sorting n items. For trees with n leaves we show: (a) $\log_2((\frac{k}{(k+1)^2}n!)) - O(\log n)$ queries are needed for trees of degree k . (b) $\log_2(n!) - O(\log n)$ queries are needed in case the tree is binary. (c) $\log_2(n!) - O(k \log k)$ queries are needed for certain classes of trees of degree k , including perfect trees with even k . The lower bounds are obtained by developing two novel techniques for a generic problem Π in the comparison-query model, and applying them to inversion minimization on trees. Both techniques can be described in terms of the Cayley graph of the symmetric group with adjacent-rank transpositions as the generating set. Consider the subgraph consisting of the edges between vertices with the same value under Π . We show that the size of any decision tree for Π must be at least: (i) the number of connected components of this subgraph, and (ii) the factorial of the average degree of the complementary subgraph, divided by n . Lower bounds on query complexity then follow by taking the base-2 logarithm.

Ivan L. Hu, Dieter Van Melkebeek, Andrew Morgan
University of Wisconsin-Madison
ilhu@wisc.edu, dieter@cs.wisc.edu, amorgan@cs.wisc.edu

CP23

Efficient Resilient Functions

An n -bit boolean function is resilient to coalitions of size q if no fixed set of q bits is likely to influence the value of the function when the other $n - q$ bits are chosen uniformly at random, even though the function is nearly balanced. We construct explicit functions resilient to coalitions of size $q = n/(\log n)^{O(\log \log n)} = n^{1-o(1)}$ computable by linear-size circuits and linear-time algorithms. We also obtain a tight size-depth tradeoff for computing such resilient functions. Constructions such as ours were not available even non-explicitly. It was known that functions resilient to coalitions of size $q = n^{0.63\dots}$ can be computed by linear-size circuits [BL85], and functions resilient to coalitions of size $q = \Theta(n/\log^2 n)$ can be computed by quadratic-size circuits [AL93]. One component of our proofs is a new composition theorem for resilient functions.

Peter Ivanov
Northeastern University
ivanov.p@northeastern.edu

Raghu Meka
UCLA
raghuvardhan@gmail.com

Emanuele Viola
Northeastern University
viola@ccs.neu.edu

CP23

Algorithmizing the Multiplicity Schwartz-Zippel Lemma

The multiplicity Schwartz-Zippel (MSZ) lemma asserts that over a field, a low-degree polynomial cannot vanish with high multiplicity very often on a sufficiently large product set. Since its discovery in a work of Dvir, Kopparty, Saraf and Sudan [SIAM J. Comput., 2013], it has found numerous applications; in particular, in the definition and properties of multiplicity codes by Kopparty,

Saraf and Yekhanin [J. ACM, 2014]. In this work, we show how to algorithmize the MSZ lemma for arbitrary product sets over any field. In other words, we give an efficient algorithm for unique decoding of multivariate multiplicity codes from half their minimum distance on arbitrary product sets over all fields. Previously, such an algorithm was known only with certain specific conditions. In particular, even unique decoding of bivariate multiplicity codes with multiplicity two from half their minimum distance was not known over arbitrary product sets over any field. Our algorithm builds upon a result of Kim and Kopparty [ToC, 2017] who gave an algorithmic version of the SZ lemma (without multiplicities) or equivalently, an efficient algorithm for unique decoding of Reed-Muller codes over arbitrary product sets. We introduce a refined notion of distance based on the MSZ lemma and design a unique decoding algorithm for this distance measure. On the way, we give an alternate analysis of Forney's classical generalized minimum distance decoder that might be of independent interest.

Siddharth Bhandari
Simons Institute for the Theory of Computing
siddharth.bhandari@tifr.res.in

Prahladh Harsha
Tata institute
prahladh@tifr.res.in

Mrinal Kumar
IIT Bombay
mrinal@cse.iitb.ac.in

Ashutosh Shankar
Tata Institute of Fundamental Research
ashushankar98@gmail.com

CP23

Approaching the Soundness Barrier: A Near Optimal Analysis of the Cube Versus Cube Test

The Cube versus Cube test is a variant of the well-known Plane versus Plane test of Raz and Safra, in which to each 3-dimensional affine subspace C of \mathbb{F}_q^n , a polynomial of degree at most d , $T(C)$, is assigned in a somewhat locally consistent manner: taking two cubes C_1, C_2 that intersect in a plane uniformly at random, the probability that $T(C_1)$ and $T(C_2)$ agree on $C_1 \cap C_2$ is at least some ϵ . An element of interest is the soundness threshold of this test, i.e. the smallest value of ϵ , such that this amount of local consistency implies a global structure; namely, that there is a global degree d function g such that $g|_C \equiv T(C)$ for at least $\Omega(\epsilon)$ fraction of the cubes. We show that the cube versus cube low degree test has soundness $\text{poly}(d)/q$. This result achieves the optimal dependence on q for soundness in low degree testing and improves upon previous soundness results of $\text{poly}(d)/q^{1/2}$ due to Bhangale, Dinur and Navon.

Kai Zhe Zheng, Dor Minzer
MIT
kzzheng@mit.edu, minzer.dor@gmail.com

CP23

The Complete Classification for Quantified Equality Constraints

We prove that $\text{QCSP}(\mathbb{N}; x = y \rightarrow y = z)$ is PSpace-

complete, settling a question open for more than ten years. This completes the complexity classification for the QCSP over equality languages as a trichotomy between Logspace, NP-complete and PSpace-complete. We additionally settle the classification for bounded alternation QCSP(Γ), for Γ an equality language. Such problems are either in Logspace, NP-complete, co-NP-complete or rise in complexity in the Polynomial Hierarchy.

Dmitriy Zhuk
Charles University
Prague
zhuk.dmitriy@gmail.com

Barnaby Martin
Durham University
barnabymartin@gmail.com

Michał Wrona
Jagiellonian university
Krakow
michal.wrona@uj.edu.pl

CP24

A Simple Combinatorial Algorithm for Robust Matroid Center

Recent progress on robust clustering led to constant-factor approximations for Robust Matroid Center. After a first combinatorial 7-approximation that is based on a matroid intersection approach, two tight LP-based 3-approximations were discovered, both relying on the Ellipsoid Method. In this paper, we show how a carefully designed, yet very simple, greedy selection algorithm gives a 5-approximation. An important ingredient of our approach is a well-chosen use of Rado matroids. This enables us to capture with a single matroid a relaxed version of the original matroid, which, as we show, is amenable to straightforward greedy selections.

Georg Anegg, Laura Vargas Koch, Rico Zenklusen
ETH Zurich
georg.anegg@ifor.math.ethz.ch,
laura.vargas@ifor.math.ethz.ch, ricoz@ethz.ch

CP24

An EPTAS for Budgeted Matroid Independent Set

We consider the budgeted matroid independent set problem. The input is a ground set, where each element has a cost and a non-negative profit, along with a matroid over the elements and a budget. The goal is to select a subset of elements which maximizes the total profit subject to the matroid and budget constraints. Several well known special cases, where we have, e.g., a uniform matroid and a budget, or no matroid constraint (i.e., the classic knapsack problem), admit a fully polynomial-time approximation scheme (FPTAS). In contrast, already a slight generalization to the multi-budgeted matroid independent set problem has a PTAS but does not admit an efficient polynomial-time approximation scheme (EPTAS). This implies a PTAS for our problem, which is the best known result prior to this work. Our main contribution is an EPTAS for the budgeted matroid independent set problem. A key idea of the scheme is to find a representative set for the instance, whose cardinality depends solely on $1/\varepsilon$, where $\varepsilon > 0$ is the accuracy parameter of the scheme. The representative set is identified via matroid basis minimization, which can be solved by a simple greedy algorithm. Our scheme enu-

merates over subsets of the representative set and extends each subset using a linear program. The notion of representative sets may be useful in solving other variants of the budgeted matroid independent set problem.

Ilan Doron-Arad
Technion - Israel Institute of Technology
Ilan Hayim
idoron-arad@cs.technion.ac.il

Ariel Kulik
CISPA Helmholtz Center for Information Security
, Saarland Informatics Campus, Germany
ariel.kulik@cispa.de

Hadas Shachnai
Computer Science Department, Technion
hadas@cs.technion.ac.il

CP24

Voting Algorithms for Unique Games on Complete Graphs

An approximation algorithm for a constraint satisfaction problem is called robust if it outputs an assignment satisfying a $(1 - f(\epsilon))$ -fraction of the constraints on any $(1 - \epsilon)$ -satisfiable instance, where the loss function f is such that $f(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$. Moreover, the runtime of a robust algorithm should not depend in any way on ϵ . In this paper, we present such an algorithm for Min-Unique-Games on complete graphs with q labels. Specifically, the loss function is $f(\epsilon) = (\epsilon + c_\epsilon \epsilon^2)$, where c_ϵ is a constant depending on ϵ such that $\lim_{\epsilon \rightarrow 0} c_\epsilon = 16$. The runtime of our algorithm is $O(qn^3)$ (with no dependence on ϵ) and can run in time $O(qn^2)$ using a randomized implementation with a slightly larger constant c_ϵ . Our algorithm is combinatorial and uses voting to find an assignment. It can furthermore be used to provide a PTAS for Min-Unique-Games on complete graphs, recovering a result of Karpinski and Schudy with a simpler algorithm and proof. We also prove NP-hardness for Min-Unique-Games on complete graphs and (using a randomized reduction) even in the case where the constraints form a cyclic permutation, which is also known as Min-Linear-Equations-mod- q on complete graphs.

Antoine Méot
Laboratoire G-SCOP (CNRS, Grenoble-INP)
Grenoble, France
antoine.meot@gmail.com

Arnaud de Mesmay
LIGM, CNRS, Univ. Gustave Eiffel, ESIEE
Paris, F-77454, Marne-la-Vallée, France
arnaud.de-mesmay@univ-eiffel.fr

Moritz Muehlenthaler
Laboratoire G-SCOP (CNRS, Grenoble-INP)
Grenoble, France
moritz.muehlenthaler@grenoble-inp.fr

Alantha Newman
CNRS, Grenoble, France
alantha.newman@grenoble-inp.fr

CP24

Simple Random Order Contention Resolution for Graphic Matroids with Almost No Prior Informa-

tion

Random order online contention resolution schemes (ROCRS) are structured online rounding algorithms with numerous applications and links to other well-known online selection problems, like the matroid secretary conjecture. We are interested in ROCRs subject to a matroid constraint, which is among the most studied constraint families. Previous ROCRs required to know upfront the full fractional point to be rounded as well as the matroid. It is unclear to what extent this is necessary. Fu, Lu, Tang, Turkieltaub, Wu, Wu, and Zhang (SOSA 2022) shed some light on this question by proving that no strong (constant-selectable) online or even offline contention resolution scheme exists if the fractional point is unknown, not even for graphic matroids. In contrast, we show, in a setting with slightly more knowledge and where the fractional point reveals one by one, that there is hope to obtain strong ROCRs by providing a simple constant-selectable ROCRs for graphic matroids that only requires to know the size of the ground set in advance. Moreover, our procedure holds in the more general adversarial order with a sample setting, where, after sampling a random constant fraction of the elements, all remaining (non-sampled) elements may come in adversarial order.

Richard Santiago
ETH Zurich
rtorres@ethz.ch

Ivan Sergeev
ETH Zürich
isergeev@ethz.ch

Rico Zenklusen
ETH Zurich
ricoz@ethz.ch

CP24

Approximation Algorithms for Matroidal and Cardinal Generalizations of Stable Matching

The Stable Marriage problem (SM), solved by the famous deferred acceptance algorithm of Gale and Shapley (GS), has many natural generalizations. If we allow ties in preferences, then the problem of finding a maximum solution becomes NP-hard, and the best known approximation ratio is 1.5 (McDermid ICALP 2009, Paluch WAOA 2011, Z. Kirly MATCH-UP 2012), achievable by running GS on a cleverly constructed modified instance. Another elegant generalization of SM is the matroid kernel problem introduced by Fleiner (IPCO 2001), which is solvable in polynomial time using an abstract matroidal version of GS. Our main result is a simple 1.5-approximation algorithm for the matroid kernel problem with ties. We also show that the algorithm works for several other versions of stability defined for cardinal preferences, by appropriately modifying the instance on which GS is executed. The latter results are new even for the stable marriage setting.

Gergely Csáji
Centre for Economic and Region
csaji.gergely@krtk.hu

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

Yu Yokoi

National Institute of Informatics
yokoi@nii.ac.jp

CP24

Capacitated Vehicle Routing in Graphic Metrics

We study the capacitated vehicle routing problem in graphic metrics (graphic CVRP). Our main contribution is a new lower bound on the cost of an optimal solution. For graphic metrics, this lower bound is tight and significantly stronger than the well-known bound for general metrics. The proof of the new lower bound is simple and combinatorial. Using this lower bound, we analyze the approximation ratio of the classical iterated tour partitioning algorithm combined with the TSP algorithms for graphic metrics of Christofides [1976], of Mmke-Svensson [JACM 2016], and of Sebő-Vygen [Combinatorica 2014]. In particular, we obtain a 1.95-approximation for the graphic CVRP.

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

Hang Zhou

Ecole Polytechnique, Institut Polytechnique de Paris
zhouhang32@gmail.com

CP25

Improved Integrality Gap in Max-Min Allocation: Or Topology at the North Pole

In the restricted max-min allocation problem (the “Santa Claus’ problem”), a set of players are to be allocated disjoint subsets of a set of indivisible resources, each of which has an intrinsic positive value, and where each player covets a subset of the resources. The aim is to find an allocation that maximizes the minimum utility among all players. This problem is NP-hard to approximate within a factor less than 2 (I. Bezakova and V. Dani, Allocating indivisible goods, SIGecom Exchanges 5(3) (2005), 1118). We show that the Configuration LP of (N. Bansal and M. Sviridenko, The Santa Claus problem, Proc. 38th ACM Symposium on Theory of Computing, (2006), 3140) has integrality gap less than 3.534, improving the previously best known bound 3.808. Our approach uses a criterion for hypergraph matching based on topological connectedness of independence complexes (R. Aharoni and P. Haxell, Hall’s theorem for hypergraphs, J. Graph Theory 35 (2000), 83–88, R. Meshulam, Domination numbers and homology, J. Combin. Theory Ser. A 102 (2003), 321–330, R. Aharoni and E. Berger, The intersection of a matroid and a simplicial complex, Trans. Amer. Math. Soc. 358 (2006), 4895–4917), instead of the combinatorial arguments used in previous work on this problem. We also give improvements on the $(1, \epsilon)$ -restricted problem, in which resources take only values 1 or ϵ , improving the known bounds in most cases.

Penny Haxell
University of Waterloo, Canada
pehaxell@math.uwaterloo.ca

Tibor Szabo
Freie Universitaet Berlin

szabo@zedat.fu-berlin.de

CP25

Improved Approximations for Unrelated Machine Scheduling

We revisit two well-studied scheduling problems in the unrelated machines setting where each job can have a different processing time on each machine. For minimizing total weighted completion time we give a 1.45-approximation, which improves upon the previous 1.488-approximation [Im and Shadloo, SODA 2020]. The key technical ingredient in this improvement lies in a new rounding scheme that gives strong negative correlation with less restrictions. For minimizing L_k -norms of machine loads, inspired by [Kalaitzis et. al., SODA 2017], we give better approximation algorithms. In particular we give a $\sqrt{4/3}$ -approximation for the L_2 -norm which improves upon the former $\sqrt{2}$ -approximation.

Sungjin Im

University of California, Merced
sim3@ucmerced.edu

Shi Li

University at Buffalo
shil@buffalo.edu

CP25

Approximating Knapsack and Partition via Dense Subset Sum

Knapsack and *Partition* are two important additive problems whose fine-grained complexities in the $(1 - \epsilon)$ -approximation setting are not yet settled. In this work, we make progress on both problems by giving improved algorithms.

- *Knapsack* can be $(1 - \epsilon)$ -approximated in $\tilde{O}(n + (1/\epsilon)^{2.2})$ time, improving the previous $\tilde{O}(n + (1/\epsilon)^{2.25})$ by Jin (ICALP'19). There is a known conditional lower bound of $(n + 1/\epsilon)^{2-o(1)}$ based on $(\min, +)$ -convolution hypothesis.
- *Partition* can be $(1 - \epsilon)$ -approximated in $\tilde{O}(n + (1/\epsilon)^{1.25})$ time, improving the previous $\tilde{O}(n + (1/\epsilon)^{1.5})$ by Bringmann and Nakos (SODA'21). There is a known conditional lower bound of $(1/\epsilon)^{1-o(1)}$ based on Strong Exponential Time Hypothesis.

Both of our new algorithms apply the additive combinatorial results on dense subset sums by Galil and Margalit (SICOMP'91), Bringmann and Wellnitz (SODA'21). Such techniques have not been explored in the context of Knapsack prior to our work. In addition, we design several new methods to speed up the divide-and-conquer steps which naturally arise in solving additive problems.

Xiao Mao

Massachusetts Institute of Technology
Massachusetts Institute of Technology
matthew99a@gmail.com

Ce Jin

EECS
MIT
cejin@mit.edu

Mingyang Deng

Massachusetts Institute of Technology
dengm@mit.edu

CP25

Generalized Unrelated Machine Scheduling Problem

We study the generalized load-balancing problem, where we are given n jobs, each of which needs to be assigned to one of m unrelated machines with processing times $\{p_{ij}\}$. Under a job assignment σ , the load of each machine i is $\psi_i(p_i[\sigma])$ where $\psi_i : \mathbb{R}^n \rightarrow \mathbb{R}_+$ is a symmetric monotone norm and $p_i[\sigma]$ is the n -dimensional vector $\{p_{ij} \cdot 1[\sigma(j) = i]\}_{j \in [n]}$. Our goal is to minimize the generalized makespan $\phi(\text{load}(\sigma))$, where $\phi : \mathbb{R}^m \rightarrow \mathbb{R}_+$ is another symmetric monotone norm and $\text{load}(\sigma)$ is the m -dimensional machine load vector. This problem significantly generalizes many classic optimization problems, e.g., makespan minimization, set cover, minimum-norm load-balancing, etc. Our main result is a polynomial time randomized algorithm that achieves an approximation factor of $O(\log n)$, matching the lower bound of set cover up to constant factor. We achieve this by rounding a novel configuration LP relaxation with exponential number of variables. To approximately solve the configuration LP, we design an approximate separation oracle for its dual program. In particular, the separation oracle can be reduced to the norm minimization with a linear constraint problem (NormLin) and we devise a polynomial time approximation scheme (PTAS) for it, which may be of independent interest. We also obtain constant factor approximation algorithms for some special cases.

Shichuan Deng

Alibaba Group
moleculecloud@gmail.com

Jian Li

Tsinghua University
lijian83@mail.tsinghua.edu.cn

Yuval Rabani

The Hebrew University of Jerusalem
yrabani@cs.huji.ac.il

CP25

On Minimizing Tardy Processing Time, Max-Min Skewed Convolution, and Triangular Structured ILPs

The starting point of this paper is the problem of scheduling n jobs with processing times and due dates on a single machine so as to minimize the total processing time of tardy jobs, i.e., $1 \parallel \sum p_j U_j$. This problem was studied by Bringmann et al. with an emphasis on subquadratic running times, obtaining an $\tilde{O}(P^{7/4})$ time algorithm through Max-Min Skewed Convolution, which they solve in $\tilde{O}(n^{7/4})$ time. Our main technical contribution is a faster and simpler convolution algorithm running in $\tilde{O}(n^{5/3})$ time, which implies a similar improvement for the scheduling problem. Inspired by recent developments for the Subset Sum and Knapsack problems, we study $1 \parallel \sum p_j U_j$ parameterized by the maximum job processing time p_{\max} . With proximity techniques, we show structural properties of the problem that lead to an $\tilde{O}(n + p_{\max}^3)$ time algorithm. Moreover, in the setting with multiple machines, we use similar techniques to get an $n \cdot p_{\max}^{O(m)}$ time algorithm for $Pm \parallel \sum p_j U_j$. Finally, we point out that the problems exhibit a triangu-

lar block structure in the constraint matrices of their ILP formulations. In light of recent ILP research, a question that arises is whether one can devise a generic algorithm for such a class of ILPs. We give a negative answer to this question: already a slight generalization of the structure of the scheduling ILP leads to a strongly NP-hard problem.

