IP1

Differential Privacy: A Gateway Concept

Differential privacy is a definition of privacy tailored to statistical analysis of very large large datasets. Invented just over one decade ago, the notion has become widely (if not deeply) deployed, and yet much remains to be done. The theoretical investigation of privacy/accuracy tradeoffs that shaped the field by delineating the boundary between possible and impossible motivate the continued search for new algorithmic techniques, as well as still meaningful relaxations of the basic definition. Differential privacy has also led to new approaches in other fields, most notably in algorithmic fairness and adaptive data analysis, in which the questions being asked of the data depend on the data themselves. We will highlight some recent algorithmic and definitional work, and focus on differential privacy as a gateway concept to these new areas of study.

Cynthia Dwork
Harvard University, USA
dwork@seas.harvard.edu

IP2

The Power of Theory in the Practice of Hashing with Focus on Similarity Estimation

Hash functions have become ubiquitous tools in modern data analysis, e.g., the construction of small randomized sketches of large data streams. We like to think of abstract hash functions, assigning independent uniformly random hash values to keys, but in practice, we have to choose a hash function that only has an element of randomness, e.g., 2-independence. While this works for sufficiently random input, the real world has structured data where such simple hash functions fail, calling for the need of more powerful hash functions. In this talk, we focus hashing for set similarity, which is an integral component in the analysis of Big Data. The basic idea is to use the same hash function to do coordinated sampling from different sets. Depending on the context, we want subsets sampled without replacement, or fixed-length vectors of samples that may be with replacement. The latter is used as input to support vector machines (SVMs) and locality sensitive hashing (LSH). The most efficient constructions require very powerful hash functions that are also needed for efficient size estimation. We discuss the interplay between the hash functions and the algorithms using them. Finally, we present experiments on both real and synthetic data where standard 2-independent hashing yield systematically poor similarity estimates, while the right theoretical choice is sharply concentrated, and faster than standard cryptographic hash functions with no proven guarantees.

Mikkel Thorup
University of Copenhagen
mikkel2thorup@gmail.com

IP3

Approximation Algorithms for Uncertain Environments

The past decade has seen considerable work on algorithms in models with uncertainty: where either the inputs to the algorithm or the algorithm's actions have some degree of uncertainty. Designing algorithms for these settings give rise to new problems and techniques. I will survey some algorithmic models that try to capture uncertainty in optimization problems, talk about some example problems, and indicate some of the techniques and ideas used to tackle the uncertainty in these problems and get provable guarantees on the performance of these algorithms.

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

IP4

Title To Be Determined - Vassilevska Williams

Not available.

Virginia Vassilevska Williams
MIT
virgi@mit.edu

CP0

A Generic Framework for Engineering Graph Canonization Algorithms

The state-of-the-art tools for practical graph canonization are all based on the individualization-refinement paradigm, and their difference is primarily in the choice of heuristics they include and in the actual tool implementation. It is thus not possible to make a direct comparison of how individual algorithmic ideas affect the performance on different graph classes. We present an algorithmic software framework that facilitates implementation of heuristics as independent extensions to a common core algorithm. It therefore becomes easy to perform a detailed comparison of the performance and behaviour of different algorithmic ideas. Implementations are provided of a range of algorithms for tree traversal, target cell selection, and node invariant, including choices from the literature and new variations. The framework readily supports extraction and visualization of detailed data from separate algorithm executions for subsequent analysis and development of new heuristics. Using collections of different graph classes we investigate the effect of varying the selections of heuristics, often revealing exactly which individual algorithmic choice is responsible for particularly good or bad performance. On several benchmark collections, including a newly proposed class of difficult instances, we additionally find that our implementation performs better than the current state-of-the-art tools.

Jakob L. Andersen
Research group Bioinformatics and Computational Biology
Faculty of Computer Science, University of Vienna
jlandersen@imada.sdu.dk

Daniel Merkle
Department of Mathematics and Computer Science
University of Southern Denmark
daniel@imada.sdu.dk

CP0

Computing Top-k Closeness Centrality in Fully-Dynamic Graphs

Closeness is a widely-studied centrality measure. Since it requires all pairwise distances, computing closeness for all nodes is infeasible for large real-world networks. However, for many applications, it is only necessary to find the k most central nodes and not all closeness values. Prior work has shown that computing the top-k nodes with highest

closeness can be done much faster than computing closeness for all nodes in real-world networks. However, for networks that evolve over time, no dynamic top-k closeness algorithm exists that improves on static recomputation. In this paper, we present several techniques that allow us to efficiently compute the k nodes with highest (harmonic) closeness after an edge insertion or an edge deletion. Our algorithms use information obtained during earlier computations to omit unnecessary work. However, they do not require asymptotically more memory than the static algorithms (i. e., linear in the number of nodes). We propose separate algorithms for complex networks (which exhibit the small-world property) and networks with large diameter such as street networks, and we compare them against static recomputation on a variety of real-world networks. On many instances, our dynamic algorithms are two orders of magnitude faster than recomputation; on some large graphs, we even reach average speedups between 10^3 and 10^4.

Patrick Bisenius
n/a
n/a

Elisabetta Bergamini, Elisabetta Bergamini
Institute of Theoretical Informatics
Karlsruhe Institute of Technology (KIT)
elisabetta.bergamini@kit.edu,
elisabetta.bergamini@kit.edu

Eugenio Angriman
Karlsruhe Institute of Technology
eangrima@uni-koeln.de

Henning Meyerhenke
Karlsruhe Institute of Technology (KIT)
meyerhenke@kit.edu

CP0

Polynomial Tuning of Multiparametric Combinatorial Samplers

Boltzmann samplers and the recursive method are prominent algorithmic frameworks for the approximate-size and exact-size random generation of large combinatorial structures, such as maps, tilings, RNA sequences or various tree-like structures. In their multiparametric variants, these samplers allow to control the profile of expected values corresponding to multiple combinatorial parameters. One can control, for instance, the number of leaves, profile of node degrees in trees or the number of certain subpatterns in strings. However, such a flexible control requires an additional non-trivial tuning procedure. In this paper, we propose an efficient polynomial-time, with respect to the number of tuned parameters, tuning algorithm based on convex optimisation techniques. Finally, we illustrate the efficiency of our approach using several applications of rational, algebraic and Pólya structures including polyomino tilings with prescribed tile frequencies, planar trees with a given specific node degree distribution, and weighted partitions.

Maciej Bendkowski
Theoretical Computer Science Department
Jagiellonian University, Kraków, Poland
maciej.bendkowski@tcs.uj.edu.pl

Olivier Bodini
LIPN, Université Paris 13
olivier.bodini@lipn.univ-paris13.fr

Sergey Dovgal
Institut Galilée, Université Paris 13
L'IRIF Université Paris 7; Moscow Inst. of Physics and Tech
dovgal@lipn.univ-paris13.fr

CP0

Scaling Up Group Closeness Maximization

Closeness is a widely-used centrality measure in social network analysis. While the identification of the k nodes with highest closeness received significant attention, many applications are actually interested in finding a group of nodes that is central as a whole. For this problem, only recently a greedy algorithm with approximation ratio $(1 - 1/e)$ has been proposed [Chen et al., ADC 2016]. Since this algorithm's running time is still expensive for large networks, a heuristic without approximation guarantee has also been proposed in the same paper. In the present paper we develop techniques to speed up the greedy algorithm without losing its guarantee. Compared to a straightforward implementation, our approach is orders of magnitude faster and, compared to the heuristic proposed by Chen et al., we always find a solution with better quality in a comparable running time in our experiments. Our method Greedy++ allows us to approximate the group with maximum closeness on networks with up to hundreds of millions of edges in at most a few hours. To have the same theoretical guarantee, the greedy approach by [Chen et al., ADC 2016] would take several days already on networks with hundreds of thousands of edges. In a comparison with the optimum, our experiments show that the solution found by Greedy++ is much better than the theoretical guarantee. Over all tested networks, the empirical approximation ratio is never lower than 0.97.

Elisabetta Bergamini, Elisabetta Bergamini
Institute of Theoretical Informatics
Karlsruhe Institute of Technology (KIT)
elisabetta.bergamini@kit.edu,
elisabetta.bergamini@kit.edu

Tanya Gonser
Karlsruhe Institute of Technology
tanya.gonser@kit.edu

Henning Meyerhenke
University of Cologne
h.meyerhenke@uni-koeln.de

CP0

Quadratic Time Algorithms Appear to Be Optimal for Sorting Evolving Data

We empirically study sorting in the *evolving data* model. In this model, a sorting algorithm maintains an approximation to the sorted order of a list of data items while simultaneously, with each comparison made by the algorithm, an adversary randomly swaps the order of adjacent items in the true sorted order. Previous work studies only two versions of quicksort, and has a gap between the lower bound of $\Omega(n)$ and the best upper bound of $O(n \log \log n)$. The experiments we perform in this paper provide empirical evidence that some quadratic-time algorithms such as insertion sort and bubble sort are asymptotically optimal for any constant rate of random swaps. In fact, these algo-

rithms perform as well as or better than algorithms such as quicksort that are more efficient in the traditional algorithm analysis model.

William Devanny, Juan Jose Besa Vial, David Eppstein, Michael Goodrich, Timothy Johnson
University of California, Irvine
Department of Computer Science
wdevanny@uci.edu, jjbesavi@uci.edu, eppstein@uci.edu, goodrich@uci.edu, tujohnso@uci.edu

CP0

Analyzing Boltzmann Samplers for Bose-Einstein Condensates with Dirichlet Generating Functions

Boltzmann sampling is commonly used to uniformly sample objects of a particular size from large combinatorial sets. For this technique to be effective, one needs to prove that (1) the sampling procedure is efficient and (2) objects of the desired size are generated with sufficiently high probability. We use this approach to give a provably efficient sampling algorithm for a class of weighted integer partitions related to Bose–Einstein condensation from statistical physics. Our sampling algorithm is a probabilistic interpretation of the ordinary generating function for these objects, derived from the symbolic method of analytic combinatorics. Using the Khintchine–Meinardus probabilistic method to bound the rejection rate of our Boltzmann sampler through singularity analysis of Dirichlet generating functions, we offer an alternative approach to analyze Boltzmann samplers for objects with multiplicative structure.

Matthew Fahrbach, Megan Bernstein, Dana Randall
Georgia Institute of Technology
matthew.fahrbach@gatech.edu, bernstein@math.gatech.edu, randall@cc.gatech.edu

CP0

Hybrid Indexing Revisited

Hybrid indexing is a recent approach to text indexing that allows the space-usage of conventional text indexes (e.g., suffix trees, suffix arrays, FM-indexes) to scale well with the text size, n, when z, the size of the Lempel-Ziv parsing of the text, is small relative to n. The price for this improved scalability is that an upper bound M on the pattern length that can be searched for must be declared at index construction time. Because the size of the resulting index contains an $O(Mz)$ term, M must be kept reasonably small, though it has been shown that $M \approx 100$ leads to acceptable performance in some genomic applications. However, despite its promise, the practical performance of hybrid indexing relative to other compressed index data structures is poorly understood. This paper addresses that need, detailing experiments that show hybrid indexing — when carefully implemented — to be significantly smaller and faster than alternative approaches on a broad range of data of different levels of compressibility. We also describe practical extensions to hybrid indexing that obviate the restriction on M, supporting search for patterns of arbitrary length.

Hector Ferrada, Dominik Kempa, Simon Puglisi
University of Helsinki
ferrada@cs.helsinki.fi, dominik.kempa@cs.helsinki.fi, simon.puglisi@cs.helsinki.fi

CP0

Trace Reconstruction with Varying Deletion Probabilities

In the trace reconstruction problem an unknown string $\mathbf{x} = (x_0, \ldots, x_{n-1}) \in \{0, 1, ..., m-1\}^n$ is observed through the deletion channel, which deletes each x_k with a certain probability, yielding a contracted string $\widetilde{\mathbf{X}}$. Earlier works have proved that if each x_k is deleted with the same probability $q \in [0, 1)$, then $\exp(O(n^{1/3}))$ independent copies of the contracted string $\widetilde{\mathbf{X}}$ suffice to reconstruct $\mathbf{x}$ with high probability. We extend this upper bound to the setting where the deletion probabilities vary, assuming certain regularity conditions. First we consider the case where x_k is deleted with some known probability q_k. Then we consider the case where each letter $\zeta \in \{0, 1, ..., m-1\}$ is associated with some possibly unknown deletion probability q_ζ.

Nina Holden
Massachusetts Institute of Technology
ninah@mit.edu

Lisa Hartung
Courant, New York University
lisa.hartung@nyu.edu

Yuval Peres
Microsoft Research, Redmond
peres@microsoft.com

CP0

Scalable Kernelization for Maximum Independent Sets

The most efficient algorithms for finding maximum independent sets use reduction rules to obtain a much smaller problem instance called a kernel. The kernel can then be solved quickly using exact or heuristic algorithms or by repeatedly kernelizing recursively in the branch-and-reduce paradigm. It is of critical importance for these algorithms that kernelization is fast and returns a small kernel. Current algorithms are either slow but produce a small kernel, or fast and give a large kernel. We attempt to accomplish both of these goals simultaneously, by giving an efficient parallel kernelization algorithm based on graph partitioning and parallel bipartite maximum matching. We combine our parallelization techniques with two techniques to accelerate kernelization further: dependency checking that prunes reductions that cannot be applied, and reduction tracking that allows us to stop kernelization when reductions become less fruitful. Our algorithm produces kernels that are orders of magnitude smaller than the fastest kernelization methods, while having a similar execution time. Furthermore, our algorithm is able to compute kernels with size comparable to the smallest known kernels, but up to two orders of magnitude faster than previously possible. Finally, we show that our kernelization algorithm can be used to accelerate existing state-of-the-art heuristic algorithms, allowing us to find larger independent sets faster on large real-world networks and synthetic instances.

Darren Strash
Department of Computer Science
Colgate University
darren.strash@gmail.com

Demian Hespe
Karlsruhe Institute of Technology
hespe@kit.edu

Christian Schulz
Institute of Theoretical Informatics
Karlsruhe Institute of Technology
christian.schulz@kit.edu

CP0

Computing 2-Connected Components and Maximal 2-Connected Subgraphs in Directed Graphs: An Experimental Study

Motivated by very recent work on 2-connectivity in directed graphs, we revisit the problem of computing the 2-edge- and 2-vertex-connected components, and the maximal 2-edge- and 2-vertex-connected subgraphs of a directed graph G. We explore the design space for efficient algorithms in practice, based on recently proposed techniques, and conduct a thorough empirical study to highlight the merits and weaknesses of each technique.

Aikaterini Karanasiou
University of Rome Tor Vergata
aikaterini.karanasiou@uniroma2.it

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

Giuseppe Italiano
Universita di Roma "Tor Vergata'
giuseppe.italiano@uniroma2.it

Nikos Parotsidis, Nilakantha Paudel
University of Rome Tor Vergata
nikos.parotsidis@uniroma2.it, nilakantha.paudel@uniroma2.it

CP0

Engineering a Delegatable and Error-Tolerant Algorithm for Counting Small Subgraphs

We study the problem of counting the number of occurrences of a given six-vertex pattern graph S in an n-vertex host graph H. We engineer an open-source GPU implementation of a distributed algorithm design of Björklund and Kaski [PODC 2016] where (i) the execution of the algorithm can be *delegated* [Goldwasser, Kalai, and Rothblum, J. ACM 2015] to produce a noninteractive probabilistically checkable proof of correctness, and (ii) the execution of the algorithm when preparing the proof tolerates a controllable number of adversarial errors. Experiments with NVIDIA Tesla K80 and Tesla P100 Accelerators demonstrate that the framework is practical for inputs of up to 512 vertices, with proof checking being several orders of magnitude more efficient than preparing the proof; however, proof preparation still carries at least one order of magnitude overhead compared with just solving the problem.

Petteri Kaski
Aalto University
Department of Computer Science
petteri.kaski@aalto.fi

CP0

Hyperbolic Embeddings for Near-Optimal Greedy Routing

Greedy routing computes paths between nodes in a network by successively moving to the neighbor closest to the target with respect to coordinates given by an embedding into some metric space. Its advantage is that only local information is used for routing decisions. We present different algorithms for generating graph embeddings into the hyperbolic plane that are well suited for greedy routing. In particular our embeddings guarantee that greedy routing always succeeds in reaching the target and we try to minimize the lengths of the resulting greedy paths. We evaluate our algorithm on multiple generated and real wold networks. For networks that are generally assumed to have a hidden underlying hyperbolic geometry, such as the Internet graph, we achieve near-optimal results, i.e., the resulting greedy paths are only slightly longer than the corresponding shortest paths. In the case of the Internet graph, they are only 6% longer when using our best algorithm, which greatly improves upon the previous best known embedding, whose creation required substantial manual intervention.

Thomas Blaesius, Tobias Friedrich,
Maximilian Katzmann, Anton Krohmer
Hasso Plattner Institute
thomas.blaesius@hpi.de, friedrich@hpi.de, maximilian.katzmann@hpi.de, anton.krohmer@gmx.de

CP0

Faster Approximation Algorithm for the k-Regret Minimizing Set and Related Problems

Efficient multi-criteria decision making often requires looking at a small set of representative objects from a large collection. A recently proposed method for finding representative objects is the k-regret minimizing set (k-RMS problem). Intuitively, given a large set of objects (points) in d dimensions, the goal is to choose a small representative subset, such that for every user preference, there is always an object in the subset whose preference score is not much worse than the score of the k-th most preferred object in the original set. We propose two new efficient approximation algorithms for the k-regret minimizing set problem with provable theoretical guarantees. Our algorithms improve on the space and time complexities of previous approximation algorithms for the k-RMS problem. In addition, we run extensive experiments on real and synthetic data sets showing that simple modifications of our theoretical algorithms run significantly faster than the previous implementations of the k-RMS problem. Finally, we present an efficient approximation algorithm with theoretical guarantees for an extension of the k-RMS problem, which is called the Top-k regret minimizing set problem.

Nirman Kumar
University of Memphis
nirman@gmail.com

Stavros Sintos
Duke University

ssintos@cs.duke.edu

CP0

Simple, Fast and Lightweight Parallel Wavelet Tree Construction

The wavelet tree (Grossi et al. [SODA, 2003]) and wavelet matrix (Claude et al. [Inf. Syst., 47:15–32, 2015]) are compact indices for texts over an alphabet $[0, \sigma)$ that support *rank*, *select* and *access* queries in $O(\lg \sigma)$ time. We first present new practical sequential and parallel algorithms for wavelet tree construction. Their unifying characteristics is that they construct the wavelet tree *bottom-up*, i. e., they compute the last level first. We also show that this bottom-up construction can easily be adapted to wavelet *matrices*. In practice, our best sequential algorithm is up to twice as fast as the currently fastest sequential wavelet tree construction algorithm (Shun [DCC, 2015]), simultaneously saving a factor of 2 in space. This scales up to 32 cores, where we are about equally fast as the currently fastest parallel wavelet tree construction algorithm (Labeit et al. [DCC, 2016]), but still use only about 75 % of the space. An additional theoretical result shows how to adapt any wavelet *tree* construction algorithm to the wavelet *matrix* in the same (asymptotic) time, using only little extra space.

Johannes Fischer, Florian Kurpicz, Marvin Löbel
TU Dortmund
johannes.fischer@cs.tu-dortmund.de, florian.kurpicz@tu-dortmund.de, marvin.loebel@tu-dortmund.de

CP0

Split-Decomposition Trees with Prime Nodes: Enumeration and Random Generation of Cactus Graphs

In this paper, we build on recent results by Chauve *et al.* and Bahrani and Lumbroso, which combined the split-decomposition, as exposed by Gioan and Paul, with analytic combinatorics, to produce new enumerative results on graphs—in particular the enumeration of several subclasses of perfect graphs (distance-hereditary, 3-leaf power, ptolemaic). Our goal was to study a simple family of graphs, of which the split-decomposition trees have prime nodes drawn from a enumerable (and manageable!) set of graphs. The cactus graphs, which we describe in more detail further down in this paper, can be thought of as trees with their vertices replaced by cycles (or arbitrary lengths). Their split-decomposition trees contain prime nodes that are cycles, making them ideal for study. We derive a characterization of the split-decomposition trees of cactus graphs, we produce a general template of symbolic grammar for cactus graphs, and then we implement random generation for these graphs, building on work by Iriza.

Maryam Bahrani, Jérémie Lumbroso
Princeton University
mbahrani@princeton.edu, lumbroso@CS.Princeton.EDU

CP0

Theoretical Analysis of Beaconless Geocast Protocols in 1D

Beaconless geocast protocols are routing protocols used to send messages in mobile ad-hoc wireless networks, in which the only information available to each node is its own location. Messages get routed in a distributed manner: each node uses local decision rules based on the message source and destination, and its own location. In this paper we analyze six different beaconless geocast protocols, focusing on two relevant 1D scenarios. The selection of protocols reflects the most relevant types of protocols proposed in the literature, including those evaluated in previous computer simulations. We present a formal and structured analysis of the maximum number of messages that a node can receive, for each protocol, in each of the two scenarios. This is a measure of the network load incurred by each protocol. Our analysis, that for some of the protocols requires an involved probabilistic analysis, confirms behaviors that had been observed only through simulations before.

Joachim Gudmundsson
University of Sydney
joachim.gudmundsson@gmail.com

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

Maarten Löffler
Utrecht University
m.loffler@uu.nl

Tobias Müller
Groningen University
tobias.muller@rug.nl

Vera Sacristán
Universitat Politecnica de Catalunya
vera.sacristan@upc.edu

Rodrigo Silveira
Universitat Politècnica de Catalunya
rodrigo.silveira@upc.edu

CP0

On Deletions in Open Addressing Hashing

Deletions in open addressing tables have often been seen as problematic. The usual solution is to use a special mark '`deleted`' so that probe sequences continue past deleted slots, as if there was an element still sitting there. Such a solution, notwithstanding is wide applicability, may involve performance degradation. In the first part of this paper we review a practical implementation of the often overlooked deletion algorithm for linear probing hash tables, analyze its properties and performance, and provide several strong arguments in favor of the Robin Hood variant. In particular, we show how a small variation can yield substantial improvements for unsuccessful search. In the second part we propose an algorithm for true deletion in open addressing hashing with secondary clustering, like quadratic hashing. As far as we know, this is the first time that such an algorithm appears in the literature. Moreover, for tables built using the Robin Hood variant the deletion algorithm strongly preserves randomness (the resulting table is identical to the table that would result if the item were not inserted at all). Although it involves some extra memory for bookkeeping, the algorithm is comparatively easy and efficient, and it might be of some practical value, besides its theoretical interest.

Conrado Martínez, Rosa Jiménez
Department of Computer Science
Univ. Politècnica de Catalunya

conrado@cs.upc.edu, jimenez@cs.upc.edu

CP0

Area-Preserving Subdivision Simplification With Topology Constraints: Exactly And In Practice

Given a planar subdivision and a set of points, we want to simplify the subdivision by removing vertices of degree 2. The vertices should be removed in such a way that all points remain in their respective faces, the area of any face changes by a factor of at most δ and the distance of the resulting lines to the original ones is at most ε. While inapproximability carries over from less general problems, we present an heuristic approach which solves continental sized instances in seconds. For city-sized instances we can compute optimal results via Integer-Linear-Programming and show, that our algorithm provides close-to-optimal results.

Thomas Mendel
University of Stuttgart
mendel@fmi.uni-stuttgart.de

CP0

Adaptive Cuckoo Filters

We introduce the adaptive cuckoo filter (ACF), a data structure for approximate set membership that extends cuckoo filters by reacting to false positives, removing them for future queries. As an example application, in packet processing queries may correspond to flow identifiers, so a search for an element is likely to be followed by repeated searches for that element. Removing false positives can therefore significantly lower the false positive rate. The ACF, like the cuckoo filter, uses a cuckoo hash table to store fingerprints. We allow fingerprint entries to be changed in response to a false positive in a manner designed to minimize the effect on the performance of the filter. We show that the ACF is able to significantly reduce the false positive rate by presenting both a theoretical model for the false positive rate and simulations using both synthetic data sets and real packet traces.

Michael Mitzenmacher
Harvard University
michaelm@eecs.harvard.edu

Salvatore Pontarelli
Consorzio Nazionale Interuniversitario
per le Telecomunicazioni (CNIT)
salvatore.pontarelli@uniroma2.it

Pedro Reviriego
Universidad Antonio de Nebrija
previrie@nebrija.es

CP0

Grid Peeling and the Affine Curve-Shortening Flow

In this paper we study an experimentally-observed connection between two seemingly unrelated processes, one from computational geometry and the other from differential geometry. The first one (which we call "grid peeling") is the convex-layer decomposition of subsets $G \subset Z^2$ of the integer grid, previously studied for the particular case $G = \{1, ..., m\}^2$ by Har-Peled and Lidicky (2013). The second one is the affine curve-shortening flow (ACSF), first studied by Alvarez et al. (1993) and Sapiro and Tannenbaum (1993). We present empirical evidence that, in a certain well-defined sense, grid peeling behaves at the limit like ACSF on convex curves. We offer some theoretical arguments in favor of this conjecture. We also pay closer attention to the simple case where $G = N^2$ is a quarter-infinite grid. This case corresponds to ACSF starting with an infinite L-shaped curve, which when transformed using the ACSF becomes a hyperbola for all times $t > 0$. We prove that, in the grid peeling of N^2, (1) the number of grid points removed up to iteration n is $\Theta(n^{3/2} \log n)$; and (2) the boundary at iteration n is sandwiched between two hyperbolas that are separated from each other by a constant factor.

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

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

Gabriel Nivasch
Ariel University
gabrieln@ariel.ac.il

CP0

Practical Minimum Cut Algorithms

The minimum cut problem for an undirected edge-weighted graph asks us to divide its set of nodes into two blocks while minimizing the weight sum of the cut edges. Here, we introduce a linear-time algorithm to compute near-minimum cuts. Our algorithm is based on cluster contraction using label propagation and Padberg and Rinaldi's contraction heuristics [SIAM Review, 1991]. We give both sequential and shared-memory parallel implementations of our algorithm. Extensive experiments on both real-world and generated instances show that our algorithm finds the optimal cut on nearly all instances significantly faster than other state-of-the-art exact algorithms, and our error rate is lower than that of other heuristic algorithms. In addition, our parallel algorithm shows good scalability.

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

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

Christian Schulz
University of Vienna
Karlsruhe Institute of Technology
christian.schulz@univie.ac.at

Darren Strash
Colgate University
dstrash@cs.colgate.edu

CP0

Linear Time Canonicalization and Enumeration of Non-Isomorphic 1-Face Embeddings

Antiparallel strong traces (ASTs) are a type of walks in

graphs which use every edge exactly twice. They correspond to 1-face embeddings in orientable surfaces and can be used to design self-assembling protein or DNA strands. Based on a novel canonical form invariant for ASTs, gap vector, we provide a linear-time isomorphism test for ASTs and thus, also for orientable 1-face embeddings of graphs. Using the canonical form, we develop an algorithm for enumerating all pairwise non-isomorphic 1-face embeddings of graphs. We compare our algorithm with an independent implementation of a recent algebraic approach (Baic et al., MATCH Commun. Math. Comput. Chem. 78 (3), 2017) on large data sets. Our results yield the first large-scale enumeration of non-isomorphic embeddings and investigation of their properties.

Marc Hellmuth
University of Greifswald
mhellmuth@mailbox.org

Anders S. Knudsen
University of Southern Denmark
antersenator@gmail.com

Michal Kotrbcik
The University of Queensland, Department of Mathematics
Brisbane St Lucia, QLD 4072
miso.kotrbcik@gmail.com

Daniel Merkle
Department of Mathematics and Computer Science
University of Southern Denmark
daniel@imada.sdu.dk

Nikolai Nøjgaard
University of Southern Denmark
nnoej10@student.sdu.dk

CP0

The Cover Time of a Biased Random Walk on $G_{n,p}$

We analyse the cover time of a biased random walk on the random graph $G_{n,p}$. The walk is biased towards visiting vertices of low degree and this makes the cover time less than in the unbiased case.

Samantha N. Petti
Georgia Tech
Georgia Tech
spetti@gatech.edu

Colin Copper
King's College London
colin.cooper@kcl.ac.uk

Alan Frieze
Carnegie Mellon University
Department of Mathematical Sciences
alan@random.math.cmu.edu

CP0

A Geometric Heuristic for Rectilinear Crossing Minimization

In this paper we consider the *rectilinear crossing minimization problem*, i.e., we seek a straight-line drawing Γ of a graph $G = (V, E)$ with a small number of edge crossings. Crossing minimization is an active field of research. While there is a lot of work on heuristics for topological drawings, these techniques are typically not transferable to the rectilinear (i.e., straight-line) setting. We introduce and evaluate three heuristics for rectilinear crossing minimization. The approaches are based on the primitive operation of moving a single vertex to its crossing-minimal position in the current drawing Γ, for which we give an $O\left((kn+m)^2 \log(kn+m)\right)$-time algorithm, where k is the degree of the vertex and n and m are the numbers of vertices and edges of the graph, respectively. In an experimental evaluation, we demonstrate that our algorithms compute straight-line drawings with fewer crossings than energy-based algorithms implemented in the Open Graph Drawing Framework on a varied set of benchmark instances. All experiments are evaluated with a statistical significance level of $\alpha = 0.05$.