Lars Rohwedder
Maastricht University
l.rohwedder@maastrichtuniversity.nl

Adam Polak
EPFL
adam.polak@epfl.ch

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

CP25

On Problems Related to Unbounded SubsetSum: A Unified Combinatorial Approach

Unbounded SubsetSum is a classical textbook problem: given integers $w_1, w_2, \dots, w_n \in [1, u]$, c, u , we need to find if there exists $m_1, m_2, \dots, m_n \in \mathbb{N}$ satisfying $c = \sum_{i=1}^n w_i m_i$. In its all-target version, $t \in \mathbb{Z}_+$ is given and the answer for all integers $c \in [0, t]$ is required. In this work, we study three generalizations of this simple problem: All-Target Unbounded Knapsack, All-Target Coin-Change, and Residue Table. With new combinatorial insights into the structures of solutions, we present a novel two-phase approach. As a result, near-optimal (up to log factors) algorithms are found for all three problems.

Mingyang Deng
Massachusetts Institute of Technology
dengm@mit.edu

Xiao Mao
Massachusetts Institute of Technology
Massachusetts Institute of Technology
matthew99a@gmail.com

Ziqian Zhong
Massachusetts Institute of Technology
zizqianz@mit.edu

CP26

Zigzagging Through Acyclic Orientations of Chordal Graphs and Hypergraphs

In 1993, Savage, Squire, and West described an inductive construction for generating every acyclic orientation of a chordal graph exactly once, flipping one arc at a time. We provide two generalizations of this result. Firstly, we describe Gray codes for acyclic orientations of hypergraphs that satisfy a simple ordering condition, which generalizes the notion of perfect elimination order of graphs. This unifies the Savage-Squire-West construction with a recent algorithm for generating elimination trees of chordal graphs (SODA 2022). Secondly, we consider quotients of lattices of acyclic orientations of chordal graphs, and we provide a Gray code for them, answering a question raised by Pilaud (FPSAC 2022). This also generalizes a recent algorithm for generating lattice congruences of the weak order on the symmetric group (SODA 2020). Our algorithms are derived from the Hartung-Hoang-Mtze-Williams combina-

torial generation framework, and they yield simple algorithms for computing Hamilton paths and cycles on large classes of polytopes, including chordal nestohedra and quotients. In particular, we derive an efficient implementation of the Savage-Squire-West construction. Along the way, we give an overview of old and recent results about the polyhedral and order-theoretic aspects of acyclic orientations of graphs and hypergraphs.

Jean Cardinal
ULB, Brussels
jean.cardinal@ulb.be

Hung Hoang
ETH Zurich
phuchung.hoang@gmail.com

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

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

CP26

Graph Classes With Few Minimal Separators. II. A Dichotomy

A class \mathcal{F} of graphs is called tame if every graph in \mathcal{F} on n vertices contains at most $n^{O(1)}$ minimal separators, quasi-tame if every graph in \mathcal{F} on n vertices contains at most $2^{\log^{O(1)}(n)}$ minimal separators, and feral if there exists a constant $c > 1$ so that \mathcal{F} contains n -vertex graphs with at least c^n minimal separators for arbitrarily large n . The classification of graph classes into (quasi-) tame or feral has numerous algorithmic consequences, and has recently received considerable attention. In this paper we precisely characterize the structure of graphs which have few minimal separators. Specifically we show that every graph which excludes certain graphs called k -creatures and k -critters as induced subgraphs has at most quasipolynomially many minimal separators. We then demonstrate that this sufficient condition for having few minimal separators is the “right” one. In particular we show that every hereditary graph class \mathcal{F} definable in CMSO logic that contains k -creatures or k -critters for every k is feral.

Peter Gartland
University of California, Santa Barbara
petergartland@ucsb.edu

Daniel Lokshantov
UCSB
daniello@ucsb.edu

CP26

Graph Classes With Few Minimal Separators. I. Finite Forbidden Induced Subgraphs

A vertex set S in a graph G is a *minimal separator* if there exist vertices u and v that are in distinct connected components of $G - S$, but in the same connected component of $G - S'$ for all $S' \subset S$. A class \mathcal{F} of graphs is called tame if there is a constant c so that every graph in \mathcal{F} on n vertices contains at most $O(n^c)$ minimal separators. If there is a constant $c > 1$ so that \mathcal{F} contains n -vertex graphs with at least c^n minimal separators for arbitrarily large n then

\mathcal{F} is called feral. The classification of graph classes into tame or feral has numerous algorithmic consequences. A key graph-theoretic object in the quest for such a classification is the notion of a k -creature. It is easy to verify that a k -creature contains at least 2^k minimal separators. On the other hand, Abrishami et al. [JCTB.'22] conjecture that every hereditary class that excludes k -creatures for some fixed constant k is tame. We give a counterexample to this conjecture. Our main result is a proof of a weaker form of their conjecture, we prove that a hereditary class \mathcal{F} is strongly quasi-tame if it excludes k -creatures for some fixed k and additionally every minimal separator can be dominated by another fixed k' number of vertices. We show this leads to a complete classification of all graph classes defined by a finite set of forbidden induced subgraphs into quasi-tame or feral.

Peter Gartland

University of California, Santa Barbara
petergartland@ucsb.edu

Daniel Lokshtanov
UCSB
daniello@ucsb.edu

CP26

Sparse Graphs with Bounded Induced Cycle Packing Number Have Logarithmic Treewidth

A graph is O_k -free if it does not contain k pairwise vertex-disjoint and non-adjacent cycles. We show that Maximum Independent Set and 3-Coloring in O_k -free graphs can be solved in quasi-polynomial time. As a main technical result, we establish that "sparse" (here, not containing large complete bipartite graphs as subgraphs) O_k -free graphs have treewidth (even, feedback vertex set number) at most logarithmic in the number of vertices. This is proven sharp as there is an infinite family of O_2 -free graphs without $K_{3,3}$ -subgraph and whose treewidth is (at least) logarithmic. Other consequences include that most of the central NP-complete problems (such as Maximum Independent Set, Minimum Vertex Cover, Minimum Dominating Set, Minimum Coloring) can be solved in polynomial time in sparse O_k -free graphs, and that deciding the O_k -freeness of sparse graphs is polynomial time solvable.

Marthe Bonamy
CNRS, LaBRI, Université de Bordeaux
marthe.bonamy@u-bordeaux.fr

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

Hugues Déprés
ENS de Lyon, LIP
Lyon, France
hugues.depres@ens-lyon.fr

louis esperet
CNRS
louis.esperet@grenoble-inp.fr

Colin Geniet
ENS Paris-Saclay
France
colin.geniet@ens-lyon.fr

Claire Hilaire

CNRS, LaBRI, Université de Bordeaux
Bordeaux, France
claire.hilaire@u-bordeaux.fr

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

Alexandra Wesolek
Simon Fraser University
Burnaby, BC, Canada
alexandrawesolek@gmail.com

CP26

A Half-Integral Erdős-Psa Theorem for Directed Odd Cycles

We prove that there exists a function $f : \mathbb{N} \rightarrow \mathbb{R}$ such that every directed graph G contains either k directed odd cycles where every vertex of G is contained in at most two of them, or a set of at most $f(k)$ vertices meeting all directed odd cycles. We also give a polynomial-time algorithm for fixed k which outputs one of the two outcomes. Using this algorithmic result, we give a polynomial-time algorithm for fixed k to decide whether such k directed odd cycles exist, or there are no k vertex-disjoint directed odd cycles. This extends the half-integral Erdős-Psa theorem for undirected odd cycles by Reed [Combinatorica 1999] to directed graphs.

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

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

O-Joung Kwon
Hanyang University
ojoungkwon@hanyang.ac.kr

Qiqin Xie
Shanghai University
qqxie@shu.edu.cn

CP26

Conflict-Free Hypergraph Matchings

A celebrated theorem of Pippenger, and Frankl and Rodl states that every almost-regular, uniform hypergraph \mathcal{H} with small maximum codegree has an almost-perfect matching. We extend this result by obtaining a conflict-free matching, where conflicts are encoded via a collection \mathcal{C} of subsets $C \subseteq E(\mathcal{H})$. We say that a matching $\mathcal{M} \subseteq E(\mathcal{H})$ is conflict-free if \mathcal{M} does not contain an element of \mathcal{C} as a subset. Under natural assumptions on \mathcal{C} , we prove that \mathcal{H} has a conflict-free, almost-perfect matching. This has many applications, one of which yields new asymptotic results for so-called "high-girth" Steiner systems. Our main tool is a random greedy algorithm which we call the "conflict-free matching process".

Felix Joos
Universität Heidelberg
joos@informatik.uni-heidelberg.de

Stefan Glock

Universität Passau
Germany
stefan.glock@uni-passau.de

Jaehoon Kim
KAIST
jaehoon.kim@kaist.ac.kr

Marcus Kühn
Universität Heidelberg
Germany
kuehn@informatik.uni-heidelberg.de

Lyuben Lichev
Université Jean Monnet and Insitut Camille Jordan
Saint-Etienne, France
lyuben.lichev@univ-st-etienne.fr

CP27

Fixed-Parameter Tractability of Directed Multicut with Three Terminal Pairs Parameterized by the Size of the Cutset: Twin-Width Meets Flow-Augmentation

We show fixed-parameter tractability of the Directed Multicut-problem with three terminal pairs (with a randomized algorithm). This problem, given a directed graph G , pairs of vertices (called terminals) (s_1, t_1) , (s_2, t_2) , and (s_3, t_3) , and an integer k , asks to find a set of at most k non-terminal vertices in G that intersect all s_1t_1 -paths, all s_2t_2 -paths, and all s_3t_3 -paths. The parameterized complexity of this case has been open since Chitnis, Hajihaghay, and Marx proved fixed-parameter tractability of the 2-terminal-pairs case at SODA 2012, and Pilipczuk and Wahlström proved the $W[1]$ -hardness of the 4-terminal-pairs case at SODA 2016. We use two recent developments in parameterized algorithms. First, using the technique of directed flow-augmentation [Kim, Kratsch, Pilipczuk, Wahlström, STOC 2022] we cast the problem as a CSP problem with few variables and constraints over a large ordered domain. This problem can then be encoded as an FO model-checking task over a structure consisting of a few 0-1 matrices. Second, we look at this problem through the lenses of twin-width, a recently introduced structural parameter [Bonnet, Kim, Thomass, Watrigant, FOCS 2020]: By a recent characterization [Bonnet, Giocanti, Ossona de Mendez, Simon, Thomass, Torunczyk, STOC 2022] the said FO model-checking task can be done in FPT time if the said matrices have bounded grid rank. To complete the proof, we show an irrelevant vertex rule.

Meike Hatzel
Technische Universität Berlin
meike.hatzel@tu-berlin.de

Lars Jaffke
University of Bergen, Norway
lars.jaffke@uib.no

Paloma Lima
IT University of Copenhagen
palt@itu.dk

Tomas Masarik, Marcin Pilipczuk
University of Warsaw
masarik@mimuw.edu.pl, m.pilipczuk@mimuw.edu.pl

Roohani Sharma

Max Planck Institute of Informatics, Saarland Informatics
Ca
rsharma@mpi-inf.mpg.de

Manuel Sorge
TU Wien
manuel.sorge@ac.tuwien.ac.at

CP27

A Logic-Based Algorithmic Meta-Theorem for Mim-Width

We introduce a logic called distance neighborhood logic with acyclicity and connectivity constraints (A&C DN for short) which extends existential MSO_1 with predicates for querying neighborhoods of vertex sets in various powers of a graph and for verifying connectivity and acyclicity of vertex sets. Building upon [Bergougnoux and Kant, ESA 2019; SIDMA 2021], we show that the model checking problem for every fixed A&C DN formula is solvable in $n^{O(w)}$ time when the input graph is given together with a branch decomposition of mim-width w . Nearly all problems that are known to be solvable in polynomial time given a branch decomposition of constant mim-width can be expressed in this framework. We add several natural problems to this list, including problems asking for diverse sets of solutions. Our model checking algorithm also unifies and extends known algorithms for tree-width, clique-width and rank-width. It has a single-exponential dependence on these three width measures and asymptotically matches run times of the fastest known algorithms for several problems. Our results are also tight in terms of the expressive power of the logic: we show that already slight extensions of our logic make the model checking problem para-NP-hard when parameterized by mim-width plus formula length.

Benjamin Bergougnoux
University of Warsaw, Poland
benjamin.bergougnoux@gmail.com

Jan Dreier
RWTH Aachen University
dreier@ac.tuwien.ac.at

Lars Jaffke
University of Bergen, Norway
lars.jaffke@uib.no

CP27

Almost Consistent Systems of Linear Equations

Checking whether a system of linear equations is consistent is a basic computational problem with ubiquitous applications. When dealing with inconsistent systems, one may seek an assignment that minimizes the number of unsatisfied equations. This problem is NP-hard and UGC-hard to approximate within any constant even for two-variable equations over the two-element field. We study this problem from the point of view of parameterized complexity, with the parameter being the number of unsatisfied equations. We consider equations defined over Euclidean domains—a family of commutative rings that generalize finite and infinite fields including the rationals, the ring of integers and many other structures. We show that if every equation contains at most two variables, the problem is fixed-parameter tractable. This generalizes many eminent graph separation problems such as Bipartization, Multiway

Cut and Multicut parameterized by the size of the cutset. To complement this, we show that the problem is $W[1]$ -hard when three or more variables are allowed in an equation, as well as for many commutative rings that are not Euclidean domains. On the technical side, we introduce the notion of important balanced subgraphs, generalizing important separators of Marx [Theor. Comput. Sci. 2006] to the setting of biased graphs. Furthermore, we use recent results on parameterized MinCSP [Kim et al., SODA 2021] to efficiently solve a generalization of Multicut with disjunctive cut requests.

Konrad K. Dabrowski
Newcastle University
konrad.dabrowski@newcastle.ac.uk

Peter Jonsson
Linköping University
peter.jonsson@liu.se

Sebastian Ordyniak
University of Sheffield, Sheffield, UK
sordyniak@gmail.com

George Osipov
Linköping University
george.osipov@pm.me

Magnus Wahlström
Royal Holloway, University of London
Magnus.Wahlstrom@rhul.ac.uk

CP27

Constant Approximating Parameterized k -SetCover is $W[2]$ -Hard

In this paper, we prove that it is $W[2]$ -hard to approximate k -SetCover within any constant ratio. Our proof is built upon the recently developed threshold graph composition technique. We propose a strong notion of threshold graphs and use a new composition method to prove this result. Our technique could also be applied to rule out polynomial time $o\left(\frac{\log n}{\log \log n}\right)$ ratio approximation algorithms for the non-parameterized k -SetCover problem with k as small as $O\left(\frac{\log n}{\log \log n}\right)^3$, assuming $W[1] \neq \text{FPT}$. We highlight that our proof does not depend on the well-known PCP theorem, and only involves simple combinatorial objects.

Xuandi Ren
Peking University
renxuandi@pku.edu.cn

Bingkai Lin
Nanjing University
lin@nju.edu.cn

Yican Sun
School of Computer Science, Peking University
sycpku@pku.edu.cn

Xiuhan Wang
Tsinghua University
wangxh19@mails.tsinghua.edu.cn

CP27

Polynomial Formulations as a Barrier for

Reduction-Based Hardness Proofs

The Strong Exponential Time Hypothesis (SETH) asserts that for every $\varepsilon > 0$ there exists k such that k -SAT requires time $(2 - \varepsilon)^n$. The field of fine-grained complexity has leveraged SETH to prove quite tight conditional lower bounds for dozens of problems in various domains and complexity classes, including Edit Distance, Graph Diameter, Hitting Set, Independent Set, and Orthogonal Vectors. Yet, it has been repeatedly asked in the literature whether SETH-hardness results can be proven for other fundamental problems such as Hamiltonian Path, Independent Set, Chromatic Number, MAX- k -SAT, and Set Cover. In this paper, we show that fine-grained reductions implying even λ^n -hardness of these problems from SETH for any $\lambda > 1$, would imply new circuit lower bounds. We also extend this barrier result to the class of parameterized problems. Namely, for every $\lambda > 1$ we conditionally rule out fine-grained reductions implying SETH-based lower bounds of λ^k for a number of problems parameterized by the solution size k . Our main technical tool is a new concept called polynomial formulations. In particular, we show that many problems can be represented by relatively succinct low-degree polynomials, and that any problem with such a representation cannot be proven SETH-hard (without proving new circuit lower bounds).

Tatiana Belova
Steklov Institute of Mathematics at St. Petersburg
yukikomodo@gmail.com

Alexander Golovnev
Georgetown University
alexgolovnev@gmail.com

Alexander Kulikov
Steklov Institute of Mathematics at St. Petersburg
Saint Petersburg University
alexander.s.kulikov@gmail.com

Ivan Mihajlin
University of California San Diego
ivmihajlin@gmail.com

Denil Sharipov
Saint Petersburg University
d.i.sharipov@yandex.ru

CP27

Flow-Augmentation III: Complexity Dichotomy for Boolean CSPs Parameterized by the Number of Unsatisfied Constraints

We study the parameterized problem of satisfying “almost all” constraints of a given formula F over a fixed, finite Boolean constraint language Γ , with or without weights. More precisely, for each finite Boolean constraint language Γ , we consider the following two problems. In $\text{MinSAT}(\Gamma)$, the input is a formula F over Γ and an integer k , and the task is to find an assignment $\alpha: V(F) \rightarrow \{0, 1\}$ that satisfies all but at most k constraints of F , or determine that no such assignment exists. In $\text{Weighted MinSAT}(\Gamma)$, the input additionally contains a weight function $\omega: F \rightarrow \mathbb{Z}_+$ and an integer W , and the task is to find an assignment α such that (1) α satisfies all but at most k constraints of F , and (2) the total weight of the violated constraints is at most W . We give a complete dichotomy for the fixed-parameter tractability of these problems: We show that for every Boolean constraint language Γ , either Weighted

MinSAT(Γ) is FPT; or Weighted MinSAT(Γ) is W[1]-hard but MinSAT(Γ) is FPT; or MinSAT(Γ) is W[1]-hard. Our result generalizes and subsumes multiple previous results, including the FPT algorithms for Weighted Almost 2-SAT, weighted and unweighted ℓ -Chain SAT and Coupled Min-Cut, as well as directed versions of the latter. The main tool used in our algorithms is the recently developed method of directed flow-augmentation.

Eun Jung Kim
LAMSADE, Paris-Dauphine University
eun-jung.kim@dauphine.fr

Stefan Kratsch
Humboldt-Universität zu Berlin
kratsch@informatik.hu-berlin.de

Marcin Pilipczuk
University of Warsaw
malcin@mimuw.edu.pl

Magnus Wahlstrom
Royal Holloway, University of London
magnus.wahlstrom@rhul.ac.uk

CP28

On the Fine-Grained Complexity of Approximating K-Center in Sparse Graphs

We study the fine-grained complexity of the k -center problem in the metric induced by a graph with n vertices and m edges. The problem is NP-hard to approximate within a factor better than 2, and several 2-approximation algorithms are known. Two of the most well-known approaches are (1) finding a maximal distance- r independent set and (2) Gonzalez's algorithm that iteratively adds the center farthest from the currently chosen centers. For the first approach, Thorup [SIAM J. Comput. '05] already gave a nearly linear time algorithm. While Thorup's algorithm is not complicated, it still requires tools such as an approximate oracle for neighborhood size by Cohen [J. Comput. Syst. Sci. '97]. Our main result is a nearly straightforward algorithm that improves the running time by an $O(\log n)$ factor. It results in a $(2 + \epsilon)$ -approximation for k -center in $O((m + n \log n) \log n \log(n/\epsilon))$ time. For Gonzalez's algorithm [Theor. Comput. Sci. 85], we show that the simple $\tilde{O}(mk)$ -time implementation is nearly optimal if we insist on an exact implementation. On the other hand, we show that a $(1 + \epsilon)$ -approximate version is efficiently implementable, leading to a $(2 + \epsilon)$ -approximation algorithm running in time $O((m + n \log n) \log^2(n)/\epsilon)$. We also show that, unlike in the first approach, the dependency on $1/\epsilon$ in the running time is essentially optimal for $(1 + \epsilon)$ -approximate Gonzalez's.