Marcel Radermacher
Karlsruhe Institute of Technology
radermacher@kit.edu

Klara Reichard
Universitaet Tuebingen
klara.reichard@uni-tuebingen.de

Ignaz Rutter
TU Eindhoven
i.rutter@tue.nl

Dorothea Wagner
Karlsruhe Institute of Technology
Institute of Theoretical Informatics
dorothea.wagner@kit.edu

CP0

Asymptotic Enumeration of Graph Classes with Many Components

We consider graph classes $\mathcal{G}$ in which every graph has components in a class $\mathcal{C}$ of connected graphs. We provide a framework for the asymptotic study of $|\mathcal{G}_{n,N}|$, the number of graphs in $\mathcal{G}$ with n vertices and $N := \lfloor \lambda n \rfloor$ components, where $\lambda \in (0,1)$. Assuming that the number of graphs with n vertices in $\mathcal{C}$ satisfies

$$|\mathcal{C}_n| \sim b n^{-(1+\alpha)} \rho^{-n} n!, \quad n \to \infty,$$

for some $b, \rho > 0$ and $\alpha > 1$ – a property commonly encountered in graph enumeration – we show that

$$|\mathcal{G}_{n,N}| \sim c(\lambda) n^{f(\lambda)} (\log n)^{g(\lambda)} \rho^{-n} h(\lambda)^N \frac{n!}{N!}, \quad n \to \infty,$$

for explicitly given $c(\lambda), f(\lambda), g(\lambda)$ and $h(\lambda)$. These functions are piecewise continuous with a discontinuity at a critical value λ^*, which we also determine. The central idea in our approach is to sample objects of $\mathcal{G}$ randomly by so-called Boltzmann generators in order to translate enumerative problems to the analysis of iid random variables. By that we are able to exploit local limit theorems and large deviation results well-known from probability theory to prove our claims. The main results are formulated for generic combinatorial classes satisfying the SET-construction.

Konstantinos Panagiotou, Leon Ramzews
University of Munich
kpanagio@math.lmu.de, ramzews@math.lmu.de

CP0

Exponential Bounds on Graph Enumerations

From Vertex Incremental Characterizations

In this paper, building on previous work by Nakano *et al.*, 2009, we develop an alternate technique which almost automatically translates (existing) vertex incremental characterizations of graph classes into asymptotics of that class. Specifically, we construct trees corresponding to the sequences of vertex incremental operations which characterize a graph class, and then use analytic combinatorics to enumerate the trees, giving an upper bound on the graph class. This technique is applicable to a wider set of graph classes compared to the tree decompositions, and we show that this technique produces accurate upper bounds. We first validate our method by applying it to the case of distance-hereditary graphs, and comparing the bound obtained by our method with that obtained by Nakano et al., 2009, and the exact enumeration obtained by Chauve *et al.*, 2014. We then illustrate its use by applying it to switch cographs, for which there are few known results: our method provide a bound of 6.301^n, to be compared with the precise exponential growth, 6.159^n, which we obtained independently through the relationship between switch cographs and bicolored cographs, first introduced by Hertz, 1999. We believe the popularity of vertex incremental characterizations might mean this may prove a fairly convenient to tool for future exploration of graph classes.

Jeremie O. Lumbroso
Princeton University
Department of Computer Science
lumbroso@princeton.edu

Jessica Shi
Princeton University
js53@princeton.edu

CP0

Computing Floods Caused by Non-Uniform Sea-Level Rise

Predicting floods caused by the rise of the sea level is a critical task for preventing large scale catastrophes. Such predictions can potentially be made using a forecast of the sea level and a detailed model of the terrain. However, since available terrain datasets can easily exceed the size of the main memory of a standard computer, I/O (rather than internal computation time) can often be the bottleneck when computing such predictions. Thus to perform predictions efficiently we need an *I/O-efficient* approach, which minimizes the transfer of data blocks between main memory and disk. Given a terrain raster T and a sea-level forecast raster S of N cells each, we examine the problem of computing the water level of the induced flood for each cell in T. We introduce an I/O-efficient algorithm for this problem that uses $O((N/B)\log_{M/B}(X/B))$ I/Os after $O((N/B)\log_{M/B}(N/B))$ I/Os of preprocessing, where X is the number of local minima in T, and M and B are the size of main memory and data block, respectively. When $X < M$ (which holds in practice) our algorithm requires optimal $O(N/B)$ I/Os after preprocessing. W e have implemented our algorithm and put considerable effort into engineering it. We present experiments that illustrate the efficiency and practicality of the algorithm, which is so efficient that work is underway to incorporate our results in the forecast services of the Danish Meteorological Institute.

Lars Arge
Madalgo, Aarhus University
large@madalgo.au.dk

Yujin Shin
MADALGO, Department of Computer Science
Aarhus University
yujinshin90@gmail.com

Constantinos Tsirogiannis
MADALGO, Aarhus University
analekta@gmail.com

CP0

The Complexity of the Multiple Pattern Matching Problem for Random Strings

We generalise a multiple string pattern matching algorithm, proposed by Fredriksson and Grabowski [J. Discr. Alg. 7, 2009], to deal with arbitrary dictionaries on an alphabet of size s. If r_m is the number of words of length m in the dictionary, and $\phi(r) = \max_m \ln(smr_m)/m$, the complexity rate for the string characters to be read by this algorithm is at most $K_{\mathrm{ub}}\phi(r)$ for some constant K_{ub}. Then, we generalise the classical lower bound of Yao [SIAM J. Comput. 8, 1979], for the problem with a single pattern, to deal with arbitrary dictionaries, and determine it to be at least $K_{\mathrm{lb}}\phi(r)$. This proves the optimality of the algorithm, improving and correcting previous claims. Furthermore, we establish a tightness result for dictionaries with the same set $\{r_m\}$: the worst-case, average-case, and best-case complexities (the latter, up to a finite fraction of the dictionaries) are all equal, up to a finite multiplicative constant.

Andrea Sportiello, Frederique Bassino, Tsinjo Rakotoarimalala
LIPN, Université Paris Nord
Villetaneuse, FRANCE
andrea.sportiello@gmail.com,
frederique.bassino@lipn.univ-paris13.fr, tsinjotony@gmail.com

CP0

Probabilistic Analysis of the Dual-Pivot Quicksort "Count"

Recently, Aumüller and Dietzfelbinger proposed a version of a dual-pivot quicksort, called "Count", which is optimal among dual-pivot versions with respect to the average number of key comparisons required. In this note we provide further probabilistic analysis of "Count". We derive an exact formula for the average number of swaps needed by "Count" as well as an asymptotic formula for the variance of the number of swaps and a limit law. Also for the number of key comparisons the asymptotic variance and a limit law are identified. We also consider both complexity measures jointly and find their asymptotic correlation.

Ralph Neininger, Jasmin Straub
J.W. Goethe University, Frankfurt
neiningr@math.uni-frankfurt.de, jstraub@math.uni-frankfurt.de

CP0

On Induced Paths, Holes and Trees in Random Graphs

The concentration of the sizes of largest induced paths and cycles (holes) are studied in the random graph model

$\mathcal{G}(n,p)$. A 2-point concentration is proved for the size of the largest induced path and cycle, for all $p = p(n)$ satisfying $p \geq n^{-1/2}(\ln n)^2$ and $p \leq 1 - \epsilon$ where $\epsilon > 0$ is any constant. No such tight concentration (within two consecutive values) was known before for induced paths and cycles. As a corollary, a significant additive improvement is obtained over a 25-year old result of Erdős and Palka [**?**]concerning the size of the largest induced tree in a dense random graph. The proofs are based on second moment calculations and an explanation as to why more powerful concentration tools cannot be employed is also provided.

C R. Subramanian
The Institute of Mathematical Sciences, Chennai, INDIA.
crs@imsc.res.in

Kunal Dutta
DataShape, INRIA Sophia Antipolis-Mediterranee,
kdutta@mpi-inf.mpg.de

CP0

A Practical Fpt Algorithm for Flow Decomposition and Transcript Assembly

The Flow Decomposition problem, which asks for the smallest set of weighted paths that covers a flow on a DAG, has recently been used as an important computational step in transcript assembly. We prove the problem is in FPT when parameterized by the number of paths by giving a practical linear fpt algorithm. Further, we implement and engineer a Flow Decomposition solver based on this algorithm, and evaluate its performance on RNA-sequence data. Crucially, our solver finds exact solutions while achieving runtimes competitive with a state-of-the-art heuristic. Finally, we contextualize our design choices with two hardness results related to preprocessing and weight recovery. Specifically, k-Flow Decomposition does not admit polynomial kernels under standard complexity assumptions, and the related problem of assigning (known) weights to a given set of paths is NP-hard.

Kyle Kloster
NCSU
Computer Science
kyle.kloster@gmail.com

Philipp Kuinke
RWTH Aachen University
kuinke@cs.rwth-aachen.de

Michael O'Brien
NC State University
mpobrie3@ncsu.edu

Felix J. Reidl
North Carolina State University
felix.reidl@gmail.com

Fernando Sanchez Villaamil
RWTH Aachen University
fernando.sanchez@cs.rwth-aachen.de

Blair Sullivan
North Carolina State University
vbsulliv@ncsu.edu

Andrew van Der Poel
NC State University
ajvande4@ncsu.edu

CP0

Efficiently Inferring Pairwise Subtree Prune-and-Regraft Adjacencies Between Phylogenetic Trees

We develop a time-optimal $O(mn^2)$-time algorithm to construct the subtree prune-regraft (SPR) graph on a collection of m phylogenetic trees with n leaves. This improves on the previous bound of $O(mn^3)$. Such graphs are used to better understand the behaviour of phylogenetic methods and recommend parameter choices and diagnostic criteria. The limiting factor in these analyses has been the difficulty in constructing such graphs for large numbers of trees. We also develop the first efficient algorithms for constructing the nearest-neighbor interchange (NNI) and tree bisection-and-reconnection (TBR) graphs. These new algorithms are enabled by a change of perspective: rather than focusing on the trees and checking for pairs of adjacencies, we enumerate the potential adjacencies themselves in the form of structures called "agreement forests.' To turn this observation into an efficient algorithm, we develop two tools: SDLNewick, the first unique string representation for agreement forests, and a new AFContainer data structure which efficiently stores tree adjacencies using such strings.

Chris Whidden, Frederick Matsen
Fred Hutchinson Cancer Research Center
cwhidden@fredhutch.org, matsen@fredhutch.org

CP0

Quicksort Is Optimal For Many Equal Keys

I prove that the average number of comparisons for median-of-k Quicksort (with fat-pivot a.k.a. three-way partitioning) is asymptotically only a constant α_k times worse than the lower bound for sorting random multisets of n elements with $\Omega(n^\varepsilon)$ duplicates of each value (for any $\varepsilon > 0$). The constant is $\alpha_k = \ln(2)/\big(H_{k+1} - H_{(k+1)/2}\big)$, which converges to 1 as $k \to \infty$, so Quicksort is asymptotically optimal for inputs with many duplicates. This partially resolves a conjecture by Sedgewick and Bentley (1999, 2002) and constitutes the first progress on the analysis of Quicksort with equal elements since Sedgewick's 1977 article.

Sebastian Wild
David R. Cheriton School of Computer Science
University of Waterloo
wild@uwaterloo.ca

CP0

Parameterizing the Hardness of Binary Search Tree Access Sequences by Inversion Counts

We study a new way of measuring the expected performance of various binary search tree algorithms that is between the worst and the average case. Our starting point is the correspondence between binary search trees and insertion sequences, and we will measure the difficulty of such sequences based on inversion counts. This measure naturally interpolates between random and sequential insertion orders. We show that if the tree is randomly picked from all trees built upon insertion length n permutations with t inversions, the height of the tree can be bounded by $O(n^2 \log n / \min\{t, \binom{n}{2} - t\})$.

Meng He
Dalhousie University

mhe@cs.dal.ca

Richard Peng
Georgia Tech
rpeng@cc.gatech.edu

Yinzhan Xu
Massachusetts Institute of Technology
xyzhan@mit.edu

CP1

Incremental Topological Sort and Cycle Detection in $\tilde{O}(m\sqrt{n})$ Expected Total Time

In the *incremental cycle detection* problem edges are inserted to a directed graph (initially empty) and the algorithm has to report once a directed cycle is formed in the graph. A closely related problem to the incremental cycle detection is that of the *incremental topological sort* problem, in which edges are inserted to an acyclic graph and the algorithm has to maintain a valid topological sort on the vertices at all times. Both incremental cycle detection and incremental topological sort have a long history. The state of the art is a recent breakthrough of Bender, Fineman, Gilbert and Tarjan [TALG 2016], with two different algorithms with respective total update times of $\tilde{O}(n^2)$ and $O(m \cdot \min\{m^{1/2}, n^{2/3}\})$. The two algorithms work for both incremental cycle detection and incremental topological sort. In this paper we introduce a novel technique that allows us to improve upon the state of the art for a wide range of graph sparsity. Our algorithms has a total expected update time of $\tilde{O}(m\sqrt{n})$ for both the incremental cycle detection and the topological sort problems.

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

Aaron Bernstein
Technical University of Berlin, Germany
bernstei@gmail.com

CP1

Dynamic Algorithms for Graph Coloring

We design fast dynamic algorithms for proper vertex and edge colorings in a graph undergoing edge insertions and deletions. In the static setting, there are simple linear time algorithms for $(\Delta + 1)$-vertex coloring and $(2\Delta - 1)$-edge coloring in a graph with maximum degree Δ. It is natural to ask if we can efficiently maintain such colorings in the dynamic setting as well. We get the following three results. (1) We present a *randomized* algorithm which maintains a $(\Delta + 1)$-vertex coloring with $O(\log \Delta)$ expected amortized update time. (2) We present a *deterministic* algorithm which maintains a $(1 + o(1))\Delta$-vertex coloring with $O(\text{polylog}\, \Delta)$ amortized update time. (3) We present a simple, deterministic algorithm which maintains a $(2\Delta - 1)$-edge coloring with $O(\log \Delta)$ *worst-case* update time. This improves the recent $O(\Delta)$-edge coloring algorithm with $\tilde{O}(\sqrt{\Delta})$ worst-case update time [BarenboimM17]

Sayan Bhattacharya
University of Vienna
Vienna, Austria
S.Bhattacharya@warwick.ac.uk

Deeparnab Chakrabarty
Dartmouth College
USA
deepc@dartmouth.edu

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

Danupon Nanongkai
KTH Royal Institute of Technology
Sweden
danupon@kth.se

CP1

Dynamic Bridge-Finding in $\tilde{O}(\log^2 n)$ Amortized Time

We present a deterministic fully-dynamic data structure for maintaining information about the bridges in a graph. We support updates in $\tilde{O}(\log^2 n)$ amortized time, and can find a bridge in the component of any given vertex, or a bridge separating any two given vertices, in $O(\log n/\log\log n)$ worst case time. Our bounds match the current best for bounds for deterministic fully-dynamic connectivity up to $\log\log n$ factors. The previous best dynamic bridge finding was an $\tilde{O}(\log^3 n)$ amortized time algorithm by Thorup [STOC2000], which was a bittrick-based improvement on the $O(\log^4 n)$ amortized time algorithm by Holm et al.[STOC98, JACM2001]. Our approach is based on a different and purely combinatorial improvement of the algorithm of Holm et al., which by itself gives a new combinatorial $\tilde{O}(\log n^3 n)$ amortized time algorithm. Combining it with Thorup's bittrick, we get down to the claimed $\tilde{O}(\log^2 n)$ amortized time. Essentially the same new trick can be applied to the biconnectivity data structure from [STOC98, JACM2001], improving the amortized update time to $\tilde{O}(\log^3 n)$. We also offer improvements in space. We describe a general trick which applies to both of our new algorithms, and to the old ones, to get down to linear space, where the previous best use $O(m + n\log n\log\log n)$. Our result yields an improved running time for deciding whether a unique perfect matching exists in a static graph.

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

Eva Rotenberg, Mikkel Thorup
University of Copenhagen
eva@rotenberg.dk, mikkel2thorup@gmail.com

CP1

Decremental Transitive Closure and Shortest Paths for Planar Digraphs and Beyond

In this paper we show that the tools used to obtain the best state-of-the-art decremental algorithms for reachability and approximate shortest paths in directed graphs can be successfully combined with the existence of small separators in certain graph classes. In particular, for graph classes admitting balanced separators of size $O(\sqrt{n})$, such as planar, bounded-genus and minor-free graphs, we show that for both transitive closure and $(1 + \epsilon)$-approximate all pairs shortest paths (where ϵ is constant), there exist decremental algorithms with $\widetilde{O}(n^{3/2})$ total update time

and $\widetilde{O}(\sqrt{n})$ worst-case query time. Additionally, for the case of planar graphs, we show that for any $t \in [1, n]$, there exists a decremental transitive closure algorithm with $\widetilde{O}(n^2/t)$ total update time and $\widetilde{O}(\sqrt{t})$ worst-case query time. In particular, for $t = n^{2/3}$, if all the edges are eventually deleted, we obtain $\widetilde{O}(n^{1/3})$ amortized update and query times. Most of the algorithms we obtain are correct with high probability against an oblivious adversary.

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

CP1

Incremental DFS Algorithms: a Theoretical and Experimental Study

Depth first search (DFS) tree is a fundamental data structure for solving graph problems, which can be built in $O(m+n)$ time for a graph with n vertices and m edges. For undirected graphs, the two prominent incremental DFS algorithms are ADFS1 and ADFS2, achieving total update time of $O(n^{3/2}\sqrt{m})$ and $O(n^2)$ respectively. For DAGs, the only non-trivial incremental DFS algorithm is FDFS, requiring total $O(mn)$ update time. However, no such algorithm with $o(m^2)$ bound exist for directed graphs. We carry out extensive experimental and theoretical analysis of incremental DFS algorithms in random and real graphs, deriving the following results.

1. For inserting $\binom{n}{2}$ edges uniformly randomly, ADFS1, ADFS2 and FDFS perform equally well taking $\Theta(n^2)$ time experimentally, which is much better than their worst case bounds. We complement this result with probabilistic analysis proving $\tilde{O}(n^2)$ bound on their time complexities.
2. We also design a simple algorithm for both undirected and directed graphs, which theoretically matches and experimentally outperforms the state-of-the-art algorithm in dense random graphs.
3. Even for real world graphs, both ADFS1 and FDFS perform much better than their known bounds. Again, we design a simple algorithm each for directed and undirected graphs, which perform very well on real graphs, and almost always matches FDFS in directed graphs.

Surender Baswana
IIT Kanpur
sbaswana@cse.iitk.ac.in

Ayush Goel
Indian Institute of Technology Kanpur
India
ayushgoel529@gmail.com

Shahbaz Khan
IIT Kanpur
shahbaz.khan@univie.ac.at

CP2

Tightening Curves on Surfaces Via Local Moves

We prove new upper and lower bounds on the number of homotopy moves required to tighten a closed curve on a compact orientable surface (with or without boundary) as much as possible. First we prove that $\Omega(n^2)$ moves are required in the worst case to simplify a contractible closed curve on a surface with non-positive Euler characteristic, where n is the number of self-intersection points. Results of Hass and Scott imply a matching $O(n^2)$ upper bound for contractible curves on orientable surfaces. Second, we prove that any closed curve on any orientable surface can be tightened as much as possible using at most $O(n^4)$ homotopy moves. Except for a few special cases, only naïve exponential upper bounds were previously known for this problem.

Hsien-Chih Chang
University of Illinois, Urbana-Champaign
hchang17@illinois.edu

Jeff Erickson
Department of Computer Science
University of Illinois at Urbana-Champaign
jeffe@illinois.edu

David Letscher
Saint Louis University
letscher@slu.edu

Arnaud de Mesmay
CNRS, GIPSA-lab
Université Grenoble-Alpes
arnaud.de-mesmay@gipsa-lab.fr

Saul Schleimer
Warwick Mathematics Institute
University of Warwick
s.schleimer@warwick.ac.uk

Eric Sedgwick
DePaul University
esedgwick@cdm.depaul.edu

Dylan Thurston
Department of Mathematics
Indiana University
dpthurst@indiana.edu

Stephan Tillmann
School of Mathematics and Statistics
The University of Sydney
stephan.tillmann@sydney.edu.au

CP2

The Classical Complexity of Boson Sampling

We study the classical complexity of the exact Boson Sampling problem where the objective is to produce provably correct random samples from a particular quantum mechanical distribution. The computational framework was proposed in STOC '11 by Aaronson and Arkhipov in 2011 as an attainable demonstration of 'quantum supremacy'. Since its introduction Boson Sampling has been the subject of intense international research in the world of quantum computing. On the face of it, the problem is challenging for classical computation. Aaronson and Arkhipov show that exact Boson Sampling is not efficiently solvable by a classical computer unless $P^{\#P} = BPP^{NP}$ and the polynomial hierarchy collapses to the third level. The fastest known exact classical algorithm for the standard Boson Sampling problem requires $O(\binom{m}{n})n2^n)$ time to produce samples for a system with input size n and m output modes, making it infeasible for anything but the smallest values of n and m. We give an algorithm that is much faster, running in

$O(n2^n + poly(m,n))$ time and O(m) additional space. The algorithm is simple to implement and has low constant factor overheads. As a consequence our classical algorithm is able to solve the exact Boson Sampling problem for system sizes far beyond current photonic quantum computing experimentation, thereby significantly reducing the likelihood of achieving near-term quantum supremacy in the context of Boson Sampling.

Raphael Clifford
University of Bristol, UK
raphael.clifford@bristol.ac.uk

Peter Clifford
University of Oxford
peter.clifford@jesus.ox.ac.uk

CP2

A Near-Quadratic Lower Bound for the Size of Quantum Circuits of Constant Treewidth

We show that any quantum circuit of treewidth t, built from r-qubit gates, requires at least $\Omega(\frac{n^2}{2^{O(r\cdot t)}\cdot \log^4 n})$ gates to compute the element distinctness function. Our result generalizes a near-quadratic lower bound for quantum formula size obtained by Roychowdhury and Vatan [SIAM J. on Computing, 2001]. The proof of our lower bound follows by an extension of Nečiporuk's method to the context of quantum circuits of constant treewidth. This extension is made via a combination of techniques from structural graph theory, tensor-network theory, and the connected-component counting method, which is a classic tool in algebraic geometry. In particular, an essential key to proving our lower bound is the development of a new algorithm for tensor network contraction which may be of independent interest.

Mateus De Oliveira Oliveira
University of Bergen
mateus.oliveira@gmail.com

CP2

A Grid-Based Approximation Algorithm for the Minimum Weight Triangulation Problem

Given a set of n points on a plane, in the *Minimum Weight Triangulation* problem, we wish to find a triangulation that minimizes the sum of Euclidean length of its edges. This incredibly challenging problem was only recently shown to be NP-Hard. In this paper we present a novel polynomial-time algorithm that computes an expected 14-approximation of the minimum weight triangulation—a constant that is significantly smaller than what has been previously known. In our algorithm, we use grids to partition the edges into levels where shorter edges appear at smaller levels and edges with similar lengths appear at the same level. We then triangulate the point set incrementally by introducing edges in increasing order of their levels. Edges of any level $i+1$ are added in two steps. In the first step, we partition the boundary of any non-triangulated face into reflex chains and add edges between successive chains using a variant of the well-known ring heuristic to generate a partial triangulation $\hat{A}_i$. In the second step, we greedily add non-intersecting level $i+1$ edges to $\hat{A}_i$ in increasing order of their length and obtain a partial triangulation A_{i+1}. The ring heuristic yields only an $O(\log n)$-approximation and the greedy heuristic achieves only a $\Theta(\sqrt{n})$-approximation. Therefore, it is surprising that their combination leads to an improved approximation ratio of 14.

Sharath Raghvendra, Mariette Wessels
Virginia Tech
sharathr@vt.edu, wmari11@vt.edu

CP2

Polycubes with Small Perimeter Defect

A polycube is a face-connected set of cubical cells on $\mathbb{Z}^3$. to-date, no formulae enumerating polycubes by volume (number of cubes) or perimeter (number of empty cubes neighboring the polycube) are known. We present a few formulae enumerating polycubes with a fixed deviation from the maximum possible perimeter.

Gill Barequet
Technion - Israel Institute of Technology
barequet@cs.technion.ac.il

Andrei Asinowski
Vienna University of Technology
andrei.asinowski@tuwien.ac.at

Yufei Zheng
Technion - Israel Institute of Technology
yufei@cs.technion.ac.il

CP3

Metric Violation Distance: Hardness and Approximation

Metric data plays an important role in various settings, for example, in metric-based indexing, clustering, classification, and approximation algorithms in general. Due to measurement error, noise, or an inability to completely gather all the data, a collection of distances may not satisfy the basic metric requirements, most notably the triangle inequality. In this paper we initiate the study of the *metric violation distance* problem: given a set of pairwise distances, modify the minimum number of distances such that the resulting set forms a metric. Three variants of the problem are considered, based on whether distances are allowed to only decrease, only increase, or the general case which allows both decreases and increases. We show that while the decrease only variant is polynomial time solvable, the increase only and general variants are NP-Complete, and moreover cannot in polynomial time be approximated to any ratio better than the minimum vertex cover problem. We then provide approximation algorithms for the increase only and general variants of the problem, by proving interesting necessary and sufficient conditions on the optimal solution, which are used to approximately reduce to a purely combinatorial problem for which we provide matching asymptotic upper and lower bounds.

Chenglin Fan, Benjamin A. Raichel, Gregory Van Buskirk
University of Texas at Dallas
cxf160130@utdallas.edu, benjamin.raichel@utdallas.edu, greg.vanbuskirk@utdallas.edu

CP3

Race Detection and Reachability in Nearly Series-Parallel Dags

A program is said to have a determinacy race if logically parallel parts of a program access the same memory location and one of the accesses is a write. These races are

generally bugs in the program, as different schedules of the program can lead to different results. Most prior work on detecting these races focuses on a subclass of programs with series-parallel or nested parallelism. This paper presents a race-detection algorithm for detecting races in a more general class of programs, namely programs that include arbitrary ordering constraints in additional to the series-parallel constructs. Our race-detection algorithm performs a serial execution of the program, augmented to detect races, in $O(T_1 + k^2)$ time, where T_1 is the sequential running time of the original program and k is the number of non series-parallel constraints. The main technical novelty is a new data structure for answering reachability queries in nearly series-parallel (SP) directed acyclic graphs (DAGs). Given as input a graph comprising an n-node series parallel graph and k additional non-SP edges, the total construction time of the data structure is $O(n + k^2)$, and each reachability query can be answered in $O(1)$ time. The data structure is *traversally incremental*, meaning that it supports the insertion of nodes/edges, but only as they are discovered through a graph traversal.

Jeremy Fineman
Georgetown University
jfineman@cs.georgetown.edu

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

Joseph Devietti
University of Pennsylvania
devietti@cis.upenn.edu

I-Ting Angelina Lee, Robert Utterback, Changming Xu
Washington University in St. Louis
angelee@wustl.edu, robert.utterback@wustl.edu, c.xu@wustl.edu

CP3

Planar Graphs As L-Intersection Or L-Contact Graphs

The L-*intersection graphs* are the graphs that have a representation as intersection graphs of axis-parallel L shapes in the plane. A subfamily of these graphs are $\{\mathsf{L}, |, -\}$-*contact graphs* which are the contact graphs of axis parallel L, $|$, and $-$ shapes in the plane. We prove here two results that were conjectured by Chaplick and Ueckerdt in 2013. We show that planar graphs are L-intersection graphs, and that triangle-free planar graphs are $\{\mathsf{L}, |, -\}$-contact graphs. These results are obtained by a new and simple decomposition technique for 4-connected triangulations. Our results also provide a much simpler proof of the known fact that planar graphs are segment intersection graphs.

Daniel Goncalves, Lucas Isenmann, Claire Pennarun
LIRMM (CNRS & Univ. Montpellier)
goncalves@lirmm.fr, lucas.isenmann@lirmm.fr, pennarun@lirmm.fr

CP3

A Submodular Measure and Approximate Gomory-Hu Theorem for Packing Odd Trails

Motivated by a problem about totally odd immersions of graphs, we define the *odd edge-connectivity* $\lambda_o(u, v)$ as the maximum number of edge-disjoint trails of odd length from u to v. It was recently discovered that $\lambda_o(u, v)$ can be approximated up to a constant multiplicative factor using edge-connectivity between u and v and the minimum value of another parameter that measures "how far from a bipartite graph' the part of the graph around u and v is. In this paper, we formalize this second ingredient and call it the *perimeter*. It is proved that it is a submodular function on the vertex-sets of a graph, and using this fact, we obtain a version of the Gomory-Hu Theorem in which minimum edge-cuts are replaced by sets of minimum perimeter. We construct (in polynomial time) a rooted forest structure, analogous to the Gomory-Hu tree of a graph, which encodes a collection of minimum-perimeter vertex-sets. Although the classical Gomory-Hu Theorem extends to arbitrary *symmetric* submodular functions, this result is novel and indicates a possibility for further generalizations. These results have implications to the study of path and trail systems with parity constraints. Two such applications are presented: an efficient data structure for storing approximate odd edge-connectivities for all pairs of vertices in a graph, and a rough structure theorem for graphs with no "totally odd' immersion of a large complete graph.

Bojan Mohar, Ross Churchley
Simon Fraser University
mohar@sfu.ca, ross@rosschurchley.com

CP3

The Complexity of Independent Set Reconfiguration on Bipartite Graphs

We settle the complexity of the Independent Set Reconfiguration problem on bipartite graphs under all three commonly studied reconfiguration models. We show that under the token jumping or token addition/removal model the problem is NP-complete. For the token sliding model, we show that the problem remains PSPACE-complete.

Daniel Lokshtanov
University of Bergen
daniello@ii.uib.no

Amer Mouawad
University of Bergen, Norway
amer.mouawad@gmail.com

CP4

Recognizing Weak Embeddings of Graphs

We present an efficient algorithm for a problem in the interface between clustering and graph embeddings. An **embedding** $\varphi : G \to M$ of a graph G into a 2-manifold M maps the vertices in $V(G)$ to distinct points and the edges in $E(G)$ to interior-disjoint Jordan arcs between the corresponding vertices. In applications in clustering, cartography, and visualization, nearby vertices and edges are often bundled to a common node or arc, due to data compression or low resolution. This raises the computational problem of deciding whether a given map $\varphi : G \to M$ comes from an embedding. A map $\varphi : G \to M$ is a **weak embedding** if it can be perturbed into an embedding $\psi_\varepsilon : G \to M$ with $\|\varphi - \psi_\varepsilon\| < \varepsilon$ for every $\varepsilon > 0$. A polynomial-time algorithm for the problem was recently found by [Fulek and Kynčl, 2017], which reduces to solving a system of linear equations over Z_2. It runs in $O(n^{2\omega}) \leq O(n^{4.75})$ time, where $\omega \approx 2.373$ is the matrix multiplication exponent and n is the number of vertices and edges of G. We improve the running time to $O(n \log n)$ by generalizing a technique developed for the case that G is a cycle and the embedding

is a simple polygon [Akitaya et al., 2016], which combines local constraints on the orientation of subgraphs directly, thereby eliminating the need for solving large systems of linear equations.

Hugo A. Akitaya
Tufts University
hugo.alves_akitaya@tufts.edu

Radoslav Fulek
Institute of Science and Technology Austria
radoslav.fulek@gmail.com

Csaba D. Tóth
California State University Northridge
csaba.toth@csun.edu

CP4

A Polynomial Excluded-Minor Approximation of Treedepth

Treedepth is a well-studied graph invariant in the family of "width measures' that includes treewidth and pathwidth. Understanding these invariants in terms of excluded minors has been an active area of research. The recent Grid Minor Theorem of Chekuri and Chuzhoy (2014) establishes that treewidth is polynomially approximated by the largest $k \times k$ grid minor. In this paper, we give a similar polynomial excluded-minor approximation for treedepth in terms of three basic obstructions: grids, tree, and paths. Specifically, we show that there is a constant c such that every graph of treedepth $\geq k^c$ contains one of the following minors (each of treedepth $\geq k$):

- the $k \times k$ grid,
- the complete binary tree of height k,
- the path of order 2^k.

Let us point out that we cannot drop any of the above graphs for our purpose. Moreover, given a graph G we can, in polynomial time, find either an embedding of one of these minors or conclude that treedepth of G is at most k^c. This result has potential applications in a variety of settings where bounded treedepth plays a role. In addition to some graph structural applications, we describe a surprising application in circuit complexity and finite model theory from recent work of the second author.

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

Benjamin Rossman
University of Toronto
ben.rossman@utoronto.ca

CP4

Beating Brute Force for (Quantified) Satisfiability of Circuits of Bounded Treewidth

We investigate the algorithmic properties of circuits of bounded treewidth. Here the treewidth of a circuit C is defined as the treewidth of the underlying undirected graph of C, after the vertices corresponding to input gates have been removed. Thus, boolean formulae correspond to circuits of treewidth 1.

- Our first main result is an algorithm for counting the number of satisfying assignments of circuits with n input gates, treewidth ω, and at most $s \cdot n$ gates, which runs in time $2^{n(1-\frac{1}{O(s\cdot\omega\cdot 4^\omega)})}$. This is the first algorithm to achieve exponential speed-up over brute force for the satisfiability of linear size circuits with treewidth bounded by a constant greater than 1. For treewidth 1, our algorithm significantly outperforms the previously fastest $2^{n(1-1/O(s^2))}$ time satisfiability algorithm by Santhanam.

- Our second main result is an algorithm for True Quantified Boolean Circuit Satisfiability for circuits of treewidth ω, in which every input gate has fanout at most s, which runs in time $2^{n(1-1/O(s\cdot\omega\cdot 4^\omega)})$. Our algorithm is the first to achieve exponential speedup over brute force for such circuits.

- We also show that the number of wires of any constant treewidth circuit that computes the majority function must be super-linear by utilizing the structural properties of low treewidth circuits.

Daniel Lokshtanov
University of Bergen
daniello@ii.uib.no

Ivan Mikhailin
UC San Diego
imikhail@eng.ucsd.edu

Ramamohan Paturi
University of California, San Diego
paturi@cs.ucsd.edu

Pavel Pudlak
Institute of Mathematics of the Czech Academy of Sciences
pudlak@math.cas.cz

CP4

Minor-Matching Hypertree Width

In this paper we present a new width measure for a tree decomposition, *minor-matching hypertree width*, μ-*tw*, for graphs and hypergraphs, such that bounding the width guarantees that set of maximal independent sets has a polynomially-sized restriction to each decomposition bag. The relaxed conditions of the decomposition allow a much wider class of graphs and hypergraphs to have bounded width compared to other tree decompositions. We show that, for fixed k, there are $2^{(1-\frac{1}{k}+o(1))\binom{n}{2}}$ n-vertex graphs of minor-matching hypertree width at most k. A number of problems including Maximum Independence Set, k-Colouring, and Homomorphism of uniform hypergraphs permit polynomial-time solutions for hypergraphs with bounded minor-matching hypertree width and bounded rank. We show that for any given k and any graph G, it is possible to construct a decomposition of minor-matching hypertree width at most $O(k^3)$, or to prove that μ-$tw(G) > k$ in time $n^{O(k^3)}$. This is done by presenting a general algorithm for approximating the hypertree width of well-behaved measures, and reducing μ-*tw* to such measure. The result relating the restriction of the maximal independent sets to a set S with the set of induced matchings intersecting S in graphs, and minor matchings intersecting S in hypergraphs, might be of independent interest.

Nikola G. Yolov
University of Oxford

Nikola.Yolov@cs.ox.ac.uk

CP4

Cliquewidth III: The Odd Case of Graph Coloring Parameterized by Cliquewidth

Max-Cut (MC), Edge Dominating Set (DS), Graph Coloring (GC) and Hamiltonian Path (HP) on graphs of bounded cliquewidth can be formulated in MSO_2 (and therefore have linear-time algorithms on bounded treewidth graphs), but cannot be formulated in MSO_1 (which would have yielded linear-time algorithms on bounded cliquewidth graphs). Each of these problems can be solved in time $g(k)n^{f(k)}$ on graphs of cliquewidth k. Fomin et al. [SICOMP 2010] showed that the running times cannot be improved to $g(k)n^{O(1)}$ assuming W[1]≠FPT. However, this does not rule out non-trivial improvements to the exponent $f(k)$ in the running times. Fomin et al. [SICOMP 2014] improved the running times for EDS and MC to $n^{\mathcal{O}(k)}$, and proved $g(k)n^{o(k)}$ lower bounds for EDS, MC and HP assuming the ETH. Recently, Bergougnoux, Kanté and Kwon [WADS 2017] gave an $n^{O(k)}$-time algorithm for HP. Prior to this work, EDS, MC and HP were known to have tight $n^{\Theta(k)}$ algorithmic upper and lower bounds. In contrast, GC has an upper bound of $n^{O(2^k)}$ and a lower bound of merely $n^{o(\sqrt[4]{k})}$ (implicit from the W[1]-hardness proof). We close the gap for GC by proving a lower bound of $n^{2^{o(k)}}$. This shows that GC behaves qualitatively different from the other three problems. To the best of our knowledge, GC is the first natural problem known to require exponential dependence on the parameter in the exponent of n.

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

Daniel Lokshtanov
University of Bergen
daniello@ii.uib.no

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

Meirav Zehavi
Ben-Gurion University
zehavimeirav@gmail.com

CP5

Subquadratic Kernels for Implicit 3-Hitting Set and 3-Set Packing Problems

We consider four well-studied NP-complete packing/covering problems on graphs: Feedback Vertex Set in Tournaments (FVST), Cluster Vertex Deletion (CVD), Triangle Packing in Tournaments (TPT) and Induced P_3-Packing. For these four problems kernels with $O(k^2)$ vertices have been known for a long time. In fact, such kernels can be obtained by interpreting these problems as finding either a packing of k pairwise disjoint sets of size 3 (3-Set-Packing) or a hitting set of size at most k for a family of sets of size at most 3 (3-Hitting-Set). In this paper, we give the first kernels for FVST, CVD, TPT and Induced P_3-Packing with a subquadratic number of vertices. Specifically, we obtain the following results.

- FVST and TPT admits a kernel with $O(k^{\frac{3}{2}})$ vertices.
- CVD and Induced P_3-Packing admits a kernel with $O(k^{\frac{5}{3}})$ vertices.

Our results resolve an open problem from WorKer 2010 on the existence of kernels with $O(k^{2-\epsilon})$ vertices for FVST and CVD. All of our results are based on novel uses of old and new "expansion lemmas', and a weak form of crown decomposition.

Tien-Nam Le
ENS de Lyon
tien-nam.le@ens-lyon.fr

Daniel Lokshtanov
University of Bergen
daniello@ii.uib.no

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

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

Meirav Zehavi
University of Bergen
meirav.zehavi@ii.uib.no

CP5

Stochastic Packing Integer Programs with Few Queries

We consider a stochastic variant of the packing-type integer linear programming problem, which contains random variables in the objective vector. We are allowed to reveal each entry of the objective vector by conducting a query, and the task is to find a good solution by conducting a small number of queries. We propose a general framework of adaptive and non-adaptive algorithms for this problem, and provide a unified methodology for analyzing the performance of those algorithms. We also demonstrate our framework by applying it to a variety of stochastic combinatorial optimization problems such as matching, matroid, and stable set problems.

Takanori Maehara
RIKEN Center for Advanced Intelligence Project
Discrete Optimization Unit
takanori.maehara@riken.jp

Yutaro Yamaguchi
University of Tokyo
Department of Mathematical Informatics
yutaro_yamaguchi@ist.osaka-u.ac.jp

CP5

Randomized MWU for Positive LPs

We describe and analyze a simple randomized multiplicative weight update (MWU) based algorithm for approximately solving positive linear programming problems, in particular, mixed packing and covering LPs. Given m explicit linear packing and covering constraints over n variables specified by N nonzero entries, Young [2014] gave a deterministic algorithm returning an $(1+\epsilon)$-approximate feasible solution (if a feasible solution exists) in $\tilde{O}(N/\epsilon^2)$

time. We show that a simple randomized implementation matches this bound, and that randomization can be further exploited to improve the running time to $\tilde{O}(N/\epsilon + m/\epsilon^2 + n/\epsilon^3)$ (both with high probability). For instances that are not very sparse (with at least $\tilde{\omega}(1/\epsilon)$ nonzeroes per column on average), this improves the running time of $\tilde{O}(N/\epsilon^2)$. The randomized algorithm also gives improved running times for several implicitly defined problems that arise in combinatorial and geometric optimization. The analysis of the general algorithm and the improvements rely on techniques that we believe are of independent interest.

Chandra Chekuri, Kent Quanrud
University of Illinois at Urbana-Champaign
chekuri@illinois.edu, quanrud2@illinois.edu

CP5

Algorithms to Approximate Column-Sparse Packing Problems

Column-sparse packing problems arise in several contexts in both deterministic and stochastic discrete optimization. We present two unifying ideas, *(non-uniform) attenuation* and *multiple-chance algorithms*, to obtain improved approximation algorithms for some well-known families of such problems. As three main examples, we attain the integrality gap, up to lower-order terms, for known LP relaxations for k-column sparse packing integer programs (Bansal *et al.*, *Theory of Computing*, 2012) and stochastic k-set packing (Bansal *et al.*, *Algorithmica*, 2012), and go "half the remaining distance" to optimal for a major integrality-gap conjecture of Füredi, Kahn and Seymour on hypergraph matching (*Combinatorica*, 1993).

Brian Brubach, Karthik Abinav Sankararaman
University of Maryland College Park
bbrubach@cs.umd.edu, kabinav@cs.umd.edu

Aravind Srinivasan
University of Maryland
srin@cs.umd.edu

Pan Xu
University of Maryland College Park
panxu@cs.umd.edu

CP5

Near Optimal Jointly Private Packing Algorithm Via Dual Multiplicative Weight Update

We present an improved (ϵ, δ)-jointly differentially private algorithm for packing problems. Our algorithm gives a feasible output that is approximately optimal up to an αn additive factor as long as the supply of each resource is at least $\tilde{O}(\sqrt{m}/\alpha\epsilon)$, where m is the number of resources. This improves the previous result by Hsu et al. (SODA '16), which requires the total supply to be at least $\tilde{O}(m^2/\alpha\epsilon)$, and only guarantees approximate feasibility in terms of total violation. Further, we complement our algorithm with an almost matching hardness result, showing that $\Omega(\sqrt{m\ln(1/\delta)}/\alpha\epsilon)$ supply is necessary for any (ϵ, δ)-jointly differentially private algorithm to compute an approximately optimal packing solution. Finally, we introduce an alternative approach that runs in linear time, is exactly truthful, can be implemented online, and can be ϵ-jointly differentially private, but requires a larger supply of each resource.

Xue Zhu, Zhiyi Huang
The University of Hong Kong
xuezhu26@hku.hk, zhiyi@cs.hku.hk

CP6

A Fast Approximation Scheme for Low-Dimensional k-Means

We consider the popular k-means problem in d-dimensional Euclidean space. Recently Friggstad, Rezapour, Salavatipour [FOCS'16] and Cohen-Addad, Klein, Mathieu [FOCS'16] showed that the standard local search algorithm yields a $(1+\epsilon)$-approximation in time $(n \cdot k)^{1/\epsilon^{O(d)}}$, giving the first polynomial-time approximation scheme for the problem in low-dimensional Euclidean space. While local search achieves optimal approximation guarantees, it is not competitive with the state-of-the-art heuristics such as the famous k-means++ and D^2-sampling algorithms. In this paper, we aim at bridging the gap between theory and practice by giving a $(1+\epsilon)$-approximation algorithm for low-dimensional k-means running in time $n \cdot k \cdot (\log n)^{(d\epsilon^{-1})^{O(d)}}$, and so matching the running time of the k-means++ and D^2-sampling heuristics up to polylogarithmic factors. We speed-up the local search approach by making a non-standard use of randomized dissections that allows to find the best local move efficiently using a quite simple dynamic program. We hope that our techniques could help design better local search heuristics for geometric problems.

Vincent Cohen-Addad
CNRS, LIP6, Université Pierre et Marie Curie
Paris, France
vincent.cohen.addad@ens-lyon.org

CP6

Hierarchical Clustering: Objective Functions and Algorithms

Hierarchical clustering is a recursive partitioning of a dataset into clusters at an increasingly finer gran- ularity. Motivated by the fact that most work on hierarchical clustering was based on providing algo- rithms, rather than optimizing a specific objective, [18] framed similarity-based hierarchical clustering as a combinatorial optimization problem, where a good hierarchical clustering is one that minimizes some cost function. He showed that this cost function has certain desirable properties, such as in order to achieve optimal cost, disconnected components must be separated first and that in structureless graphs, i.e., cliques, all clusterings achieve the same cost. We take an axiomatic approach to defining good objective functions for both similarity and dissimilarity-based hierarchical clustering. We char- acterize a set of admissible objective functions (that includes the one introduced by Dasgupta) that have the property that when the input admits a natu- ral ground-truth hierarchical clustering, the ground-truth clustering has an optimal value.

Vincent Cohen-Addad
CNRS, LIP6, Université Pierre et Marie Curie
Paris, France
vincent.cohen.addad@ens-lyon.org

Varun Kanade
University of Oxford

varunk@cs.ox.ac.uk

Claire Mathieu
CNRS, Ecole normale supérieure
clairemmathieu@gmail.com

Frederik Mallmann-Trenn
MIT
fmallman@sfu.ca

CP6

Adaptive Hierarchical Clustering Using Ordinal Queries

In many applications of clustering (for example, ontologies or clusterings of animal or plant species), hierarchical clusterings are more descriptive than a flat clustering. A hierarchical clustering over n elements is represented by a rooted binary tree with n leaves, each corresponding to one element. The subtrees rooted at interior nodes capture the clusters. In this paper, we study active learning of a hierarchical clustering using only ordinal queries. An ordinal query consists of a set of three elements, and the response to a query reveals the two elements (among the three elements in the query) which are "closer' to each other than to the third one. We say elements x and y are closer to each other than z if there exists a cluster containing x and y, but not z. Our main result is a deterministic algorithm that learns the underlying hierarchical clustering using at most $n \log_2 n$ adaptive ordinal queries. We generalize our algorithm to be robust in a model in which each query response is correct independently with probability $p > 1/2$, and adversarially incorrect with probability $1 - p$. We show that in the presence of noise, our algorithm outputs the correct hierarchical clustering with probability at least $1-\delta$, using $O(n(\log n + \log(1/\delta)))$ adaptive ordinal queries. For our results, adaptivity is crucial: we prove that even in the absence of noise, every non-adaptive algorithm requires $\Omega(n^3)$ ordinal queries in the worst case.

Ehsan Emamjomeh-Zadeh, David Kempe
University of Southern California
emamjome@usc.edu, dkempe@usc.edu

CP6

Approximation Schemes for Clustering with Outliers

Clustering problems are well-studied in a variety of fields such as data science, operations research, and computer science. Such problems include variants of centre location problems, k-Median, and k-Means to name a few. In some cases, not all data points need to be clustered; some may be discarded for various reasons. For instance, some points may arise from noise in a data set or one might be willing to discard a certain fraction of the points to avoid incurring unnecessary overhead in the cost of a clustering solution. We study some clustering problems with outliers: Uncapaciated Facility Location (UFL) with uniform opening costs, k-Median, and k-Means. Our main focus is when the metric is a doubling metric (including fixed-dimensional Euclidean metrics) or is the shortest path metrics of a graph from a minor-closed family of graphs. For Uncapacitated Facility Location with outliers on such metrics we show that a multiswap simple local search heuristic yields a PTAS. With a bit more work, we extend this to bicriteria approximations for k-Median and k-Means where, for any constant $\epsilon > 0$, we can find a solution using $(1+\epsilon)k$ centres whose cost is at most a $(1 + \epsilon)$-factor of the optimum (while still using at most z outliers). Furthermore, we show how our analysis can be extended to general metrics for k-Means with outliers to obtain a $(25 + \epsilon, 1 + \epsilon)$ bicriteria.

Zachary Friggstad
University of Alberta, Canada
zacharyf@ualberta.ca

Kamyar Khodamoradi, Mohsen Rezapour, Mohammad Salavatipour
University of Alberta
khodamor@ualberta.ca, rezapour@ualberta.ca, mrs@ualberta.ca

CP6

The Bane of Low-Dimensionality Clustering

In this paper, we give a conditional lower bound of $n^{\Omega(k)}$ for the classic k-median and k-means clustering objectives (where n is the size of the input), even in low-dimensional Euclidean space of dimension four, assuming the Exponential Time Hypothesis (ETH). We also consider k-median (and k-means) with penalties where each point need not be assigned to a center, in which case it must pay a penalty, and extend our lower bound to at least three-dimensional Euclidean space. This stands in stark contrast to many other geometric problems such as the traveling salesman problem, or computing an independent set of unit spheres. While these problems benefit from the so-called (limited) blessing of dimensionality, as they can be solved in time $n^{O(k^{1-1/d})}$ or $2^{n^{1-1/d}}$ in d dimensions, our work shows that widely-used clustering objectives have a lower bound of $n^{\Omega(k)}$, even in dimension four. We complete the picture by considering the two-dimensional case: we show that there is no algorithm that solves the penalized version in time less than $n^{o(\sqrt{k})}$, and provide a matching upper bound of $n^{O(\sqrt{k})}$. The main tool we use to establish these lower bounds is the placement of points on the moment curve, which takes its inspiration from constructions of point sets yielding Delaunay complexes of high complexity.

Vincent Cohen-Addad
CNRS, LIP6, Université Pierre et Marie Curie
Paris, France
vincent.cohen.addad@ens-lyon.org

Arnaud de Mesmay
Ecole Normale Superieur
ademesmay@gmail.com

Eva Rotenberg
Department of Applied Mathematics and Computer Science
Technical University of Denmark, Denmark
eva.rotenberg@gmail.com

Alan Roytman
Department of Computer Science
University of Copenhagen, Denmark
alanr@di.ku.dk

CP7

Near-Optimal Compression for the Planar Graph Metric

The Planar Graph Metric Compression Problem is to com-

pactly encode the distances among k nodes in a planar graph of size n. Two naive solutions are to store the graph using $O(n)$ bits, or to explicitly store the distance matrix with $O(k^2 \log n)$ bits. The only lower bounds are from the seminal work of Gavoille, Peleg, Prennes, and Raz [SODA'01], who rule out compressions into a polynomially smaller number of bits, for *weighted* planar graphs, but leave a large gap for unweighted planar graphs. For example, when $k = \sqrt{n}$, the upper bound is $O(n)$ and their constructions imply an $\Omega(n^{3/4})$ lower bound. Our main result is a new compression of the planar graph metric into $\tilde{O}(\min(k^2, \sqrt{kn}))$ bits, which is optimal up to log factors. Our data structure circumvents a lower bound for compression using minors, and the lower bound of Gavoille et al. for weighted planar graphs. This is an unexpected and decisive proof that weights can make planar graphs inherently more complex. Moreover, we design a new Subset Distance Oracle for planar graphs with $\tilde{O}(\sqrt{kn})$ space, and $\tilde{O}(n^{3/4})$ query time. Our work carries strong messages to related fields. In particular, the $O(n^{1/2})$ vs. $\Omega(n^{1/3})$ gap for distance labeling schemes in planar graphs cannot be resolved with current lower bound techniques. On the positive side, we introduce the powerful tool of unit-monge to planar graph algorithms.

Amir Abboud
Stanford University
abboud@cs.stanford.edu

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

Shay Mozes
IDC Herzliya, Israel
smozes@idc.ac.il

Oren Weimann
University of Haifa
oren@cs.haifa.ac.il

CP7

Better Tradeoffs for Exact Distance Oracles in Planar Graphs

We present an $O(n^{1.5})$-space distance oracle for directed planar graphs that answers distance queries in $O(\log n)$ time. Our oracle both significantly simplifies and significantly improves the recent oracle of Cohen-Addad, Dahlgaard and Wulff-Nilsen [FOCS 2017], which uses $O(n^{5/3})$-space and answers queries in $O(\log n)$ time. We achieve this by designing an elegant and efficient point location data structure for Voronoi diagrams on planar graphs. We further show a smooth tradeoff between space and query-time. For any $S \in [n, n^2]$, we show an oracle of size S that answers queries in $\tilde{O}(\max\{1, n^{1.5}/S\})$ time. This new tradeoff is currently the best (up to polylogarithmic factors) for the entire range of S and improves by polynomial factors over all previously known tradeoffs for the range $S \in [n, n^{5/3}]$.

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

Shay Mozes
IDC Herzliya, Israel
smozes@idc.ac.il

Oren Weimann
University of Haifa
oren@cs.haifa.ac.il

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

CP7

A Faster Algorithm for Minimum-Cost Bipartite Perfect Matching in Planar Graphs

Given a weighted planar bipartite graph $G(A \cup B, E)$ where each edge has a positive integer edge cost, we give an $\tilde{O}(n^{4/3} \log nC)$ time algorithm to compute minimum-cost perfect matching; here C is the maximum edge cost in the graph. The previous best-known planarity exploiting algorithm has a running time of $O(n^{3/2} \log n)$ and is achieved by using planar separators (Lipton and Tarjan '78). Our algorithm is based on the bit-scaling paradigm (Gabow and Tarjan '89). For each scale, our algorithm executes $O(n^{1/3})$ iterations of Gabow and Tarjan's algorithm in $O(n^{4/3})$ time. Next, it constructs a compressed residual graph H with $O(n^{2/3})$ vertices and $O(n)$ edges. This is achieved by constructing an r-division with $r = n^{2/3}$. For each partition of the r-division, there is an edge between two vertices of H if and only if they are connected by a directed path inside the partition. Using existing efficient shortest-path data structures, the remaining $O(n^{2/3})$ vertices are matched by iteratively computing a minimum-cost augmenting path each taking $\tilde{O}(n^{2/3})$ time. Augmentation changes the residual graph, so the algorithm updates the compressed representation for each affected partition in $\tilde{O}(n^{2/3})$ time. We bound the total number of affected partitions over all the augmenting paths by $O(n^{2/3} \log n)$. Therefore, the total time taken by the algorithm is $\tilde{O}(n^{4/3})$.

Mudabir Kabir Asathulla
Virginia Tech
mudabir@vt.edu

Sanjeev Khanna
University of Pennsylvania
sanjeev@cis.upenn.edu

Nathaniel Lahn, Sharath Raghvendra
Virginia Tech
lahnn@vt.edu, sharathr@vt.edu

CP7

Voronoi Diagrams on Planar Graphs, and Computing the Diameter in Deterministic $\tilde{O}(n^{5/3})$ Time

We present an explicit and efficient construction of additively weighted Voronoi diagrams on planar graphs. Let G be a planar graph with n vertices and b sites that lie on a constant number of faces. We show how to preprocess G in $\tilde{O}(nb^2)$ time so that one can compute any additively weighted Voronoi diagram for these sites in $\tilde{O}(b)$ time. We use this construction to compute the diameter of a directed planar graph with real arc lengths in $\tilde{O}(n^{5/3})$ time. This improves the recent breakthrough result of Cabello (SODA'17), both by improving the running time (from $\tilde{O}(n^{11/6})$), and by using a deterministic algorithm.

It is in fact the first truly subquadratic *deterministic* algorithm for this problem. Our use of Voronoi diagrams to compute the diameter follows that of Cabello, but he used abstract Voronoi diagrams, which makes his diameter algorithm more involved, more expensive, and randomized. As in Cabello's work, our algorithm can also compute the Wiener index of a planar graph (i.e., the sum of all pairwise distances) within the same bounds. Our construction of Voronoi diagrams for planar graphs is of independent interest. It has already been used to obtain fast exact distance oracles for planar graphs [Cohen-Addad et al. FOCS'17].

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

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

Shay Mozes
IDC Herzliya, Israel
smozes@idc.ac.il

Micha Sharir
Tel Aviv University & New York University
michas@post.tau.ac.il

Oren Weimann
University of Haifa
oren@cs.haifa.ac.il

CP7

Minimum Cut of Directed Planar Graphs in $O(n \log\log n)$ Time

We give an $O(n \log\log n)$ time algorithm for computing the minimum cut (or equivalently, the shortest cycle) of a weighted directed planar graph. This improves the previous fastest $O(n \log^3 n)$ solution. Interestingly, while in undirected planar graphs both min cut and min *st*-cut have $O(n \log\log n)$ solutions, in directed planar graphs our result makes min cut faster than min *st*-cut, which currently requires $O(n \log n)$.

Shay Mozes
IDC Herzliya, Israel
smozes@idc.ac.il

Cyril Nikolaev
University of Haifa, Israel
cyril.nikolaev@gmail.com

Yahav Nussbaum
Tel-Aviv University
yahav.nussbaum@cs.tau.ac.il

Oren Weimann
University of Haifa
oren@cs.haifa.ac.il

CP8

Discrete Choice, Permutations, and Reconstruction

In this paper we study the well-known family of Random Utility Models, developed over 50 years ago to codify rational user behavior in choosing one item from a finite set of options. In this setting each user draws i.i.d. from some distribution a utility function mapping each item in the universe to a real-valued utility. The user is then offered a subset of the items, and selects the one of maximum utility. A max-distribution oracle for this choice model takes any subset of items and returns the probability (over the distribution of utility functions) that each will be selected. A discrete choice algorithm, given access to a max-distribution oracle, must return a function that approximates the oracle. We show three primary results. First, we show that any algorithm exactly reproducing the oracle must make exponentially many queries. Second, we show an equivalent representation of the distribution over utility functions, based on permutations, and show that if this distribution has support size k, then it is possible to approximate the oracle using $O(nk)$ queries. Finally, we consider settings in which the subset of items is always small. We give an algorithm that makes less than $n^{(1-\epsilon/2)K}$ queries, each to sets of size at most $(1-\epsilon/2)K$, in order to approximate the max-distribution oracle on every set of size $|T| \leq K$ with statistical error at most ϵ.

Flavio Chierichetti
Sapienza University
Dipartimento di Informatica
flavio@chierichetti.name

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

Andrew Tomkins
Google
atomkins@gmail.com

CP8

Nearly Tight Bounds for Sandpile Transience on the Grid

We use techniques from the theory of electrical networks to give nearly tight bounds for the transience class of the Abelian sandpile model on the two-dimensional grid up to polylogarithmic factors. The Abelian sandpile model is a discrete process on graphs that is intimately related to the phenomenon of self-organized criticality. In this process, vertices receive grains of sand, and once the number of grains exceeds their degree, they topple by sending grains to their neighbors. The transience class of a model is the maximum number of grains that can be added to the system before it necessarily reaches its steady-state behavior or, equivalently, a recurrent state. Through a more refined and global analysis of electrical potentials and random walks, we give an $O(n^4 \log^4 n)$ upper bound and an $\Omega(n^4)$ lower bound for the transience class of the $n \times n$ grid. Our methods naturally extend to n^d-sized d-dimensional grids to give $O(n^{3d-2} \log^{d+2} n)$ upper bounds and $\Omega(n^{3d-2})$ lower bounds.

David Durfee, Matthew Fahrbach, Yu Gao
Georgia Institute of Technology
ddurfee@gatech.edu, matthew.fahrbach@gatech.edu,
ygao380@gatech.edu

Tao Xiao
Shanghai Jiao Tong University

xt_1992@sjtu.edu.cn

CP8

Time and Space Efficient Representations of Distributive Lattices

We present a space-efficient data structure using $O(n \log n)$ bits that represents a distributive lattice on n elements and supports finding meets and joins in $O(\log n)$ time. Our data structure extends the ideal tree structure of Habib and Nourine which occupies $O(n \log n)$ bits of space and requires $O(m)$ time to compute a meet or join, where m depends on the specific lattice and may be as large as $n-1$. We also give an encoding of a distributive lattice using $\frac{10}{7}n + O(\log n)$ bits, which is very close to the information theoretic lower bound. This encoding can be created or decompressed in $O(n \log n)$ time.

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

CP8

Consistent Hashing with Bounded Loads

In dynamic load balancing, we wish to allocate a set of clients (balls) to a set of servers (bins) with the goal of minimizing the maximum load of any server and also minimizing the number of moves after adding or removing a server or a client. We want a hashing-style solution where we given the ID of a client can efficiently find its server in a distributed dynamic environment. In such a dynamic environment, both servers and clients may be added and/or removed from the system in any order. In particular, we consider a problem with balls and bins, and given a user-specified balancing parameter $c = 1 + \epsilon > 1$, we aim to find a hashing scheme with no load above $c/$, referred to as the *capacity* of the bins. We show that in our hashing scheme when a ball or bin is inserted or deleted, the expected number of balls that have to be moved is within a multiplicative factor of $O(\frac{1}{\epsilon^2})$ of the optimum for $\epsilon \leq 1$ (Theorem **??**) and within a factor $1 + O(\frac{\log c}{c})$ of the optimum for $\epsilon \geq 1$ (Theorem **??**). Technically, the latter bound is the most challenging to prove. It implies that we for superconstant c, we only pay a negligible cost in extra moves. We also get the same bounds for the simpler problem where we instead of a user specified balancing parameter have a fixed bin capacity C for all bins, and define $c = 1 + \epsilon = C/$.