Amir Abboud
Weizmann Institute of Science
Israel
amir.abboud@weizmann.ac.il

Vincent Cohen-Addad
Google Research
cohenaddad@google.com

Euiwoong Lee
University of Michigan
euiwoong@umich.edu

Pasin Manurangsi
Google Research
pasin@google.com

CP28

Faster Walsh-Hadamard Transform and Matrix Multiplication over Finite Fields Using Lookup Tables

We use lookup tables to design faster algorithms for important algebraic problems over finite fields. These faster algorithms, which only use arithmetic operations and lookup table operations, may help to explain the difficulty of determining the complexities of these important problems. Our results over a constant-sized finite field are as follows. The Walsh-Hadamard transform of a vector of length N can be computed using $O(N \log N / \log \log N)$ bit operations. This generalizes to any transform defined as a Kronecker power of a fixed matrix. By comparison, the Fast Walsh-Hadamard transform (similar to the Fast Fourier transform) uses $O(N \log N)$ arithmetic operations, which is believed to be optimal up to constant factors. Any algebraic algorithm for multiplying two $N \times N$ matrices using $O(N^\omega)$ operations can be converted into an algorithm using $O(N^\omega / (\log N)^{\omega/2-1})$ bit operations. For example, Strassen's algorithm can be converted into an algorithm using $O(N^{2.81} / (\log N)^{0.4})$ bit operations. It remains an open problem with practical implications to determine the smallest constant c such that Strassen's algorithm can be implemented to use $c \cdot N^{2.81} + o(N^{2.81})$ arithmetic operations; using a lookup table allows one to save a super-constant factor in bit operations.

Josh Alman
Columbia University
josh@cs.columbia.edu

CP28

A Simple Deterministic Distributed Low-Diameter Clustering

We give a simple, local process for nodes in an undirected graph to form non-adjacent clusters that (1) have at most a polylogarithmic diameter and (2) contain at least half of all vertices. Efficient deterministic distributed clustering algorithms for computing strong-diameter network decompositions and other key tools follow immediately. Overall, our process is a direct and drastically simplified way for computing these fundamental objects.

Christoph Grunau
ETH Zurich
cgrunau@student.ethz.ch

Václav Rozhon
Carnegie Mellon University & ETH Zurich
rozhonv@inf.ethz.ch

Bernhard Haeupler
ETH Zurich
bernhard.haeupler@inf.ethz.ch

CP28

Sinkless Orientation Made Simple

The sinkless orientation problem plays a key role in understanding the foundations of distributed computing. The problem can be used to separate two fundamental models

of distributed graph algorithms, LOCAL and SLOCAL: the locality of sinkless orientation is $\Omega(\log n)$ in the deterministic LOCAL model and $O(\log \log n)$ in the deterministic SLOCAL model. Both of these results are known by prior work, but here we give new simple, self-contained proofs for them.

Alkida Balliu
Gran Sasso Science Institute
alkida.balliu@gssi.it

Janne Korhonen
TU Berlin
janne.h.korhonen@gmail.com

Fabian Kuhn
University of Freiburg, Germany
kuhn@cs.uni-freiburg.de

Henrik Lievenen
Aalto University
henrik.lievenen@aalto.fi

Dennis Olivetti
Gran Sasso Science Institute
dennis.olivetti@gssi.it

Shreyas Pai
Aalto University
shreyas.pai@aalto.fi

Ami Paz
LISN, CNRS
ami.paz@lisn.fr

Joel Rybicki
IST Austria
joel.rybicki@ist.ac.a

Stefan Schmid
TU Berlin
stefan.schmid@tu-berlin.de

Jan Studen, Jukka Suomela, Jara Uitto
Aalto University
jan.studen@aalto.fi, jukka.suomela@aalto.fi,
jara.uitto@aalto.fi

CP28

Simpler and Faster Algorithms for Detours in Planar Digraphs

In the Directed Detour problem one is given a digraph G and a pair of vertices s and t , and the task is to decide whether there is a directed simple path from s to t in G whose length is larger than $\text{dist}_G(s, t)$. The more general parameterized variant, Directed Long Detour, asks for a simple s -to- t path of length at least $\text{dist}_G(s, t) + k$, for a given parameter k . For planar digraphs, Wu and Wang [Networks, '15] proposed an $\mathcal{O}(n^3)$ -time algorithm for Directed Detour, while Fomin et al. [STACS 2022] gave a $2^{\mathcal{O}(k)} \cdot n^{\mathcal{O}(1)}$ -time algorithm for Directed Long Detour. The algorithm of Wu and Wang relies on a nontrivial analysis of how short planar detours may look like, while the algorithm of Fomin et al. is based on a reduction to the 3-Disjoint Paths problem on planar digraphs. This latter problem is solvable in polynomial time using the algebraic machinery of Schrijver [SIAM J. Comp., '94], but the degree of the ob-

tained polynomial factor is huge. In this paper we propose two simple algorithms: we show how to solve, in planar digraphs, Directed Detour in time $\mathcal{O}(n^2)$ and Directed Long Detour in time $2^{\mathcal{O}(k)} \cdot n^4 \log n$. In both cases, the idea is to reduce to the 2-Disjoint Paths problem in a planar digraph, and to observe that the obtained instances of this problem have a certain topological structure.

Konrad Majewski
University of Warsaw
k.majewski@mimuw.edu.pl

Meike Hatzel
Technische Universität Berlin
meike.hatzel@tu-berlin.de

Michał Pilipczuk, Marek Sokolowski
University of Warsaw
michal.pilipczuk@mimuw.edu.pl,
marek.sokolowski@mimuw.edu.pl

CP28

Min-Max Optimization Made Simple: Approximating the Proximal Point Method via Contraction Maps

In this paper we present a first-order method that admits near-optimal convergence rates for convex/concave min-max problems while requiring a simple and intuitive analysis. Similarly to the seminal work of Nemirovski [25] and the recent approach of Piliouras et al. [27] in normal form games, our work is based on the fact that the update rule of the *Proximal Point method* (PP) can be approximated up to accuracy ϵ with only $\mathcal{O}(\log 1/\epsilon)$ additional gradient-calls through the iterations of a contraction map. Then combining the analysis of (PP) method with an error-propagation analysis we establish that the resulting first order method, called *Clairvoyant Extra Gradient*, admits near-optimal time-average convergence for general domains and last-iterate convergence in the unconstrained case.

Volkan Cevher
EPFL
volkan.cevher@epfl.ch

Georgios Piliouras, Ryann Sim
Singapore University of Technology and Design
georgios.piliouras@gmail.com, ryann_sim@mymail.sutd.sg

Stratis Skoulakis
Ecole Polytechnique de Lausanne
efstratis.skoulakis@epfl.ch

CP29

Moser-Tardos Algorithm: Beyond Shearer's Bound

In a seminal paper, Moser and Tardos developed a simple and powerful algorithm to find solutions for constraint satisfaction problems. Kolipaka and Szegedy proved that the Moser-Tardos algorithm is efficient up to the tight condition of the abstract Lovasz Local Lemma, known as Shearer's bound. A fundamental problem around the LLL is whether the efficient region of the Moser-Tardos algorithm can be further extended. In this paper, we give a positive answer to this problem. We show that the efficient region of the Moser-Tardos algorithm indeed goes beyond the Shearer's bound of the underlying dependency graph,

if the graph is not chordal. This chordal condition is sufficient and necessary, since it has been shown that Shearer's bound exactly characterizes the efficient region for chordal dependency graph. Moreover, we demonstrate that the efficient region can exceed Shearer's bound by a constant amount by explicitly calculating the gaps on several infinite lattices. The core of our proof is a new criterion on the efficiency of the Moser-Tardos algorithm which takes the intersection between dependent events into consideration. Our criterion is strictly larger than Shearer's bound whenever there exist two dependent events with non-empty intersection. Meanwhile, if any two dependent events are mutually exclusive, our criterion becomes the Shearer's bound, which is known to be tight in this situation for the Moser-Tardos algorithm.

Kun He

Institute of Computing Technology
Chinese Academy of Sciences
hekun.threebody@foxmail.com

Qian Li

Institute of Computing Technology Chinese Academy of Sciences
liqian@ict.ac.cn

Xiaoming Sun

Institute of Computing Technology
Chinese Academy of Sciences
sunxiaoming@ict.ac.cn

CP29

Instability of Backoff Protocols with Arbitrary Arrival Rates

In contention resolution, multiple processors are trying to coordinate to send discrete messages through a shared channel with sharply limited communication. If two processors inadvertently send at the same time, the messages collide and are not transmitted successfully. An important case is acknowledgement-based contention resolution, in which processors cannot listen to the channel at all; all they know is whether or not their own messages have got through. This situation arises frequently in both networking and cloud computing, with variants of binary exponential backoff being used in both Ethernet and TCP/IP. In queueing models, where each processor has a queue of messages, stable acknowledgement-based protocols are already known (Hstad et al., SICOMP 1996). In queue-free models, where each processor has a single message but processors arrive randomly, it is widely conjectured that no stable acknowledgement-based protocols exist for any positive arrival rate of processors. Despite exciting recent results for full-sensing protocols which assume greater listening capabilities of the processors (see e.g. Bender et al. STOC 2020 or Chen et al. PODC 2021), this foundational question remains open even for backoff protocols unless the arrival rate of processors is at least 0.42 (Goldberg et al. SICOMP 2004). We prove the conjecture for all backoff protocols outside of a tightly-constrained special case, and set out the remaining technical obstacles to a full proof.

Leslie Ann Goldberg

University of Oxford
leslie.goldberg@seh.ox.ac.uk

John Lapinskas

University of Birmingham

john.lapinskas@bristol.ac.uk

CP29

From Algorithms to Connectivity and Back: Finding a Giant Component in Random k -SAT

We take an algorithmic approach to studying the solution space geometry of relatively sparse random and bounded degree k -CNFs for large k . In the course of doing so, we establish that with high probability, a random k -CNF Φ with n variables and clause density $\alpha = m/n \lesssim 2^{k/6}$ has a giant component of solutions that are connected in a graph where solutions are adjacent if they have Hamming distance $O_k(\log n)$ and that a similar result holds for bounded degree k -CNFs at similar densities. We are also able to deduce looseness results for random and bounded degree k -CNFs in a similar regime. Our methods also have algorithmic implications; we construct an idealized block dynamics that samples solutions from a random k -CNF Φ with density $\alpha = m/n \lesssim 2^{k/52}$. We show this Markov chain can with high probability be implemented in polynomial time and by leveraging spectral independence, we also observe that it mixes relatively fast, giving a polynomial time algorithm to with high probability sample a uniformly random solution to a random k -CNF.

Zongchen Chen

Massachusetts Institute of Technology
zongchen@mit.edu

Nitya Mani

Massachusetts Institute of Technology, U.S.
nmani@mit.edu

Ankur Moitra

MIT
moitra@mit.edu

CP29

Subexponential Mixing for Partition Chains on Grid-Like Graphs

We consider the problem of generating uniformly random partitions of the vertex set of a graph such that every piece induces a connected subgraph. For the case where we want to have partitions with linearly many pieces of bounded size, we obtain approximate sampling algorithms based on Glauber dynamics which are fixed-parameter tractable with respect to the bandwidth of G , with simple-exponential dependence on the bandwidth. For example, for rectangles of constant or logarithmic width this gives polynomial-time sampling algorithms. More generally, this gives sub-exponential algorithms for bounded-degree graphs without large expander subgraphs (for example, we obtain $2^{O(\sqrt{n})}$ time algorithms for square grids). In the case where we instead want partitions with a small number of pieces of linear size, we show that Glauber dynamics can have exponential mixing time, even just for the case of 2 pieces, and even for 2-connected subgraphs of the grid with bounded bandwidth.

Wesley Pegden

Carnegie Mellon University
wes@math.cmu.edu

Alan Frieze

Carnegie Mellon University
Department of Mathematical Sciences

alan@random.math.cmu.edu

CP29

Deterministic Counting Lovsz Local Lemma Beyond Linear Programming

We give a simple combinatorial algorithm to deterministically approximately count the number of satisfying assignments of general constraint satisfaction problems (CSPs). Suppose that the CSP has domain size $q = O(1)$, each constraint contains at most $k = O(1)$ variables, shares variables with at most $\Delta = O(1)$ constraints, and is violated with probability at most p by a uniform random assignment. The algorithm returns in polynomial time in an improved local lemma regime:

$$q^2 \cdot k \cdot p \cdot \Delta^5 \leq C_0 \quad \text{for a suitably small absolute constant } C_0.$$

Here the key term Δ^5 improves the previously best known Δ^7 for general CSPs [JPV21b] and $\Delta^{5.714}$ for the special case of k -CNF [JPV21a, HSW21]. Our deterministic counting algorithm is a derandomization of the very recent fast sampling algorithm in [HWY22]. It departs substantially from all previous deterministic counting Lovsz local lemma algorithms which relied on linear programming, and for the first time, gives a deterministic approximate counting algorithm that straightforwardly derandomizes a fast sampling algorithm, hence unifying the fast sampling and deterministic approximate counting in the same algorithmic framework. To obtain the improved regime, in our analysis we develop a refinement of the $\{2, 3\}$ -trees that were used in the previous analyses of counting/sampling LLL, which may be of independent interest.

Chunyang Wang
Nanjing University
Nanjing University
wcysai@mail.nju.edu.cn

Kun He
Institute of Computing Technology
Chinese Academy of Sciences
hekun@ict.ac.cn

Yitong Yin
Nanjing University
yinyt@nju.edu.cn

CP29

Improved Bounds for Sampling Solutions of Random CNF Formulas

Let Φ be a random k -CNF formula on n variables and m clauses, where each clause is a disjunction of k literals chosen independently and uniformly. Our goal is, for most Φ , to (approximately) uniformly sample from its solution space. Let $\alpha = m/n$ be the density. The previous best algorithm runs in time $n^{\text{poly}(k, \alpha)}$ for any $\alpha \lesssim 2^{k/300}$ [Galanis, Goldberg, Guo, and Yang, SIAM J. Comput. '21]. In contrast, our algorithm runs in almost-linear time for any $\alpha \lesssim 2^{k/3}$.

Kun He
Institute of Computing Technology
Chinese Academy of Sciences
hekun.threebody@foxmail.com

Kewen Wu
University of California at Berkeley

shlw_kevin@hotmail.com

Kuan Yang
John Hopcroft Center for Computer Science
Shanghai Jiao Tong University
kuan.yang@sjtu.edu.cn

CP30

Higher Degree Sum-Of-Squares Relaxations Robust Against Oblivious Outliers

We consider estimation models of the form $Y = X^* + N$, where X^* is some m -dimensional signal we wish to recover, and N is symmetrically distributed noise that may be unbounded in all but a small α fraction of the entries. We introduce a family of algorithms that under mild assumptions recover the signal in all problems for which there exists a sum-of-squares algorithm that recovers the signal when the noise N is Gaussian. Our framework extends far beyond previous results on symmetric noise models and is even robust to adversarial corruptions. As concrete examples, we investigate two problems for which no efficient algorithms were known to work for heavy-tailed noise: tensor PCA and sparse PCA. For the former, our algorithm runs in polynomial time and matches (up to log factors) current best known guarantees for polynomial time algorithms for tensor PCA with Gaussian noise. For the latter, our algorithm runs in quasipolynomial time and matches current best known guarantees for quasipolynomial time algorithms for sparse PCA with Gaussian noise. We provide evidence that the quasipolynomial time is likely to be necessary for sparse PCA with symmetric noise. In our proofs we use bounds on the covering numbers of sets of pseudo-expectations, which we obtain by certifying in sum-of-squares upper bounds on the Gaussian complexities of sets of solutions. This approach may be interesting in its own right and may find other application in future works.

Tommaso D'Orsi
ETH Zurich
tommaso.dorsi@inf.ethz.ch

Rajai Nasser
Google Zurich; This work was done at ETH Zurich
rajai.nasser@inf.ethz.ch

Gleb Novikov, David Steurer
ETH Zurich
gleb.novikov@inf.ethz.ch, dsteur@inf.ethz.ch

CP30

Interactive Coding with Small Memory

In this work, we design an interactive coding scheme that converts any two party interactive protocol Π into another interactive protocol Π' , such that even if errors are introduced during the execution of Π' , the parties are able to determine what the outcome of running Π would be in an error-free setting. Importantly, our scheme preserves the space complexity of the protocol, in addition to the communication and computational complexities. Specifically, if the protocol Π has communication complexity T , computational complexity t , and space complexity s , the resulting protocol Π' is resilient to a constant $\epsilon > 0$ fraction of adversarial errors, and has communication complexity approaching T as ϵ approaches 0, computational complexity $\text{poly}(t)$, and space complexity $\mathcal{O}(s \log T)$. Prior to this work, all known interactive coding schemes required the

parties to use at least $\Omega(T)$ space, as the parties were required to remember the transcript of the conversation thus far, or considered weaker error models.

Klim Efremenko
Ben Gurion University
klimefrem@gmail.com

Bernhard Haeupler
ETH Zurich
bernhard.haeupler@inf.ethz.ch

Yael Kalai
Microsoft Research
yael@microsoft.com

Gillat Kol
IAS
gillat.kol@gmail.com

Nicolas Resch
University of Amsterdam
n.a.resch@uva.nl

Raghuvansh Saxena
Microsoft Research
raghuvansh.saxena@gmail.com

CP30

Positivity of the Symmetric Group Characters is as Hard as the Polynomial Time Hierarchy

We prove that deciding the vanishing of the character of the symmetric group is C=P-complete. We use this hardness result to prove that the absolute value and also the square of the character are not contained in #P, unless the polynomial hierarchy collapses to the second level. This rules out the existence of any (unsigned) combinatorial description for the square of the characters. As a byproduct of our proof we conclude that deciding positivity of the character is PP-complete under many-one reductions, and hence PH-hard under Turing-reductions.

Christian Ikenmeyer
University of Liverpool
christian.ikenmeyer@warwick.ac.uk

Igor Pak
UCLA
pak@math.ucla.edu

Greta Panova
University of Southern California
gpanova@usc.edu

CP30

Robust Voting Rules from Algorithmic Robust Statistics

Maximum likelihood estimation furnishes powerful insights into voting theory, and the design of voting rules. However the MLE can usually be badly corrupted by a single outlying sample. This means that a single voter or a group of colluding voters can vote strategically and drastically affect the outcome. Motivated by recent progress in algorithmic robust statistics, we revisit the fundamental problem of estimating the central ranking in a Mallows model, but ask for an estimator that is provably robust, unlike the MLE.

Our main result is an efficiently computable estimator that achieves nearly optimal robustness guarantees. In particular the robustness guarantees are dimension-independent in the sense that our overall accuracy does not depend on the number of alternatives being ranked. As an immediate consequence, we show that while the landmark Gibbard-Satterthwaite theorem tells us a strong impossibility result about designing strategy-proof voting rules, there are quantitatively strong ways to protect against large coalitions if we assume that the remaining voters are honest and their preferences are sampled from a Mallows model. Our work also makes technical contributions to algorithmic robust statistics by designing new spectral filtering techniques that can exploit the intricate combinatorial dependencies in the Mallows model.

Allen X. Liu, Ankur Moitra
MIT
cliu568@mit.edu, moitra@mit.edu

CP30

Non-Stochastic Cdf Estimation Using Threshold Queries

We present an algorithm to estimate the empirical distribution of a scalar-valued data set, when the data are adversarially generated and the algorithm may only ask a limited number of threshold queries about each data point rather than observing it directly. This estimation problem models, for example, a seller experimenting with posted prices to estimate the distribution of consumers' willingness to pay for a product: offering a price and observing a consumer's purchase decision is equivalent to asking a single threshold query about their value, and the distribution of consumers' values may be non-stationary over time, as early adopters may differ markedly from late adopters. Our main result quantifies, to within a constant factor, the sample complexity of estimating the empirical CDF of a sequence of elements of $[n]$, up to ϵ additive error, using one threshold query per sample. The complexity depends only logarithmically on n , hence our result extends existing logarithmic-complexity results for noisy binary search to the more challenging setting where noise is non-stochastic. Along the way to designing our algorithm, we solve a more general problem with a limited number of simultaneous threshold queries per sample, using Blackwell's Approachability Theorem and the exponential weights method. As a side result of independent interest, we characterize the minimum number of simultaneous threshold queries required by deterministic CDF estimation algorithms.