Vahab Mirrokni
Google Inc.
mirrokni@google.com

Mikkel Thorup
University of Copenhagen
mikkel2thorup@gmail.com

Morteza Zadimoghaddam
Google
zadim@google.com

CP8

A Hamilton Path for the Sigma-Tau Problem

Nijenhuis and Wilf asked the following question in their *Combinatorial Algorithms* textbook from 1975: Can the permutations of $\{1, 2, \ldots, n\}$ be ordered so that each permutation is transformed into the next by applying either the operation σ, a rotation to the left, or τ, a transposition of the first two symbols? Knuth rated the challenge of finding a cyclic solution for odd n (cycles do not exist for even $n > 2$) at 48/50 in *The Art of Computer Programming*, which makes it Volume 4's hardest open problem since the 'middle levels' problem was solved by Mütze. In this paper we solve the 40 year-old question by Nijenhuis and Wilf, by providing a simple successor rule to generate each successive permutation. We also present insights into how our solution can be modified to find a Hamilton cycle for odd n.

Joe Sawada
University of Guelph
School of Computer Science
jsawada@uoguelph.ca

Aaron Williams
Simon's Rock
awilliams@simons-rock.edu

CP9

The Gotsman-Linial Conjecture is False

In 1991, Craig Gotsman and Nathan Linial conjectured that for all n and d, the average sensitivity of a degree-d polynomial threshold function on n variables is maximized by the degree-d symmetric polynomial which computes the parity function on the d layers of the hypercube with Hamming weight closest to $n/2$. We refute the conjecture for almost all d and for almost all n, and we confirm the conjecture in many of the remaining cases.

Brynmor Chapman
MIT
brynmor@mit.edu

CP9

Approximate Local Decoding of Cubic Reed-Muller Codes Beyond the List Decoding Radius

We consider the question of decoding Reed-Muller codes over $\mathbf{F}_2^n$ beyond their list-decoding radius. Since, by definition, in this regime one cannot demand a list-decoder efficient in the message length, we seek an approximate decoder: Given a word F and radii $r' > r > 0$, the goal is to output a codeword within radius r' of F, if there exists a codeword within distance r. As opposed to the list decoding problem, it suffices here to output any codeword with this property, since the list may be too large if r exceeds the list decoding radius. Prior to our work, such decoders were known for Reed-Muller codes of degree 2, due to works of Wolf and the second author [FOCS 2011]. In this work we make the first progress on this problem for the degree 3 where the list decoding radius is 1/8. We show that there is a constant $\delta = 1/2 - \sqrt{1/8} > 1/8$ and an efficient approximate decoder, that given query access to a function $F : \mathbf{F}_2^n \to \mathbf{F}_2$, such that F is within distance $r = \delta - \epsilon$ from a cubic polynomial, runs in time polynomial in message length and outputs with high probability a cubic polynomial which is at distance at most $r' = 1/2 - \epsilon'$ from F, where ϵ' is a quasi polynomial function of ϵ.

Pooya Hatami
UT Austin
pooyahat@gmail.com

Madhur Tulsiani

TTI Chicago
madhurt@ttic.edu

CP9

Coding Against Deletions in Oblivious and Online Models

We consider binary error correcting codes when errors are deletions. A basic challenge concerning deletion codes is determining $p_0^{(adv)}$, the zero-rate threshold of adversarial deletions, defined to be the supremum of all p for which there exists a code family with rate bounded away from 0 capable of correcting a fraction p of adversarial deletions. A recent construction shows $p_0^{(adv)} \geq \sqrt{2}-1$, and the trivial upper bound, $p_0^{(adv)} \leq \frac{1}{2}$, is the best known. Perhaps surprisingly, we do not know whether $p_0^{(adv)} = 1/2$. In this work, we explore two related error models: oblivious deletions and online deletions, which are in between random and adversarial deletions in power. In the oblivious model, the channel can inflict an arbitrary pattern of pn deletions, picked without knowledge of the codeword. We prove the existence of binary codes of positive rate that can correct any fraction $p < 1$ of oblivious deletions, establishing that the associated zero-rate threshold $p_0^{(obliv)}$ equals 1. For online deletions, where the channel decides whether to delete bit x_i based only on knowledge of bits $x_1 x_2 \ldots x_i$, define the deterministic zero-rate threshold for online deletions $p_0^{(on,d)}$ to be the supremum of p for which there exist deterministic codes against an online channel causing pn deletions with low average probability of error. We prove $p_0^{(adv)} = \frac{1}{2}$ if and only if $p_0^{(on,d)} = \frac{1}{2}$.

Venkatesan Guruswami
Carnegie Mellon University
venkatg@cs.cmu.edu

Ray Li
Stanford University
rayyli@stanford.edu

CP9

Syndrome Decoding of Reed-Muller Codes and Tensor Decomposition over Finite Fields

Reed-Muller codes are some of the oldest and most widely studied error-correcting codes, of interest for both their algebraic structure as well as their many algorithmic properties. A recent beautiful result of Saptharishi, Shpilka and Volk showed that for binary Reed-Muller codes of length n and distance $d = O(1)$, one can correct polylog(n) random errors in poly(n) time. In this paper, we consider the problem of deciding Reed-Muller codes from the polylog(n)-bit long syndrome vector of a codeword corrupted in polylog(n) random coordinates. This problem turns out to be equivalent to a basic question about computing tensor decomposition of random low-rank tensors over finite fields. Our main result is that syndrome decoding of Reed-Muller codes (and the equivalent tensor decomposition problem) can be solved efficiently, i.e., in polylog(n) time. We give two algorithms for this problem:

1. The first algorithm is a finite field variant of a classical algorithm for tensor decomposition over real numbers due to Jennrich. This also gives an alternate proof for the main result by Saptharshi et al.
2. The second algorithm is obtained by implementing the steps of the Berlekamp-Welch-style decoding algorithm of Saptharshi et al. in sublinear-time. A new ingredient is an algorithm for solving certain kinds of systems of polynomial equations.

Aditya Potukuchi, Swastik Kopparty
Rutgers University
apotu.57@gmail.com, swastik.kopparty@gmail.com

CP9

Average-Radius List-Recoverability of Random Linear Codes

We analyze the list-decodability, and related notions, of random linear codes. This has been studied extensively before: there are many different parameter regimes and many different variants. Previous works have used complementary styles of arguments—which each work in their own parameter regimes but not in others—and moreover have left some gaps in our understanding of the list-decodability of random linear codes. In particular, none of these arguments work well for list-recovery, a useful generalization of list-decoding. In this work, we present a new approach, which works across parameter regimes and further generalizes to list-recovery. This argument unifies the landscape of this problem, and can establish the following new results: (*) Random linear codes over large fields are list-recoverable and list-decodable up to near-optimal rates (within a multiplicative factor of 0.99), with list sizes that depend quasi-polynomially on the gap-to-capacity. (*) First quasipolynomial list sizes for high-rate list-recovery of random linear codes. (*) Our results exponentially improve the list size bounds in the best known results on linear-time list-recoverable codes and the near-linear time list decodable codes (with optimal rate).

Atri Rudra
Univ. of Washington
atri@buffalo.edu

Mary Wootters
Stanford University
marykw@stanford.edu

CP10

Prophet Secretary for Combinatorial Auctions and Matroids

The secretary and the prophet inequality problems are central to the field of Stopping Theory. Recently, there has been a lot of work in generalizing these models to multiple items because of their applications in mechanism design. The most important of these generalizations are to matroids and to combinatorial auctions (extends bipartite matching). Kleinberg-Weinberg [KW12] and Feldman et al. [FGL15] show that for adversarial arrival order of random variables the optimal prophet inequalities give a 1/2-approximation. For many settings, however, it's conceivable that the arrival order is chosen uniformly at random, akin to the secretary problem. For such a random arrival model, we improve upon the 1/2-approximation and obtain (1-1/e)- approximation prophet inequalities for both matroids and combinatorial auctions. This also gives improvements to the results of Yan [Yan11] and Esfandiari et al. [EHLM17] who worked in the special cases where we can fully control the arrival order or when there is only a single item. Our techniques are threshold based. We convert our discrete problem into a continuous setting and then give a generic template on how to dynamically adjust these thresholds to lower bound the expected total welfare.

Soheil Ehsani
University of Maryland at College Park
ehsani@umd.edu

MohammadTaghi Hajiaghayi
University of Maryland
hajiaghayi@gmail.com

Thomas Kesselheim
TU Dortmund
thomas.kesselheim@tu-dortmund.de

Sahil Singla
CMU
ssingla@cmu.edu

CP10

A Framework for the Secretary Problem on the Intersection of Matroids

The secretary problem became one of the most prominent online selection problems due to its numerous applications in online mechanism design. The task is to select a maximum weight subset of elements subject to given constraints, where elements arrive one-by-one in random order, revealing a weight upon arrival. The decision whether to select an element has to be taken immediately after its arrival. The different applications that map to the secretary problem ask for different constraint families to be handled. The most prominent ones are matroid constraints, which both capture many relevant settings and admit strongly competitive secretary algorithms. However, dealing with more involved constraints proved to be much more difficult, and strong algorithms are known only for a few specific settings. In this paper, we present a general framework for dealing with the secretary problem over the intersection of several matroids. This framework allows us to combine and exploit the large set of matroid secretary algorithms known in the literature. As one consequence, we get constant-competitive secretary algorithms over the intersection of any constant number of matroids whose corresponding (single-)matroid secretary problems are currently known to have a constant-competitive algorithm. Moreover, we show that our results extend to submodular objectives.

Moran Feldman
Open University of Israel
moranfe@openu.ac.il

Ola Svensson
EPFL
ola.svensson@epfl.ch

Rico Zenklusen
ETH Zurich
ricoz@math.ethz.ch

CP10

Truthful Multi-Parameter Auctions with Online Supply: An Impossible Combination

We study a basic auction design problem with online supply. There are two unit-demand bidders and two types of items. The first item type will arrive first for sure, and the second item type may or may not arrive. The auctioneer has to decide the allocation of an item immediately after each item arrives, but is allowed to compute payments after knowing how many items arrived. For this problem we show that there is no deterministic truthful and individually rational mechanism that, even with unbounded computational resources, gets any finite approximation factor to the optimal social welfare.

Vasilis Syrgkanis
Microsoft Research
vasy@microsoft.com

Nikhil R. Devanur
Microsoft Research, Redmond
nikdev@microsoft.com

Balasubramanian Sivan
Google Research
balusivan@google.com

CP10

Strong Algorithms for the Ordinal Matroid Secretary Problem

In the ordinal Matroid Secretary Problem (MSP), elements from a weighted matroid are presented in random order to an algorithm that must incrementallty select a large weight independent set. However, the algorithm can only compare pairs of revealed elements without using its numerical value. An algorithm is α probability-competitive (PC) if every element from the optimum appears with probability $1/\alpha$ in the output.
We present a technique to design algorithms with strong PC ratios, improving the guarantees for almost every matroid class considered in the literature: e.g., we get PC ratios of 4 for graphic matroids (improving on $2e$ by Korula and Pál [ICALP 2009]) and of 5.19 for laminar matroids (improving on 9.6 by Ma et al. [THEOR COMPUT SYST 2016]). We also obtain new results for superclasses of k column sparse matroids, for hypergraphic matroids, certain gammoids and graph packing matroids, and a $1 + O(\sqrt{\log \rho/\rho})$ PC algorithm for uniform matroids of rank ρ based on Kleinberg's $1 + O(\sqrt{1/\rho})$ utility-competitive algorithm [SODA 2005] for that class.
Our second contribution are algorithms for the ordinal MSP on arbitrary matroids of rank ρ. We devise an $O(\log \rho)$ PC algorithm and an $O(\log \log \rho)$ ordinal-competitive algorithm, a weaker notion of competitiveness but stronger than the utility variant. These are based on the $O(\log \log \rho)$ utility-competitive algorithm by Feldman et al. [SODA 2015].

Jose A. Soto
Universidad de Chile
jsoto@dim.uchile.cl

Abner Turkieltaub
Universidad de chile
abnerturkieltaub@gmail.com

Victor Verdugo
Universidad de Chile &
École normale supérieure,
vverdugo@dii.uchile.cl

CP10

Variance Reduced Value Iteration and Faster Al-

gorithms for Solving Markov Decision Processes

In this paper we provide faster algorithms for approximately solving discounted Markov Decision Processes in multiple parameter regimes. Given a discounted Markov Decision Process (DMDP) with $|S|$ states, $|A|$ actions, discount factor $\gamma \in (0,1)$, and rewards in the range $[-M, M]$, we show how to compute an ϵ-optimal policy, with probability $1-\delta$ in time

$$\widetilde{O}\left(\left(|S|^2|A| + \frac{|S||A|}{(1-\gamma)^3}\right)\log\left(\frac{M}{\epsilon}\right)\log\left(\frac{1}{\delta}\right)\right) .$$

This contribution reflects the first nearly linear time, nearly linearly convergent algorithm for solving DMDP's for intermediate values of γ. We also show how to obtain improved sublinear time algorithms and provide an algorithm which computes an ϵ-optimal policy with probability $1-\delta$ in time

$$\widetilde{O}\left(\frac{|S||A|M^2}{(1-\gamma)^4\epsilon^2}\log\left(\frac{1}{\delta}\right)\right)$$

provided we can sample from the transition function in $O(1)$ time. We obtain our results by combining approximate value iteration with new techniques in variance reduction. Our fastest algorithms leverage further insights to ensure that our algorithms make monotonic progress towards the optimal value. This paper is one of few instances in using sampling to obtain a linearly convergent linear programming algorithm and we hope that the analysis may be useful more broadly.

Xian Wu, Aaron Sidford
Stanford University
xwu20@stanford.edu, sidford@stanford.edu

Mengdi Wang
Department of Operations Research and Financial Engineering
Princeton University
mengdiw@princeton.edu

Yinyu Ye
Stanford University
yyye@stanford.edu

CP11

Lifting Linear Extension Complexity Bounds to the Mixed-Integer Setting

Mixed-integer mathematical programs are ubiquitous in Operations Research and related fields. However, there is still very little known about what can be expressed by *small* mixed-integer programs. In particular, prior to this work, it was open whether some classical problems, like the minimum odd-cut problem, can be expressed by a compact mixed-integer program with few (even constantly many) integer variables. We provide a general framework for lifting inapproximability results of extended formulations to the setting of mixed-integer extended formulations, and obtain almost tight lower bounds on the number of integer variables needed to describe a variety of classical combinatorial optimization problems. Among the implications we obtain, we show that any mixed-integer extended formulation of sub-exponential size for the matching polytope, cut polytope, or dominant of the odd-cut polytope, needs $\Omega(n/\log n)$ many integer variables, where n is the number of vertices of the underlying graph. Conversely, the above-mentioned polyhedra admit polynomial-size mixed-integer formulations with only $O(n)$ or $O(n \log n)$ (for the traveling salesman polytope) many integer variables. Our results build upon a new decomposition technique that, for any convex set C, allows for approximating any mixed-integer description of C by the intersection of C with the union of a small number of affine subspaces.

Alfonso Cevallos, Stefan Weltge, Rico Zenklusen
ETH Zurich
alfonso.cevallos@ifor.math.ethz.ch, stefan.weltge@ifor.math.ethz.ch, ricoz@math.ethz.ch

CP11

Proximity Results and Faster Algorithms for Integer Programming Using the Steinitz Lemma

We consider integer programming problems in standard form $\max\{c^T x : Ax = b,\ x \geq 0,\ x \in Z^n\}$ where $A \in Z^{m \times n}$, $b \in Z^m$ and $c \in Z^n$. We show that such an integer program can be solved in time $m^{O(m)} \cdot \Delta^{O(m)}$, where Δ is an upper bound on each absolute value of an entry in A and b. This improves upon the longstanding best bound of Papadimitriou (1981) of $m^{O(m^2)} \cdot \Delta^{O(m^2)}$ and addresses an open problem raised by Fomin. Our result relies on a lemma of Steinitz that states that a set of vectors in R^m that is contained in the unit ball and that sum up to zero can be ordered such that all partial sums are of norm bounded by m. We also use the Steinitz lemma to show that the ℓ_1-distance of an optimal integer and fractional solution of the integer program, also under the presence of upper bounds on the variables, is bounded by $m \cdot (2 \cdot m \cdot \Delta_A + 1)^m$. Here Δ_A is an upper bound on the absolute values of the entries of A only. The novel strength of our bound is that it is independent of n. We provide evidence for the significance of our bound by applying it to general knapsack problems where we obtain structural and algorithmic results that improve upon the recent literature.

Friedrich Eisenbrand
Ecole Polytechnique Federale de Lausanne
friedrich.eisenbrand@epfl.ch

Robert Weismantel
ETH Zuerich
robert.weismantel@ifor.math.ethz.ch

CP11

Approximating Weighted Tree Augmentation Via Chvátal-Gomory Cuts

In the weighted tree augmentation problem (WTAP), we are given an undirected tree $G = (V, E)$ and an additional set of edges L called links, with costs $c \geq 1$ for all links. The goal is to choose a minimum cost subset $S \subseteq L$ such that $G = (V, E \cup S)$ is 2-edge-connected. In the unweighted case, with $c = 1$ for all links, the problem is called the tree augmentation problem (TAP). Both problems are known to be APX-hard, and the best known approximation factors are 2 for WTAP [Frederickson and JáJá] and 1.5 for TAP [Kortsarz and Nutov]. Adjiashvili (SODA '17) gave a $1.96418 + \epsilon$-approximation algorithm for WTAP under the assumption that all link costs are bounded by a constant. This is the first approximation with a better guarantee than 2 not requiring a special structure of the tree or the links. In this paper, we improve Adjiashvili's approximation to a $1.5 + \epsilon$-approximation for WTAP under the bounded cost assumption. We achieve this by introducing a strong LP that combines $\{0, 0.5\}$-Chvátal-Gomory cuts for the standard LP for the problem with bundle constraints from Adjiashvili. We show that our LP can be solved efficiently and that it is exact for some instances that arise

at the core of Adjiashvili's approach. This results in an improved performance guarantee of $1.5 + \epsilon$.

Martin Gro
University of Waterloo
martin.gross@mailbox.org

Samuel Fiorini
Université Libre de Bruxelles
sfiorini@ulb.ac.be

Jochen Könemann, Laura Sanità
University of Waterloo
jochen@uwaterloo.ca, laura.sanita@uwaterloo.ca

CP11

Submodular Minimization Under Congruency Constraints

Submodular function minimization (SFM) is a fundamental and efficiently solvable problem class in combinatorial optimization with a multitude of applications in various fields. Surprisingly, there is only very little known about constraint types under which SFM remains efficiently solvable. The arguably most relevant non-trivial constraint class for which polynomial SFM algorithms are known are parity constraints. Parity constraints capture classical combinatorial optimization problems like the odd-cut problem, and they are a key tool in a recent technique to efficiently solve integer programs with a constraint matrix whose subdeterminants are bounded by two in absolute value. We show that efficient SFM is possible even for a significantly larger class than parity constraints, by introducing a new approach that combines techniques from Combinatorial Optimization, Combinatorics, and Number Theory. In particular, we can show that efficient SFM is possible over all sets of cardinality r mod m, as long as m is a constant prime power. This covers generalizations of the odd-cut problem with open complexity status, and with relevance in the context of integer programming with higher subdeterminants. Moreover, our results settle two open questions raised by Geelen and Kapadia [Combinatorica, 2017] in the context of computing the girth and cogirth of certain types of binary matroids.

Martin Nägele, Benny Sudakov, Rico Zenklusen
ETH Zurich
martin.naegele@ifor.math.ethz.ch, benjamin.sudakov@math.ethz.ch, ricoz@math.ethz.ch

CP11

Geometric Rescaling Algorithms for Submodular Function Minimization

We present a new class of polynomial-time algorithms for submodular function minimization (SFM), as well as a unified framework to obtain strongly polynomial SFM algorithms. Our new algorithms are based on simple iterative methods for the minimum-norm problem. We exhibit two techniques to turn simple iterative methods into polynomial-time algorithms. Firstly, we use the geometric rescaling technique, which has recently gained attention in linear programming. We adapt this technique to SFM and obtain a weakly polynomial bound $O((n^4 EO + n^5)\log(nL))$. Secondly, we exhibit a general combinatorial black-box approach to turn any strongly polynomial εL-approximate submodular function minimization oracle into a strongly polynomial algorithm. This framework can be applied to a wide range of combinatorial and continuous algorithms, including pseudo-polynomial ones. In particular, we can obtain strongly polynomial algorithms by a repeated application of the conditional gradient or of the Fujishige-Wolfe algorithm. Combined with the geometric rescaling technique, the black-box approach provides a $O((n^5 EO + n^6)\log^2 n)$ algorithm. Finally, we show that our technique can be used to obtain a simplifed variant of the $O(n^3 \log^2 nEO + n^4 \log^{O(1)} n)$ cutting-plane SFM algorithm of Lee, Sidford, and Wong.

Daniel Dadush
New York University
dadush@cwi.nl

Laszlo A. Vegh, Giacomo Zambelli
London School of Economics
l.vegh@lse.ac.uk, g.zambelli@lse.ac.uk

CP12

More Logarithmic-Factor Speedups for 3SUM, (median,+)-Convolution, and Some Geometric 3SUM-Hard Problems

We present an algorithm that solves the 3SUM problem for n real numbers in $O((n^2/\log^2 n)(\log\log n)^{O(1)})$ time, improving previous solutions by about a logarithmic factor. Our framework for shaving off two logarithmic factors can be applied to other problems, such as (median,+)-convolution/matrix multiplication and algebraic generalizations of 3SUM. We also obtain the first subquadratic results on some 3SUM-hard problems in computational geometry, for example, deciding whether (the interiors of) a constant number of simple polygons have a common intersection.

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

CP12

Voronoi Tessellations in the Crt and Continuum Random Maps of Finite Excess

Given a large graph G and k agents on this graph, we consider the Voronoi tessellation induced by the graph distance. Each agent gets control of the portion of the graph that is closer to itself than to any other agent. We study the limit law of the vector $Vor := (V_1/n, V_2/n, ..., V_k/n)$, whose i'th coordinate records the fraction of vertices of G controlled by the i'th agent, as n tends to infinity. We show that if G is a uniform random tree, and the agents are placed uniformly at random, the limit law of vor is uniform on the $(k-1)$-dimensional simplex. In particular, when $k = 2$, the two agents each get a uniform random fraction of the territory. In fact, we prove the result directly on the Brownian continuum random tree (CRT), and we also prove the same result for a "higher genus" analogue of the CRT that we call the continuum random unicellular map, indexed by a genus parameter $g \geq 0$. As a key step of independent interest, we study the case when G is a random planar embedded graph with a finite number of faces. The main idea of the proof is to show that vor has the same distribution as another partition of mass $Int := (I_1/n, I_2/n, ..., I_k/n)$ where I_j is the contour length separating the i-th agent from the first one following it clockwise around the graph.

Louigi Addario-Berry
McGill University

louigi.addario@mcgill.ca

Omer Angel
University of British Columbia
angel@math.ubc.ca

Guillaume Chapuy
CNRS, IRIF, Universite Paris Diderot
France
guillaume.chapuy@liafa.univ-paris-diderot.fr

Christina Goldschmidt
University of Oxford
goldschm@stats.ox.ac.uk

Éric Fusy
CNRS, LIX, École Polytechnique
fusy@lix.polytechnique.fr

CP12

On the Complexity of Range Searching Among Curves

Modern tracking technology has made the collection of large numbers of densely sampled trajectories of moving objects widely available. We consider a fundamental problem encountered when analysing such data: Given n polygonal curves S in $\mathbb{R}^d$, preprocess S into a data structure that answers queries with a query curve q and radius ρ for the curves of S that have Fréchet distance at most ρ to q. We initiate a comprehensive analysis of the space/query-time trade-off for this data structuring problem. Our lower bounds imply that any data structure in the pointer model model that achieves $Q(n) + O(k)$ query time, where k is the output size, has to use roughly $\Omega\left((n/Q(n))^2\right)$ space in the worst case, even if queries are mere points or line segments. More importantly, we show that the space/query-time trade-off worsens by an *exponential* factor of input and query complexity. This behaviour addresses an open question in the range searching literature: whether it is possible to avoid the additional logarithmic factors in the space and query time of a multilevel partition tree. We answer this question negatively. On the positive side, we show we can build data structures for the Fréchet distance by using semialgebraic range searching. Here, our solution is in line with our lower bound in terms of the asymptotic number of levels in the data structure.

Peyman Afshani
MADALGO, Aarhus University
peyman@madalgo.au.dk

Anne Driemel
TU Eindhoven, The Netherlands
Department of Mathematics and Computer Science
a.driemel@tue.nl

CP12

On Separating Points by Lines

Given a set of n points in the plane, its *separability* is the minimum number of lines needed to separate all its pairs of points from each other. We show that the minimum number of lines needed to separate n points, picked randomly (and uniformly) in the unit square, is $\widetilde{\Theta}(n^{2/3})$, where $\widetilde{\Theta}$ hides polylogarithmic factors. In addition, we provide a fast approximation algorithm for computing the separability of a given point set in the plane.

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

Mitchell F. Jones
University of Illinois at Urbana-Champaign
mfjones2@illinois.edu

CP12

The Entropy of Backwards Analysis

Backwards analysis, first popularized by Seidel, is often the simplest most elegant way of analyzing a randomized algorithm. It applies to incremental algorithms where elements are added incrementally, following some random permutation, e.g., incremental Delauney triangulation of a pointset, where points are added one by one, and where we always maintain the Delauney triangulation of the points added thus far. For backwards analysis, we think of the permutation as generated backwards, implying that the ith point in the permutation is picked uniformly at random from the i points not picked yet in the backwards direction. Backwards analysis has also been applied elegantly by Chan to the randomized linear time minimum spanning tree algorithm of Karger, Klein, and Tarjan. The question considered in this paper is how much randomness we need in order to trust the expected bounds obtained using backwards analysis, exactly and approximately. For the exact case, it turns out that a random permutation works if and only if it is minwise, that is, for any given subset, each element has the same chance of being first.Minwise permutations are known to have $\Theta(n)$ entropy, and this is then also what we need for exact backwards analysis. However, when it comes to approximation, the two concepts diverge dramatically. To get backwards analysis to hold within a factor α, the random permutation needs entropy $\Omega(n/\alpha)$. This contra ... Cut off due to character limit

Mathias Knudsen, Mikkel Thorup
University of Copenhagen
mathias@tejs.dk, mikkel2thorup@gmail.com

CP13

Online Bipartite Matching with Amortized $O(\log^2 N)$ Replacements

In the online bipartite matching problem with replacements, all the vertices on one side of the bipartition are given, and the vertices on the other side arrive one by one with all their incident edges. The goal is to maintain a maximum matching while minimizing the number of changes (replacements) to the matching. We show that the greedy algorithm that always takes a shortest augmenting path from the newly inserted vertex (denoted SAP) uses at most amortized $O(\log^2 n)$ replacements per insertion, where n is the total number of vertices inserted. This is the first analysis to achieve a polylogarithmic number of replacements for *any* strategy, almost matching the $\Omega(\log n)$ lower bound. The previous best strategy achieved amortized $O(\sqrt{n})$ replacements [Bosek, Leniowski, Sankowski, Zych, FOCS 2014]. For SAP in particular, nothing better than the trivial $O(n)$ bound was known except in special cases. Our analysis immediately implies the same upper bound of $O(\log^2 n)$ reassignments for the capacitated assignment problem where each vertex on the static side of the bipartition is initialized with the capacity to serve a

number of vertices. We also analyze the problem of minimizing the maximum server load. We show that if the final graph has maximum server load L, then SAP makes amortized $O(\min\{L \log^2 n, \sqrt{n} \log n\})$ reassignments. We also show that this is close to tight because $\Omega(\min\{L, \sqrt{n}\})$ reassignments can be necessary.

Aaron Bernstein
Technical University of Berlin, Germany
bernstei@gmail.com

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

Eva Rotenberg
University of Copenhagen
eva@rotenberg.dk

CP13

Randomized Algorithms for Online Vector Load Balancing

We study randomized algorithms for the online vector bin packing and vector scheduling problems. For vector bin packing, we achieve a competitive ratio of $\tilde{O}(d^{1/B})$, where d is the number of dimensions and B the size of a bin. This improves the previous bound of $\tilde{O}(d^{1/(B-1)})$ by a polynomial factor, and is tight up to logarithmic factors. For vector scheduling, we show a lower bound of $\Omega(\frac{\log d}{\log \log d})$ on the competitive ratio of randomized algorithms, which is the first result for randomized algorithms and is asymptotically tight. Finally, we analyze the widely used "power of two choices" algorithm for vector scheduling, and show that its competitive ratio is $O(\log \log n + \frac{\log d}{\log \log d})$, which is optimal up to the additive $O(\log \log n)$ term that also appears in the scalar version of this algorithm.

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

Ilan R. Cohen
Tel Aviv University
Tel Aviv University
ilanrcohen@gmail.com

Debmalya Panigrahi
Duke University
debmalya@cs.duke.edu

CP13

Competitive Algorithms for Generalized k-Server in Uniform Metrics

The generalized k-server problem is a far-reaching extension of the k-server problem with several applications. Here, each server s_i lies in its own metric space M_i. A request is a k-tuple $r = (r_1, r_2, \ldots, r_k)$ and to serve it, we need to move some server s_i to the point $r_i \in M_i$, and the goal is to minimize the total distance traveled by the servers. Despite much work, no $f(k)$-competitive algorithm is known for the problem for $k > 2$ servers, even for special cases such as uniform metrics and lines. Here, we consider the problem in uniform metrics and give the first $f(k)$-competitive algorithms for general k. In particular, we obtain deterministic and randomized algorithms with competitive ratio $O(k2^k)$ and $O(k^3 \log k)$ respectively. Our deterministic bound is based on a novel application of the polynomial method to online algorithms, and essentially matches the long-known lower bound of $2^k - 1$. We also give a $2^{2^{O(k)}}$-competitive deterministic algorithm for weighted uniform metrics, which also essentially matches the recent doubly exponential lower bound for the problem.