Princewill Okoroafor, Vaishnavi Gupta, Robert Kleinberg, Eleanor Goh
Cornell University
pco9@cornell.edu, vg222@cornell.edu, rdk@cs.cornell.edu, eg552@cornell.edu

CP30

Concentration of Polynomial Random Matrices via Efron-Stein Inequalities

Analyzing concentration of large random matrices is a common task in a wide variety of fields. Given independent random variables, many tools are available to analyze random matrices whose entries are linear in the variables, e.g. the matrix-Bernstein inequality. However, in many applications, we need to analyze random matrices whose entries are polynomials in the variables. These arise naturally in the analysis of spectral algorithms and in lower bounds for

semidefinite programs based on the Sum of Squares hierarchy. In this work, we present a general framework to obtain such bounds, based on the matrix Efron-Stein inequalities developed by Paulin-Mackey-Tropp [Annals of Probability 2016]. The Efron-Stein inequality bounds the norm of a random matrix by the norm of another simpler (but still random) matrix, which we view as arising by "differentiating" the starting matrix. By recursively differentiating, our framework reduces the main task to analyzing far simpler matrices. For Rademacher variables, these simpler matrices are in fact deterministic. For general non-Rademacher variables, the task reduces to scalar concentration. Using this, we recover known bounds for "tensor networks", "dense graph matrices" and "sparse graph matrices", the last of which were obtained only recently by Jones et al. [FOCS 2021] using a nontrivial application of the trace power method. We expect our framework to be broadly helpful to analyze concentration phenomena for nonlinear random matrices.

Goutham Rajendran
University of Chicago
goutham@uchicago.edu

Madhur Tulsiani
TTI Chicago
madhurt@ttic.edu

CP31

Kernelization for Graph Packing Problems via Rainbow Matching

We introduce a new kernelization tool, called rainbow matching technique, that is appropriate for the design of polynomial kernels for packing problems. Our technique capitalizes on the powerful combinatorial results of [Graf, Harris, Haxell, Algorithms for weighted independent transversals and strong colouring, SODA 2021]. We consider two (di)graph packing problems, namely the Triangle-Packing in Tournament problem (TPT), where we ask for a packing of k directed triangles in a tournament, and the Induced 2-Path-Packing (IPP) where we ask for a packing of k induced paths of length two in a graph. The existence of sub-quadratic kernels for these problems was proven for the first time in [Fomin, Le, Lokshtanov, Saurabh, Thomass, Zehavi, Subquadratic kernels for implicit 3-hitting set and 3-set packing problems, ACM Trans. Algorithms, 2019], where authors provided a kernel of $\mathcal{O}(k^{3/2})$ vertices and $\mathcal{O}(k^{5/3})$ vertices respectively, and asked if these bounds can be (optimally) improved to linear ones. We answer this question and prove using the rainbow matching technique that TPT admits an (almost linear) kernel of $k^{1+\frac{\mathcal{O}(1)}{\sqrt{\log k}}}$ vertices and that IPP admits a kernel of $\mathcal{O}(k)$ vertices. In our presentation we will focus on the application on IPP, which is simpler than TPT, and give insights on where are the crucial points when applying this technique to an arbitrary packing problem.

Stéphane Bessy
LIRMM, University of Montpellier, France
bessy@lirmm.fr

Marin Bougeret
LIRMM, University of Montpellier
marin.bougeret@lirmm.fr

Dimitrios Thilikos
CNRS, LIRMM, Montpellier

sedthilk@thilikos.info

Sebastian Wiederrecht
Institute for Basic Science, DIMAG
sebastian.wiederrecht@gmail.com

CP31

Parameterized Algorithm for the Planar Disjoint Paths Problem: Exponential in k^2 , and Linear in n

In this paper, we study the Planar Disjoint Paths problem: Given an undirected planar graph G with n vertices and a set T of k pairs $(s_i, t_i)_{i=1}^k$ of vertices, the goal is to find a set \mathcal{P} of k pairwise vertex-disjoint paths connecting s_i and t_i for all indices $i \in \{1, \dots, k\}$. We present a $2^{O(k^2)}n$ -time algorithm for the Planar Disjoint Paths problem. This improves the two previously best-known algorithms: $2^{2^{O(k)}}n$ -time algorithm [Discrete Applied Mathematics 1995] and $2^{O(k^2)}n^6$ -time algorithm [STOC 2020].

Kyungjin Cho
POSTECH
Computer Science and Technology
kyungjincho@postech.ac.kr

Eunjin Oh
Max Planck Institute for Informatics
eoh@mpi-inf.mpg.de

Seunghyeok Oh
POSTECH
Computer Science and Technology
seunghyeokoh@postech.ac.kr

CP31

Fixed-Parameter Tractability of Maximum Colored Path and Beyond

We introduce a general method for obtaining fixed-parameter algorithms for problems about finding paths in undirected graphs, where the length of the path could be unbounded in the parameter. The first application of our method is as follows. We give a randomized algorithm, that given a colored n -vertex undirected graph, vertices s and t , and an integer k , finds an (s, t) -path containing at least k different colors in time $2^k n^{O(1)}$. This is the first FPT algorithm for this problem, and it generalizes the algorithm of Björklund, Husfeldt, and Taslaman [SODA 2012] on finding a path through k specified vertices. It also implies the first $2^k n^{O(1)}$ time algorithm for finding an (s, t) -path of length at least k . Our method yields FPT algorithms for even more general problems. For example, we consider the problem where the input consists of an n -vertex undirected graph G , a matroid M whose elements correspond to the vertices of G and which is represented over a finite field of order q , a positive integer weight function on the vertices of G , two sets of vertices $S, T \subseteq V(G)$, and integers p, k, w , and the task is to find p vertex-disjoint paths from S to T so that the union of the vertices of these paths contains an independent set of M of cardinality k and weight w , while minimizing the sum of the lengths of the paths. We give a $2^{p+O(k^2 \log(q+k))} n^{O(1)} w$ time randomized algorithm for this problem.

Fedor Fomin
University of Bergen
fomin@ii.uib.no

Petr Golovach
University of Bergen, Norway
petr.golovach@uib.no

Tuukka Korhonen
University of Bergen
tuukka.korhonen@uib.no

Kirill Simonov
TU Wien
kirillsimonov@gmail.com

Giannos Stamoulis
University of Montpellier and the National Center
for Scientific Research
giannos.stamoulis@lirmm.fr

CP31

Tight Complexity Bounds for Counting Generalized Dominating Sets in Bounded-Treewidth Graphs

For sets σ, ρ of non-negative integers, a (σ, ρ) -set of a graph G is a set S of vertices such that $|N(u) \cap S| \in \sigma$ for every $u \in S$, and $|N(v) \cap S| \in \rho$ for every $v \notin S$. This unifies standard problems including Independent Set, Dominating Set, Independent Dominating Set. For all pairs of finite or cofinite sets (σ, ρ) , we determine (under standard complexity assumptions) the best possible value $c_{\sigma, \rho}$ to count (σ, ρ) -sets in time $O^*(c_{\sigma, \rho}^{tw})$ (if a tree decomposition of width tw is given in the input). Surprisingly, $c_{\sigma, \rho}$ is often significantly smaller than the natural bound achieved by existing algorithms [van Rooij, 2020]. For example, for the Perfect Code problem ($\sigma = \{0\}$, $\rho = \{1\}$), this improves the $O^*(3^{tw})$ algorithm to $O^*(2^{tw})$. We show that our algorithms are most likely optimal, i.e., for any (non-trivial) pair (σ, ρ) of finite or cofinite sets, and any $\epsilon > 0$, a $O^*((c_{\sigma, \rho} - \epsilon)^{tw})$ -algorithm counting the number of (σ, ρ) -sets would violate #SETH. For finite sets σ and ρ , our lower bounds also extend to the decision version, showing that our algorithms are optimal in this setting as well. In contrast, for many cofinite sets, we show that further significant improvements for the decision and optimization versions are possible using the technique of representative sets.

Jacob Focke, Dániel Marx, Fionn Mc Inerney, Daniel Neuen
CISPA Helmholtz Center for Information Security
jacob.focke@cispa.de, marx@cispa.de,
fmcinern@gmail.com, dneuen@sfu.ca

Govind Sankar
Duke University
govind.subash.sankar@duke.edu

Philipp Schepper
CISPA Helmholtz Center for Information Security
philipp.schepper@cispa.de

Philip Wellnitz
Max Planck Institute for Informatics,
Saarland Informatics Campus, Germany
wellnitz@mpi-inf.mpg.de

CP31

Model-Checking for First-Order Logic with Disjoint Paths Predicates in Proper Minor-Closed

Graph Classes

The disjoint paths logic, FOL+DP, is an extension of First Order Logic (FOL) with the extra atomic predicate $\text{dp}_k(x_1, y_1, \dots, x_k, y_k)$, expressing the existence of internally vertex-disjoint paths between x_i and y_i , for $i \in \{1, \dots, k\}$. This logic can express a wide variety of problems that escape the expressibility potential of FOL. We prove that for every minor-closed graph class, model-checking for FOL+DP can be done in quadratic time. We also introduce an extension of FOL+DP, namely the scattered disjoint paths logic, FOL+SDP, where we further consider the atomic predicate $s\text{-sdp}_k(x_1, y_1, \dots, x_k, y_k)$, demanding that the disjoint paths are within distance bigger than some fixed value s . Using the same technique we prove that model-checking for FOL+SDP can be done in quadratic time on classes of graphs with bounded Euler genus.

Giannos Stamoulis

University of Montpellier and the National Center
for Scientific Research
giannos.stamoulis@lirmm.fr

Dimitrios Thilikos
CNRS, LIRMM, Montpellier
sedthilk@thilikos.info

Petr Golovach
University of Bergen, Norway
petr.golovach@uib.no

CP31

Parameterized Approximation Scheme for Biclique-Free Max K-Weight Sat and Max Coverage

Max-SAT with cardinality constraint (CC-Max-Sat) is one of the classical NP-complete problems, that generalizes Maximum Coverage, Partial Vertex Cover, Max-2-SAT with bisection constraints, and has been extensively studied across all algorithmic paradigms. In this problem, we are given a CNF-formula Φ , and a positive integer k , and the goal is to find an assignment β with at most k variables set to true (also called a weight k -assignment) such that the number of clauses satisfied by β is maximized. We consider $K_{d,d}$ -free formulas (that is, the clause-variable incidence bipartite graph of the formula excludes $K_{d,d}$ as an induced subgraph). We show that for every $\epsilon > 0$, there exists an algorithm for CC-Max-Sat on $K_{d,d}$ -free formulas with approximation ratio $(1 - \epsilon)$ and running in time $2^{O((\frac{dk}{\epsilon})^d)}(n + m)^{O(1)}$ (these algorithms are called FPT-AS). For, Maximum Coverage on $K_{d,d}$ -free set families, we obtain FPT-AS with running time $(\frac{dk}{\epsilon})^{O(dk)}n^{O(1)}$. Our second result considers “optimizing k ”, with fixed covering constraint for the Maximum Coverage problem. In particular, if k denotes the minimum size of $S \subseteq A$, such that $|N(S)| \geq t$, then our algorithm runs in time $(kd)^{O(kd)}n^{O(1)}$ and returns a set S' such that $|N(S')| \geq t$ and $|S'| \leq k + 1$.

Anannya Upasana
The Institute of Mathematical Sciences
anannya.upasana23@gmail.com

Pallavi Jain, Lawqueen Kanesh
Indian Institute of Technology Jodhpur
pallavijain.t.cms@gmail.com,
lawqueenkanesh091@gmail.com

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

Souvik Saha
The Institute of Mathematical Sciences, HBNI
souviksaha41@gmail.com

Abhishek Sahu
National Institute of Science Education and Research,
HBNI
asahuiitkgp@gmail.com

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

CP32

An Optimal Algorithm for Certifying Monotone Functions

Given query access to a monotone function $f: \{0,1\}^n \rightarrow \{0,1\}$ with certificate complexity $C(f)$ and an input x^* , we design an algorithm that outputs a size- $C(f)$ subset of x^* certifying the value of $f(x^*)$. Our algorithm makes $O(C(f) \cdot \log n)$ queries to f , which matches the information-theoretic lower bound for this problem and resolves the main open question posed in the STOC 2022 paper of Blanc, Koch, Lange, and Tan [BKLT22]. We extend this result to an algorithm that finds a size- $2C(f)$ certificate for a real-valued monotone function with $O(C(f) \cdot \log n)$ queries. We also complement our algorithms with a hardness result, in which we show that finding the shortest possible certificate for x^* may require $\Omega\left(\binom{n}{C(f)}\right)$ queries in the worst case.

Meghal Gupta
Microsoft Research
meghal@mit.edu

Naren S. Manoj
Toyota Technological Institute Chicago
nsm@ttic.edu

CP32

Fully-Dynamic-to-Incremental Reductions with Known Deletion Order (E.G. Sliding Window)

Dynamic algorithms come in three main flavors: incremental (insertions-only), decremental (deletions-only), or fully dynamic (both insertions and deletions). Fully dynamic is the holy grail of dynamic algorithm design; it is obviously more general than the other two, but is it strictly harder? Several works managed to reduce fully dynamic to the incremental or decremental models by taking advantage of either specific structure of the incremental/decremental algorithms (e.g. [HK99, HLT01, BKS12, ADKKP16]), or specific order of insertions/deletions (e.g. [AW14, HKNS15, KPP16]). Our goal in this work is to get a black-box fully-to-incremental reduction that is as general as possible. We find that the following conditions are necessary: • The incremental algorithm must have a worst-case (rather than amortized) running time guarantee. • The reduction must work in what we call the *deletions-look-ahead model*, where the order of deletions among current elements is known in advance. A notable practical example is the “sliding window” (FIFO) order of updates. Under those

conditions, we design: • A simple, practical, amortized-fully-dynamic to worst-case-incremental reduction with a $\log(T)$ -factor overhead on the running time, where T is the total number of updates. • A theoretical worst-case-fully-dynamic to worst-case-incremental reduction with a $\text{polylog}(T)$ -factor overhead on the running time.

Binghui Peng
Columbia University
bp2601@columbia.edu

Aviad Rubinfeld
Stanford
aviad@cs.stanford.edu

CP32

Tight Bounds for Vertex Connectivity in Dynamic Streams

We present a streaming algorithm for the vertex connectivity problem in dynamic streams with a (nearly) optimal space bound: for any n -vertex graph G and any integer $k \geq 1$, our algorithm with high probability outputs whether or not G is k -vertex-connected in a single pass using $\tilde{O}(kn)$ space. Our upper bound matches the known $\Omega(kn)$ lower bound for this problem even in insertion-only streams—which we extend to multi-pass algorithms in this paper—and closes one of the last remaining gaps in our understanding of dynamic versus insertion-only streams. Our result is obtained via a novel analysis of the previous best dynamic streaming algorithm of Guha, McGregor, and Tench [PODS 2015] who obtained an $\tilde{O}(k^2n)$ space algorithm for this problem. This also gives a model-independent algorithm for computing a “certificate” of k -vertex-connectivity as a union of $O(k^2 \log n)$ spanning forests, each on a random subset of $O(n/k)$ vertices, which may be of independent interest.

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

Vihan Shah
Rutgers University
vihan.shah98@rutgers.edu

CP32

Estimating the Effective Support Size in Constant Query Complexity

Estimating the support size of a distribution is a well-studied problem in statistics. Motivated by the fact that this problem is highly nonrobust (as small perturbations in the distributions can drastically affect the support size) and thus hard to estimate, Goldreich [ECCC 2019] studied the notion of ϵ -effective support size Ess_ϵ of a distribution \mathcal{P} , which is equal to the smallest support size of a distribution that is ϵ far in total variation distance from \mathcal{P} . In his paper, he shows an algorithm in the dual access setting (where we may both receive random samples and query the sampling probability $p(x)$ for any x) for a bicriteria approximation, giving an answer in $[\text{Ess}_{(1+\beta)\epsilon}, (1+\gamma)\text{Ess}_\epsilon]$ for some values β and γ . However, his algorithm has either super-constant complexity in the support size or super-constant approximation ratio $1+\gamma = \omega(1)$. He then asked if this is necessary, or if it is possible to get a constant-factor approximation in the number of samples independent of the support size. We answer his question by showing that not only is com-

plexity independent of n possible for $\gamma > 0$, but also for $\gamma = 0$, that is, that the bicriteria relaxation is not necessary. Our algorithm is very simple, and has 4 short lines of pseudocode.

Shyam Narayanan
Massachusetts Institute of Technology
shyamsn@mit.edu

Jakub Tetek
Basic Algorithms Research Copenhagen
University of Copenhagen
j.tetek@gmail.com

CP32

Sampling an Edge in Sublinear Time Exactly and Optimally

Sampling edges from a graph in sublinear time is a fundamental problem and a powerful subroutine for designing sublinear-time algorithms. Suppose we have access to the vertices of the graph and know a constant-factor approximation to the number of edges. An algorithm for pointwise ε -approximate edge sampling with complexity $O(n/\sqrt{\varepsilon m})$ has been given by Eden and Rosenbaum [SOSA 2018]. This has been later improved by Tetek and Thorup [STOC 2022] to $O(n \log(\varepsilon^{-1})/\sqrt{m})$. At the same time, $\Omega(n/\sqrt{m})$ time is necessary. We close the problem, under the assumption of knowing m up to constant factor, for all but very dense graphs by giving an algorithm with complexity $O(n/\sqrt{m} + \log \varepsilon^{-1})$. Our algorithm is based on a new technique that we call *Bernoulli trial simulation*. We believe this technique could also be useful for other problems. Given access to trials of the form $Bern(p)$, this technique allows us to simulate a Bernoulli trial $Bern(f(p) \pm \varepsilon)$ (without knowing p), in time complexity $O(\log \varepsilon^{-1})$ for some functions f . We specifically use this for $f(p) = 1/(2p)$ for $p \geq 2/3$. Therefore, we can perform rejection sampling, without the algorithm having to know the desired rejection probability.

Talya Eden
MIT and Boston University
talyaa01@gmail.com

Shyam Narayanan
Massachusetts Institute of Technology
shyamsn@mit.edu

Jakub Tetek
Basic Algorithms Research Copenhagen
University of Copenhagen
j.tetek@gmail.com

CP32

Simple Set Sketching

Imagine handling collisions in a hash table by storing, in each cell, the bit-wise exclusive-or of the set of keys hashing there. This appears to be a terrible idea: For αn keys and n buckets, where α is constant, we expect that a constant fraction of the keys will be unrecoverable due to collisions. We show that if this collision resolution strategy is repeated three times independently the situation reverses: If α is below a threshold of ≈ 0.81 then we can recover the set of all inserted keys in linear time with high probability. Even though the description of our data structure is simple, its analysis is nontrivial. Our approach can be seen as a variant of the Invertible Bloom Filter (IBF) of Eppstein

and Goodrich. While IBFs involve an explicit checksum per bucket to decide whether the bucket stores a single key, we exploit the idea of quotienting, namely that some bits of the key are implicit in the location where it is stored. We let those serve as an implicit checksum. These bits are not quite enough to ensure that no errors occur and the main technical challenge is to show that decoding can recover from these errors.

Houen Jakob Bæk Tejs, Rasmus Pagh
University of Copenhagen
jahn@di.ku.dk, pagh@di.ku.dk

Stefan Walzer
Cologne University
walzer@cs.uni-koeln.de

CP33

An Optimal Lower Bound for Simplex Range Reporting

We give a simplified and improved lower bound for the simplex range reporting problem. We show that given a set P of n points in \mathbb{R}^d , any data structure that uses $S(n)$ space to answer such queries must have $Q(n) = \Omega((n^2/S(n))^{(d-1)/d} + k)$ query time, where k is the output size. For near-linear space data structures, i.e., $S(n) = O(n \log^{O(1)} n)$, this improves the previous lower bounds by Chazelle and Rosenberg [CR96] and Afshani [A12] but perhaps more importantly, it is the first ever tight lower bound for any variant of simplex range searching for $d \geq 3$ dimensions. We obtain our lower bound by making a simple connection to well-studied problems in incident geometry which allows us to use known constructions in the area. We observe that a small modification of a simple already existing construction can lead to our lower bound. We believe that our proof is accessible to a much wider audience, at least compared to the previous intricate probabilistic proofs based on measure arguments by Chazelle and Rosenberg [CR96] and Afshani [A12]. The lack of tight or almost-tight (up to polylogarithmic factor) lower bounds for near-linear space data structures is a major bottleneck in making progress on problems such as proving lower bounds for multilevel data structures. It is our hope that this new line of attack based on incidence geometry can lead to further progress in this area.