Nikhil Bansal
Department of Mathematics and Computer Science
Eindhoven University of Technology, Eindhoven,
Netherlands
n.bansal@tue.nl

Marek Elias
Eindhoven University of Technology, Netherlands
m.elias@tue.nl

Grigorios Koumoutsos, Jesper Nederlof
TU Eindhoven
g.koumoutsos@tue.nl, j.nederlof@tue.nl

CP13

Online Facility Location Against a t-Bounded Adversary

In the streaming model, the order of the stream can significantly affect the difficulty of a problem. A t-semirandom stream was introduced as an interpolation between random-order ($t = 1$) and adversarial-order ($t = n$) streams where an adversary intercepts a random-order stream and can delay up to t elements at a time. IITK Sublinear Open Problem #15 asks to find algorithms whose performance degrades smoothly as t increases. We show that the celebrated online facility location algorithm achieves an expected competitive ratio of $O(\frac{\log t}{\log \log t})$. We present a matching lower bound that any randomized algorithm has an expected competitive ratio of $\Omega(\frac{\log t}{\log \log t})$. We use this result to construct an $O(1)$-approximate streaming algorithm for k-median clustering that stores $O(k \log t)$ points and has $O(k \log t)$ worst-case update time. Our technique generalizes to any dissimilarity measure that satisfies a weak triangle inequality, including k-means, M-estimators, and ℓ_p norms. The special case $t = 1$ yields an optimal $O(k)$ space algorithm for random-order streams as well as an optimal $O(nk)$ time algorithm in the RAM model, closing a long line of research on this problem.

Harry Lang
Inria Saclay
Johns Hopkins University
hlang08@gmail.com

CP13

Randomized Online Matching in Regular Graphs

In this paper we study the classic online matching problem, introduced in the seminal work of Karp, Vazirani and Vazirani (STOC 1990), in *regular* graphs. For such graphs, an optimal deterministic algorithm as well as efficient algorithms under stochastic input assumptions were known. In this work, we present a novel randomized algorithm with competitive ratio tending to *one* on this family of graphs, under *adversarial* arrival order. Our main contribution is a novel algorithm which achieves competitive ratio $1-O\big(\sqrt{\log d}/\sqrt{d}\big)$ in expectation on d-regular graphs. In contrast, we show that all previously-studied online algo-

rithms have competitive ratio strictly bounded away from one. Moreover, we show the convergence rate of our algorithm's competitive ratio to one is nearly tight, as no algorithm achieves competitive ratio better than $1-O\big(1/\sqrt{d}\big)$. Finally, we show that our algorithm yields a similar competitive ratio with high probability, as well as guaranteeing each vertex a probability of being matched tending to *one*.

David Wajc
Carnegie Mellon University
dwajc@cs.cmu.edu

Ilan R. Cohen
Tel Aviv University
Tel Aviv University
ilanrcohen@gmail.com

CP14

Approximating the Largest Root and Applications to Interlacing Families

Abstract not available at time of publication

Nima Anari
Stanford University
anari@stanford.edu

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

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

Nikhil Srivastava
U.C. Berkeley
srivastn@berkeley.edu

CP14

A Two-Pronged Progress in Structured Dense Matrix Vector Multiplication

We address the broad question of identifying classes of *structured dense* square matrices that can be represented with a linear (in dimension) number of parameters, and for which operations such as matrix-vector multiplication can be performed in a near-linear number of operations. In this paper, we make progress on two fronts: First, we identify a notion of *recurrence width* of matrices. For matrices $\mathbf{A}$ with constant recurrence width, we design algorithms to compute $\mathbf{Ab}$, $\mathbf{A}^T\mathbf{b}$, and more with a near-linear number of operations. This notion of width is more fine-grained than classic structures such as orthogonal polynomial transforms and Toeplitz/Cauchy/Vandermonde matrices, and thus we can compute superfast matrix-vector multiplication for all of them using the same core algorithm. Second, we adapt this algorithm to a matrix-vector multiplication algorithm for a much more general class of matrices with displacement structure: those with low displacement rank with respect to quasiseparable matrices. This result is a novel connection between matrices with displacement structure and those with rank structure, two large but previously separate classes of structured matrices. This class includes Toeplitz-plus-Hankel-like matrices, the Discrete Trigonometric Transforms, and more, and captures all previously known matrices with displacement structure under a unified parametrization and algorithm.

Christopher De Sa
Cornell University
cdesa@cs.cornell.edu

Albert Gu
Stanford University
albertgu@stanford.edu

Rohan Puttagunta
Instagram
rohanp@instagram.com

Christopher Re
Stanford University
chrismre@stanford.edu

Atri Rudra
Univ. of Washington
atri@buffalo.edu

CP14

Improved Rectangular Matrix Multiplication Using Powers of the Coppersmith-Winograd Tensor

In the past few years, successive improvements of the asymptotic complexity of square matrix multiplication have been obtained by developing methods to analyze the powers of the *Coppersmith-Winograd tensor*, a construction introduced thirty years ago. We show how to generalize this approach to make progress on the complexity of rectangular matrix multiplication as well, by developing a framework to analyze powers of tensors in an asymmetric way. By applying this methodology to the fourth power of the Coppersmith-Winograd tensor, we improve the complexity of rectangular matrix multiplication. Let α denote the maximum value such that the product of an $n\times n^\alpha$ matrix by an $n^\alpha \times n$ matrix can be computed with $O(n^{2+\epsilon})$ arithmetic operations for any $\epsilon > 0$. By analyzing the fourth power of the Coppersmith-Winograd tensor, we obtain the new lower bound $\alpha > 0.31389$, which improves the previous lower bound $\alpha > 0.30298$ obtained five years ago by Le Gall (FOCS'12) from the analysis of the second power of the Coppersmith-Winograd tensor. More generally, we give faster algorithms computing the product of an $n \times n^k$ matrix by an $n^k \times n$ matrix for any $k \neq 1$. These improvements lead to improvements in the complexity of a multitude of fundamental problems for which the bottleneck is rectangular matrix multiplication, such as computing the all-pair shortest paths in directed graphs with bounded weights.

François Le Gall
Kyoto University
legall@i.kyoto-u.ac.jp

Florent Urrutia
IRIF
Université Paris Diderot
urrutia@irif.fr

CP14

A Tight Lower Bound for Counting Hamiltonian Cycles Via Matrix Rank

For even k, the matchings connectivity matrix M_k is a binary matrix that is indexed by the perfect matchings on

k vertices; it encodes whether the union of two matchings forms a single cycle. In [Cygan et al., Fast hamiltonicity checking via bases of perfect matchings, STOC 2013], it was shown that the rank of M_k over the integers modulo 2 is $\Theta(\sqrt{2}^k)$. This was used to give an $O^*((2+\sqrt{2})^w)$ time algorithm for counting Hamiltonian cycles modulo 2 on graphs of pathwidth w. Furthermore, a tight lower bound under the Strong Exponential Time Hypothesis (SETH) was shown; this relied on finding a large permutation submatrix within M_k. We present a new technique to obtain similar complexity lower bounds when only a black-box lower bound on the rank of M_k is given. To apply this technique, we prove that the rank of M_k over the rationals is 4^k, up to polynomial factors in k, using the representation theory of the symmetric group. We also show that the rank of M_k over the integers mod p is $\Omega(1.56^k)$ for any prime $p \neq 2$. As a consequence, we obtain that Hamiltonian cycles cannot be counted in time $O^*((6-\epsilon)^w)$ for any $\epsilon > 0$ unless SETH fails. This is essentially tight. We also obtain that Hamiltonian cycles cannot be counted modulo primes $p \neq 2$ in time $O^*(3.56^w)$.

Radu Curticapean
Simons Institute, Berkeley
radu.curticapean@gmail.com

Nathan Lindzey
Department of Combinatorics and Optimization
University of Waterloo
nlindzey@uwaterloo.ca

Jesper Nederlof
TU Eindhoven
j.nederlof@tue.nl

CP14

A Fast Generalized DFT for Finite Groups of Lie Type

We give an arithmetic algorithm using $O(|G|^{\omega/2+o(1)})$ operations to compute the generalized Discrete Fourier Transform (DFT) over group G for finite groups of Lie type, including the linear, orthogonal, and symplectic families and their variants, as well as all finite simple groups of Lie type. Here ω is the exponent of matrix multiplication, so the exponent $\omega/2$ is optimal if $\omega = 2$. Previously, "exponent one' algorithms were known for supersolvable groups and the symmetric and alternating groups. No exponent one algorithms were known (even under the assumption $\omega = 2$) for families of linear groups of fixed dimension, and indeed the previous best-known algorithm for $\mathrm{SL}_2(F_q)$ had exponent 4/3 despite being the focus of significant effort. We unconditionally achieve exponent at most 1.19 for this group, and exponent one if $\omega = 2$. We also show that $\omega = 2$ implies a $\sqrt{2}$ exponent for general finite groups G, which beats the longstanding previous best upper bound (assuming $\omega = 2$) of 3/2.

Chris Umans, Chloe Hsu
California Institute of Technology
umans@cs.caltech.edu, chhsu@caltech.edu

CP15

Persistent Path Homology of Directed Networks

While standard persistent homology has been successful in extracting information from metric datasets, its applicability to more general data, e.g. directed networks, is hindered by its natural insensitivity to asymmetry. We extend a construction of homology of digraphs due to Grigoryan, Lin, Muranov and Yau to the persistent framework. The result, which we call persistent path homology or PPH, encodes a rich level of detail about the asymmetric structure of the input directed network. For example, we prove that PPH identifies a class of directed cyclic networks as directed analogues of the circle. In general, PPH produces signatures that differ from natural extensions of Rips or Čech persistence to the directed setting, but we prove that PPH agrees with Čech persistence on symmetric spaces. Additionally, we prove that PPH agrees with Čech persistence on directed networks satisfying a local condition that we call square-freeness. We prove stability of PPH by utilizing a separate theory of homotopy of digraphs that is compatible with path homology. Finally, we study computational aspects of PPH, and derive an algorithm showing that over field coefficients, computing PPH requires the same worst case running time as standard persistent homology.

Samir Chowdhury
The Ohio State University
chowdhury.57@osu.edu

Facundo Memoli
Ohio State University
memoli@math.osu.edu

CP15

On the Decidability of the Frechet Distance Between Surfaces

We show that the Frechet distance between two piecewise linear surfaces can be decided in finite time, hence, the problem is decidable. For the special case that one of the surfaces is a triangle, we show that the problem is in PSAPCE. In both cases, our computational model is a Turing Machine, and our algorithms rely on Cannys result [STOC 1988] that the existential theory of the real numbers is decidable in PSPACE.

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

CP15

On the Complexity of Optimal Homotopies

In this article, we provide new structural results and algorithms for the Homotopy Height problem. In broad terms, this problem quantifies how much a curve on a surface needs to be stretched to sweep continuously between two positions. More precisely, given two homotopic curves γ_1 and γ_2 on a combinatorial (say, triangulated) surface, we investigate the problem of computing a homotopy between γ_1 and γ_2 where the length of the longest intermediate curve is minimized. Such optimal homotopies are relevant for a wide range of purposes, from very theoretical questions in quantitative homotopy theory to more practical applications such as similarity measures on meshes and graph searching problems. We prove that Homotopy Height is in the complexity class NP, and the corresponding exponential algorithm is the best one known for this problem. This result builds on a structural theorem on monotonicity of optimal homotopies, which is proved in a companion paper. Then we show that this problem encompasses the Homotopic Fréchet distance problem which we therefore also establish to be in NP, answering a question

which has previously been considered in several different settings. We also provide an $O(\log n)$-approximation algorithm for Homotopy Height on surfaces by adapting an earlier algorithm of Har-Peled, Nayyeri, Salvatipour and Sidiropoulos in the planar setting.

Erin Chambers
Saint Louis University
echambe5@slu.edu

Arnaud de Mesmay
Ecole Normale Superieur
ademesmay@gmail.com

Tim Ophelders
Department of Mathematics and Computer Science
TU Eindhoven
t.a.e.ophelders@tue.nl

CP15

Frechet-Stable Signatures Using Persistence Homology

For a metric space Y, the Fréchet distance is a metric on trajectories $f, g : [0,1] \to Y$ that minimizes over continuous reparameterizations h of time of $\max_{t\in[0,1]} d_Y(f(t), g(h(t)))$. One can define the generalized Fréchet distance between more complex objects, functions $f : X \to Y$ where X is some topological space that minimizes over homeomorphisms from $X \to X$. This more general definition has been studied for surfaces and often leads to computationally hard problems. We show how to compute in polynomial-time signatures for these functions for which the resulting metric on the signatures can also be computed in polynomial-time and provides a meaningful lower bound on the generalized Fréchet distance. Our approach uses persistent homology and exploits the natural invariance of persistence diagrams of functions to homeomorphisms of the domain. Our algorithm for computing the signatures in Euclidean spaces uses a new method for computing persistent homology of convex functions on simplicial complexes which may be of independent interest.

Donald Sheehy
University of Connecticut
don.r.sheehy@gmail.com

CP15

Computing Simplicial Representatives of Homotopy Group Elements

A central problem of algebraic topology is to understand the *homotopy groups* $\pi_d(X)$ of a topological space X. For the computational version of the problem, it is well known that there is no algorithm to decide whether the *fundamental group* $\pi_1(X)$ of a given finite simplicial complex X is trivial. On the other hand, there are several algorithms that, given a finite simplicial complex X that is *simply connected* (i.e., with $\pi_1(X)$ trivial), compute the higher homotopy group $\pi_d(X)$ for any given $d \geq 2$. However, these algorithms come with a caveat: They compute the isomorphism type of $\pi_d(X)$, $d \geq 2$ as an *abstract* finitely generated abelian group given by generators and relations, but they work with very implicit representations of the elements of $\pi_d(X)$. Here we present an algorithm that, given a simply connected simplicial complex X, computes $\pi_d(X)$ and represents its elements as simplicial maps from a suitable triangulation of the d-sphere S^d to X. For fixed d, the algorithm runs in time exponential in $size(X)$, the number of simplices of X. Moreover, we prove that this is optimal: For every fixed $d \geq 2$, we construct a family of simply connected simplicial complexes X such that for any simplicial map representing a generator of $\pi_d(X)$, the size of the triangulation of S^d on which the map is defined is exponential in $size(X)$.

Peter Franek
Institute of Computer Science
Academy of Sciences of the Czech Republic
peter.franek@gmail.com

Marek Filakovsky
Masaryk University
Brno
m.filakovsky@seznam.cz

Uli Wagner, Stephan Zhechev
IST Austria
uli@ist.ac.at, stephan.zhechev@ist.ac.at

CP16

Multivariate Fine-Grained Complexity of Longest Common Subsequence

We revisit the classic combinatorial pattern matching problem of finding a longest common subsequence (LCS). For strings x and y of length n, a textbook algorithm solves LCS in time $O(n^2)$, but despite much effort, no $O(n^{2-\varepsilon})$-time algorithm is known. Recent work indeed shows that such an algorithm would refute the Strong Exponential Time Hypothesis (SETH). In spite of the quadratic-time barrier, an enduring scientific interest produced strongly subquadratic time algorithms for special cases of interest, e.g., differential file comparison. In this paper, using the lens of fine-grained complexity, our goal is to (1) justify the lack of faster specialized algorithms since 1990 and (2) determine whether some special cases of LCS admit faster algorithms than currently known. To this end, we provide a systematic study of the multivariate complexity of LCS, taking into account all parameters previously discussed in the literature: the input size $n := \max\{|x|, |y|\}$, the length of the shorter string $m := \min\{|x|, |y|\}$, the length L of an LCS of x and y, the numbers of deletions $\delta := m - L$ and $\Delta := n - L$, the alphabet size, as well as the numbers of matching pairs M and dominant pairs d. For any class of instances defined by fixing each parameter individually to a polynomial in terms of the input size, we determine the optimal running time under SETH as $(n + \min\{d, \delta\Delta, \delta m\})^{1\pm o(1)}$, up to lower order factors.

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

Marvin Künnemann
Max-Planck-Institut für Informatik, Saarbrücken,
Germany
marvin@mpi-inf.mpg.de

CP16

Tight Hardness for Shortest Cycles and Paths in Sparse Graphs

Fine-grained reductions have established equivalences between many core problems with $\tilde{O}(n^3)$-time algorithms on n-node weighted graphs, such as Shortest Cycle, All-

Pairs Shortest Paths (APSP), Radius, Replacement Paths, Second Shortest Paths, and so on. These problems also have $\tilde{O}(mn)$-time algorithms on m-edge n-node weighted graphs, and such algorithms have wider applicability. Are these mn bounds optimal when $m \ll n^2$? Starting from the hypothesis that the minimum weight $(2\ell+1)$-Clique problem in edge weighted graphs requires $n^{2\ell+1-o(1)}$ time, we prove that for all sparsities of the form $m = \Theta(n^{1+1/\ell})$, there is no $O(n^2 + mn^{1-\epsilon})$ time algorithm for $\epsilon > 0$ for *any* of the below problems:

- Minimum Weight $(2\ell+1)$-Cycle in a directed weighted graph,
- Shortest Cycle in a directed weighted graph,
- APSP in a directed or undirected weighted graph,
- Radius (or Eccentricities) in a directed or undirected weighted graph,
- Wiener index, Replacement Paths, Second Shortest Path, and Betweenness Centrality in directed or undirected weighted graphs.

That is, we prove hardness for a variety of sparse graph problems from the hardness of a dense graph problem. Our results also lead to new conditional lower bounds from several hypothesis for unweighted sparse graph problems including k-cycle, shortest cycle, Radius, Wiener index and APSP.

Andrea Lincoln, Virginia Vassilevska Williams
MIT
andreali@mit.edu, virgi@mit.edu

Ryan Williams
Stanford University
rrw@mit.edu

CP16

Approximating Edit Distance in Truly Subquadratic Time: Quantum and MapReduce

The *edit distance* between two strings is defined as the smallest number of *insertions*, *deletions*, and *substitutions* that need to be made to transform one of the strings to another one. Approximating edit distance in subquadratic time is "one of the biggest unsolved problems in the field of combinatorial pattern matching." Our main result is a quantum constant approximation algorithm for computing the edit distance in truly subquadratic time. More precisely, we give an $O(n^{1.858})$ quantum algorithm that approximates the edit distance within a factor of 7. We further extend this result to an $O(n^{1.781})$ quantum algorithm that approximates the edit distance within a larger constant factor. Our solutions are based on a framework for approximating edit distance in parallel settings. This framework requires as black box an algorithm that computes the distances of several smaller strings all at once. For a quantum algorithm, we reduce the black box to *metric estimation* and provide efficient algorithms for approximating it. We further show that this framework enables us to approximate edit distance in distributed settings. To this end, we provide a MapReduce algorithm to approximate edit distance within a factor of 3, with sublinearly many machines and sublinear memory. Also, our algorithm runs in a logarithmic number of rounds.

Mahdi Boroujeni
Sharif University of Technology
safarnejad@ce.sharif.edu

Soheil Ehsani
University of Maryland
ehsani@cs.umd.edu

Mohammad Ghodsi
Sharif University of Technology
Institute for Research in Fundamental Sciences (IPM)
ghodsi@sharif.edu

MohammadTaghi Hajiaghayi, Saeed Seddighin
University of Maryland, College Park
hajiagha@cs.umd.edu, saeedreza.seddighin@gmail.com

CP16

Tree Edit Distance Cannot Be Computed in Strongly Subcubic Time (unless Apsp Can)

The edit distance between two rooted ordered trees with n nodes labeled from an alphabet Σ is the minimum cost of transforming one tree into the other by a sequence of elementary operations consisting of deleting and relabeling existing nodes, as well as inserting new nodes. Tree edit distance is a well known generalization of string edit distance. The fastest known algorithm for tree edit distance runs in cubic $O(n^3)$ time and is based on a similar dynamic programming solution as string edit distance. In this paper we show that a truly subcubic $O(n^{3-\varepsilon})$ time algorithm for tree edit distance is unlikely: For $|\Sigma| = \Omega(n)$, a truly subcubic algorithm for tree edit distance implies a truly subcubic algorithm for the all pairs shortest paths problem. For $|\Sigma| = O(1)$, a truly subcubic algorithm for tree edit distance implies an $O(n^{k-\varepsilon})$ algorithm for finding a maximum weight k-clique. Thus, while in terms of upper bounds string edit distance and tree edit distance are highly related, in terms of lower bounds string edit distance exhibits the hardness of the strong exponential time hypothesis [Backurs, Indyk STOC'15] whereas tree edit distance exhibits the hardness of all pairs shortest paths. Our result provides a matching conditional lower bound for one of the last remaining classic dynamic programming problems.

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

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

Shay Mozes
IDC Herzliya, Israel
smozes@idc.ac.il

Oren Weimann
University of Haifa
oren@cs.haifa.ac.il

CP16

On the Difference Between Closest, Furthest, and Orthogonal Pairs: Nearly-Linear Vs Barely-Subquadratic Complexity in Computational Geometry

Pair-finding problems for n points in d-dimensional Euclidean space (and ℓ_p spaces more generally) have typically had two kinds of running-time solutions: (Nearly-Linear) less than $d^{poly(d)} \cdot n \log^{O(d)} n$ time, or (Barely-

Subquadratic) $f(d) \cdot n^{2-1/\Theta(d)}$ time, for various f. For example, in Euclidean space, finding a Closest Pair among n points in R^d is nearly-linear, while known algorithms for finding a Furthest Pair are only barely-subquadratic. Is there a barrier to obtaining nearly-linear algorithms for problems which are currently only barely-subquadratic? We give a novel exact and deterministic self-reduction for the Orthogonal Vectors problem on n vectors in $\{0,1\}^d$ to n vectors in $Z^{\omega(\log d)}$ that runs in $2^{o(d)}$ time. As a consequence, barely-subquadratic problems such as Euclidean Furthest Pair, Euclidean Bichromatic Closest Pair, and Incidence Detection do not have $O(n^{2-\epsilon})$ time algorithms (in Turing models of computation) for dimensionality $d = \omega(\log\log n)^2$, unless the Orthogonal Vectors Conjecture and the Strong Exponential Time Hypothesis are false. That is, while the poly-log-log-dimensional case of Closest Pair is in $n^{1+o(1)}$ time, the poly-log-log-dimensional case of Furthest Pair can encode larger-dimensional problems conjectured to require $n^{2-o(1)}$ time. Other related results are shown.

Ryan Williams
Stanford University
rrw@mit.edu

CP17

Boolean Function Analysis Meets Stochastic Optimization: An Approximation Scheme for Stochastic Knapsack

The stochastic knapsack problem is the stochastic variant of the classical knapsack problem in which the algorithm designer is given a a knapsack with a given capacity and a collection of items where each item is associated with a profit and a probability distribution on its size. The goal is to select a subset of items with maximum profit and violate the capacity constraint with probability at most p (referred to as the overflow probability). In this paper, we design efficient approximation schemes for this problem without relaxing the capacity constraint.

- Our first result is in the case when item sizes are Bernoulli random variables. In this case, we design a (nearly) fully polynomial time approximation scheme (FPTAS) which only relaxes the overflow probability.
- Our second result generalizes the first result to the case when all the item sizes are supported on a (common) set of constant size.
- Our third result is in the case when item sizes are so-called hypercontractive random variables i.e., random variables whose second and fourth moments are within constant factors of each other. In other words, the kurtosis of the random variable is upper bounded by a constant. In this case, we design a polynomial time approximation scheme which relaxes both the overflow probability and maximum profit.

Anindya De
Northwestern University
anindya.de1@northwestern.edu

CP17

Nested Convex Bodies Are Chaseable

In the Convex Body Chasing problem, we are given an initial point $v_0 \in R^d$ and an online sequence of n convex bodies $F_1, \ldots, F_n$. When we receive F_i, we are required to move inside F_i. Our goal is to minimize the total distance travelled. This fundamental online problem was first studied by Friedman and Linial (DCG 1993). They proved an $\Omega(\sqrt{d})$ lower bound on the competitive ratio, and conjectured that a competitive ratio depending only on d is possible. However, despite much interest in the problem, the conjecture remains wide open. We consider the setting in which the convex bodies are nested: $F_1 \supset \ldots \supset F_n$. The nested setting is closely related to extending the online LP framework of Buchbinder and Naor (ESA 2005) to arbitrary linear constraints. Moreover, this setting retains much of the difficulty of the general setting and captures an essential obstacle in resolving Friedman and Linial's conjecture. In this work, we give the first $f(d)$-competitive algorithm for chasing nested convex bodies in R^d.

Nikhil Bansal
Department of Mathematics and Computer Science
Eindhoven University of Technology, Eindhoven, Netherlands
n.bansal@tue.nl

Martin Bohm
Charles University
bohm@iuuk.mff.cuni.cz

Marek Elias
Eindhoven University of Technology, Netherlands
m.elias@tue.nl

Grigorios Koumoutsos
TU Eindhoven
g.koumoutsos@tue.nl

Seeun W. Umboh
University of Wisconsin-Madison
seeun.umboh@gmail.com

CP17

Stochastic Load Balancing on Unrelated Machines

We consider the problem of makespan minimization: i.e., scheduling jobs on machines to minimize the maximum load. For the deterministic case, good approximations are known even when the machines are unrelated. However, the problem is not well-understood when there is uncertainty in the job sizes. In our setting the job sizes are stochastic, i.e., the size of a job j on machine i is a random variable X_{ij}, whose distribution is known. The goal is to find a fixed assignment of jobs to machines, to minimize the expected makespan. For the identical machines special case when the size of a job is the same across all machines, a constant-factor approximation algorithm has long been known. However, the problem has remained open even for the related machines case. Our main result is a constant-factor approximation for the most general case of unrelated machines. The main technical challenge we overcome is obtaining an efficiently computable lower bound for the optimal solution. We give an exponential-sized LP that we argue gives a strong lower bound. Then we show how to round any fractional solution to satisfy only a small subset of the constraints, which are enough to bound the expected makespan of our solution. We then consider two generalizations. The first is the budgeted makespan minimization problem. We extend our above result to a constant-factor. The second problem is the q-norm minimization problem. Here we give an $O(q/\log q)$-approximation algorithm.

Xiangkun Shen
University of Michigan
xkshen@umich.edu

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

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

Viswanath Nagarajan
University of Michigan
viswa@umich.edu

CP17

An Alon-Boppana Type Bound for Weighted Graphs and Lowerbounds for Spectral Sparsification

We prove the following Alon-Boppana type theorem for general (not necessarily regular) weighted graphs: if G is an n-node weighted undirected graph with $dn/2$ edges and girth $g > 2d^{1/8} + 1$, and if $\lambda_1 \leq \lambda_2 \leq \cdots \lambda_n$ are the eigenvalues of the (non-normalized) Laplacian of G, then

$$\frac{\lambda_n}{\lambda_2} \geq 1 + \frac{4}{\sqrt{d}} - O\left(\frac{1}{d^{\frac{5}{8}}}\right)$$

(The Alon-Boppana theorem gives $\frac{\lambda_n}{\lambda_2} \geq 1 + \frac{4}{\sqrt{d}} - O\left(\frac{1}{d}\right)$ in unweighted d-regular graphs of diameter $> d^{1.5}$.) Our result implies a lower bound for spectral sparsifiers. Batson, Spielman and Srivastava proved that for every G there is an ϵ-spectral-sparsifier H of average degree d where $\epsilon \approx \frac{4\sqrt{2}}{\sqrt{d}}$ and the edges of H are a (weighted) subset of the edges of G. Batson, Spielman and Srivastava also show that the bound on ϵ cannot be reduced below $\approx \frac{2}{\sqrt{d}}$ when G is a clique; our result implies that ϵ cannot be reduced below $\approx \frac{4}{\sqrt{d}}$. The method of Batson, Spielman and Srivastava proves a more general result, about sparsifying sums of rank-one matrices, and their method applies to an "online' setting. We show that for the online matrix setting the $4\sqrt{2}/\sqrt{d}$ bound is tight, up to lower order terms.

Luca Trevisan
UC Berkeley
luca@berkeley.edu

Nikhil Srivastava
U.C. Berkeley
srivastn@berkeley.edu

CP17

Scheduling When You Don't Know the Number of Machines

Often in a scheduling problem, there is uncertainty about the jobs to be processed. The issue of uncertainty regarding the machines has been much less studied. In this paper, we study a scheduling environment in which jobs first need to be grouped into some sets before the number of machines is known, and then the sets need to be scheduled on machines without being separated. In order to evaluate algorithms in such an environment, we introduce the idea of an α-robust algorithm, one which is guaranteed to return a schedule on any number m of machines that is within an α factor of the optimal schedule on m machine, where the optimum is not subject to the restriction that the sets cannot be separated. Under such environment, we give a $(\frac{5}{3} + \epsilon)$-robust algorithm for scheduling on parallel machines to minimize makespan, and show a lower bound $\frac{4}{3}$. For the special case when the jobs are infinitesimal, we give a 1.233-robust algorithm with an asymptotic lower bound of 1.207. We also study a case of fair allocation, where the objective is to minimize the difference between the maximum and minimum machine load.

Clifford Stein, Mingxian Zhong
Columbia University
cliff@ieor.columbia.edu, mz2325@columbia.edu

CP18

Steiner Point Removal — Distant Terminals Don't (Really) Bother

Given a weighted graph $G = (V, E, w)$ with a set of k terminals $T \subset V$, the Steiner Point Removal problem seeks for a minor of the graph with vertex set T, such that the distance between every pair of terminals is preserved within a small multiplicative distortion. Kamma, Krauthgamer and Nguyen used a ball-growing algorithm to show that the distortion is at most $O(\log^5 k)$ for general graphs. In this paper, we improve the distortion bound to $O(\log^2 k)$. The improvement is achieved based on a known algorithm that constructs terminal-distance exact-preservation minor with $O(k^4)$ (which is independent of $|V|$) vertices, and also two tail bounds on the sums of independent exponential random variables, which allow us to show that it is unlikely for a non-terminal being contracted to a distant terminal.