Peyman Afshani
Aarhus University
peyman@cs.au.dk

Pingan Cheng
Department of Computer Science
Aarhus University
pingancheng@cs.au.dk

CP33

Derandomization of Cell Sampling

Since 1989, the best known lower bound on static data structures was Siegel's classical cell sampling lower bound. Siegel showed an explicit problem with n inputs and m possible queries such that every data structure that answers queries by probing t memory cells requires space $s \geq \tilde{\Omega}\left(n \cdot \left(\frac{m}{n}\right)^{1/t}\right)$. In this work, we improve this bound for non-adaptive data structures to $s \geq \tilde{\Omega}\left(n \cdot \left(\frac{m}{n}\right)^{1/(t-1)}\right)$ for all $t \geq 2$. For $t = 2$, we give a lower bound of

$s > m - o(m)$, improving on the bound $s > m/2$ recently proved by Viola over \mathbb{F}_2 and Siegel's bound $s \geq \tilde{\Omega}(\sqrt{mn})$ over other finite fields.

Alexander Golovnev
Georgetown University
alexgolovnev@gmail.com

Tom Gur
University of Warwick
tom.gur@warwick.ac.uk

Igor Shinkar
Simon Fraser University
ishinkar@sfu.ca

CP33

Splay Top Trees

The top tree data structure is an important and fundamental tool in dynamic graph algorithms. Top trees have existed for decades, and today serve as an ingredient in many state-of-the-art algorithms for dynamic graphs. In this work, we give a new direct proof of the existence of top trees, facilitating simpler and more direct implementations of top trees, based on ideas from splay trees. This result hinges on new insights into the structure of top trees, and in particular the structure of each root path in a top tree. In amortized analysis, our top trees match the asymptotic bounds of the state of the art.

Jacob Holm
University of Copenhagen
University of Copenhagen
jaho@di.ku.dk

Eva Rotenberg, Alice Ryhl
Technical University of Denmark
eva@rotenberg.dk, alice@ryhl.io

CP33

Proximity Search in the Greedy Tree

Over the last 50 years, there have been many data structures proposed to perform proximity search problems on metric data. Perhaps the simplest of these is the ball tree, which was independently discovered multiple times over the years. However, there is a lack of strong theoretical guarantees for standard ball trees, often leading to more complicated structures when guarantees are required. In this paper, we present the *greedy tree*, a simple ball tree construction for which we can prove strong theoretical guarantees for proximity search queries, matching the state of the art under reasonable assumptions. To our knowledge, this is the first ball tree construction providing such guarantees. Like a standard ball tree, it is a binary tree with the points stored in the leaves. Only a point, a radius, and an integer are stored for each node. The asymptotic running times of search algorithms in the greedy tree match those of more complicated structures regularly used in practice.

Oliver Chubet, Parth Parikh
North Carolina State University
oachubet@ncsu.edu, pmparikh@ncsu.edu

Donald Sheehy
University of Connecticut

don.r.sheehy@gmail.com

Siddharth S. Sheth
North Carolina State University
ssheth4@ncsu.edu

CP33

Optimal Resizable Arrays

A *resizable array* is an array that can *grow* and *shrink* by the addition or removal of items from its end while still supporting constant-time *access* to each stored item given its *index*. Since the size of an array varies over time, space-efficient maintenance of a resizable array requires dynamic memory management. A standard doubling technique allows the maintenance of an array of size N using only $O(N)$ space, with $O(1)$ amortized time, or even $O(1)$ worst-case time, per operation. Sitarski and Brodnik et al. describe much better solutions that maintain a resizable array of size N using only $N + O(\sqrt{N})$ space, still with $O(1)$ time per operation. Brodnik et al. give a simple proof that this is best possible. We distinguish between the space needed for *storing* a resizable array, and accessing its items, and the *temporary* space that may be needed while growing or shrinking the array. For every $r \geq 2$, we show that $N + O(N^{1/r})$ space is sufficient for storing and accessing an array of size N , if $N + O(N^{1-1/r})$ space can be used briefly during grow and shrink operations. Accessing an item by index takes $O(1)$ worst-case time while grow and shrink operations take $O(r)$ amortized time. Using an exact analysis of a *growth game*, we show that for any data structure from a wide class of data structures that uses only $N + O(N^{1/r})$ space to store the array, the amortized cost of grow is $\Omega(r)$.

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

Robert Tarjan
Princeton University
ret@princeton.edu

CP34

Sampling Equilibria: Fast No-Regret Learning in Structured Games

Learning and equilibrium computation in games are fundamental problems across computer science and economics. Much of the work in this area revolves around a simple algorithm termed *randomized weighted majority* (RWM), also known as “Hedge” or “Multiplicative Weights Update,” which is well known to achieve statistically optimal rates in adversarial settings (Freund and Schapire 99). Unfortunately, RWM comes with an inherent computational barrier: it requires maintaining and sampling from a distribution over all possible actions. In typical settings of interest the action space is exponentially large, seemingly rendering RWM useless in practice. In this work, we refute this notion for a broad variety of *structured* games, showing it is possible to efficiently (approximately) sample the action space in RWM in *polylogarithmic* time. This gives the first efficient no-regret algorithms for problems such as the (*discrete*) *Colonel Blotto game*, *matroid congestion*, *matroid security*, and basic *dueling games*. As an immediate corollary, we give a polylogarithmic time meta-algorithm to compute approximate Nash Equilibria for these games that is exponentially faster than prior methods in several

important settings. Further, our algorithm is the first to efficiently compute equilibria for more involved variants of these games with general sums, more than two players, and, for Colonel Blotto, multiple resource types.

Daniel Beaglehole
UC San Diego
dbeaglehole@ucsd.edu

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

Daniel Kane, Sihan Liu
UC San Diego
dakane@eng.ucsd.edu, sil046@ucsd.edu

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

CP34

Bidder Subset Selection Problem in Auction Design

Motivated by practical concerns in the online advertising industry, we study a bidder subset selection problem in single-item auctions. In this problem, a large pool of candidate bidders has independent values sampled from known prior distributions. The seller needs to pick a subset of bidders and run a given auction format on the selected subset to maximize her expected revenue. We propose two frameworks for the subset restrictions: (i) capacity constraint on the set of selected bidders; and (ii) incurred costs for the bidders invited to the auction. For the second-price auction with anonymous reserve (SPA-AR), we give constant approximation polynomial time algorithms in both frameworks (in the latter framework under mild assumptions about the market). Our results are in stark contrast to the previous work of Mehta, Nadav, Psomas, Rubinstein [NeurIPS 2020], who showed hardness of approximation for the SPA without a reserve price. We also give complementary approximation results for other well-studied auction formats such as anonymous posted pricing and sequential posted pricing. On a technical level, we find that the revenue of SPA-AR as a set function $f(S)$ of its bidders S is fractionally-subadditive but not submodular. Our bidder selection problem with invitation costs is a natural question about (approximately) answering a demand oracle for $f(\cdot)$ under a given vector of costs, a common computational assumption in the literature on combinatorial auctions.

Xiaohui Bei
Nanyang Technological University
xhbei@ntu.edu.sg

Nick Gravin, Pinyan Lu, Zhihao Gavin Tang
Shanghai University of Finance and Economics
nikolai@mail.shufe.edu.cn, lu.pinyan@mail.shufe.edu.cn, gavintang3@gmail.com

CP34

Foundations of Transaction Fee Mechanism Design

In blockchains such as Bitcoin and Ethereum, users compete in a transaction fee auction to get their transactions confirmed in the next block. A line of recent works set forth the desiderata for a "dream" transaction fee mechanism

(TFM), and explored whether such a mechanism existed. A dream TFM should satisfy 1) user incentive compatibility (UIC), i.e., truthful bidding should be a user's dominant strategy; 2) miner incentive compatibility (MIC), i.e., the miner's dominant strategy is to faithfully implement the prescribed mechanism; and 3) miner-user side contract proofness (SCP), i.e., no coalition of the miner and one or more user(s) can increase their joint utility by deviating from the honest behavior. The weakest form of SCP is called 1-SCP, where we only aim to provide resilience against the collusion of the miner and a single user. Sadly, despite the various attempts, to the best of knowledge, no existing mechanism can satisfy all three properties in all situations. Since the TFM departs from classical mechanism design in modeling and assumptions, to date, our understanding of the design space is relatively little. In this paper, we further unravel the mathematical structure of transaction fee mechanism design by proving the following results: - Can we have a dream TFM? - Rethinking the incentive compatibility notions. - Do the new design elements make a difference?

Hao Chung
Carnegie Mellon University
haochung@andrew.cmu.edu

Elaine Shi
Cornell University
runting@gmail.com

CP34

"Who Is Next in Line?" On the Significance of Knowing the Arrival Order in Bayesian Online Settings

We introduce a new measure for the performance of online algorithms in Bayesian settings, where the input is drawn from a known prior, but the realizations are revealed one-by-one in an online fashion. Our new measure is called order-competitive ratio. It is defined as the worst case (over all distribution sequences) ratio between the performance of the best order-unaware and order-aware algorithms, and quantifies the loss that is incurred due to lack of knowledge of the arrival order. Despite the growing interest in the role of the arrival order on the performance of online algorithms, this loss has been overlooked thus far. We study the order-competitive ratio in the paradigmatic prophet inequality problem, for the two common objective functions of (i) maximizing the expected value, and (ii) maximizing the probability of obtaining the largest value; and with respect to two families of algorithms, namely (i) adaptive algorithms, and (ii) single-threshold algorithms. We provide tight bounds for all four combinations, with respect to deterministic algorithms. Our analysis requires new ideas and departs from standard techniques. In particular, our adaptive algorithms inevitably go beyond single-threshold algorithms. The results with respect to the order-competitive ratio measure capture the intuition that adaptive algorithms are stronger than single-threshold ones, and may lead to a better algorithmic advice than the classical competitive ratio measure.

Zhihao Gavin Tang
Shanghai University of Finance and Economics
gavintang3@gmail.com

Tomer Ezra
Sapienza University of Rome
ezra@diag.uniroma1.it

Michal Feldman
Tel-Aviv University
mfeldman@tau.ac.il

Nick Gravin
Shanghai University of Finance and Economics
nikolai@mail.shufe.edu.cn

CP34

Simple Mechanism for Non-Linear Agents

We show that economic conclusions derived from Bulow and Roberts (1989) for linear utility models approximately extend to non-linear utility models. Specifically, we quantify the extent to which agents with non-linear utilities resemble agents with linear utilities, and we show that the approximation of mechanisms for agents with linear utilities approximately extend for agents with non-linear utilities. We illustrate the framework for the objectives of revenue and welfare on non-linear models that include agents with budget constraints, agents with risk aversion, and agents with endogenous valuations. We derive bounds on how much these models resemble the linear utility model and combine these bounds with well-studied approximation results for linear utility models. We conclude that simple mechanisms are approximately optimal for these non-linear agent models.

Yiding Feng
Northwestern University
Department of Electrical Engineering & Computer Science
yidingfeng2021@u.northwestern.edu

Jason Hartline
Northwestern University
hartline@northwestern.edu

Yingkai Li
Yale University
yingkai.li@yale.edu

CP34

A Polynomial-Time Algorithm for 1/2-Well-Supported Nash Equilibria in Bimatrix Games

Since the seminal PPAD-completeness result for computing a Nash equilibrium even in two-player games, an important line of research has focused on relaxations achievable in polynomial time. In this paper, we consider the notion of ϵ -well-supported Nash equilibrium, where $\epsilon \in [0, 1]$ corresponds to the approximation guarantee. Put simply, in an ϵ -well-supported equilibrium, every player chooses with positive probability actions that are within ϵ of the maximum achievable payoff, against the other player's strategy. Ever since the initial approximation guarantee of $2/3$ for well-supported equilibria, which was established more than a decade ago, the progress on this problem has been extremely slow and incremental. Notably, the small improvements to 0.6608, and finally to 0.6528, were achieved by algorithms of growing complexity. Our main result is a simple and intuitive algorithm, that improves the approximation guarantee to $1/2$. Our algorithm is based on linear programming and in particular on exploiting suitably defined zero-sum games that arise from the payoff matrices of the two players. As a byproduct, we show how to achieve the same approximation guarantee in a query-efficient way.

Argyrios Deligkas
Royal Holloway, University of London
argyrios.deligkas@rhul.ac.uk

Michail Fasoulakis
Foundation for Research and Technology-Hellas
Athens University of Economics and Business
mfasoul@ics.forth.gr

Evangelos Markakis

Athens University of Economics and Business (AUEB)
and IOG
Athens University of Economics and Business - Research Center
markakis@gmail.com

CP35

Beating Greedy Matching in Sublinear Time

We study sublinear time algorithms for estimating the size of maximum matching in graphs. Our main result is a $(\frac{1}{2} + \Omega(1))$ -approximation algorithm which can be implemented in $O(n^{1+\epsilon})$ time, where n is the number of vertices and the constant $\epsilon > 0$ can be made arbitrarily small. The best known lower bound for the problem is $\Omega(n)$, which holds for any constant approximation. Existing algorithms either obtain the greedy bound of $\frac{1}{2}$ -

approximation [Behnezhad FOCS'21], or require some assumption on the maximum degree to run in $o(n^2)$ -time [Yoshida, Yamamoto, and Ito STOC'09]. We improve over these by designing a less 'adaptive' augmentation algorithm for maximum matching that might be of independent interest.

Soheil Behnezhad, Mohammad Roghani
Stanford University
soheil.behnezhad@gmail.com, roghani@stanford.edu

Aviad Rubinfeld
Stanford
aviad@cs.stanford.edu

Amin Saberi
Stanford University
amin.saberi@gmail.com

CP35

Near-Linear Sample Complexity for L_p Polynomial Regression

We study L_p polynomial regression. Given query access to a function $f : [-1, 1] \rightarrow \mathbb{R}$, the goal is to find a degree d polynomial \hat{q} such that, for a given parameter $\epsilon > 0$,

$$\|\hat{q} - f\|_p \leq (1 + \epsilon) \cdot \min_{q: \deg(q) \leq d} \|q - f\|_p$$

Here $\|\cdot\|_p$ is the L_p norm. We show that querying f at points randomly drawn from the Chebyshev measure on $[-1, 1]$ is a near-optimal strategy for polynomial regression in all L_p norms. To find \hat{q} , it suffices to sample $O(d \frac{\text{polylog } d}{\text{poly } \epsilon})$ points from $[-1, 1]$ with probabilities proportional to this measure. While the optimal sample complexity for polynomial regression was understood for L_2 and L_∞ , ours is the first result that achieves sample complexity linear in

d and error $(1 + \epsilon)$ for other values of p without any assumptions. Our result requires two main technical contributions. The first is for $p \leq 2$, where we explicitly bound the L_p Lewis weight function of the infinite linear operator underlying polynomial regression. Using tools from the orthogonal polynomial literature, we bound this function by the Chebyshev density. Our second key contribution is to take advantage of the structure of polynomials to reduce the $p > 2$ case to the $p \leq 2$ case. By doing so, we obtain a better sample complexity than what is possible for general p -norm linear regression problems, for which $\Omega(d^{p/2})$ samples are required.

Raphael A. Meyer
New York University
ram900@nyu.edu

Cameron Musco
University of Massachusetts Amherst
cmusco@cs.umass.edu

Christopher Musco
New York University
Tandon School of Engineering
cmusco@nyu.edu

David Woodruff, Samson Zhou
Carnegie Mellon University
dwoodruf@cs.cmu.edu, samsonzhou@gmail.com

CP35

Testing Convex Truncation

We study the basic statistical problem of testing whether normally distributed n -dimensional data has been *truncated*, i.e. altered by only retaining points that lie in some unknown truncation set $S \subseteq \mathbb{R}^n$. As our main algorithmic results, 1. We give a computationally efficient $O(n)$ -sample algorithm that can distinguish the standard normal distribution $N(0, I_n)$ from $N(0, I_n)$ conditioned on an unknown and arbitrary convex set S . 2. We give a different computationally efficient $O(n)$ -sample algorithm that can distinguish $N(0, I_n)$ from $N(0, I_n)$ conditioned on an unknown and arbitrary mixture of symmetric convex sets. These results stand in sharp contrast with known results for learning or testing convex bodies with respect to the normal distribution or learning convex-truncated normal distributions, where state-of-the-art algorithms require essentially $n^{\sqrt{n}}$ samples. An easy argument shows that no finite number of samples suffices to distinguish $N(0, I_n)$ from an unknown and arbitrary mixture of general (not necessarily symmetric) convex sets, so no common generalization of results (1) and (2) above is possible. We also prove lower bounds on the sample complexity of distinguishing algorithms (computationally efficient or otherwise) for various classes of convex truncations; in some cases these lower bounds match our algorithms up to logarithmic or even constant factors.

Shivam Nadimpalli, Rocco Servedio
Columbia University
sn2855@columbia.edu, ras2105@columbia.edu

Anindya De
University of Pennsylvania

de.anindya@gmail.com

CP35

Spencer's Theorem in Nearly Input-Sparsity Time

A celebrated theorem of Spencer states that for every set system $S_1, \dots, S_m \subseteq [n]$, there is a coloring of the ground set with $\{\pm 1\}$ with discrepancy $O(\sqrt{n \log(m/n+2)})$. We provide an algorithm to find such a coloring in near input-sparsity time $\tilde{O}(n + \sum_{i=1}^m |S_i|)$.

Mehtaab Sawhney
Massachusetts Institute of Technology
msawhney@mit.edu

Vishesh Jain
University of Illinois Chicago
visheshj@uic.edu

Ashwin Sah
MIT
asah@mit.edu

CP35

Optimal Algorithms for Linear Algebra in the Current Matrix Multiplication Time

We study fundamental problems in linear algebra, such as finding a maximal linearly independent subset of rows or columns, solving linear regression, or computing a subspace embedding. For these problems, we consider input matrices $A \in \mathbb{R}^{n \times d}$ with $n > d$. The input can be read in $\text{nnz}(A)$ time, which denotes the number of nonzero entries of A . In this paper, we show that beyond the time required to read the input matrix, these fundamental linear algebra problems can be solved in d^ω time, where ω is the current matrix-multiplication exponent. To do so, we introduce a constant-factor subspace embedding with the optimal $m = O(d)$ number of rows, and which can be applied in time $O\left(\frac{\text{nnz}(A)}{\alpha}\right) + d^{2+\alpha} \text{poly}(\log d)$ for any trade-off parameter $\alpha > 0$, tightening a recent result by Chepurko et. al. that achieves an $\exp(\text{poly}(\log \log n))$ distortion with $m = d \cdot \text{poly}(\log \log d)$ rows in $O\left(\frac{\text{nnz}(A)}{\alpha} + d^{2+\alpha+o(1)}\right)$ time.

Our subspace embedding uses a recently shown property of stacked Subsampled Randomized Hadamard Transforms, which actually increase the input dimension, to “spread” the mass of an input vector among a large number of coordinates, followed by random sampling. We then use our constant-factor subspace embedding to give the first optimal runtime algorithms for finding a maximal linearly independent subset of columns, regression, and leverage score sampling.

Yeshwanth Cherapanamjeri
UC Berkeley
yeshwanth@berkeley.edu

Sandeep Silwal
MIT
silwal@mit.edu

David Woodruff, Samson Zhou
Carnegie Mellon University

dwoodruf@cs.cmu.edu, samsonzhou@gmail.com

CP35

Streaming Complexity of CSPs with Randomly Ordered Constraints

We initiate a study of the streaming complexity of constraint satisfaction problems (CSPs) when the constraints arrive in a random order. We show that there exists a CSP, namely Max-DICUT, for which random ordering makes a provable difference. Whereas a $4/9 \approx 0.445$ approximation of DICUT requires $\Omega(\sqrt{n})$ space with adversarial ordering, we show that with random ordering of constraints there exists a 0.48-approximation algorithm that only needs $O(\log n)$ space. We also give new algorithms for Max-DICUT in variants of the adversarial ordering setting. On the negative side, we prove that CSPs where the satisfying assignments of the constraints support a one-wise independent distribution require $\Omega(\sqrt{n})$ -space for any non-trivial approximation, even when the constraints are randomly ordered. This was previously known only for adversarially ordered constraints. The only CSP to have been considered previously with random ordering is Max-CUT where the ordering is known not to change the approximability. Specifically it is known to be as hard to approximate with random ordering as with adversarial ordering, for $o(\sqrt{n})$ space algorithms. Our results show a richer variety of possibilities and motivate further study of CSPs with randomly ordered constraints.