Yun Kuen Cheung
Max Planck Institute for Informatics
Saarland Informatics Campus
ycheung@mpi-inf.mpg.de

CP18

Fast, Deterministic and Sparse Dimensionality Reduction

We provide a deterministic construction of the sparse Johnson-Lindenstrauss transform of Kane & Nelson (J.ACM 2014) which runs, under a mild restriction, in the time necessary to apply the sparse embedding matrix to the input vectors. Specifically, given a set of n vectors in $\mathbb{R}^d$ and target error ε, we give a deterministic algorithm to compute a $\{-1, 0, 1\}$ embedding matrix of rank $O((\log n)/\varepsilon^2)$ with $O((\log n)/\varepsilon)$ entries per column which preserves the norms of the vectors to within $1 \pm \varepsilon$. If NNZ, the number of non-zero entries in the input set of vectors, is $\Omega(d^2)$, our algorithm runs in time $O(\text{NNZ} \cdot \log n/\epsilon)$. One ingredient in our construction is an extremely simple proof of the Hanson-Wright inequality for subgaussian random variables, which is more amenable to derandomization. As an interesting byproduct, we are able to derive the essentially optimal form of the inequality in terms of its functional dependence on the parameters.

Daniel Dadush
New York University
dadush@cwi.nl

CP18

Steiner Point Removal with Distortion O(log K)

In the Steiner point removal (SPR) problem, we are given a weighted graph $G = (V, E)$ and a set of terminals $K \subset V$ of size k. The objective is to find a minor M of G with only the terminals as its vertex set, such that the distance between the terminals will be preserved up to a

small multiplicative distortion. Kamma, Krauthgamer and Nguyen [KKN15] used a ball-growing algorithm with exponential distributions to show that the distortion is at most $O(\log^5 k)$. Cheung [Cheung18] improved the analysis of the same algorithm, bounding the distortion by $O(\log^2 k)$. We improve the analysis of this ball-growing algorithm even further, bounding the distortion by $O(\log k)$.

Arnold Filtser
Ben gurion university of the negev
arnoldf@cs.bgu.ac.il

CP18

Impossibility of Dimension Reduction in the Nuclear Norm

Let S_1 be the Banach space of compact linear operators $T : \ell_2 \to \ell_2$ whose nuclear norm $\|T\|_{S_1} = \sum_{j=1}^{\infty} \sigma_j(T)$ is finite, where $\{\sigma_j(T)\}_{j=1}^{\infty}$ are the singular values of T. We prove that there exists arbitrarily large $C \subset S_1$ that cannot be embedded with distortion $O(1)$ into any $|C|^{o(1)}$-dimensional linear subspace of S_1. C is not even a $O(1)$-Lipschitz quotient of any subset of any $|C|^{o(1)}$-dimensional linear subspace of S_1. Thus, S_1 does not admit a dimension reduction result á la Johnson and Lindenstrauss (1984), which complements the work of Harrow, Montanaro and Short (2011) on the limitations of quantum dimension reduction under the assumption that the embedding into low dimensions is a quantum channel. Such a statement was previously known with S_1 replaced by ℓ_1 via the work of Brinkman and Charikar (2003). In fact, C can be taken to be the same set as the one that Brinkman and Charikar considered, viewed as a collection of diagonal matrices in S_1. The challenge is to demonstrate that C cannot be faithfully realized in an arbitrary low-dimensional subspace of S_1, while Brinkman and Charikar obtained such an assertion only for subspaces of S_1 that consist of diagonal operators. We establish this by proving that the Markov 2-convexity constant of any finite dimensional linear subspace X of S_1 is $O(\sqrt{\log \dim(X)})$.

Assaf Naor
Theory group
Microsoft Research
naor@math.princeton.edu

Gilles Pisier
Texas A&M University
pisier@math.tamu.edu

Gideon Schechtman
Weizmann Institute
gideon@weizmann.ac.il

CP18

Embeddability in R^3 is NP-hard

We prove that the problem of deciding whether a 2- or 3-dimensional simplicial complex embeds into $\mathbb{R}^3$ is NP-hard. This stands in contrast with the lower dimensional cases which can be solved in linear time, and a variety of computational problems in $\mathbb{R}^3$ like unknot or 3-sphere recognition which are in NP $\cap$ co-NP (assuming the generalized Riemann hypothesis). Our reduction encodes a satisfiability instance into the embeddability problem of a 3-manifold with boundary tori, and relies extensively on techniques from low-dimensional topology, most importantly Dehn fillings on link complements.

Arnaud de Mesmay
Ecole Normale Superieur
ademesmay@gmail.com

Yo'av Rieck
Department of Mathematical Sciences
University of Arkansas
yoav.rieck@gmail.com

Eric Sedgwick
DePaul University
esedgwick@cdm.depaul.edu

Martin Tancer
IST Austria
tancer@kam.mff.cuni.cz

CP19

A Tight $\sqrt{2}$-Approximation for Linear 3-Cut

We investigate the approximability of the linear 3-cut problem in directed graphs, which is the simplest unsolved case of the linear k-cut problem. The input here is a directed graph $D = (V, E)$ with node weights and three specified terminal nodes $s, r, t \in V$, and the goal is to find a minimum weight subset of non-terminal nodes whose removal ensures that s cannot reach r and t, and r cannot reach t. The problem is approximation-equivalent to the problem of blocking rooted in- and out-arborescences, and it also has applications in network coding and security. The approximability of linear 3-cut has been wide open until now: the best known lower bound under the Unique Games Conjecture (UGC) was 4/3, while the best known upper bound was 2 using a trivial algorithm. In this work we completely close this gap: we present a $\sqrt{2}$-approximation algorithm and show that this factor is tight assuming UGC. Our contributions are twofold: (1) we analyze a natural two-step deterministic rounding scheme through the lens of a single-step randomized rounding scheme with *non-trivial* distributions, and (2) we construct integrality gap instances that meet the upper bound of $\sqrt{2}$. Our gap instances can be viewed as a weighted graph sequence converging to a "graph limit structure'.

Vivek Madan
University of illinois Urbana-Champaign
vmadan2@illinois.edu

Kristof Berczi
Eötvös Loránd University
Hungary
berkri@cs.elte.hu

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

Tamas Kiraly
Eotvos Lorand University, Budapest
tkiraly@cs.elte.hu

CP19

Near-Optimal Approximation Algorithm for Simul-

taneous Max-Cut

In the simultaneous Max-Cut problem, we are given k weighted graphs on the same set of n vertices, and the goal is to find a cut of the vertex set so that the minimum, over the k graphs, of the cut value is as large as possible. Previous work of Bhangale et.al. [BKS15] gave a polynomial time algorithm which achieved an approximation factor of $1/2 - o(1)$ for this problem (and an approximation factor of $1/2 + \epsilon_k$ in the unweighted case, where ϵ_k goes to 0 as k goes to infinity). In this work, we give an approximation algorithm for simultaneous Max-Cut with an approximation factor of 0.878 (for all constant k). The natural SDP formulation for simultaneous Max-Cut was shown to have an integrality gap of $1/2 + \epsilon_k$ in Bhangale et.al. [BKS15]. In achieving the better approximation guarantee, we use a stronger Sum-of-Squares hierarchy SDP relaxation and a rounding algorithm based on Raghavendra-Tan [RT12], in addition to techniques from [BKS15].

Amey Bhangale
Weizmann Institute
ameyrbh@gmail.com

Subhash Khot
New York University
khot@cs.nyu.edu

Swastik Kopparty
Rutgers University
swastik.kopparty@gmail.com

Sushant Sachdeva
Google, Mountain View CA
sachdevasushant@gmail.com

Devanathan Thiruvenkatachari
New York University
dt1295@nyu.edu

CP19

Approximating Cycles in Directed Graphs

The girth of a graph is a fundamental graph parameter. Unfortunately all known algorithms for computing, even approximately, the girth and girth-related structures in directed weighted m-edge and n-node graphs require $\Omega(minn^{\omega}, mn)$ time (for $2 \leq \omega < 2.373$). In this paper, we drastically improve these runtimes as follows: * Multiplicative Approximations in Nearly Linear Time: We give an algorithm that in $\widetilde{O}(m)$ time computes an $\widetilde{O}(1)$-multiplicative approximation of the girth as well as an $\widetilde{O}(1)$-multiplicative roundtrip spanner with $\widetilde{O}(n)$ edges with high probability * Nearly Tight Additive Approximations: For unweighted graphs and any $\alpha \in (0, 1)$ we give an algorithm that in $\widetilde{O}(mn^{1-\alpha})$ time computes an $O(n^{a})$-additive approximation of the girth, w.h.p. We show that the runtime of our algorithm cannot be significantly improved without a breakthrough in combinatorial boolean matrix multiplication. If the girth is $O(n^{a})$, then the algorithm is deterministic. Our main technical contribution to achieve these results is the first nearly linear time algorithm for computing roundtrip covers, a directed graph decomposition concept key to previous roundtrip spanner constructions. Previously it was not known how to compute these significantly faster than $\Omega(mn)$ time. Given the traditional difficulty in efficiently processing directed graphs, we hope our techniques may find further applications.

Jakub Pachocki
OpenAI
merettm@gmail.com

Liam Roditty
Bar Ilan University
liam.roditty@gmail.com

Aaron Sidford
Stanford University
aaron.sidford@gmail.com

Roe Tov
Bar Ilan University
roei81@gmail.com

Virginia Vassilevska Williams
MIT
virgi@mit.edu

CP19

Hypergraph k-Cut in Randomized Polynomial Time

In the hypergraph k-cut problem, the input is a hypergraph, and the goal is to find a smallest subset of hyperedges whose removal ensures that the remaining hypergraph has at least k connected components. This problem is known to be at least as hard as the densest k-subgraph problem when k is part of the input (Chekuri-Li, 2015). We present a randomized polynomial time algorithm to solve the hypergraph k-cut problem for constant k. Our algorithm solves the more general hedge k-cut problem when the subgraph induced by every hedge has a constant number of connected components. In the hedge k-cut problem, the input is a *hedgegraph* specified by a vertex set and a disjoint set of *hedges*, where each *hedge* is a subset of edges defined over the vertices. The goal is to find a smallest subset of hedges whose removal ensures that the number of connected components in the remaining underlying (multi-)graph is at least k. Our algorithm is based on random contractions akin to Karger's min cut algorithm. Our main technical contribution is a distribution over the hedges (hyperedges) so that random contraction of hedges (hyperedges) chosen from the distribution succeeds in returning an optimum solution with large probability.

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

Chao Xu
Department of Computer Science
University of Illinois, Urbana-Champaign
chaoxu3@illinois.edu

Xilin Yu
Department of Computer Science
University of Illinois at UrbanaChampaign
yuxilin51@gmail.com

CP19

A Near-Linear Approximation Scheme for Multicuts of Embedded Graphs with a Fixed Number of

Terminals

For an undirected edge-weighted graph G and a set R of pairs of vertices called pairs of *terminals*, a multicut is a set of edges such that removing these edges from G disconnects each pair in R. We provide an algorithm computing a $(1+\varepsilon)$-approximation of the minimum multicut of a graph G in time $(g+t)^{(O(g+t)^3)} \cdot (1/\varepsilon)^{O(g+t)} \cdot n \log n$, where g is the genus of G and t is the number of terminals. This is tight in several aspects, as the minimum multicut problem is both APX-hard and W[1]-hard (parameterized by the number of terminals), even on planar graphs (equivalently, when $g = 0$). Our result, in the field of fixed-parameter approximation algorithms, mostly relies on concepts borrowed from computational topology of graphs on surfaces. In particular, we use and extend various recent techniques concerning homotopy, homology, and covering spaces (even in the planar case). We also exploit classical ideas stemming from approximation schemes for planar graphs and low-dimensional geometric inputs. A key insight towards our result is a novel characterization of a minimum multicut as the union of some Steiner trees in the universal cover of the surface in which G is embedded.

Vincent Cohen-Addad
CNRS, LIP6, Université Pierre et Marie Curie
Paris, France
vincent.cohen.addad@ens-lyon.org

Éric Colin De Verdière
CNRS, LIGM
Université Paris-Est Marne-la-Vallée
eric.colindeverdiere@u-pem.fr

Arnaud de Mesmay
Ecole Normale Superieur
ademesmay@gmail.com

CP20
Optimal Dynamic Strings

In this paper, we study the fundamental problem of maintaining a dynamic collection of strings under the following operations:

- `make_string` – add a string of constant length,
- `concat` – concatenate two strings,
- `split` – split a string into two at a given position,
- `compare` – find the lexicographical order (less, equal, greater) between two strings,
- `LCP` – calculate the longest common prefix of two strings.

We develop a generic framework for dynamizing the recompression method recently introduced by Jeż [J. ACM, 2016]. It allows us to present an efficient data structure for the above problem, where an update requires only $O(\log n)$ worst-case time with high probability, with n being the total length of all strings in the collection, and a query takes constant worst-case time. On the lower bound side, we prove that even if the only possible query is checking equality of two strings, either updates or queries must take amortized $\Omega(\log n)$ time; hence our implementation is optimal.

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

Adam Karczmarz, Tomasz Kociumaka
University of Warsaw
a.karczmarz@mimuw.edu.pl, kociumaka@mimuw.edu.pl

Jakub Łącki
Google Research, New York
jlacki@google.com

Piotr Sankowski
University of Warsaw
p.sankowski@mimuw.edu.pl

CP20
Lempel-Ziv: A "One-Bit Catastrophe" But Not a Tragedy

The so-called "one-bit catastrophe" for the compression algorithm LZ'78 asks whether the compression ratio of an infinite word can change when a single bit is added in front of it. We answer positively this open question raised by Lutz and others: we show that there exists an infinite word w such that $\rho_{\sup}(w) = 0$ but $\rho_{\inf}(0w) > 0$, where $\rho_{\sup}$ and $\rho_{\inf}$ are respectively the lim sup and the lim inf of the compression ratios ρ of the prefixes. To that purpose we explore the behaviour of LZ'78 on finite words and show the following results:

- There is a constant $C > 0$ such that, for any finite word w and any letter a, $\rho(aw) \leq C\sqrt{\rho(w)\log|w|}$. Thus, sufficiently compressible words ($\rho(w) = o(1/\log|w|)$) remain compressible with a letter in front;
- The previous result is tight up to a multiplicative constant for any compression ratio $\rho(w) = O(1/\log|w|)$. In particular, there are infinitely many words w satisfying $\rho(w) = O(1/\log|w|)$ but $\rho(0w) = \Omega(1)$.

Guillaume Lagarde, Sylvain Perifel
IRIF, Univ. Paris Diderot
glagarde@irif.fr, sylvain.perifel@irif.fr

CP20
Optimal-Time Text Indexing in BWT-runs Bounded Space

Indexing highly repetitive texts has become an important problem since the turn of the millennium. A relevant compressibility measure for repetitive texts is r, the number of runs in their Burrows-Wheeler Transform (BWT). One of the earliest indexes for repetitive collections, the Run-Length FM-index, used $O(r)$ space and was able to efficiently count the number of occurrences of a pattern in the text. However, it was unable to locate the positions of those occurrences efficiently within a space bounded in terms of r. Since then, a number of other indexes with space bounded by other measures of repetitiveness — the number of phrases in the Lempel-Ziv parse, the size of the smallest grammar generating the text, the size of the smallest automaton recognizing the text factors — have been proposed for efficiently locating, but not directly counting, the occurrences of a pattern. In this paper we close this long-standing problem, showing how to extend the Run-Length FM-index so that it can locate the occurrences efficiently within $O(r)$ space (in loglogarithmic time each), and reaching optimal time within $O(r\log(n/r))$ space. We also describe a structure using $O(r\log(n/r))$ space that replaces the text and efficiently extracts any text substring, with an $O(\log(n/r))$ additive time penalty over the optimum. Preliminary experiments show that our new struc-

ture outperforms the alternatives by orders of magnitude in the space/time tradeoff map.

Nicola Prezza
Technical University of Denmark
nicola.prezza@gmail.com

Gonzalo Navarro
University of Chile
gnavarro@dcc.uchile.cl

Travis Gagie
EIT, Diego Portales University, Chile
travis.gagie@gmail.com

CP20

In-Place Sparse Suffix Sorting

Suffix arrays encode the lexicographical order of all suffixes of a text and are often combined with the Longest Common Prefix array (LCP) to simulate navigational queries on the suffix tree in reduced space. In space-critical applications such as sparse and compressed text indexing, only information regarding the lexicographical order of a size-b subset of all n text suffixes is often needed. Such information can be stored space-efficiently (in b words) in the sparse suffix array (SSA). The SSA and its relative sparse LCP array (SLCP) can be used as a space-efficient substitute of the sparse suffix tree. Very recently, Gawrychowski and Kociumaka showed that the sparse suffix tree (and therefore SSA and SLCP) can be built in asymptotically optimal $O(b)$ space with a Monte Carlo algorithm running in $O(n)$ time. The main reason for using the SSA and SLCP arrays in place of the sparse suffix tree is, however, their reduced space of b words each. This leads naturally to the quest for in-place algorithms building these arrays. Franceschini and Muthukrishnan showed that the full suffix array can be built in-place and in optimal running time. On the other hand, finding sub-quadratic in-place algorithms for building the SSA and SLCP for *general* subsets of suffixes has been an elusive task for decades. In this paper, we give the first solution to this problem.

Nicola Prezza
Technical University of Denmark
nicola.prezza@gmail.com

CP20

Improved Bounds for Testing Dyck Languages

We consider the problem of deciding membership in Dyck languages, a fundamental family of context-free languages, comprised of well-balanced strings of parentheses. In this problem we are given a string of length n in the alphabet of parentheses of m types and must decide if it is well-balanced. We consider this problem in the property testing setting, where one would like to make the decision while querying as few characters of the input as possible. Property testing of strings for Dyck language membership for $m = 1$, with a number of queries independent of the input size n, was provided in [Alon, Krivelevich, Newman and Szegedy, SICOMP 2001]. Property testing of strings for Dyck language membership for $m \geq 2$ was first investigated in [Parnas, Ron and Rubinfeld, RSA 2003]. They showed an upper bound and a lower bound for distinguishing strings belonging to the language from strings that are far (in terms of the Hamming distance) from the language, which are respectively (up to polylogarithmic factors) the 2/3 power and the 1/11 power of the input size n. Here we improve the power of n in both bounds. For the upper bound, we introduce a recursion technique, that provides a test for any power of n larger than 2/5. For the lower bound, we introduce a new problem called Truestring Equivalence, which is easily reducible to the 2-type Dyck language property testing problem. For this new problem, we show a lower bound of n to the power of 1/5.

Tatiana Starikovskaya
University of Bristol, UK
tat.starikovskaya@gmail.com

Eldar Fischer
Technion Israel Institute of Technology,
Haifa, Israel
eldar@cs.technion.ac.il

Frédéric Magniez
CNRS, IRIF, Univ Paris Diderot
France
magniez@irif.fr

CP21

Computing the Independence Polynomial: from the Tree Threshold Down to the Roots

Abstract not available at time of publication

Nicholas Harvey
University of British Columbia
nickhar@cs.ubc.ca

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

Jan Vondrak
Stanford University
n/a

CP21

Probabilistic Existence of Large Sets of Designs

A new probabilistic technique for establishing the existence of certain regular combinatorial structures has been recently introduced by Kuperberg, Lovett, and Peled (STOC 2012). Using this technique, it can be shown that under certain conditions, a randomly chosen structure has the required properties of a t-(n, k, λ) combinatorial design with tiny, yet positive,probability. Herein, we strengthen both the method and the result of Kuperberg, Lovett, and Peled as follows. We modify the random choice and the analysis to show that, under the same conditions, not only does a t-(n, k, λ) design exist but, in fact, with positive probability there exists a *large set* of such designs — that is, a partition of the set of k-subsets of $[n]$ into t-(n, k, λ) designs. Specifically, using the probabilistic approach derived herein, we prove that for all sufficiently large n,large sets of t-(n, k, λ) designs exist whenever $k > 9t$ and the necessary divisibility conditions are satisfied. This resolves the existence conjecture for large sets of designs for all $k > 9t$.

Sankeerth Rao Karingula
PhD Student UCSD
sankeerth1729@gmail.com

Shachar Lovett, Alexander Vardy
University of California San Diego

slovett@ucsd.edu, avardy@ucsd.edu

CP21

Stability of the Lanczos Method for Matrix Function Approximation

The ubiquitous Lanczos method can approximate $f(A)x$ for any symmetric matrix A, vector x, and function f. In exact arithmetic, the method's error after k iterations is bounded by the error of the best degree-k polynomial approximating $f(x)$ on the range $[\lambda_{min}(A), \lambda_{max}(A)]$. However, despite decades of work, it has been unclear if this powerful guarantee holds in finite precision. We resolve this problem, proving that when $\max_{x\in[\lambda_{min},\lambda_{max}]} |f(x)| \leq C$, Lanczos essentially matches the exact arithmetic guarantee if computations use roughly $\log(C\|A\|)$ bits of precision. Our proof leverages stability of the Chebyshev recurrence to bound the stability of any polynomial approximating f. We further study the special case of $f(A) = A^{-1}$, where exact arithmetic Lanczos matches the best polynomial approximating f at each of A's eigenvalues, rather than on the full range. In seminal work, Greenbaum extends this bound to finite precision: Lanczos matches any polynomial approximating $1/x$ in a tiny range around each eigenvalue. While this bound is stronger than ours, we exhibit matrices with condition number κ where exact arithmetic Lanczos converges in $polylog(\kappa)$ iterations, but Greenbaum's bound predicts $\Omega(\kappa^{1/5})$. Our analysis raises the question of if less than $poly(\kappa)$ iterations can be expected in finite precision, even for favorable eigenvalue distributions.

Christopher Musco, Cameron Musco
Massachusetts Institute of Technology
cpmusco@mit.edu, cnmusco@mit.edu

Aaron Sidford
Stanford University
sidford@stanford.edu

CP21

Localization of Electrical Flows

We show that in any graph, the average length of a flow path in an electrical flow between the endpoints of a random edge is $O(\log^2 n)$. This is a consequence of a more general result which shows that the spectral norm of the entrywise absolute value of the transfer impedance matrix of a graph is $O(\log^2 n)$. This result implies a simple oblivious routing scheme based on electrical flows in the case of transitive graphs.

Nikhil Srivastava
Microsoft Research India
niksri@microsoft.com

Aaron Schild, Satish Rao
University of California, Berkeley
aschild@berkeley.edu, satishr@berkeley.edu

CP21

Lower Bounds for Approximating the Matching Polytope

We prove that any linear program that approximates the matching polytope on n-vertex graphs up to a factor of $(1+\epsilon)$ for any $\frac{2}{n} \leq \epsilon \leq 1$ must have at least $\binom{n}{\alpha/\epsilon}$ inequalities where $0 < \alpha < 1$ is an absolute constant. This is tight as exhibited by the $(1+\epsilon)$ approximating linear program obtained by dropping the odd set constraints of size larger than $(1+\epsilon)/\epsilon$ from the description of the matching polytope. Previously, a tight lower bound of $2^{\Omega(n)}$ was only known for $\epsilon = O\left(\frac{1}{n}\right)$ [Rothvoss, STOC '14; Braun and Pokutta, IEEE Trans. Information Theory '15] whereas for $\frac{2}{n} \leq \epsilon \leq 1$, the best lower bound was $2^{\Omega(1/\epsilon)}$ [Rothvoss, STOC '14]. The key new ingredient in our proof is a close connection to the non-negative rank of a lopsided version of the unique disjointness matrix.

Makrand Sinha
Paul G. Allen School of Computer Science & Engineering
University of Washington, Seattle
makrand@cs.washington.edu

CP22

Thin Graph Classes and Polynomial-Time Approximation Schemes

Baker (1994) devised a powerful technique to obtain approximation schemes for various problems restricted to planar graphs. Her technique can be directly extended to various other graph classes, among the most general ones the graphs avoiding a fixed apex graph as a minor. Further generalizations (e.g., to all proper minor closed graph classes) are known, but they use a combination of techniques and usually focus on somewhat restricted classes of problems. We present a new type of graph decompositions (thin systems of overlays) generalizing Baker's technique and leading to straightforward polynomial-time approximation schemes. We also show that many graph classes (all proper minor-closed classes, and all subgraph-closed classes with bounded maximum degree and strongly sublinear separators) admit such decompositions.

Zdenek Dvorak
Computer Science Institute
Charles University, Prague
rakdver@iuuk.mff.cuni.cz

CP22

Ramsey Spanning Trees and Their Applications

The metric Ramsey problem asks for the largest subset S of a metric space that can be embedded into an ultrametric (more generally into a Hilbert space) with a given distortion. Mendel and Naor devised the so called Ramsey Partitions to address this problem, and showed the algorithmic applications to approximate distance oracles. We study the natural extension of the metric Ramsey problem to graphs, and introduce the notion of Ramsey Spanning Trees. We ask for the largest subset $S \subseteq V$ of a given graph $G = (V, E)$, such that there exists a spanning tree of G that has small stretch for S. Applied iteratively, this provides a small collection of spanning trees, such that each vertex has a tree providing low stretch paths to all other vertices. We use this collection to devise the first compact stateless routing scheme with $O(1)$ routing decision time, and labels which are much shorter than in all previous schemes. We first revisit the metric Ramsey problem, and provide a new deterministic construction. We prove that for every k, any n-point metric space has a subset S of size at least $n^{1-1/k}$ which embeds into an ultrametric with distortion $8k$, providing the state-of-the-art deterministic construction of distance oracles. Next, we prove that for every k, any n-vertex graph $G = (V, E)$ has a subset S of size at least $n^{1-1/k}$, and a spanning tree of G, that has stretch

$O(k \log \log n)$ between any point in S and any point in V.

Ittai Abraham
Vmware Research
iabraham@vmware.com

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

Michael Elkin
Ben-Gurion University
elkinm@cs.bgu.ac.il

Arnold Filtser
Ben gurion university of the negev
arnoldf@cs.bgu.ac.il

Ofer Neiman
Ben-Gurion University of the Negev
neimano@cs.bgu.ac.il

CP22

Quasi-Regular Sequences and Optimal Schedules for Security Games

We study security games in which a defender commits to a mixed strategy for protecting a finite set of targets of different values. An attacker, knowing the defender's strategy, chooses which target to attack and for how long. If the attacker spends time t at a target i of value a_i without interruption by the defender, his utility is ta_i; if interrupted, his utility is 0. The defender aims to minimize the attacker's utility. The defender's strategy consists of a schedule for visiting the targets; it takes her unit time to switch between targets. Optimal defender play, although occurring in continuous time, reduces to a combinatorial question regarding the existence of infinite sequences over a finite alphabet, with the following properties for each symbol i: (1) i constitutes a prescribed fraction p_i of the sequence. (2) The occurrences of i are spread apart close to evenly: the ratio of the longest to shortest interval between consecutive occurrences is bounded by a parameter K. We call such sequences K-quasi-regular. A 1-quasi-regular sequence ensures defender optimality. However, for $K < 2$, K-quasi-regular sequences may not exist. Surprisingly, randomized 2-quasi-regular sequences also suffice for defender optimality. We show that such sequences always exist, and can be calculated efficiently. We provide several additional results on the existence of K-quasi regular and approximately optimal sequences.

David Kempe
University of Southern California
dkempe@usc.edu

Leonard Schulman
Caltech
schulman@caltech.edu

Omer Tamuz
California Institute of Technology
omertamuz@gmail.com

CP22

Erdős-Pósa Property of Chordless Cycles and Its Applications

A chordless cycle in a graph G is an induced subgraph of G which is a cycle of length at least four. We prove that the Erdős-Pósa property holds for chordless cycles, which resolves the major open question concerning the Erdős-Pósa property. Our proof for chordless cycles is constructive: in polynomial time, one can either find $k+1$ vertex-disjoint chordless cycles, or $ck^2 \log k$ vertices hitting every chordless cycle for some constant c. It immediately implies an approximation algorithm of factor $\mathcal{O}(\mathsf{opt} \log \mathsf{opt})$ for Chordal Vertex Deletion. We complement our main result by showing that the class of all chordless cycles of length at least ℓ for any fixed $\ell \geq 5$ does not have the Erdős-Pósa property. As a corollary, for a non-negative integral function w defined on the vertex set of a graph G, the minimum value $\sum_{x \in S} w(x)$ over all vertex sets S where $G - S$ is a forest is at most $\mathcal{O}(k^2 \log k)$ where k is the maximum number of cycles (not necessarily vertex-disjoint) in G such that each vertex v is used at most $w(v)$ times.

O-Joung Kwon
Technische Universität Berlin
ojoungkwon@gmail.com

Eun Jung Kim
CNRS, LAMSADE-Universite Paris Dauphine
eunjungkim78@gmail.com

CP22

A Hamiltonian Cycle in the Square of a 2-Connected Graph in Linear Time

Fleischner's theorem says that the square of every 2-connected graph contains a Hamiltonian cycle. We present a proof resulting in an $O(|E|)$ algorithm for producing a Hamiltonian cycle in the square G^2 of a 2-connected graph $G = (V, E)$. The previous best was $O(|V|^2)$ by Lau in 1980. More generally, we get an $O(|E|)$ algorithm for producing a Hamiltonian path between any two prescribed vertices, and we get an $O(|V|^2)$ algorithm for producing cycles $C_3, C_4, \ldots, C_{|V|}$ in G^2 of lengths $3, 4, \ldots, |V|$, respectively.

Stephen Alstrup
Department of Computer Science, University of Copenhagen
Denmark
s.alstrup@di.ku.dk

Agelos Georgakopoulos
Mathematics Institute, University of Warwick
UK
a.georgakopoulos@warwick.ac.uk

Eva Rotenberg
University of Copenhagen
eva@rotenberg.dk

Carsten Thomassen
Department of Applied Mathematics and Computer Science
Technical University of Denmark
ctho@dtu.dk

CP23

Sampling Random Colorings of Sparse Random

Graphs

We study the mixing properties of the Glauber dynamics for sampling k-colorings of a random graph $G(n, d/n)$ for constant d. The best known rapid mixing results for general graphs are in terms of the maximum degree Δ of the input graph G and hold when $k > 11\Delta/6$. Improved results hold when $k > \alpha\Delta$ for graphs with girth ≥ 5 and Δ sufficiently large where $\alpha = 1.7632\ldots$; further improvements on the constant α hold with stronger assumptions. For $G(n, d/n)$ the maximum degree is a function of n and the goal is to obtain results in terms of the expected degree d. The following rapid mixing results for $G(n, d/n)$ hold with high probability over the instances of $G(n, d/n)$ for large d. Mossel and Sly (2009) proved rapid mixing for constant k, and Efthymiou (2014) improved this to k linear in d. The condition was improved to $k > 3d$ by Yin and Zhang (2016) using non-MCMC methods. Here we prove rapid mixing when $k > \alpha d$ where $\alpha = 1.7632\ldots$ is the same constant as above. Moreover we obtain $O(n^3)$ mixing time of the Glauber dynamics, while in previous rapid mixing results the exponent was an increasing function in d. Our proof analyzes an appropriately defined block dynamics to "hide' high-degree vertices. One new aspect in our improved approach is utilizing so-called local uniformity properties for the analysis of block dynamics.

Charilaos Efthymiou
Goethe University Frankfurt
efthymiou@math.uni-frankfurt.de

Thomas Hayes
University of New Mexico, USA
hayes@cs.unm.edu

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

Eric Vigoda
Georgia Institute of Technology
ericvigoda@gmail.com

CP23

Uniform Generation of Random Graphs with Power-Law Degree Sequences

We give a linear-time algorithm that approximately uniformly generates a random simple graph with a power-law degree sequence whose exponent is at least 2.8811. While sampling graphs with power-law degree sequence of exponent at least 3 is fairly easy, and many samplers work efficiently in this case, the problem becomes dramatically more difficult when the exponent drops below 3; ours is the first provably practicable sampler for this case. We also show that with an appropriate rejection scheme, our algorithm can be tuned into an exact uniform sampler. The running time of the exact sampler is $O(n^{2.107})$ with high probability, and $O(n^{4.081})$ in expectation.

Pu Gao
University of Toronto
jane.gao@monash.edu

Nick Wormald
Monash University
nick.wormald@monash.edu

CP23

Tight Bounds for Coalescing-Branching Random Walks on Regular Graphs

A *Coalescing-Branching Random Walk* (COBRA) is a natural extension to the standard random walk on a graph. The process starts with one pebble at an arbitrary node. In each round every pebble splits into k pebbles, which are sent to k random neighbors. At the end of the round all pebbles at the same node coalesce into a single pebble. Besides its mathematical interest, this process is relevant as an information dissemination primitive and a basic model for the spread of epidemics. We study the *cover time* of COBRA, which is the time until each node has seen at least one pebble. Our main result is a bound of $O(\phi^{-1} \log n)$ rounds with high probability on the cover time of COBRA with $k = 2$, on any regular graph with n nodes and conductance ϕ. This improves upon all previous bounds for the problem in term of graph expansion parameters (Dutta et al. 2015, Mitzenmacher et al. 2016, Cooper et al. 2016/17). We also show that for any connected regular graph, the cover time is $O(n \log n)$ with high probability. Both bounds are asymptotically tight. Since our bounds coincide with the worst-case time bounds for push rumor spreading on regular graphs until all nodes are informed, this raises the question whether COBRA and push rumor spreading perform similarly in general. We answer this negatively by separating the cover time of COBRA and the rumor spreading time of push by a super-polylogarithmic factor on a family of tree-like regular graphs.

Petra Berenbrink
University of Hamburg
berenbrink@informatik.uni-hamburg.de

George Giakkoupis
INRIA Rennes, France
george.giakkoupis@inria.fr

Peter Kling
Universität Hamburg
peter.kling@uni-hamburg.de

CP23

Estimating Graph Parameters Via Random Walks with Restarts

In this paper we discuss the problem of estimating graph parameters from a random walk with restarts. An algorithm observes the trajectory of a random walk over an unknown graph G, starting from a vertex x. The algorithm also sees the degrees along the trajectory. The only other power that the algorithm has is to request that the random walk be reset to its initial state at any given time, based on what it has seen so far. Our main results are as follows. For *regular graphs* G, one can estimate the number of vertices n_G and the ℓ^2 mixing time of G from x in $\widetilde{O}(\sqrt{n_G}(t^G_{\text{unif}})^{3/4})$ steps, where t^G_{unif} is the uniform mixing time of the random walk. The algorithm is based on the number of intersections of random walk paths and improves on previous methods which only consider collisions. We also show that the time complexity of our algorithm is optimal (up to log factors) for 3-regular graphs with prescribed mixing times. For *general graphs*, we adapt the intersections algorithm to compute the number of edges m_G and the ℓ^2 mixing time from the starting vertex x in

$\widetilde{O}(\sqrt{m_G}\,(t^G_{\text{unif}})^{3/4})$ steps. Under mild additional assumptions, the number of vertices can also be estimated by this time. We complement these results by showing that if either m_G or the mixing time is known, the other parameter can be estimated with few steps via a self-stopping algorithm.

Anna Ben-Hamou
Université Pierre et Marie Curie
benhamou@math.univ-paris-diderot.fr

Roberto I. Oliveira
IMPA
rimfo@impa.br

Yuval Peres
Microsoft Research, Redmond
peres@microsoft.com

CP23

Comparing Mixing Times on Sparse Random Graphs

It is natural to expect that nonbacktracking random walk will mix faster than simple random walks, but so far this has only been proved in regular graphs. To analyze typical irregular graphs, let G be a random graph on n vertices with minimum degree 3 and a degree distribution that has exponential tails. We determine the precise worst-case mixing time for simple random walk on G, and show that, with high probability, it exhibits cutoff at time $\mathbf{h}^{-1}\log n$, where $\mathbf{h}$ is the asymptotic entropy for simple random walk on a Galton–Watson tree that approximates G locally. (Previously this was only known for typical starting points.) Furthermore, we show this asymptotic mixing time is strictly larger than the mixing time of nonbacktracking walk, via a delicate comparison of entropies on the Galton–Watson tree.

Anna Ben-Hamou
Université Pierre et Marie Curie
benhamou@math.univ-paris-diderot.fr

Eyal Lubetzky
Courant Institute
New York University
eyal@courant.nyu.edu

Yuval Peres
Microsoft Research, Redmond
peres@microsoft.com

CP24

Promise Constraint Satisfaction: Structure Theory and a Symmetric Boolean Dichotomy

A classic result of Schaefer [STOC, 1978] classifies all constraint satisfaction problems (CSPs) over the Boolean domain to be either in P or NP-hard. This paper considers a promise-problem variant of CSPs called PCSPs. Many problems such as approximate graph and hypergraph coloring, the $(2+\epsilon)$-SAT problem due to Austrin, Guruswami, and Håstad [SIAM Journal on Computing, 2017], and the digraph homomorphism problem can be placed in this framework. This paper is motivated by the pursuit of understanding the computational complexity of Boolean PCSPs, determining which PCSPs are polynomial-time tractable or NP-hard. As our main result, we show that PCSPs exhibits a dichotomy (it is either polynomial-time tractable or NP-hard) when the clauses are symmetric and allow for negations of variables. In particular, we show that every such polynomial-time tractable instance can be solved via either Gaussian elimination over $\mathbb{F}_2$ or a linear programming relaxation. We achieve our dichotomy theorem by extending the weak polymorphism framework of AGH which itself is a generalization of the algebraic approach used by polymorphisms to study CSPs. In both the algorithm and hardness portions of our proof, we incorporate new ideas and techniques not utilized in the CSP case.

Joshua Brakensiek, Venkatesan Guruswami
Carnegie Mellon University
jbrakens@andrew.cmu.edu, venkatg@cs.cmu.edu

CP24

The Complexity of Counting Surjective Homomorphisms and Compactions

A homomorphism from a graph G to a graph H is a function from the vertices of G to the vertices of H that preserves edges. A homomorphism is *surjective* if it uses all of the vertices of H and it is a *compaction* if it uses all of the vertices of H and all of the non-loop edges of H. Hell and Nešetřil gave a complete characterisation of the complexity of deciding whether there is a homomorphism from an input graph G to a fixed graph H. A complete characterisation is not known for surjective homomorphisms or for compactions, though there are many interesting results. Dyer and Greenhill gave a complete characterisation of the complexity of counting homomorphisms from an input graph G to a fixed graph H. In this paper, we give a complete characterisation of the complexity of counting surjective homomorphisms from an input graph G to a fixed graph H and we also give a complete characterisation of the complexity of counting compactions from an input graph G to a fixed graph H.

Jacob Focke, Leslie Ann Goldberg, Standa Živný
University of Oxford
jacob.focke@cs.ox.ac.uk, leslie.goldberg@cs.ox.ac.uk, standa.zivny@cs.ox.ac.uk

CP24

Resource-Efficient Common Randomness and Secret-Key Schemes

We study *common randomness* where two parties have access to i.i.d. samples from a known random source, and wish to generate a shared random key using limited (or no) communication with the largest possible probability of agreement. This problem is at the core of secret key generation in cryptography, with connections to communication under uncertainty and locality sensitive hashing. We take the approach of treating correlated sources as a critical resource, and ask whether common randomness can be generated *resource-efficiently*. We consider two notable sources in this setup arising from correlated bits and correlated Gaussians. We design the first *explicit* schemes that use only a *polynomial* number of samples (in the key length) so that the players can generate shared keys that agree with constant probability using optimal communication. The best previously known schemes were both nonconstructive and used an exponential number of samples. In the *amortized* setting, we characterize the largest achievable ratio of key length to communication in terms of the *external* and *internal* information costs, two well-studied

quantities in theoretical computer science. Our schemes reveal a new connection between common randomness and *unbiased error-correcting codes*, e.g., dual-BCH codes and their analogues in Euclidean space.

Badih Ghazi
MIT
badih@mit.edu

T.S. Jayram
IBM Research - Almaden
jayram@us.ibm.com

CP24

Dichotomy for Real Holant^c Problems

Holant problems capture a class of Sum-of-Product computations such as counting matchings. It is inspired by holographic algorithms and is equivalent to tensor networks, with counting CSP being a special case. A classification for Holant problems is more difficult to prove, not only because it logically implies a classification for counting CSP, but also due to the deeper reason that there exist more intricate polynomial time tractable problems in the broader framework. We discover a new family of constraint functions $\mathcal{L}$ which define polynomial time computable counting problems. These do not appear in counting CSP, and no newly discovered tractable constraints can be symmetric. It has a delicate support structure related to error-correcting codes. Local holographic transformations is fundamental in its tractability. We prove a complexity dichotomy theorem for all Holant problems defined by any real valued constraint function set on Boolean variables and contains two 0-1 pinning functions. Previously, dichotomy for the same framework was only known for *symmetric* constraint functions. The set $\mathcal{L}$ supplies the last piece of tractability. We also prove a dichotomy for a variant of counting CSP as a technical component toward this Holant dichotomy.

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

Jin-Yi Cai
University of Wisconsin, Madison
jyc@cs.wisc.edu

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

CP24

The Robust Sensitivity of Boolean Functions

The sensitivity conjecture is one of the central open problems in Boolean complexity. A recent work of Gopalan et al. [CCC 2016] conjectured a robust analog of the sensitivity conjecture, which relates the decay of the Fourier mass of a Boolean function to moments of its sensitivity. We prove the robust sensitivity conjecture in this work with near optimal parameters.

Shachar Lovett
University of California, San Diego
shachar.lovett@gmail.com

Avishay Tal
Stanford University
avishay.tal@gmail.com

Jiapeng Zhang
University of California, San Diego
jpeng.zhang@gmail.com

CP25

Reachability Preservers: New Extremal Bounds and Approximation Algorithms

Say we are given a directed graph $G = (V, E)$ on n nodes, a set of sources $S \subseteq V$ of size $|S| = n^{1/3}$, and a subset $P \subseteq S \times V$ of pairs (s, t) where $s \in S$, of size $O(n^{2/3})$, such that for all pairs $(s, t) \in P$, there is a path from s to t. Our goal is to remove as many edges from G as possible while maintaining the reachability of all pairs in P. How many edges will we have to keep? In this paper, we make polynomial progress in both the upper and lower bounds for these *Reachability Preservers* over bounds that were implicit in the literature. We show that in the above scenario, $O(n)$ edges will always be sufficient, and in general one is even guaranteed a subgraph on $O(n + \sqrt{n \cdot |P| \cdot |S|})$ edges that preserves the reachability of all pairs in P. We complement this with a lower bound graph construction, establishing that the above result fully characterizes the settings in which we are guaranteed a preserver of size $O(n)$. We also design an efficient algorithm that can always compute a preserver of existentially optimal size. The second contribution of this paper is a new connection between extremal graph sparsification results and classical Steiner Network Design problems. We improve the state of the art approximation algorithms for the most basic Steiner-type problem in directed graphs from the $O(n^{0.6+\epsilon})$ of Chlamatac, Dinitz, Kortsarz, and Laekhanukit (SODA'17) to $O(n^{0.577+\epsilon})$.

Amir Abboud
Stanford University
abboud@cs.stanford.edu

Greg Bodwin
MIT
greg.bodwin@gmail.com

CP25

Optimal Vertex Fault Tolerant Spanners (for fixed stretch)

A *k-spanner* of a graph G is a sparse subgraph H whose shortest path distances match those of G up to a multiplicative error k. In this paper we study spanners that are resistant to faults. A subgraph $H \subseteq G$ is an *f vertex fault tolerant (VFT)* k-spanner if $H \setminus F$ is a k-spanner of $G \setminus F$ for any small set F of f vertices that might "fail.' One of the main questions in the area is: what is the minimum size of an f fault tolerant k-spanner that holds for all n node graphs (as a function of f, k and n)? This question was first studied in the context of geometric graphs [Levcopoulos et al. STOC '98, Czumaj and Zhao SoCG '03] and has more recently been considered in general undirected graphs [Chechik et al. STOC '09, Dinitz and Krauthgamer PODC '11]. In this paper, we settle the question of the optimal size of a VFT spanner, in the setting where the stretch factor k is fixed. Specifically, we prove that every (undirected, possibly weighted) n-node graph G has a *$(2k-1)$-spanner* resilient to f vertex faults with $O_k(f^{1-1/k}n^{1+1/k})$ edges, and this is fully optimal (unless the famous Erdös Girth Conjecture is false). Our lower bound even generalizes

to imply that no *data structure* capable of approximating $dist_{G \setminus F}(s,t)$ similarly can beat the space usage of our spanner in the worst case.

Greg Bodwin
MIT
greg.bodwin@gmail.com

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

Merav Parter
The Weizmann Institute of Science, Rehovot, Israel.
merav.parter@weizmann.ac.il

Virginia Vassilevska Williams
MIT
virgi@mit.edu

CP25

Approximate Single Source Fault Tolerant Shortest Path

Let $G = (V, E)$ be an n-vertices m-edges directed graph with edge weights in the range $[1, W]$ and $L = \log(W)$. Let $s \in V$ be a designated source. In this paper we address several variants of the problem of maintaining the $(1+\epsilon)$-approximate shortest path from s to each $v \in V \setminus \{s\}$ in the presence of a failure of an edge or a vertex. We show that G has a subgraph H with $\widetilde{O}(nL/\epsilon)$ edges such that for any $x, v \in V$, the graph $H \setminus x$ contains a path whose length is a $(1+\epsilon)$-approximation of the length of the shortest path from s to v in $G \setminus x$. We show that the size of the subgraph H is optimal (up to logarithmic factors) by proving a lower bound of $\Omega(nL/\epsilon)$ edges. We show that there exists an $\widetilde{O}(nL/\epsilon)$ size oracle that for any $v \in V$ reports a $(1+\epsilon)$-approximate distance of v from s on a failure of any $x \in V$ in $O(\log\log_{1+\epsilon}(nW))$ time. We show that the size of the oracle is optimal (up to logarithmic factors) by proving a lower bound of $\Omega(nL/\epsilon \log n)$. We also present two *distributed algorithms*, namely single source *routing scheme* and single source *labeling scheme*.

Keerti Choudhary, Surender Baswana
IIT Kanpur
keerti@cse.iitk.ac.in, sbaswana@gmail.com

Moazzam Hussain
WorldQuant Research, India
moazzamh94@gmail.com

Liam Roditty
Bar Ilan University
liam.roditty@gmail.com

CP25

When Recursion Is Better Than Iteration: A Linear-Time Algorithm for Acyclicity with Few Error Vertices

Planarity, bipartiteness and (directed) acyclicity are basic graph properties with classic linear time recognition algorithms. However, the problems of testing whether a given (di)graph has k vertices whose deletion makes it planar, bipartite or a directed acyclic graph (DAG) are all fundamental NP-complete problems when k is part of the input. While we now know that for any fixed k, we can test in linear time whether a graph is k vertices away from being planar [FOCS 2009, SODA 2014] or bipartite [SODA 2014, SICOMP 2016], the best known algorithms in the case of directed acyclicity are the algorithm of Garey and Tarjan [IPL 78] which runs in time $O(n^{k-1}m)$ and the algorithm of Chen, Liu, Lu, O'Sullivan and Razgon [JACM 2008] which runs in time $O(k!4^k k^4 nm)$. We settle this question by giving an algorithm that decides whether a given graph is k vertices away from being acyclic, in time $O(k!4^k k^5(n+m))$. Our algorithm is designed via a general methodology that shaves off a factor of n from some algorithms that use the powerful technique of iterative compression. The two main features of our methodology are: (i) This is the first generic technique for designing linear time algorithms for *directed cut-problems* and (ii) it can be used in combination with future improvements in algorithms for the *compression* version of other well-studied cut-problems such as Multicut and Directed Subset Feedback Vertex Set.

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

Daniel Lokshtanov
University of Bergen
daniello@ii.uib.no

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

CP25

Approaching $\frac{3}{2}$ for the s-t-path TSP

We show that there is a polynomial-time algorithm with approximation guarantee $\frac{3}{2} + \epsilon$ for the s-t-path TSP, for any fixed $\epsilon > 0$. It is well known that Wolsey's analysis of Christofides' algorithm also works for the s-t-path TSP with its natural LP relaxation except for the *narrow cuts* (in which the LP solution has value less than two). A fixed optimum tour has either a single edge in a narrow cut (then call the edge and the cut *lonely*) or at least three (then call the cut *busy*). Our algorithm "guesses' (by dynamic programming) lonely cuts and edges. Then we partition the instance into smaller instances and strengthen the LP, requiring value at least three for busy cuts. By setting up a k-stage recursive dynamic program, we can compute a spanning tree (V, S) and an LP solution y such that $(\frac{1}{2} + O(2^{-k}))y$ is in the T-join polyhedron, where T is the set of vertices whose degree in S has the wrong parity.

Vera Traub, Jens Vygen
University of Bonn
traub@or.uni-bonn.de, vygen@or.uni-bonn.de

CP26

Spatial Mixing and Non-Local Markov Chains

We consider spin systems with nearest-neighbor interactions on an n-vertex d-dimensional cube of the integer lattice graph $\mathbb{Z}^d$. We study the effects that the strong spatial mixing condition (SSM) has on the rate of convergence to equilibrium distribution of *non-local* Markov chains. We prove that SSM implies $O(\log n)$ mixing of a *block dynamics* whose steps can be implemented efficiently. We then develop a comparison methodology that allows us to extend this result to other non-local dynamics. As a first application of our method we prove that, if SSM holds, then the

relaxation time (i.e., the inverse spectral gap) of general block dynamics is $O(r)$, where r is the number of blocks. As a second application of our technology we show that SSM implies an $O(1)$ bound for the relaxation time of the Swendsen-Wang dynamics for the ferromagnetic Ising and Potts models. We also prove that for *monotone* spin systems SSM implies that the mixing time of systematic scan dynamics is $O(\log n(\log\log n)^2)$. Our proofs use a variety of techniques for the analysis of Markov chains including coupling, functional analysis and linear algebra.

Antonio Blanca
U.C. Berkeley
ablanca@cc.gatech.edu

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

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

Eric Vigoda
Georgia Institute of Technology
vigoda@cc.gatech.edu

CP26
Testing Ising Models

Given samples from an unknown multivariate distribution p, is it possible to distinguish whether p is the product of its marginals versus p being far from every product distribution? Similarly, is it possible to distinguish whether p equals a given distribution q versus p and q being far from each other? These problems of testing independence and goodness-of-fit have received enormous attention in statistics, information theory, and theoretical computer science, with sample-optimal algorithms known in several interesting regimes of parameters. Unfortunately, it has also been understood that these problems become intractable in large dimensions, necessitating exponential samples. Motivated by the exponential lower bounds for general distributions as well as the ubiquity of Markov Random Fields (MRFs) in the modeling of high-dimensional distributions, we initiate the study of distribution testing on structured multivariate distributions, and in particular the prototypical example of MRFs: the Ising Model. We demonstrate that, in this structured setting, we can avoid the curse of dimensionality, obtaining sample and time efficient testers for independence and goodness-of-fit. Along the way, we develop new tools for bounding the variance of functions of the Ising model, using and improving upon the exchangeable pairs framework developed by Chatterjee. In particular, we prove variance bounds for multi-linear functions of the Ising model in the high-temperature regime.

Constantinos Daskalakis, Nishanth Dikkala, Gautam Kamath
MIT
costis@csail.mit.edu, nishanthd@csail.mit.edu, g@csail.mit.edu

CP26
Exponentially Slow Mixing in the Mean-Field Swendsen-Wang Dynamics

Swendsen–Wang dynamics for the Potts model was proposed in the late 1980's as an alternative to single-site heat-bath dynamics, in which global updates allow this MCMC sampler to switch between metastable states and ideally mix faster. Gore and Jerrum (1997) found that this dynamics may in fact exhibit slow mixing: they showed that, for the Potts model with $q \geq 3$ colors on the complete graph on n vertices at the critical point $\beta_c(q)$, Swendsen–Wang dynamics has $t_{\text{MIX}} \geq \exp(c\sqrt{n})$. Galanis *et al.* (2015) showed that $t_{\text{MIX}} \geq \exp(cn^{1/3})$ throughout the critical window (β_s, β_S) around β_c, and Blanca and Sinclair (2015) established that $t_{\text{MIX}} \geq \exp(c\sqrt{n})$ in the critical window for corresponding mean-field FK model, which implied the same bound for Swendsen–Wang via known comparison estimates. In both cases, an upper bound of $t_{\text{MIX}} \leq \exp(c'n)$ was known. Here we show that the mixing time is truly exponential in n: namely, $t_{\text{MIX}} \geq \exp(cn)$ for Swendsen–Wang dynamics when $q \geq 3$ and $\beta \in (\beta_s, \beta_S)$, and the same bound holds for the related MCMC samplers for the mean-field FK model when $q > 2$.

Reza Gheissari, Eyal Lubetzky
Courant Institute
New York University
reza@cims.nyu.edu, eyal@courant.nyu.edu

Yuval Peres
Microsoft Research, Redmond
peres@microsoft.com

CP26
The Diameter of Dense Random Regular Graphs

There is a tight upper bound on the order (the number of vertices) of d-regular graphs of diameter D, known as the Moore bound in graph theory. This bound yields the lower bound $D_0(n,d)$ of the diameter of d-regular graphs of order n. Actually, the diameter $\text{diam}(G_{n,d})$ of a random d-regular graph $G_{n,d}$ of order n is known to be the same as $D_0(n,d)$ up to a factor $1+o(1)$ for $3 \leq d = O(1)$, whereas there exists a gap $\text{diam}(G_{n,d}) - D_0(n,d) = \Omega(\log\log n)$. In this paper, we investigate the gap $\text{diam}(G_{n,d}) - D_0(n,d)$ for $d = (\beta + o(1))n^\alpha$ where $\alpha \in (0,1)$ and $\beta > 0$ are any constants. We show that for such a d, $\text{diam}(G_{n,d}) = \lfloor \alpha^{-1} \rfloor + 1$ with high probability. Our result yields that the gap is 1 if α^{-1} is an integer and $d \geq n^\alpha$, and is 0 otherwise. The upper bound of $\text{diam}(G_{n,d})$ follows from the embedding theorem due to Dudek et al. We obtain the lower bound of $\text{diam}(G_{n,d})$ by the careful analysis of the shortest path lengths between fixed vertex pairs.

Nobutaka Shimizu
The University of Tokyo
nobutaka_shimizu@mist.i.u-tokyo.ac.jp

CP26
Consensus of Interacting Particle Systems on Erdös-Rényi Graphs

Interacting Particle Systems—exemplified by the voter model, iterative majority, and iterative $k-$majority processes—have found use in many disciplines including distributed systems, statistical physics, social networks, and Markov chain theory. In these processes, nodes update their "opinion' according to the frequency of opinions amongst their neighbors. We propose a family of models parameterized by an update function that we call Node Dynamics: every node initially has a binary opinion. At each round a node is uniformly chosen and randomly updates its opinion with the probability distribution specified by the value of the update function applied to the frequencies of its

neighbors' opinions. In this work, we prove that the Node Dynamics converge to consensus in time $\Theta(n \log n)$ in complete graphs and dense Erdös-Rényi random graphs when the update function is from a large family of "majority-like' functions. Our technical contribution is a general framework that upper bounds the consensus time. In contrast to previous work that relies on handcrafted potential functions, our framework systematically constructs a potential function based the state space structure.

Grant Schoenebeck, Fang-Yi Yu
University of Michigan
schoeneb@umich.edu, fayu@umich.edu

CP27

Separation in Correlation-Robust Monopolist Problem with Budget

We consider a monopolist seller that has n heterogeneous items to sell to a single buyer. The seller's goal is to maximize her revenue. We study this problem in the correlation-robust framework recently proposed by Carroll [Econometrica 2017]. In this framework, the seller only knows marginal distributions for each separate item but has no information about correlation across different items in the joint distribution. Any mechanism is then evaluated according to its expected profit in the worst-case, over all possible joint distributions with given marginal distributions. Carroll's main result states that in multi-item monopoly problem with buyer, whose value for a set of items is additive, the optimal correlation-robust mechanism should sell items separately. We use alternative dual Linear Programming formulation for the optimal correlation-robust mechanism design problem. This LP can be used to compute optimal mechanisms in general settings. We give an alternative proof for the additive monopoly problem without constructing worst-case distribution. As a surprising byproduct of our approach we get that separation result continues to hold even when buyer has a budget constraint on her total payment. Namely, the optimal robust mechanism splits the total budget in a fixed way across different items independent of the bids, and then sells each item separately with a respective per item budget constraint.

Pinyan Lu, Nick Gravin
Shanghai University of Finance and Economics
lu.pinyan@mail.shufe.edu.cn, nikolai@mail.sufe.edu.cn

CP27

On the Competition Complexity of Dynamic Mechanism Design

The Competition Complexity of an auction measures how much competition is needed for the revenue of a simple auction to surpass the optimal revenue. A classic result by Bulow and Klemperer, states that the Competition Complexity of VCG, in the case of n i.i.d. buyers and one item, is 1. In other words, it is better to invest in recruiting one extra buyer and run VCG than to invest in learning the buyers' underlying distribution and run the revenue maximizing auction tailored to this distribution. In this paper we study the Competition Complexity of dynamic auctions. Consider the following problem: a monopolist is selling m items in m consecutive stages to n interested buyers. A buyer realizes her value for item k in the beginning of stage k. How many additional buyers are necessary and sufficient for VCG at each stage to extract revenue at least that of the optimal dynamic auction? We prove that the Competition Complexity of dynamic auctions is at most 3n and at least linear in n, even when the buyers' values are correlated across stages, under a monotone hazard rate assumption on the stage distributions. We also prove results on the number of additional buyers necessary for VCG at every stage to be an α-approximation of the optimal revenue. As a corollary we provide the first results on prior-independent dynamic auctions. This is, to the best of our knowledge, the first non-trivial positive guarantees for simple ex-post IR dynamic auctions for correlated stages.

Siqi Liu
UC Berkeley
sliu18@berkeley.edu

Alexandros Psomas
Carnegie Mellon University
cpsomas@cs.cmu.edu

CP27

Revenue Maximization with an Uncertainty-Averse Buyer

Most work in mechanism design assumes that buyers are risk neutral; some considers risk aversion arising due to a non-linear utility for money. Yet behavioral studies have established that real agents exhibit risk attitudes which cannot be captured by any expected utility model. We initiate the study of revenue-optimal mechanisms under behavioral models beyond expected utility theory. We adopt a model from prospect theory which arose to explain these discrepancies and incorporates agents under-weighting uncertain outcomes. In our model, an event occurring with probability $x < 1$ is worth strictly less to the agent than x times the value of the event when it occurs with certainty. We present three main results. First, we characterize optimal mechanisms as menus of two-outcome lotteries. Second, we show that under a reasonable bounded-risk-aversion assumption, posted pricing obtains a constant approximation to the optimal revenue. Notably, this result is "risk-robust' in that it does not depend on the details of the buyer's risk attitude. Third, we consider dynamic settings in which the buyers uncertainty about his future value may allow the seller to extract more revenue. In contrast to the positive result above, here we show it is not possible to achieve any constant-factor approximation to revenue using deterministic mechanisms in a risk-robust manner.

Shuchi Chawla
Department of Computer Sciences
University of Wisconsin-Madison
shuchi@cs.wisc.edu

Kira Goldner
Oberlin College
kgoldner@cs.washington.edu

J. Benjamin Miller
University of Wisconsin-Madison
bmiller@cs.wisc.edu

Emmanouil Pountourakis
University of Texas at Austin
manolis@utexas.edu

CP27

On the Complexity of Simple and Optimal Deter-

ministic Mechanisms for An Additive Buyer

We show that the problem of implementing a Revenue-Optimal Deterministic Mechanism Design problem for a single additive buyer is #P-hard, even when the distributions have support size 2 for each item and, more importantly, even when the optimal solution is guaranteed to be of a very simple kind: the seller picks a price for each individual item and a price for the grand bundle of all the items; the buyer can purchase either the grand bundle at its given price or any subset of items at their total individual prices. The following problems are also #P-hard, as immediate corollaries of the proof:

1. determining if individual item pricing is optimal for a given instance,
2. determining if grand bundle pricing is optimal, and
3. computing the optimal (deterministic) revenue.