Noah Singer
Carnegie Mellon University
ngsinger@cs.cmu.edu

Raghuvansh Saxena
Microsoft Research
raghuvansh.saxena@gmail.com

Madhu Sudan
Harvard
madhu@cs.harvard.edu

Santhoshini Velusamy
Harvard University
svelusamy@g.harvard.edu

CP36

Smaller Low-Depth Circuits for Kronecker Powers

We give new, smaller constructions of constant-depth linear circuits for computing any matrix which is the Kronecker power of a fixed matrix. A standard argument (e.g., the mixed product property of Kronecker products, or a generalization of the Fast Walsh-Hadamard transform) shows that any such $N \times N$ matrix has a depth-2 circuit of size $O(N^{1.5})$. We improve on this for all such matrices, and especially for some such matrices of particular interest: For any integer $q > 1$ and any matrix which is the Kronecker power of a fixed $q \times q$ matrix, we construct a depth-2 circuit of size $O(N^{1.5-a_q})$, where $a_q > 0$ is a positive constant depending only on q . No bound beating size $O(N^{1.5})$ was previously known for any $q > 2$. For the case $q = 2$, i.e., for any matrix which is the Kronecker power of a fixed 2×2 matrix, we construct a depth-2 circuit of size $O(N^{1.446})$, improving the prior best size $O(N^{1.493})$ [Alman, 2021]. For the Walsh-Hadamard transform, we construct a depth-2 circuit of size $O(N^{1.443})$, improving the prior best size $O(N^{1.476})$ [Alman, 2021]. For the disjointness matrix

(the communication matrix of set disjointness, or equivalently, the matrix for the linear transform that evaluates a multilinear polynomial on all 0/1 inputs), we construct a depth-2 circuit of size $O(N^{1.258})$, improving the prior best size $O(N^{1.272})$ [Jukna and Sergeev, 2013].

Yunfeng Guan, Josh Alman, Ashwin Padaki
Columbia University
yunfeng.guan@columbia.edu, josh@cs.columbia.edu,
ap4025@columbia.edu

CP36

Algebraic Algorithms for Fractional Linear Matroid Parity via Non-Commutative Rank

Matrix representations are a powerful tool for designing efficient algorithms for combinatorial optimization problems such as matching, and linear matroid intersection and parity. In this paper, we initiate the study of matrix representations using the concept of non-commutative rank (nc-rank), which has recently attracted attention in the research of Edmonds' problem. We reveal that the nc-rank of the matrix representation of linear matroid parity corresponds to the optimal value of fractional linear matroid parity: a half-integral relaxation of linear matroid parity. Based on our representation, we present an algebraic algorithm for the fractional linear matroid parity problem by building a new technique to incorporate the search-to-decision reduction into the half-integral problem represented via the nc-rank. We further present a faster divide-and-conquer algorithm for finding a maximum fractional matroid matching and an algebraic algorithm for finding a dual optimal solution. They together lead to an algebraic algorithm for the weighted fractional linear matroid parity problem. Our algorithms are significantly simpler and faster than the existing algorithms.

Taihei Oki
The University of Tokyo
oki@mist.i.u-tokyo.ac.jp

Tasuku Soma
MIT
tasuku@mit.edu

CP36

On the Orbit Closure Intersection Problems for Matrix Tuples under Conjugation and Left-Right Actions

Let G be a linear algebraic group acting on the vector space V . Given $v, v' \in V$, the orbit closure intersection problem asks to decide if the orbit closures of v and v' under G intersect. Due to connections with polynomial identity testing, the orbit closure intersection problems for the conjugation and left-right actions on matrix tuples received considerable attention in computational complexity and computational invariant theory, as seen in the works of Forbes-Shpilka (RANDOM 2013), Allen-Zhu-Garg-Li-Oliveira-Wigderson (STOC 2018), and Derksen-Makam (Algebra & Number Theory 2020). In this paper, we present new algorithms for the orbit closure problem for the conjugation and left-right actions on matrix tuples. The main novel feature is that in the case of intersecting orbit closures, our algorithm outputs cosets of one-parameter subgroups that drive the matrix tuples to a tuple in the intersection of the orbit closures.

Youning Qiao

Centre for Quantum Software and Information
University of Technology Sydney
jimmyqiao86@gmail.com

Gábor Ivanyos
Institute for Computer Science and Control
Hungarian Academy of Sciences
gabor.ivanyos@sztaki.mta.hu

CP36

Toeplitz Low-Rank Approximation with Sublinear Query Complexity

We present a sublinear query algorithm for outputting a near-optimal low-rank approximation to any positive semidefinite Toeplitz matrix $T \in \mathbb{R}^{d \times d}$. In particular, for any integer rank $k \leq d$ and $\epsilon, \delta > 0$, our algorithm makes $\tilde{O}(k^2 \cdot \log(1/\delta) \cdot \text{poly}(1/\epsilon))$ queries to the entries of T and outputs a rank $\tilde{O}(k \cdot \log(1/\delta)/\epsilon)$ matrix $\tilde{T} \in \mathbb{R}^{d \times d}$ such that $\|T - \tilde{T}\|_F \leq (1 + \epsilon) \cdot \|T - T_k\|_F + \delta \|T\|_F$. Here, $\|\cdot\|_F$ is the Frobenius norm and T_k is the optimal rank- k approximation to T , given by projection onto its top k eigenvectors. $\tilde{O}(\cdot)$ hides $\text{poly} \log(d)$ factors. Our algorithm is structure-preserving, in that the approximation \tilde{T} is also Toeplitz. A key technical contribution is a proof that any positive semidefinite Toeplitz matrix in fact has a near-optimal low-rank approximation which is itself Toeplitz. Surprisingly, this basic existence result was not previously known. Building on this result, along with the well-established off-grid Fourier structure of Toeplitz matrices [Cybenko'82], we show that Toeplitz \tilde{T} with near optimal error can be recovered with a small number of random queries via a leverage-score-based off-grid sparse Fourier sampling scheme.

Michael Kapralov
EPFL
michael.kapralov@epfl.ch

Hannah Lawrence
CSAIL,
Massachusetts Institute of Technology
hanlaw@mit.edu

Mikhail Makarov
EPFL
mikhail.makarov@epfl.ch

Cameron Musco
University of Massachusetts Amherst
cmusco@cs.umass.edu

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

CP36

Equivalence Test For Read-Once Arithmetic Formulas

We study the polynomial equivalence problem for orbits of read-once arithmetic formulas (ROFs). ROFs have received considerable attention in both algebraic and Boolean complexity and have served as a testbed for developing techniques to analyze circuits. Two n -variate polynomials $f, g \in \mathbb{F}[x]$ are equivalent if there is an $A \in GL(n, \mathbb{F})$

s.t. $f = g(Ax)$. The orbit of f is the set of all polynomials equivalent to f . We study the following two natural problems on ROFs: 1. Equivalence test for ROFs: Given black-box access to f , check if it is in the orbit of an ROF. If yes, output an ROF C and an $A \in GL(n, \mathbb{F})$ s.t. $f = C(Ax)$. 2. Polynomial equivalence for orbits of ROFs: Given black-box access to f and g in the orbits of two unknown ROFs, check if they are equivalent. If yes, output an $A \in GL(n, \mathbb{F})$ s.t. $f = g(Ax)$. These problems are significant generalizations of two well-studied problems in algebraic complexity, viz. reconstruction of ROFs and quadratic form equivalence. In this work, we give the first randomized polynomial-time algorithms (with oracle access to quadratic form equivalence) to solve these problems. The equivalence test works for general ROFs; it also implies an efficient learning algorithm for random arithmetic formulas of unbounded depth and fan-in (in the high number of variables setting). The algorithm for the second problem works for additive-constant-free ROFs.

Nikhil Gupta, Chandan Saha, Bhargav Thankey
Indian Institute of Science, Bangalore
nikhilg@iisc.ac.in, chandan@iisc.ac.in,
thankeyd@iisc.ac.in

CP36

Nonlinear Codes Exceeding the Gilbert-Varshamov and Tsfasman-Vlăduț-Zink Bounds

The Gilbert-Varshamov (GV for short) bound has been a benchmark for good Hamming-metric codes. It was even conjectured by some coding theorists that the asymptotic Gilbert-Varshamov bound is tight. The GV bound had remained to be the best asymptotic lower bound for thirty years before it was broken by the Tsfasman-Vlăduț-Zink bound via algebraic geometry codes. The discovery of algebraic geometry codes by Goppa was a breakthrough in coding theory. After another twenty years, no any improvements on the Tsfasman-Vlăduț-Zink bound took place before the work by Xing-Elkies [?, ?, ?, ?] in the early of 2000 via tools from algebraic geometry. By using the similar ideas as in [?, ?, ?], some further improvements were given in [?, ?]. Since then, no further progress on asymptotic lower bounds has been made. The main result of this paper is to show that all previous asymptotic lower bounds can be improved in an interval. We present two types of constructions of Hamming-metric codes. Both constructions involve algebraic geometry. In order to obtain good codes, one construction requires a larger number of positive divisors of fixed degree, while other construction requires a smaller number of positive divisors of fixed degree. As a result, no matter how large the number of positive divisors of fixed degree is, we can always obtain codes with good parameters. It turns out that all previous asymptotic lower bounds are improved.

Chaoping Xing
Shanghai Jiao Tong University
xingcp@sjtu.edu.cn

Shu Liu
University of Electronic Science and Technology of China
shuliu@uestc.edu.cn

Ting Yi Wu
Huawei Technology Co. Ltd.

wu.ting.yi@huawei.com

CP37

Improved Distributed Algorithms for the Lovász Local Lemma and Edge Coloring

The Lovsz Local Lemma is a classic result in probability theory. It states that if we have n ‘bad events’, each of which occurs with probability at most p and is independent of all but d other events, then under certain criteria on p and d , all bad events can be avoided. While the original proof was existential, there has been much study on the algorithmic Lovsz Local Lemma: that is, on finding an assignment of the underlying random variables such that all the bad events are indeed avoided. Notably, the celebrated result of Moser and Tardos [JACM ’10] implied an $O(\log^2 n)$ -round distributed algorithm for the problem. For instances with low d , this was improved to $O(d^2 + \log^{O(1)} \log n)$ by Fischer and Ghaffari [DISC ’17], a result that proved highly important in distributed complexity theory. We give an improved algorithm for the Lovsz Local Lemma, providing a trade-off between the strength of the criterion relating p and d , and the distributed round complexity. In particular, in the same regime as Fischer and Ghaffari’s algorithm, we improve the round complexity to $O(\frac{d}{\log d} + \log^{O(1)} \log n)$. As our main application, we also give the first $\log^{O(1)} \log n$ -round distributed algorithm for the problem of $\Delta + o(\Delta)$ -edge coloring a graph of maximum degree Δ . This is an almost exponential improvement over previous results: no prior $\log^{o(1)} n$ -round algorithm was known even for $2\Delta - 2$ -edge coloring.

Peter W. Davies
Durham University
peter.w.davies@durham.ac.uk

CP37

A Nearly Time-Optimal Distributed Approximation of Minimum Cost k -Edge-Connected Spanning Subgraph

The minimum-cost k -edge-connected spanning subgraph (k -ECSS) problem is a generalization and strengthening of the well-studied minimum-cost spanning tree (MST) problem. While the round complexity of distributedly computing the latter has been well-understood, the former remains mostly open, especially as soon as $k \geq 3$. In this paper, we present the first distributed algorithm that computes an approximation of k -ECSS in sublinear time for general k . Concretely, we describe a randomized distributed algorithm that, in $\tilde{O}(k(D + k\sqrt{n}))$ rounds, computes a k -edge-connected spanning subgraph whose cost is within an $O(\log n \log k)$ factor of optimal. Here n is the number of vertices and D is the diameter of the graph. This time complexity is nearly optimal for any $k = \text{poly}(\log n)$, almost matching an $\tilde{\Omega}(D + \sqrt{n/k})$ lower bound. Our algorithm is the first to achieve a sublinear round complexity for $k \geq 3$. We note that this case is considerably more challenging than the well-studied and well-understood $k = 1$ case and the closely related $k = 2$ case. One key ingredient in our algorithm is a novel structural lemma that allows us to compress the information about all minimum cuts in a graph into a succinct representation, which is computed in a decentralized fashion. We hope that this succinct representation may find applications in other computational settings or for other problems.

Michal Dory

University of Haifa
mdory@ds.haifa.ac.il

Mohsen Ghaffari
MIT
ghaffari@mit.edu

CP37

Local Distributed Rounding: Generalized to MIS, Matching, Set Cover, and Beyond

We develop a general deterministic distributed method for locally rounding fractional solutions of graph problems for which the analysis can be broken down into analyzing pairs of vertices. Roughly speaking, the method can transform fractional/probabilistic label assignments of the vertices into integral/deterministic label assignments for the vertices, while approximately preserving a potential function that is a linear combination of functions, each of which depends on at most two vertices (subject to some conditions usually satisfied in pairwise analyses). The method unifies and significantly generalizes prior work on deterministic local rounding techniques [Ghaffari, Kuhn FOCS’21; Harris FOCS’19; Fischer, Ghaffari, Kuhn FOCS’17; Fischer DISC’17] to obtain polylogarithmic-time deterministic distributed solutions for combinatorial graph problems. Our general rounding result enables us to locally and efficiently derandomize a range of distributed algorithms for local graph problems, including maximal independent set (MIS), maximum-weight independent set approximation, and minimum-cost set cover approximation.

Salwa Faour
University of Freiburg
Germany
salwa.faour@cs.uni-freiburg.de

Mohsen Ghaffari
MIT
ghaffari@mit.edu

Christoph Grunau
ETH Zurich
Switzerland
cgrunau@inf.ethz.ch

Fabian Kuhn
University of Freiburg, Germany
kuhn@cs.uni-freiburg.de

Vaclav Rozhon
ETH Zurich
Switzerland
rozhonv@inf.ethz.ch

CP37

Parallel Exact Shortest Paths in Almost Linear Work and Square Root Depth

We devise a randomized parallel exact single-source shortest paths (SSSP) algorithm for directed graphs with non-negative integer edge weights in $\tilde{O}(m)$ work and $n^{1/2+o(1)}$ span with high probability. All previous exact SSSP algorithms with nearly linear work have linear span, even for undirected unweighted graphs. Our main technical contribution is to show a reduction from the exact SSSP to directed hopsets using the iterative gradual rounding technique. A (h, ϵ) -hopset is a set of weighted edges (some-

times called shortcuts) that when added to the graph admit h -hop paths with weights no more than $(1 + \epsilon)$ times the true shortest path distances. Furthermore, we show how to combine this algorithm with Forster and Nanongkai's framework to improve the distributed exact SSSP algorithm. Specifically, we obtain an $\tilde{O}(\sqrt{n} + D + n^{2/5+o(1)}D^{2/5})$ rounds exact SSSP algorithm for directed graphs with non-negative integer edge weights in the CONGEST model, where D is the diameter of the underlying undirected unweighted graph.

Nairen Cao, Jeremy Fineman
Georgetown University
nc645@georgetown.edu, jf474@georgetown.edu

CP37

Byzantine Agreement with Optimal Resilience via Statistical Fraud Detection

Since the mid-1980s it has been known that Byzantine Agreement can be solved with probability 1 asynchronously, even against an omniscient, computationally unbounded adversary that can adaptively *corrupt* up to $f < n/3$ parties. Moreover, the problem is insoluble with $f \geq n/3$ corruptions. However, Bracha's [Bra87] 1984 protocol (see also Ben-Or [Ben83]) achieved $f < n/3$ resilience at the cost of *exponential* expected latency $2^{\Theta(n)}$, a bound that has never been improved in this model with $f = (n - 1)/3$ corruptions. In this paper we prove that Byzantine Agreement in the asynchronous, full information model can be solved with probability 1 against an adaptive adversary that can corrupt $f < n/3$ parties, while incurring only *polynomial latency with high probability*. Our protocol follows earlier polynomial latency protocols of King and Saia [KS16, KS18] and Huang, Pettie, and Zhu [HPZ22], which had *suboptimal* resilience, namely $f \approx n/10^9$ and $f < n/4$, respectively. The core technical problem we solve is to design a collective coin-flipping protocol that *eventually* lets us flip a coin with an unambiguous outcome. After a polynomial number of executions of our coin-flipping protocol, either (a) the game ends or (b) we can "blacklist" players at different rates. The blacklisting criterion is based on a simple statistical test of *fraud detection*.

Shang-En Huang
University of Michigan
sehuang@umich.edu

Seth Pettie
University of Michigan, Ann Arbor
pettie@umich.edu

Leqi Zhu
University of Michigan
lezhu@umich.edu

CP37

Massively Parallel Computation on Embedded Planar Graphs

Many of the classic graph problems cannot be solved in the Massively Parallel Computation setting (MPC) with strongly sublinear space per machine and $o(\log n)$ rounds, unless the 1-vs-2 cycles conjecture is false. This is true even on planar graphs. Such problems include, for example, counting connected components, bipartition, minimum spanning tree problem, (approximate) shortest paths, and

(approximate) diameter/radius. In this paper, we show a way to get around this limitation. Specifically, we show that if we have a "nice" (for example, straight-line) embedding of the input graph, all the mentioned problems can be solved with $O(n^{2/3+\epsilon})$ space per machine in $O(1)$ rounds. In conjunction with existing algorithms for computing the Delaunay triangulation, our results imply an MPC algorithm for exact Euclidean minimum spanning tree (EMST) that uses $O(n^{2/3+\epsilon})$ space per machine and finishes in $O(1)$ rounds. This is the first improvement over a straightforward use of the standard Boruvka's algorithm with the Delaunay triangulation algorithm of Goodrich [SODA 1997] which results in $\Theta(\log n)$ rounds. This also partially negatively answers a question of Andoni, Nikolov, Onak, and Yaroslavtsev [STOC 2014], asking for lower bounds for exact EMST.

Jakob Holm, Jakub Tetek
Basic Algorithms Research Copenhagen
University of Copenhagen
jaho@di.ku.dk, j.tetek@gmail.com

CP38

Spatial Mixing and the Random-Cluster Dynamics on Lattices

An important paradigm in the understanding of mixing times of Glauber dynamics for spin systems is the correspondence between spatial mixing properties of the models and bounds on the mixing time of the dynamics. This includes, in particular, the classical notions of weak and strong spatial mixing, which have been used to show the best known mixing time bounds in the high-temperature regime for the Glauber dynamics for the Ising and Potts models. Glauber dynamics for the random-cluster model does not naturally fit into this spin systems framework because its transition rules are not local. In this paper, we present various implications between weak spatial mixing, strong spatial mixing, and the newer notion of spatial mixing within a phase, and mixing time bounds for the random-cluster dynamics in finite subsets of \mathbb{Z}^d for general $d \geq 2$. These imply a host of new results, including optimal $O(N \log N)$ mixing for the random cluster dynamics on torii and boxes on N vertices in \mathbb{Z}^d at all high temperatures and at sufficiently low temperatures, and for large values of q quasi-polynomial (or quasi-linear when $d = 2$) mixing time bounds from random phase initializations on torii at the critical point (where by contrast the mixing time from worst-case initializations is exponentially large). In the same parameter regimes, these results translate to fast sampling algorithms for the Potts model on \mathbb{Z}^d for general d .

Reza Gheissari
Berkeley
gheissari@northwestern.edu

Alistair Sinclair
University of California, Berkeley
sinclair@cs.berkeley.edu

CP38

Balanced Allocations with Heterogeneous Bins: The Power of Memory

We consider the allocation of m balls (jobs) into n bins (servers). In the standard Two-Choice process, at each step $t = 1, 2, \dots, m$ we first sample two bins uniformly at random and place a ball in the least loaded bin. It

is well-known that for any $m \geq n$, this results in a gap (difference between the maximum and average load) of $\log_2 \log n + \Theta(1)$ (with high probability). In the Memory process instead of two choices, we only sample one bin per step but we have access to a cache which can store the location of one bin. Mitzenmacher, Prabhakar and Shah (2002) showed that in the lightly loaded case ($m = n$), the Memory process achieves a gap of $O(\log \log n)$. Extending the setting of Mitzenmacher et al. in two ways, we first allow the number of balls m to be arbitrary, which includes the challenging heavily loaded case where $m \geq n$. Secondly, we follow the heterogeneous bins model of Wieder (2007), where the sampling distribution of bins can be biased up to some arbitrary multiplicative constant. Somewhat surprisingly, we prove that even in this setting, the Memory process still achieves a $O(\log \log n)$ gap bound. This is in stark contrast with the Two-Choice (or any d -Choice with $d = O(1)$) process, where it is known that the gap diverges, by Wieder (2007). Finally, we prove gap bounds for Memory in two more relaxed settings: one with heterogeneous (weighted) balls, and the other with arbitrary sampling distributions.

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

John Sylvester
University of Liverpool
john.sylvester@liverpool.ac.uk

CP38

The Need for Seed (in the Abstract Tile Assembly Model)

In the abstract Tile Assembly Model (aTAM) square tiles self-assemble, autonomously binding via glues on their edges, to form structures. Algorithmic aTAM systems can be designed in which the patterns of tile attachments are forced to follow the execution of targeted algorithms. Such systems have been proven to be computationally universal as well as intrinsically universal (IU), a notion borrowed and adapted from cellular automata showing that a single tile set exists which is capable of simulating all aTAM systems (FOCS 2012). The input to an algorithmic aTAM system can be provided in a variety of ways, with a common method being via the “seed” assembly, which is a pre-formed assembly from which all growth propagates. In this paper we present a series of results which investigate the trade-offs of using seeds consisting of a single tile, versus those containing multiple tiles. We show that arbitrary systems with multi-tile seeds cannot be converted to functionally equivalent systems with single-tile seeds without using a scale factor ≥ 1 . We prove tight bounds on the scale factor required, and also present a construction which uses a large scale factor but an optimal number of unique tile types. That construction is then used to develop a construction that performs simultaneous simulation of all aTAM systems in parallel, as well as to display a connection to other tile-based self-assembly models via the notion of intrinsic universality.

Matthew Patitz, Andrew Alseth
University of Arkansas
patitz@uark.edu, andrew.althseth@gmail.com

CP38

Faster Algorithm for Turn-Based Stochastic Games

with Bounded Treewidth

Turn-based stochastic games (aka simple stochastic games) are two-player zero-sum games played on directed graphs with probabilistic transitions. The goal of the max-player is to maximize the probability to reach a target state against the adversarial min-player. These games lie in $NP \cap coNP$ and are among the rare combinatorial problems that belong to this complexity class for which the existence of a polynomial-time algorithm is a major open question. While randomized sub-exponential time algorithm exists, all known deterministic algorithms require exponential time in the worst case. An important open question has been whether faster algorithms can be obtained parametrized by the treewidth of the game graph. Even deterministic sub-exponential time algorithm for constant treewidth turn-based stochastic games has remain elusive. In this work, our main result is a deterministic algorithm to solve turn-based stochastic games that, given a game with n states, treewidth at most t , and the bit-complexity of the probabilistic transition function $\log D$, has running time $O((tn^2 \log D)^{t \log n})$. In particular, our algorithm is quasi-polynomial time for games with constant or poly-logarithmic treewidth.

Jakub Svoboda, Raimundo Julian Saona Urmeneta,
Tobias Meggendorfer
Institute of Science and Technology Austria
jakub.svoboda@ist.ac.at,
raimundojulian.saonaermeneta@ist.ac.at,
tobias.meggendorfer@ist.ac.at

Krishnendu Chatterjee
IST Austria
krishnendu.chatterjee@ist.ac.at

CP38

Almost-Linear Planted Cliques Elude the Metropolis Process

A seminal work of Jerrum (1992) showed that large cliques elude the Metropolis process. More specifically, Jerrum showed that the Metropolis algorithm cannot find a clique of size $k = \Theta(n^\alpha)$ for $\alpha \in (0, 1/2)$, which is planted in the Erdős-Rényi random graph $G(n, 1/2)$, in polynomial time. Information theoretically it is possible to find such planted cliques as soon as $k \geq (2 + \epsilon) \log n$. Since the work of Jerrum, the computational problem of finding a planted clique in $G(n, 1/2)$ was studied extensively and many polynomial time algorithms were shown to find the planted clique if it is of size $k = \Omega(\sqrt{n})$, while no polynomial-time algorithm is known to work when $k = o(\sqrt{n})$. The computational problem of finding a planted clique of size $k = o(\sqrt{n})$ is now widely considered a foundational problem in the study of computational-statistical gaps. Notably, the first evidence of the problem’s algorithmic hardness is commonly attributed to the result of Jerrum from 1992. In this paper, we revisit the original Metropolis algorithm suggested by Jerrum. Perhaps contrary to a common belief in our community, we prove that the Metropolis algorithm actually *fails* to recover a planted clique of size $k = \Theta(n^\alpha)$ for *any* constant $\alpha \in (0, 1)$. Our results confirm recent predictions by Gamarnik and Zadik (2019) and Angelini, Fachin, de Feo (2021).

Ilias Zadik
MIT
izadik@mit.edu

Zongchen Chen

Massachusetts Institute of Technology
zongchen@mit.edu

Elchanan Mossel
MIT
elmos@mit.edu

CP38

A Near-Linear Time Sampler for the Ising Model with External Field

We give a near-linear time sampler for the Gibbs distribution of the ferromagnetic Ising models with edge activities $\beta > 1$ and external fields $\lambda < 1$ (or symmetrically, $\lambda > 1$) on general graphs with bounded or unbounded maximum degree. Our algorithm is based on the field dynamics given in [CFYZ21]. We prove the correctness and efficiency of our algorithm by establishing spectral independence of distribution of the random cluster model and the rapid mixing of Glauber dynamics on the random cluster model in a low-temperature regime, which may be of independent interest.

Xiaoyu Chen, Xinyuan Zhang
Nanjing University
chenxiaoyu233@smail.nju.edu.cn,
zhangxy@smail.nju.edu.cn

CP39

Query Complexity of the Metric Steiner Tree Problem

We study the query complexity of the metric Steiner Tree problem, where we are given an $n \times n$ metric on a set V of vertices along with a set $T \subseteq V$ of k terminals, and the goal is to find a tree of minimum cost that contains all terminals in T . The query complexity for the related MST problem is well-understood. For any $\varepsilon > 0$, one can estimate the MST cost to within a $(1 + \varepsilon)$ -factor using only $\tilde{O}(n)$ queries. This implies that a $(2 + \varepsilon)$ -approximate estimate of Steiner Tree cost can be obtained with $\tilde{O}(k)$ queries by simply applying the MST cost estimation algorithm on the metric induced by the terminals. Our first result shows that any (randomized) algorithm that estimates the Steiner Tree cost to within a $(5/3 - \varepsilon)$ -factor requires $\Omega(n^2)$ queries, even if k is a constant. This lower bound is in sharp contrast to an upper bound of $O(nk)$ queries for computing a $(5/3)$ -approximate Steiner Tree, which follows from previous work by Du and Zelikovsky. Our second main result, and the main technical contribution of this work, is a sublinear query algorithm for estimating the Steiner Tree cost to within a strictly better-than-2 factor, with query complexity $\tilde{O}(n^{12/7} + n^{6/7} \cdot k)$. We complement this result by showing an $\tilde{\Omega}(n + k^{6/5})$ query lower bound for any algorithm that estimates Steiner Tree cost to a strictly better-than-2 factor.

Yu Chen, Sanjeev Khanna
University of Pennsylvania
yu.chen@epfl.ch, sanjeev@cis.upenn.edu

Zihan Tan
Rutgers University

zihantan1993@gmail.com

CP39

Cubic Goldreich Levin

We give a cubic Goldreich-Levin algorithm which makes polynomially-many queries to a function $f: \mathbb{F}_p^n \rightarrow \mathbb{C}$ and produces a decomposition of f as a sum of cubic phases and a small error term. This is a natural higher-order generalization of the (linear) Goldreich-Levin algorithm which has found many applications in computer science. The only known result in this direction prior to this work is the quadratic Goldreich-Levin theorem, proved by Tulsiani and Wolf in 2011. The main step of their result involves an algorithmic version of the U^3 inverse theorem. More complications appear in the inverse theory of the U^4 and higher norms. Our cubic Goldreich-Levin algorithm is based on algorithmizing recent work by Gowers and Milicevic who proved new quantitative bounds for the U^4 inverse theorem. Our cubic Goldreich-Levin algorithm is constructed from two main tools: an algorithmic U^4 inverse theorem and an arithmetic decomposition result in the style of the Frieze-Kannan graph regularity lemma. As one application of our main theorem we solve the problem of self-correction for cubic Reed-Muller codes beyond the list decoding radius.

Anqi Li
MIT
angili@mit.edu

Jonathan Tidor
Stanford University
jtidor@stanford.edu

Dain Kim
MIT
dain0327@mit.edu

CP39

Super-Resolution and Robust Sparse Continuous Fourier Transform in Any Constant Dimension: Nearly Linear Time and Sample Complexity

The ability to resolve detail in the object that is being imaged, named by resolution, is the core parameter of an imaging system. Super-resolution is a class of techniques that can enhance the resolution of an imaging system and even transcend the diffraction limit of systems. Despite huge success in the application, super-resolution is not well understood on the theoretical side, especially for any dimension $d \geq 2$. In particular, in order to recover a k -sparse signal, all previous results suffer from either/both $\text{poly}(k)$ samples or running time. We design robust algorithms for any (constant) dimension under a strong noise model based on developing some new techniques in Sparse Fourier transform (Sparse FT), such as inverting a robust linear system, “eggshell” sampling schemes, and partition and voting methods in high dimension. These algorithms are the first to achieve running time and sample complexity (nearly) linear in the number of source points and logarithmic in bandwidth for any constant dimension, and we believe the techniques developed in the work can find their further applications on the Super-resolution and Sparse FT problem.

Yaonan Jin
Columbia University
yj2552@columbia.edu

Daogao Liu

The University of Washington
dgliu@uw.edu

Zhao Song

Adobe Research
zsong@adobe.com

CP39

Traversing the Fft Computation Tree for Dimension-Independent Sparse Fourier Transforms

We are interested in the well studied Sparse Fourier transform problem, where one aims to quickly recover an approximately Fourier k -sparse domain vector $\hat{x} \in \mathbb{C}^{n^d}$ from observing its time domain representation x . In the exact k -sparse case the best known dimension-independent algorithm runs in near cubic time in k . Beyond that, all known approaches either suffer from an exponential dependence of their runtime on the dimension d or can only tolerate a trivial amount of noise. Our work aims to address the above issues. First, we provide a translation/reduction of the exactly k -sparse Sparse FT problem to a concrete tree exploration task which asks to recover k leaves in a full binary tree under certain exploration rules. Subsequently we provide (a) an almost quadratic in k time algorithm for the latter task, and (b) evidence that obtaining a strongly subquadratic time for Sparse FT via this approach is likely to be impossible. Thus, our results combined can be viewed as an almost complete understanding of this approach, which is the only known approach that yields sublinear time dimension-independent Sparse FT algorithms. Subsequently, we provide a robustification of our algorithm, yielding a robust cubic time algorithm under bounded ℓ_2 noise. This requires proving new structural properties of the recently introduced adaptive aliasing filters combined with a variety of new techniques and ideas.

Vasileios Nakos

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

Mikhail Makarov

EPFL
mikhail.makarov@epfl.ch

Karl Bringmann

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

Michael Kapralov, Amir Zandieh

EPFL
michael.kapralov@epfl.ch, amir.zed512@gmail.com

Amir Yagudin

MIPT
yagudin.amir@gmail.com

CP39

Online Lewis Weight Sampling

The seminal work of Cohen and Peng (STOC 2015) introduced Lewis weight sampling, which yields fast row

sampling algorithms for approximating d -dimensional subspaces of ℓ_p up to $(1 + \epsilon)$ relative error. Several works have extended this primitive to other settings, including the online coresot, sliding window, and the adversarial streaming models. However, these results are only for $p \in \{1, 2\}$, and results for $p = 1$ require a suboptimal $\tilde{O}(d^2/\epsilon^2)$ samples. In this work, we design the first nearly optimal ℓ_p subspace embeddings for all $p \in (0, \infty)$ in the online coresot, sliding window, and the adversarial streaming models. In all three models, our algorithms store $\tilde{O}(d/\epsilon^2)$ rows for $p \in (0, 2)$ and $\tilde{O}(d^{p/2}/\epsilon^2)$ rows for $p \in (2, \infty)$. This answers the main open question of (Braverman et al., 2020) and is nearly optimal for all p . Towards our result, we give the first analysis of "one-shot" Lewis weight sampling of sampling rows proportionally to their Lewis weights, with sample complexity $\tilde{O}(d^{p/2}/\epsilon^2)$ for $p > 2$. Previously, such a sampling scheme was only known to have a sample complexity of $\tilde{O}(d^{p/2}/\epsilon^5)$ (Cohen and Peng, 2015). Perhaps surprisingly, our analysis crucially uses a novel connection to online numerical linear algebra, even for offline Lewis weight sampling.

David Woodruff, Taisuke Yasuda

Carnegie Mellon University
dwoodruf@cs.cmu.edu, taisukey@cs.cmu.edu

CP39

Sublinear-Time Algorithms for Max Cut, Max E2Lin(q), and Unique Label Cover on Expanders

We show sublinear-time algorithms for Max Cut and Max E2Lin(q) on expanders in the adjacency list model that distinguishes instances with the optimal value more than $1 - \epsilon$ from those with the optimal value less than $1 - \rho$ for $\rho \gg \epsilon$. The time complexities for Max Cut and Max 2Lin(q) are $\tilde{O}(\frac{1}{\phi^2 \rho} \cdot m^{1/2+O(\epsilon/(\phi^2 \rho))})$ and $\tilde{O}(\text{poly}(\frac{q}{\phi \rho}) \cdot (mq)^{1/2+O(q^6 \epsilon / \phi^2 \rho^2)})$, respectively, where m is the number of edges in the underlying graph and ϕ is its conductance. Then, we show a sublinear-time algorithm for Unique Label Cover on expanders with $\phi \gg \epsilon$ in the bounded-degree model. The time complexity of our algorithm is $\tilde{O}_d(2^{q^{O(1)} \cdot \phi^{1/q} \cdot \epsilon^{-1/2}} \cdot n^{1/2+q^{O(q)} \cdot \epsilon^{4^{1.5-q}} \cdot \phi^{-2}})$, where n is the number of variables. We complement these algorithmic results by showing that testing 3-colorability requires $\Omega(n)$ queries even on expanders.

Yuichi Yoshida

National Institute of Informatics
yyoshida@nii.ac.jp

Pan Peng

University of Science and Technology of China
ppeng@ustc.edu.cn

CP40

Simplified Prophet Inequalities for Combinatorial Auctions

We consider prophet inequalities for XOS and MPH- k combinatorial auctions and give a simplified proof for the existence of static and anonymous item prices which recover the state-of-the-art competitive ratios. Our proofs make use of a linear programming formulation which has a non-negative objective value if there are prices which admit a given competitive ratio $\alpha \geq 1$. Changing our perspective to dual space by an application of strong LP duality, we use an interpretation of the dual variables as probabili-

ties to directly obtain our result. In contrast to previous work, our proofs do not require to argue about specific values of buyers for bundles, but only about the presence or absence of items. As a side remark, for any $k \geq 2$, this simplification also leads to a tiny improvement in the best competitive ratio for MPH- k combinatorial auctions from $4k - 2$ to $2k + 2\sqrt{k(k-1)} - 1$.

Alexander Braun, Thomas Kesselheim
University of Bonn
alexander.braun@uni-bonn.de, thomas.kesselheim@uni-bonn.de

CP40

A Tight Analysis of Hutchinsons Diagonal Estimator

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a matrix with diagonal $\text{diag}(\mathbf{A}) \in \mathbb{R}^n$. We show that the simple and practically popular Hutchinson's estimator, run for m trials, returns a diagonal estimate $\tilde{d} \in \mathbb{R}^n$ such that with probability $1 - \delta$,

$$\|\tilde{d} - \text{diag}(\mathbf{A})\|_2 \leq c \sqrt{\frac{\log(2/\delta)}{m}} \|\bar{\mathbf{A}}\|_F.$$

Above c is a fixed constant and $\bar{\mathbf{A}}$ equals \mathbf{A} with its diagonal set to zero. This result improves on recent work in [Baston and Nakatsukasa, 2022] by a $\log(n)$ factor, yielding a bound that is independent of the matrix dimension, n . We show a similar bound for variants of Hutchinson's estimator that use non-Rademacher random vectors.

Prathamesh Dharangutte
Rutgers University
ptd39@rutgers.edu

Christopher Musco
New York University
Tandon School of Engineering
cmusco@nyu.edu

CP40

An Improved Online Reduction from Pac Learning to Mistake-Bounded Learning

A basic result in learning theory is that mistake-bounded learnability implies PAC learnability. It was shown by Littlestone that, if a problem can be learned with M mistakes, it can be (ϵ, δ) -PAC-learned from $O(\frac{1}{\epsilon}(M + \log \frac{1}{\delta}))$ samples. However, this reduction needs to store either $O(\frac{1}{\epsilon} \log \frac{1}{\delta})$ samples or $O(M)$ hypotheses. A different reduction, given by Kearns et. al., only needs to store $O(1)$ samples and hypotheses but was only shown to work with $O(\frac{1}{\epsilon}(M \cdot \log \frac{M}{\delta}))$ samples. We give a refined analysis of this reduction, showing that it only uses $O(\frac{M}{\epsilon})$ samples with probability $1 - M^{-O(1)}$. This gives the optimal sample complexity with only $O(1)$ space overhead, for $\delta > M^{-O(1)}$.

Lucas Gretta
UC Berkeley
lucas_gretta@berkeley.edu

Eric Price
The University of Texas at Austin

ecprice@cs.utexas.edu

CP40

A Simple Optimal Algorithm for the 2-Arm Bandit Problem

In the bandit problem, an agent faces a slot machine with two arms; at each round they choose one to pull and receive a reward sampled randomly according to the distribution of that arm. Their goal is to maximise the total reward. In the analysis, one typically looks at the missed reward called the regret. It is not difficult to see that the expected regret is at least $\Omega(\sqrt{T})$ where T is the number of rounds played. Using Azuma's inequality one can easily create an algorithm achieving $O(\sqrt{T} \log T)$ regret; the additional log-factor was removed by Audibert and Bubeck '10 by using a cleverly adapted multiplicative-weight approach. In this paper we consider the non-stationary version in which the underlying reward distributions may change arbitrarily. The known lower bound for this problem is $\Omega(\sqrt{LT})$, where L denotes the total number changes until time T . Using a multiplicative-weight approach, Auer et al. '02 presented an algorithm achieving regret $O(\sqrt{LT} \log T)$ when the agent knows the value of L (but not when the changes happen). An algorithm with similar regret bound but based on Azuma's inequality was later presented by Garivier and Moulines '11. We present a new algorithm using random walks with asymmetric stopping boundaries. The analysis is simple and shows that our algorithm achieves regret $O(\sqrt{LT})$, thus matching the lower bound. This is the first optimal algorithm for the bandit problem with changes.

Maxime Larcher
ETH Zurich
larcherm@inf.ethz.ch

Robert Meier
ETH Zürich
romeier@inf.ethz.ch

Angelika Steger
Institute of Theoretical Computer Science
ETH, Zürich
steger@inf.ethz.ch

CP40

Gaussian Mean Testing Made Simple

We study the following fundamental hypothesis testing problem, which we term Gaussian mean testing. Given i.i.d. samples from a distribution p on \mathbb{R}^d , the task is to distinguish, with high probability, between the following cases: (i) p is the standard Gaussian distribution, $\mathcal{N}(0, I_d)$, and (ii) p is a Gaussian $\mathcal{N}(\mu, \Sigma)$ for some unknown covariance Σ and mean $\mu \in \mathbb{R}^d$ satisfying $\|\mu\|_2 \geq \epsilon$. Recent work gave an algorithm for this testing problem with the optimal sample complexity of $\Theta(\sqrt{d}/\epsilon^2)$. Both the previous algorithm and its analysis are quite complicated. Here we give an extremely simple algorithm for Gaussian mean testing with a one-page analysis. Our algorithm is sample optimal and runs in sample linear time.

Ilias Diakonikolas
UW-Madison
ilias@cs.wisc.edu

Daniel Kane

UC San Diego
dakane@cs.ucsd.edu

Ankit Pensia
UW-Madison
ankitp@cs.wisc.edu

CP40

A Simple Polynomial-Time Approximation Algorithm for the Total Variation Distance Between Two Product Distributions

We give a simple polynomial-time approximation algorithm for the total variation distance between two product distributions.

Weiming Feng, Heng Guo
University of Edinburgh
wfeng@ed.ac.uk, hguo@inf.ed.ac.uk

Mark Jerrum
Queen Mary University of London
m.jerrum@qmul.ac.uk

Jiaheng Wang
University of Edinburgh
jiaheng.wang@ed.ac.uk

CP41

Differentially Private All-Pairs Shortest Path Distances: Improved Algorithms and Lower Bounds

We study the problem of releasing the weights of all-pairs shortest paths in a weighted undirected graph with differential privacy (DP). In this setting, the underlying graph is fixed and two graphs are neighbors if their edge weights differ by at most 1 in the ℓ_1 -distance. We give an algorithm with additive error $\tilde{O}(n^{2/3}/\epsilon)$ in the ϵ -DP case and an algorithm with additive error $\tilde{O}(\sqrt{n}/\epsilon)$ in the (ϵ, δ) -DP case, where n denotes the number of vertices. This positively answers a question of Sealfon [Sea16, Sea20], who asked whether a $o(n)$ -error algorithm exists. We also show that an additive error of $\Omega(n^{1/6})$ is necessary for any sufficiently small $\epsilon, \delta > 0$. Furthermore, we show that if the graph is promised to have reasonably bounded weights, one can improve the error further to roughly $n^{(\sqrt{17}-3)/2+o(1)}/\epsilon$ in the ϵ -DP case and roughly $n^{\sqrt{2}-1+o(1)}/\epsilon$ in the (ϵ, δ) -DP case, improving on previous work on bounded-weight graphs [Sea16]. Finally, we consider a relaxation where a multiplicative approximation is allowed. We show that, with a multiplicative approximation factor k , the additive error can be reduced to $\tilde{O}(n^{1/2+O(1/k)}/\epsilon)$ in the ϵ -DP case and $\tilde{O}(n^{1/3+O(1/k)}/\epsilon)$ in the (ϵ, δ) -DP case.

Justin Y. Chen
MIT
justc@mit.edu

Badih Ghazi
Google Research
badihghazi@gmail.com

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

Pasin Manurangsi
Google Research
pasin@google.com

Shyam Narayanan
Massachusetts Institute of Technology
shyamsn@mit.edu

Jelani Nelson
UC Berkeley
Google Research
minilek@alum.mit.edu

Yinzhan Xu
Massachusetts Institute of Technology
xyzhan@mit.edu

CP41

Private Convex Optimization in General Norms

We propose a new framework for differentially private optimization of convex functions which are Lipschitz in an arbitrary norm $\|\cdot\|$. Our algorithms are based on a regularized exponential mechanism which samples from the density $\propto \exp(-k(F + \mu r))$ where F is the empirical loss and r is a regularizer which is strongly convex with respect to $\|\cdot\|$, generalizing a recent work of [Gopi, Lee and Liu COLT 2022] to non-Euclidean settings. We show that this mechanism satisfies Gaussian differential privacy and solves both DP-ERM (empirical risk minimization) and DP-SCO (stochastic convex optimization), by using localization tools from convex geometry. Our framework is the first to apply to private convex optimization in general normed spaces, and directly recovers non-private SCO rates achieved by mirror descent, as the privacy parameter $\epsilon \rightarrow \infty$. As applications, for Lipschitz optimization in ℓ_p norms for all $p \in (1, 2)$, we obtain the first optimal privacy-utility tradeoffs; for $p = 1$, we improve tradeoffs obtained by the recent works [Asi, Feldman, Koren and Tawal ICML 2021] and [Bassily, Guzman and Nandi COLT 2021] by at least a logarithmic factor. Our ℓ_p norm and Schatten- p norm optimization frameworks are complemented with polynomial-time samplers whose query complexity we explicitly bound.

Sivakanth Gopi
Microsoft Research
sigopi@microsoft.com

Yin Tat Lee, Daogao Liu
University of Washington
yintat@uw.edu, liudaogao@gmail.com

Ruoqi Shen
University of Washington, U.S.
shenr3@cs.washington.edu

Kevin Tian
Stanford University
tiankevin@microsoft.com

CP41

Private Query Release via the Johnson Lindenstrauss Transform

We introduce a new method for releasing answers to statistical queries with differential privacy, based on the Johnson-Lindenstrauss lemma. The key idea is to ran-

domly project the query answers to a lower dimensional space so that the distance between any two vectors of feasible query answers is preserved up to an additive error. Then we answer the projected queries using a simple noise-adding mechanism, and lift the answers up to the original dimension. Using this method, we give, for the first time, purely differentially private mechanisms with optimal worst case sample complexity under average error for answering a workload of k queries over a universe of size N . As other applications, we give the first purely private efficient mechanisms with optimal sample complexity for computing the covariance of a bounded high-dimensional distribution, and for answering 2-way marginal queries. We also show that, up to the dependence on the error, a variant of our mechanism is nearly optimal for every given query workload.

Aleksandar Nikolov
University of Toronto
anikolov@cs.toronto.edu

CP41

Stronger Privacy Amplification by Shuffling for Rnyi and Approximate Differential Privacy

The shuffle model of differential privacy has gained significant interest as an intermediate trust model between the standard local and central models [Erlingsson, Feldman, Mironov, Raghunathan, Talwar, and Thakurta, 2019; Cheu, Smith, Ullman, Zeber, and Zhi, 2019]. A key result in this model is that randomly shuffling locally randomized data amplifies differential privacy guarantees. Such amplification implies substantially stronger privacy guarantees for systems in which data is contributed anonymously [Bittau et al., 2017]. In this work, we improve the state of the art privacy amplification by shuffling results both theoretically and numerically. Our first contribution is the first asymptotically optimal analysis of the Rnyi differential privacy parameters for the shuffled outputs of LDP randomizers. Our second contribution is a new analysis of privacy amplification by shuffling. This analysis improves on the techniques of Feldman, McMillan and Talwar (2020) and leads to tighter numerical bounds in all parameter settings.

Vitaly Feldman, Audra Mcmillan, Kunal Talwar
Apple
vitalyf@apple.com, audra.mcmillan@apple.com, ktalwar@apple.com

CP41

Almost Tight Error Bounds on Differentially Private Continual Counting

The first large-scale deployment of private federated learning uses differentially private counting in the continual release model as a subroutine (Google AI blog titled “Federated Learning with Formal Differential Privacy Guarantees”). In this case, a concrete bound on the error is very relevant to reduce the privacy parameter. The standard mechanism for continual counting is the binary mechanism. We present a novel mechanism and show that its mean squared error is both asymptotically optimal and a factor 10 smaller than the error of the binary mechanism. We also show that the constants in our analysis are almost tight by giving non-asymptotic lower and upper bounds that differ only in the constants of lower-order terms. Our algorithm is a matrix mechanism for the counting matrix and takes constant time per release. We also use our explicit factor-

ization of the counting matrix to give an upper bound on the excess risk of the private learning algorithm of Denisov et al. (NeurIPS 2022). Our lower bound for any continual counting mechanism is the first tight lower bound on continual counting under approximate differential privacy. It is achieved using a new lower bound on a factorization norm in terms of the singular values of the matrix. We believe this technique will be useful in proving lower bounds for a larger class of linear queries. To illustrate the power of this technique, we show the first lower bound on the mean squared error for answering parity queries.

Jalaj Upadhyay
Rutgers University
jalaj.upadhyay@rutgers.edu

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

Sarvagya Upadhyay
Fujitsu Research of America
supadhyay@us.fujitsu.com

CP42

Optimal Square Detection Over General Alphabets

Squares (fragments of the form xx , for some string x) are arguably the most natural type of repetition in strings. The basic algorithmic question concerning squares is to check if a given string of length n is square-free, that is, does not contain a fragment of such form. Main and Lorentz [J. Algorithms 1984] designed an $O(n \log n)$ time algorithm for this problem, and proved a matching lower bound assuming the so-called general alphabet, meaning that the algorithm is only allowed to check if two characters are equal. As an open question, they asked if there is a faster algorithm if one restricts the size of the alphabet. Crochemore [Theor. Comput. Sci. 1986] designed a linear-time algorithm for constant-size alphabets, and combined with the more recent results his approach in fact implies such an algorithm for linearly-sortable alphabets. Very recently, Ellert and Fischer [ICALP 2021] significantly relaxed this assumption by designing a linear-time algorithm for general ordered alphabets, that is, assuming a linear order on the characters. However, the open question of Main and Lorentz from 1984 remained unresolved for general (unordered) alphabets. In this paper, we show that testing square-freeness of a length- n string over general alphabet of size σ can be done with $O(n \log \sigma)$ comparisons, and cannot be done with $o(n \log \sigma)$ comparisons. We complement this result with an $O(n \log \sigma)$ time algorithm in the Word RAM model.

Jonas Ellert
TU Dortmund University
jonas.ellert@tu-dortmund.de

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

Garance Gourdel
DI/ENS, PSL Research University
IRISA Inria Rennes

garance.gourdell@irisa.fr

msaks30@gmail.com

CP42

Breaking the $O(n)$ -Barrier in the Construction of Compressed Suffix Arrays and Suffix Trees

The suffix array and the suffix tree are the two most fundamental data structures for string processing. For a length- n text, however, they use $\Theta(n \log n)$ bits of space, which is often too costly. To address this, Grossi and Vitter [STOC 2000] and, independently, Ferragina and Manzini [FOCS 2000] introduced space-efficient versions of the suffix array, known as the compressed suffix array (CSA) and the FM-index. Sadakane [SODA 2002] then showed how to augment them to obtain the compressed suffix tree (CST). For a length- n text over an alphabet of size σ , these structures use only $\mathcal{O}(n \log \sigma)$ bits. The biggest remaining open question is how efficiently they can be constructed. After two decades, the fastest algorithms still run in $\mathcal{O}(n)$ time [Hon et al., FOCS 2003]. In this paper, we make the first in 20 years improvement in n for this problem by proposing a new compressed suffix array and a new compressed suffix tree which admit $o(n)$ -time construction algorithms while matching the space bounds and the query times of the original CSA/CST and the FM-index. Our structures take $\mathcal{O}(n \log \sigma)$ bits, support SA queries and full suffix tree functionality in $\mathcal{O}(\log^\epsilon n)$ time per operation, and can be constructed in $\mathcal{O}(n \min(1, \log \sigma / \sqrt{\log n}))$ time using $\mathcal{O}(n \log \sigma)$ bits of working space. Based on the new techniques, we also develop a new index for pattern matching.

Dominik Kempa
Stony Brook University
kempa@cs.stonybrook.edu

Tomasz Kociumaka
Max Planck Institute for Informatics
tomasz.kociumaka@mpi-inf.mpg.de

CP42

Simple, Deterministic, Fast (but Weak) Approximations to Edit Distance and Dyck Edit Distance

We consider the problem of obtaining approximation algorithms for standard edit distance and Dyck edit distance that are simple, deterministic and fast, but whose approximation factor may be high. For the standard edit distance of two strings, we introduce a class of simple and fast algorithms called basic single-pass algorithms. Saha (2014) gave a randomized algorithm in this class that achieves an $\mathcal{O}(d)$ approximation on inputs x, y whose edit distance is $\mathcal{O}(d)$. In this paper, we (1) present a deterministic algorithm in this class that achieves similar performance and (2) prove that no algorithm (even randomized) in this class can give a better approximation factor. For the Dyck edit distance problem, Saha gave a randomized reduction from Dyck edit distance to standard two string edit distance at a cost of a $\mathcal{O}(\log d)$ factor where d is the Dyck edit distance. We give a deterministic reduction whose description and proof are very simple.

Michal Koucky
Czech Academy of Sciences
koucky@iuuk.mff.cuni.cz

Michael Saks
Rutgers University

CP42

Time-Space Tradeoffs for Element Distinctness and Set Intersection via Pseudorandomness

In the Element Distinctness problem, one is given an array a_1, \dots, a_n of integers from $\text{poly}(n)$ and is tasked to decide if $\{a_i\}$ are mutually distinct. Beame, Clifford and Machmouchi (FOCS 2013) gave a low-space algorithm for this problem running in space S_n and time $T(n)$ where $T(n) \leq \tilde{O}(n^{3/2}/S(n)^{1/2})$, assuming a random oracle (i.e., random access to polynomially many random bits). A recent breakthrough by Chen, Jin, Williams and Wu (SODA 2022) showed how to remove the random oracle assumption in the regime $S(n) = \text{polylog}(n)$ and $T(n) = \tilde{O}(n^{3/2})$. They designed the first truly $\text{polylog}(n)$ -space, $\tilde{O}(n^{3/2})$ -time algorithm by constructing a small family of hash functions $H \subseteq \{h : h : [m] \rightarrow [n]\}$ with a certain pseudorandom property. In this paper, we give a significantly simplified analysis of the pseudorandom hash family by Chen et al. Our analysis clearly identifies the key pseudorandom property required to fool the BCM algorithm, allowing us to explore the full potential of this construction. As our main result, we show a time-space tradeoff for Element Distinctness without random oracle. Namely, for every $S(n), T(n)$ such that $T(n) = \tilde{O}(n^{3/2}/S(n)^{1/2})$. Our algorithm can solve the problem in space $S(n)$ and time $T(n)$. Our algorithm also works for a related problem Set Intersection, for which this tradeoff is tight due to a matching lower bound by Dinur (Eurocrypt 2020).

Xin Lyu
UC Berkeley
xinlyu@berkeley.edu

Weihao Zhu
Shanghai Jiao Tong University
zhuweihao@sjtu.edu.cn

CP42

Quantum Speed-Ups for String Synchronizing Sets, Longest Common Substring, and k -Mismatch Matching

Longest Common Substring (LCS) is an important text processing problem, which has recently been investigated in the quantum query model. We show that LCS with threshold d has a quantum algorithm in $n^{2/3+o(1)}/d^{1/6}$ query complexity and time complexity, and requires at least $\Omega(n^{2/3}/d^{1/6})$ quantum query complexity. Our result improves upon previous upper bounds $\tilde{O}(\min\{n/d^{1/2}, n^{2/3}\})$ (Le Gall and Seddighin ITCS 2022, Akmal and Jin SODA 2022), and answers an open question of Akmal and Jin. Our main technical contribution is a quantum speed-up of the powerful String Synchronizing Set technique introduced by Kempa and Kociumaka (STOC 2019). It consistently samples $n/\tau^{1-o(1)}$ synchronizing positions in the string depending on their length- $\Theta(\tau)$ contexts, and each synchronizing position can be reported by a quantum algorithm in $\tilde{O}(\tau^{1/2+o(1)})$ time. Our quantum string synchronizing set also yields a near-optimal LCE data structure in the quantum setting. As another application of our quantum string synchronizing set, we study the k -mismatch Matching problem. Using a structural result of Charalampopoulos, Kociumaka, and Wellnitz (FOCS 2020), we obtain that k -mismatch matching has a quan-

tum algorithm with $k^{3/4}n^{1/2+o(1)}$ query complexity and $\tilde{O}(kn^{1/2})$ time complexity. We also observe a non-matching quantum query lower bound of $\Omega(\sqrt{kn})$.

Jakob Nogler
ETH Zurich
jnogler@ethz.ch

Ce Jin
EECS
MIT
cejin@mit.edu

CP43

Distance and Time Sensitive Filters for Similarity Search in Trajectory Datasets

Similarity searching is a well-studied problem in trajectory processing. Several works have focused on developing near neighbor data structures for trajectories under various metrics (e.g., Hausdorff distance or Fréchet distance). However, these measures ignore the temporal aspect of time-stamped trajectories. Furthermore, these data structures use space that is super-linear in the size of the trajectory database. This work focuses on optimizing near neighbor data structures by addressing the two issues above. We present a data structure that preprocesses a database \mathcal{T} of trajectories so that, given a query trajectory Q and a similarity factor $0 \leq \rho \leq 1$, it answers Yes or No to whether there is a trajectory in \mathcal{T} that is ρ -similar to Q . Similarity between two trajectories here is defined as the Hamming distance between their binary sketches obtained from certain 3D lifts of the trajectories, where the third dimension incorporates the temporal aspect. With respect to this similarity measure, the data structure has only false positives and no false negatives. We show via experiments that the data structure has fast construction and query time, and gives up to 95% space savings on the original datasets. We show a low mismatch rates between our similarity measure and the Hausdorff and SketchMin distances, showing that the new measure is meaningful. Such a data structure can be used in a two-tiered data structure.

Paul Cesaretti
Graduate Center, CUNY
pcesaretti@gradcenter.cuny.edu

Madhav Narayan Bhat
Columbia University
tmnbhat@gmail.com

Mayank Goswami
Queens College, CUNY and Graduate Center, CUNY
mayank.goswami@qc.cuny.edu

Prashant Pandey
University of Utah
prashant.prashn@gmail.com

CP43

Analyzing and Implementing GPU Hash Tables

We revisit the problem of building static hash tables on the GPU and present an efficient implementation of bucketed hash tables. By decoupling the probing scheme from the hash table in-memory representation, we offer an implementation where the number of probes and the bucket size are the only factors limiting performance. Our analy-

sis sweeps through the hash table parameter space for two probing schemes: cuckoo and iceberg hashing. We show that a bucketed cuckoo hash table (BCHT) that uses three hash functions outperforms alternative methods that use iceberg hashing and a cuckoo hash table that uses a bucket size of one. At load factors as high as 0.99, BCHT enjoys an average probe count of 1.43 during insertion. Using three hash functions only, positive and negative queries require at most 1.39 and 2.8 average probes per key, respectively.

Muhammad A. Awad
UC Davis
mawad@ucdavis.edu

Saman Ashkiani
NVIDIA
sa.ashkiani@gmail.com

Serban Porumbescu
UC Davis
sdporumbescu@ucdavis.edu

Martin Farach-Colton
Rutgers University
farach@cs.rutgers.edu

John Owens
University of California, Davis
jowens@ece.ucdavis.edu

CP43

Efficient Algorithms for Parallel Bi-Core Decomposition

We present new shared-memory parallel algorithms for the bi-core decomposition problem, which discovers dense subgraphs in bipartite graphs and is the bipartite analogue of the classic k -core decomposition problem. We develop a theoretically-efficient parallel bi-core decomposition algorithm that discovers a hierarchy by peeling vertices from the graph in parallel. Our algorithm improves the span (parallel running time) over the state-of-the-art parallel bi-core decomposition algorithm, while matching the state-of-the-art sequential algorithm in work. We additionally prove the bi-core decomposition problem to be P-complete, meaning that a polylogarithmic span solution is unlikely under standard assumptions. We also devise a theoretically-efficient parallel bi-core index structure to allow for fast parallel queries of vertices in given cores. Finally, we propose a novel practical optimization that prunes unnecessary computations, and we provide optimized parallel implementations of our bi-core decomposition algorithms that are scalable and fast. Using 30 cores with two-way hyper-threading, our implementation achieves up to a 4.9x speedup over the state-of-the-art parallel algorithm. Our parallel index structure can be constructed up to 27.7x faster than the state-of-the-art sequential counterpart. Due to the improved storage format of our index structure, our parallel queries are up to 116.3x faster than the state-of-the-art sequential queries.

Yihao Huang, Claire Wang
Phillips Academy
yhuang23@andover.edu, cwang23@andover.edu

Jessica Shi
MIT CSAIL
jeshi@mit.edu

Julian Shun
MIT
jshun@mit.edu

CP43

Extending Optimal Oblivious Reconfigurable Networks to All N

Reconfigurable networks have traditionally suffered from long reconfiguration times due to hardware limitations. With the emergence of new technologies that can reconfigure on the order of nanoseconds, the Oblivious Reconfigurable Network (ORN) design paradigm has been proposed to take advantage of this new capability. Two of the most important performance metrics for network operators considering ORNs are latency and throughput, which are inherently opposed and thus lead to a tradeoff. Previously, we constructed two families of ORN designs, called EBS and VBS, and proved that together they achieve optimal maximum latency (up to a constant factor) for any given throughput value r . Unfortunately, both families are defined only for very restrictive network sizes, N . This poses a challenge for implementing networks based on these designs in practice, as it is unlikely that a given system will have a suitable size. In this work, we extend both EBS and VBS to any network size while minimizing the impact on maximum latency and throughput. We prove that the extended versions achieve optimal maximum latency for all sufficiently large N , except when the throughput r exactly equals the reciprocal of an even integer.

Tegan Wilson, Daniel Amir
Cornell University
teganwilson@cs.cornell.edu, amir@cs.cornell.edu

Vishal Shrivastav
Purdue University
vshriva@purdue.edu

Hakim Weatherspoon, Robert Kleinberg
Cornell University
hweather@cs.cornell.edu, rdk@cs.cornell.edu