On the positive side, we show that when the distributions are i.i.d. with support size 2, the optimal revenue obtainable by any mechanism, even a randomized one, can be achieved by a simple solution of the above kind (individual item pricing with a discounted price for the grand bundle) and furthermore, it can be computed in polynomial time. The problem can be solved in polynomial time too when the number of items is constant.

Dimitris Paparas, Xi Chen, Mihalis Yannakakis, George Matikas
Columbia University
paparas@cs.wisc.edu, xichen@cs.columbia.edu, mihalis@cs.columbia.edu, matikas@cs.columbia.edu

CP27

The Menu Complexity of 'One-and-a-Half-Dimensional' Mechanism Design

We study the menu complexity of optimal and approximately-optimal auctions in the context of the "FedEx' problem, a so-called "one-and-a-half-dimensional' setting where a single bidder has both a value and a deadline for receiving an item[Fiat:2016]. The menu complexity[HartN13] of an auction is equal to the number of distinct (allocation, price) pairs that a bidder might receive. We show the following when the bidder has n possible deadlines:

- **Exponential menu complexity is necessary to be exactly optimal:** There exist instances where the optimal mechanism has menu complexity $\geq 2^n - 1$. This matches *exactly* the upper bound provided by Fiat et al.'s algorithm, and resolves one of their open questions.
- **Fully polynomial menu complexity is necessary and sufficient for approximation:** For all instances, there exists a mechanism guaranteeing a multiplicative $(1-\epsilon)$-approximation to the optimal revenue with menu complexity $O(n^{3/2}\sqrt{\frac{\min\{n/\epsilon, \ln(v_{\max})\}}{\epsilon}}) = O(n^2/\epsilon)$, where $v_{\max}$ denotes the largest value in the support of integral distributions.
- There exist instances where any mechanism guaranteeing a multiplicative $(1-O(1/n^2))$-approximation to the optimal revenue requires menu complexity $\Omega(n^2)$.

Our main technique is the polygon approximation of concave functions[Rote91], and our results here should be of independent interest.

Raghuvansh R. Saxena, Ariel Schvartzman, Matt Weinberg
Princeton University
rrsaxena@princeton.edu, schvartzman.ariel@gmail.com, smweinberg@princeton.edu

CP28

Improved Bounds for Testing Forbidden Order Patterns

A sequence $f\colon \{1,\ldots,n\} \to \mathbb{R}$ contains a permutation π of length k if there exist $i_1 < \ldots < i_k$ such that, for all x, y, $f(i_x) < f(i_y)$ if and only if $\pi(x) < \pi(y)$; otherwise, f is said to be π-free. In this work, we consider the problem of testing for π-freeness with one-sided error, continuing the investigation of [Newman et al., SODA'17]. We demonstrate a surprising behavior for non-adaptive tests with one-sided error: While a trivial sampling-based approach yields an ε-test for π-freeness making $\Theta(\varepsilon^{-1/k}n^{1-1/k})$ queries, our lower bounds imply that this is almost optimal for most permutations! Specifically, for most permutations π of length k, any non-adaptive one-sided ε-test requires $\varepsilon^{-1/(k-\Theta(1))}n^{1-1/(k-\Theta(1))}$ queries; furthermore, the permutations that are hardest to test require $\Theta(\varepsilon^{-1/(k-1)}n^{1-1/(k-1)})$ queries, which is tight in n and ε. Additionally, we show two hierarchical behaviors here. First, for any k and $l \leq k-1$, there exists some π of length k that requires $\tilde{\Theta}_\varepsilon(n^{1-1/l})$ non-adaptive queries. Second, we show an adaptivity hierarchy for $\pi = (1,3,2)$ by proving upper and lower bounds for (one- and two-sided) testing of π-freeness with r rounds of adaptivity. The results answer open questions of Newman et al. and [Canonne and Gur, CCC'17].

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

Clément Canonne
Columbia University
ccanonne@cs.columbia.edu

CP28

A $o(d)\cdot$ polylog n Monotonicity Tester for Boolean Functions over the Hypergrid $[n]^d$

We study monotonicity testing of Boolean functions over the hypergrid $[n]^d$ and design a non-adaptive tester with 1-sided error whose query complexity is $\tilde{O}(d^{5/6})\cdot$ polylog n. Previous to our work, the best known testers had query complexity linear in d but independent of n. We improve upon these testers as long as $n = 2^{d^{o(1)}}$. To obtain our results, we work with what we call the *augmented hypergrid*, which adds extra edges to the hypergrid. Our main technical contribution is a Margulis-style isoperimetric result for the augmented hypergrid, and our tester (like previous testers for the hypercube domain) performs directed random walks on this structure.

Hadley Black
University of California, Santa Cruz
hablack@ucsc.edu

Deeparnab Chakrabarty
Dartmouth College
deeparnab.chakrabarty@dartmouth.edu

C. Seshadhri
University of California, Santa Cruz

sesh@ucsc.edu

CP28

Tolerant Junta Testing and the Connection to Submodular Optimization and Function Isomorphism

A function $f : \{0,1\}^n \to \{0,1\}$ is a k-junta if it depends on at most k of its variables. We consider the problem of tolerant testing of k-juntas, where the testing algorithm must accept any function that is ϵ-close to some k-junta and reject any function that is ϵ'-far from every k'-junta for some $\epsilon' = O(\epsilon)$ and $k' = O(k)$. Our first result is an algorithm that solves this problem with query complexity polynomial in k and $1/\epsilon$. This result is obtained via a new polynomial-time approximation algorithm for submodular function minimization under large cardinality constraints, which holds even when only given an approximate oracle access to the function. Our second result considers the case where $k' = k$. We show how to obtain a smooth tradeoff between the amount of tolerance and the query complexity in this setting. Specifically, we design an algorithm that given $\rho \in (0, 1/2)$ accepts any function that is $\epsilon\rho/8$-close to some k-junta and rejects any function that is ϵ-far from every k-junta. The query complexity of the algorithm is $O(\frac{k \log k}{\epsilon\rho(1-\rho)^k})$. Finally, we show how to apply the second result to the problem of tolerant isomorphism testing between two unknown Boolean functions f and g. We give an algorithm for this problem whose query complexity only depends on the (unknown) smallest k such that either f or g is close to being a k-junta.

Eric Blais
University of Waterloo
eric.blais@uwaterloo.ca

Clément Cannone
Columbia University
ccanonne@cs.columbia.edu

Talya Eden
Tel Aviv University
talyaa01@gmail.com

Amit Levi
University of Waterloo
amit.levi@uwaterloo.ca

Dana Ron
Tel-Aviv University
danaron@tau.ac.il

CP28

Testing Bounded Arboricity

In this paper we consider the problem of testing whether a graph has bounded arboricity. Graphs with bounded arboricity have been studied extensively in the past, in particular since for many problems they allow for much more efficient algorithms and/or better approximation ratios. We present a tolerant tester in the sparse-graphs model. The sparse-graphs model allows access to degree queries and neighbor queries, and the distance is defined with respect to the actual number of edges. More specifically, our algorithm distinguishes between graphs that are ϵ-close to having arboricity α and graphs that $c \cdot \epsilon$-far from having arboricity 3α, where c is an absolute small constant. The query complexity and running time of the algorithm are $\tilde{O}\left(\frac{n}{\sqrt{m}} \cdot \frac{\log(1/\epsilon)}{\epsilon} + \frac{n \cdot \alpha}{m} \cdot \left(\frac{1}{\epsilon}\right)^{O(\log(1/\epsilon))}\right)$ where n denotes the number of vertices and m denotes the number of edges. In terms of the dependence on n and m this bound is optimal up to poly-logarithmic factors. We leave it as an open question whether the dependence on $1/\epsilon$ can be improved from quasi-polynomial to polynomial. Our techniques include an efficient local simulation for approximating the outcome of a global (almost) forest-decomposition algorithm as well as a tailored procedure of edge sampling.

Reut Levi
Weizmann Institute
reuti.levi@gmail.com

Talya Eden
Tel-Aviv University
talyaa01@gmail.com.

Dana Ron
Tel-Aviv Univeristy
danaron@tau.ac.il.

CP28

Cache-Oblivious and Data-Oblivious Sorting and Applications

Although external-memory sorting has been a classical algorithms abstraction and has been heavily studied in the literature, perhaps somewhat surprisingly, when *data-obliviousness* is a requirement, even very rudimentary questions remain open. Prior to our work, it is not even known how to construct a *comparison-based*, external-memory *oblivious* sorting algorithm that is optimal in IO-cost. We make a significant step forward in our understanding of external-memory, oblivious sorting algorithms. Not only do we construct a *comparison-based*, external-memory oblivious sorting algorithm that is optimal in IO-cost, our algorithm is also *cache-agnostic* in that the algorithm need not know the storage hierarchy's internal parameters such as the cache and cache-line sizes. Our result immediately implies a *cache-agnostic* ORAM construction whose asymptotic IO-cost matches the best known *cache-aware* scheme. Last but not the least, we propose and adopt a new and stronger security notion for external-memory, oblivious algorithms and argue that this new notion is desirable for resisting possible cache-timing attacks. Thus our work also lays a foundation for the study of oblivious algorithms in the cache-agnostic model.

T-H. Hubert Chan
The University of Hong Kong
hubert@cs.hku.hk

Yue Guo, Wei-Kai Lin, Elaine Shi
Cornell University
yg393@cornell.edu, wklin@cs.cornell.edu, runting@gmail.com

CP29

Tight Analysis of Parallel Randomized Greedy Mis

We provide a tight analysis which settles the round complexity of the well-studied *parallel randomized greedy MIS* algorithm, thus answering the main open question of Blelloch, Fineman, and Shun [SPAA'12]. The parallel/distributed randomized greedy Maximal Independent Set (MIS) algorithm works as follows. An order of the

vertices is chosen uniformly at random. Then, in each round, all vertices that appear before their neighbors in the order are added to the independent set and removed from the graph along with their neighbors. The main question of interest is the number of rounds it takes until the graph is empty. This algorithm has been studied since 1987, initiated by Coppersmith, Raghavan, and Tompa [FOCS'87], and the previously best known bounds were $O(\log n)$ rounds in expectation for Erdős-Rényi random graphs by Calkin and Frieze [Random Struc. & Alg. '90] and $O(\log^2 n)$ rounds with high probability for general graphs by Blelloch, Fineman, and Shun [SPAA'12]. We prove a high probability upper bound of $O(\log n)$ on the round complexity of this algorithm in general graphs, and that this bound is tight. This also shows that parallel randomized greedy MIS is as fast as the celebrated algorithm of Luby [STOC'85, JALG'86]

Manuela Fischer, Andreas Noever
ETH Zurich
manuela.fischer@inf.ethz.ch, anoever@inf.ethz.ch

CP29

Space-Optimal Majority in Population Protocols

Population protocols are a model of distributed computing, in which n agents with limited local state interact randomly, and cooperate to collectively compute global predicates. Majority is a central task, in which agents need to collectively reach a decision as to which one of two states A or B had a higher initial count. Two important complexity metrics are the time that a protocol requires to stabilize to an output decision, and the state space size that each agent requires. It is known that majority requires $\Omega(\log \log n)$ states per agent to allow for poly-logarithmic time stabilization, and that $O(\log^2 n)$ states are sufficient. Thus, there is an exponential gap between the upper and lower bounds. We provide a new lower bound of $\Omega(\log n)$ states for any protocol which stabilizes in $O(n^{1-c})$ time, for any $c > 0$. This result is conditional on basic monotonicity and output assumptions, satisfied by all known protocols. Technically, it represents a significant departure from previous lower bounds. We give an algorithm for majority which uses $O(\log n)$ states, and stabilizes in $O(\log^2 n)$ time. Central to the algorithm is a new leaderless phase clock, which allows nodes to synchronize in phases of $\Theta(n \log n)$ consecutive interactions using $O(\log n)$ states per node. We also employ our phase clock to build a leader election algorithm with $O(\log n)$ states, which stabilizes in $O(\log^2 n)$ time.

Rati Gelashvili
MIT
gelash@cs.toronto.edu

Dan Alistarh
ETH Zurich
dan.alistarh@inf.ethz.ch

James Aspnes
Yale
james.aspnes@yale.edu

CP29

Derandomized Concentration Bounds for Polynomials, and Hypergraph Maximal Independent Set

A parallel algorithm for maximal independent set (MIS) in hypergraphs has been a long-standing algorithmic challenge, dating back nearly 30 years to a survey of Karp & Ramachandran (1990). Despite its apparent simplicity, there have been no general sub-polynomial-time algorithms or hardness reductions. The best randomized parallel algorithm for hypergraphs of fixed rank r was developed by Beame & Luby (1990) and Kelsen (1992), running in time roughly $(\log n)^{r!}$. The key probabilistic tool of this algorithm is a concentration bound for low-degree polynomials applied to independent input variables; this is a natural generalization of concentration bounds for sums of independent random variables, which are ubiquitous in combinatorics and computer science. The algorithm of Kelsen cannot be derandomized in a standard way, and there are no deterministic parallel algorithms for hypergraph MIS known for any fixed rank $r > 3$. We improve the randomized algorithm of Kelsen to obtain a running time of $(\log n)^{2^r}$. We also give a method for derandomizing concentration bounds for polynomials, thus obtaining a deterministic algorithm running in $(\log n)^{2^{r+3}}$ time and $(mn)^{O(1)}$ processors. Our analysis can also apply when r is slowly growing; using this in conjunction with a strategy of Bercea et al. (2015) gives a deterministic MIS algorithm running in time $\exp(O(\frac{\log m}{\log \log m} + \log \log n)$.

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

CP29

Community Detection on Euclidean Random Graphs

We study Community Detection (CD) on a class of sparse spatial random graphs embedded in the Euclidean space. Our graph is the planted-partition version of the random connection model studied in Stochastic Geometry. Each node has two labels - an i.i.d. uniform $\{-1, +1\}$ valued community label and a $\mathbb{R}^d$ valued location label which form the support of a Poisson Point Process of intensity λ on $\mathbb{R}^d$. Conditional on the labels, edges are drawn independently at random depending on both the community and location labels of nodes. The CD problem then consists in estimating the partition of nodes into communities better than at random, based on an observation of the random graph and the spatial location labels on nodes. We establish a non-trivial phase-transition for this problem in terms of λ. We show that for small λ, there exists no algorithm for CD. For large λ, we propose an algorithm which solves CD efficiently. In certain special cases, we establish the exact threshold on λ which separates the existence of an algorithm which solves CD from the impossibility of such an algorithm. We also establish a distinguishability result which says that one can always efficiently infer the existence of a partition given the graph and the spatial locations even when one cannot identify the partition better than at random. This is a new phenomenon not observed thus far in any Erdős-Rényi based models.

Abishek Sankararaman
The University of Texas at Austin
abishek@utexas.edu

François Baccelli
The University of Texas At Austin

francois.baccelli@austin.utexas.edu

CP29

Approximate Positive Correlated Distributions and Approximation Algorithms for D-Optimal Design

Experimental design is a classical area in statistics. In the combinatorial experimental design problem, the aim is to estimate an unknown m-dimensional vector x from linear measurements where a Gaussian noise is introduced in each measurement. The goal is to pick k out of the given n experiments so as to make the most accurate estimate of the unknown parameter x. Given a set S of chosen experiments, the most likelihood estimate x' can be obtained by a least squares computation. One of the robust measures of error estimation is the D-optimality criterion which aims to minimize the generalized variance of the estimator. The problem gives rise to two natural variants depending on whether repetitions of experiments is allowed or not. We show a close connection between approximation algorithms for the D-optimal design problem and constructions of *approximately m-wise positively correlated distributions*. This connection allows us to obtain a $\frac{1}{e}$-approximation for the D-optimal design problem with and without repetitions giving the first constant factor approximation for the problem. We then consider the case when the number of experiments chosen is much larger than the dimension m and show one can obtain $(1-\epsilon)$-approximation if $k \geq \frac{2m}{\epsilon}$ when repetitions are allowed and if $k = O\left(\frac{m}{\epsilon} + \frac{1}{\epsilon^2} \cdot \log \frac{1}{\epsilon}\right)$ when no repetitions are allowed improving on previous work.

Weijun Xie
Virginia Tech
wxie@vt.edu

Mohit Singh
H. Milton Stewart School of Industrial & Systems Engineering
Georgia Institute of Technology
mohitsinghr@gmail.com

CP30

From Battlefields to Elections: Winning Strategies of Blotto and Auditing Games

Mixed strategies are often evaluated based on the expected payoff that they guarantee. This is not always desirable. In this paper, we consider games for which maximizing the expected payoff deviates from the actual goal of the players. To address this issue, we introduce the notion of a (u, p)-maxmin strategy which ensures receiving a minimum utility of u with probability at least p. We then give approximation algorithms for the problem of finding a (u, p)-maxmin strategy for these games. We consider the classic Colonel Blotto game. Two colonels divide their troops among a set of battlefields. Each battlefield is won by the colonel that puts more troops in it. The payoff of each colonel is the weighted number of battlefields that she wins. The Colonel Blotto game has found applications in the analysis of many different forms of competition: from sports to advertisement, to politics. We show that if we maximize the expected payoff of a player, it does not necessarily maximize the winning probability of that player for certain applications of Colonel Blotto. We give an exact algorithm for a natural variant of the continuous version of this game. More generally, we provide constant and logarithmic approximation algorithms that approximate the optimal strategies of the players in this game when the goal of the players is to win a race rather than maximizing the expected payoff.

Soheil Behnezhad
University of Maryland
soheil@cs.umd.edu

Avrim Blum
Carnegie Mellon University
avrim@cs.cmu.edu

Mahsa Derakhshan
University of Maryland
mah.derakhshan@gmail.com

MohammadTaghi Hajiaghayi
University of Maryland, College Park
hajiagha@cs.umd.edu

Mohammad Mahdian
Google Research
mahdian@google.com

Christos H. Papadimitriou
?University of California at Berkeley
christos@berkeley.edu

Ronald L. Rivest
Massachusetts Institute of Technology
rivest@mit.edu

Saeed Seddighin
University of Maryland, College Park
saeedreza.seddighin@gmail.com

Philip B. Stark
University of California at Berkeley
stark@stat.berkeley.edu

CP30

Approximating the Nash Social Welfare with Budget-Additive Valuations

We present the first constant-factor approximation algorithm for maximizing the Nash social welfare when allocating indivisible items to agents with budget-additive valuation functions. Budget-additive valuations represent an important class of submodular functions. They have attracted a lot of research interest in recent years due to many interesting applications. For every $\epsilon > 0$, our algorithm obtains a $(2.404 + \epsilon)$-approximation in time polynomial in the input size and $1/\epsilon$. Our algorithm relies on rounding an approximate equilibrium in a linear Fisher market where sellers have *earning limits* (upper bounds on the amount of money they want to earn) and buyers have *utility limits* (upper bounds on the amount of utility they want to achieve). In contrast to markets with *either* earning *or* utility limits, these markets have not been studied before. They turn out to have fundamentally different properties. Although the existence of equilibria is not guaranteed, we show that the market instances arising from the Nash social welfare problem always have an equilibrium. Further, we show that the set of equilibria is not convex, answering a question of Cole et al. (EC 2017). We design an FPTAS to compute an approximate equilibrium, a result that may be of independent interest.

Jugal Garg

UNIV OF ILLINOIS AT URBANA CHAMPAIGN
jugal@illinois.edu

Martin Hoefer
Goethe-Universitat Frankfurt am Main
mhoefer@mpi-inf.mpg.de

Kurt Mehlhorn
MPI for Informatics
Saarbruecken, Germany
mehlhorn@mpi-inf.mpg.de

CP30

Nash Social Welfare for Indivisible Items under Separable, Piecewise-Linear Concave Utilities

Recently Cole and Gkatzelis gave the first constant factor approximation algorithm for the problem of allocating indivisible items to agents, under additive valuations, so as to maximize the Nash Social Welfare (NSW). We give constant factor algorithms for a substantial generalization of their problem – to the case of separable, piecewise-linear concave utility functions. We give two such algorithms, the first using market equilibria and the second using the theory of real stable polynomials. Both approaches require new algorithmic ideas.

Tung Mai
Georgia Institute of Technology
tung.mai@cc.gatech.edu

Nima Anari
Stanford University
anari@stanford.edu

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

Vijay Vazirani
University of California, Irvine
vazirani@cc.gatech.edu.

CP30

On Simultaneous Two-Player Combinatorial Auctions

We consider the following communication problem: Alice and Bob each have some valuation functions $v_1(\cdot)$ and $v_2(\cdot)$ over subsets of m items, and their goal is to partition the items into $S, \bar{S}$ in a way that maximizes the *welfare*, $v_1(S) + v_2(\bar{S})$. We study both the *allocation problem*, which asks for a welfare-maximizing partition and the *decision problem*, which asks whether or not there exists a partition guaranteeing certain welfare, for binary XOS valuations. For interactive protocols with poly(m) communication, a tight 3/4-approximation is known. For interactive protocols, the allocation problem is provably *harder* than the decision problem: any solution to the allocation problem implies a solution to the decision problem with one additional round and $\log m$ additional bits of communication via a trivial reduction. Surprisingly, the allocation problem is provably *easier* for simultaneous protocols. In other words, this trivial reduction from decision to allocation problems provably requires the extra round of communication. We further discuss the implications of our results for the design of truthful combinatorial auctions in general, and extensions to general XOS valuations. In particular, our protocol for the allocation problem implies a new style of truthful mechanisms.

Mark Braverman, Jieming Mao, S. Matthew Weinberg
Princeton University
mbraverman@cs.princeton.edu, jiemingm@cs.princeton.edu, smweinberg@princeton.edu

CP30

A New Class of Combinatorial Markets with Covering Constraints: Algorithms and Applications

We introduce a new class of combinatorial markets in which agents have covering constraints over resources required and are interested in delay minimization. Our market model is applicable to several settings including scheduling and communicating over a network. This model is quite different from the traditional models, to the extent that neither do the classical equilibrium existence results seem to apply to it nor do any of the efficient algorithmic techniques developed to compute equilibria. In particular, our model does not satisfy the condition of non-satiation, which is used critically to show the existence of equilibria in traditional market models and we observe that our set of equilibrium prices could be a connected, non-convex set. We give a proof of the existence of equilibria and a polynomial time algorithm for finding one, drawing heavily on techniques from LP duality and submodular minimization. Finally, we show that our model inherits many of the fairness properties of traditional equilibrium models as well as new models, such as CEEI.

Sadra Yazdanbod
Georgia Institute of Technology
yazdanbod@gatech.edu

Nikhil R. Devanur
Microsoft Research, Redmond
nikdev@microsoft.com

Jugal Garg
UNIV OF ILLINOIS AT URBANA CHAMPAIGN
jugal@illinois.edu

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

Vijay V. Vazirani
Georgia Institute of Technology
vazirani@cc.gatech.edu

CP31

Tight Bounds on the Round Complexity of the Distributed Maximum Coverage Problem

We study the distributed maximum coverage problem in which the input sets $S_1, \ldots, S_m$ over universe $[n]$ are partitioned across p machines and the goal is to find k sets whose union covers the most number of elements. The computation proceeds in rounds. In each round, all machines simultaneously send a message to a central coordinator who then communicates back to the machines a summary to guide the computation for the next round. At the end, the coordinator outputs the answer. The main measures of efficiency in this setting are the approximation ratio of the returned solution, the communication cost of each machine, and the number of rounds of computation.

Our main result is an asymptotically tight bound on the tradeoff between these measures for the maximum coverage problem. We show that any r-round protocol for this problem either incurs a communication cost of $k \cdot m^{\Omega(1/r)}$ or only achieves an approximation factor of $k^{\Omega(1/r)}$. We then complement our lower bound by showing that there exist an r-round protocol that achieves an $\frac{e}{e-1}$-approximation with a communication cost of $k \cdot m^{O(1/r)}$ as well as an r-round protocol that achieves a $k^{O(1/r)}$-approximation with only $\tilde{O}(n)$ communication per machine. We further use our results in this setting to obtain new bounds for the maximum coverage problem in other main models of computation for massive datasets such as dynamic streams and the MapReduce model.

Sepehr Assadi, Sanjeev Khanna
University of Pennsylvania
sassadi@cis.upenn.edu, sanjeev@cis.upenn.edu

CP31

Conflict-Free Coloring of Intersection Graphs of Geometric Objects

In FOCS'2002, Even et al. introduced the notion of conflict-free (CF) colorings of geometrically defined hypergraphs. They motivated it by frequency assignment problems in cellular networks. This notion has been extensively studied since then. A CF coloring of a graph is a coloring of its vertices such that the neighborhood (pointed or closed) of each vertex contains a vertex whose color differs from the colors of all other vertices in that neighborhood. In this paper we study CF colorings of intersection graphs of geometric objects. We show that any intersection graph of n pseudo-discs in the plane admits a CF coloring with $O(\log n)$ colors, with respect to both closed and pointed neighborhoods, and that the latter bound is asymptotically sharp. Using our methods, we obtain a strengthening of the two main results of Even et al.: Any family F of n discs in the plane can be colored with $O(\log n)$ colors in such a way that for any disc B in the plane, the set of discs in F that intersect B contains a uniquely-colored element. Moreover, such a coloring can be computed deterministically in polynomial time. In view of the motivation to study such colorings, this suggests further applications to frequency assignment in wireless networks. Finally, we present bounds on the number of colors needed for CF colorings of other classes of intersection graphs, including intersection graphs of axis-parallel rectangles and of ρ-fat objects in the plane.

Chaya Keller, Shakhar Smorodinsky
Ben Gurion University
chaya.keller27@gmail.com, shakhar@math.bgu.ac.il

CP31

Kirchhoff Index As a Measure of Edge Centrality in Weighted Networks: Nearly Linear Time Algorithms

We propose to use the well-known Kirchhoff index as the measure of edge centrality in weighted networks, called θ-Kirchhoff edge centrality. The Kirchhoff index of a network is defined as the sum of effective resistances over all vertex pairs. The centrality of an edge e is reflected in the increase of Kirchhoff index of the network when the edge e is partially deactivated, characterized by a parameter θ. We define two equivalent measures for θ-Kirchhoff edge centrality. Both are global metrics and have a better discriminating power than commonly used measures, based on local or partial structural information of networks. Despite the strong advantages of Kirchhoff index as a centrality measure, computing the exact value of Kirchhoff edge centrality for each edge in a graph is computationally demanding. To solve this problem, for each of the θ-Kirchhoff edge centrality metrics, we present an efficient algorithm to compute its ϵ-approximation for all the m edges in nearly linear time in m. The proposed θ-Kirchhoff edge centrality is the first global metric of edge importance that can be provably approximated in nearly-linear time. Moreover, according to the θ-Kirchhoff edge centrality, we present a θ-Kirchhoff vertex centrality measure, as well as a fast algorithm to compute ϵ-approximate Kirchhoff vertex centrality for all the n vertices in nearly linear time in m.

Huan Li, Zhongzhi Zhang
Fudan University, China
huanli.me@gmail.com, zhangzz@fudan.edu.cn

CP31

Lower Bounds for Symbolic Computation on Graphs: Strongly Connected Components, Liveness, Safety, and Diameter

A model of computation that is widely used in the formal analysis of reactive systems is symbolic algorithms. In this model the access to the input graph is restricted to symbolic operations, which are expensive in comparison to the standard RAM operations. We give lower bounds on the number of symbolic operations for basic graph problems such as the computation of the strongly connected components (SCCs) and of the approximate diameter as well as for fundamental problems in model checking such as safety, liveness, and co-liveness. Our lower bounds are linear in the number of vertices of the graph, even for constant-diameter graphs. For none of these problems lower bounds on the number of symbolic operations were known before. The lower bounds show an interesting separation of these problems from the reachability problem, which can be solved with $O(D)$ symbolic operations, where D is the diameter of the graph. Additionally we present an approximation algorithm for the graph diameter which requires $\tilde{O}(n\sqrt{D})$ symbolic steps to achieve a $(1+\epsilon)$-approximation for any constant $\epsilon > 0$. This compares to $O(n \cdot D)$ symbolic steps for the (naive) exact algorithm and $O(D)$ symbolic steps for a 2-approximation. Finally we also give a refined analysis of the SCC algorithm of [Gentilini et al. 2008], showing that it uses an optimal number of symbolic steps that is proportional to the sum of the diameters of the SCCs.

Veronika Loitzenbauer
University of Vienna
veronika.loitzenbauer@univie.ac.at

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

Wolfgang Dvorák
TU Wien
dvorak@dbai.tuwien.ac.at

Monika Henzinger
Univeristy of Vienna, Austria

monika.henzinger@univie.ac.at

CP31

Algorithms Based on *-Algebras, and Their Applications to Isomorphism of Polynomials with One Secret, Group Isomorphism, and Polynomial Identity Testing

We consider two basic algorithmic problems concerning (skew-)symmetric matrix tuples. The first problem asks to decide, given two tuples of (skew-)symmetric matrices $(B_1, \ldots, B_m)$ and $(C_1, \ldots, C_m)$, whether there exists an invertible matrix A such that for every $i \in \{1, \ldots, m\}$, $A^t B_i A = C_i$. We show that this problem can be solved in randomized polynomial time over finite fields of odd size, the real field, and the complex field. The second problem asks to decide, given a tuple of square matrices $(B_1, \ldots, B_m)$, whether there exist invertible matrices A and D, such that for every $i \in \{1, \ldots, m\}$, AB_iD is (skew-)symmetric. We show that this problem can be solved in deterministic polynomial time over fields of characteristic not 2. For both problems we exploit the structure of the underlying $*$-algebras (algebras with an involutive anti-automorphism). Applications of our results range from multivariate cryptography, group isomorphism, to polynomial identity testing: (1) Test isomorphism of quadratic forms with one secret over a finite field of odd size. (2) Test isomorphism of p-groups of class 2 and exponent p (p odd) with order p^k in time polynomial in the group order, when the commutator subgroup is of order $p^{O(\sqrt{k})}$. (3) Deterministically reveal two families of singularity witnesses for the symbolic determinant identity testing problem.

Youming 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

CP32

Optimal Streaming and Tracking Distinct Elements with High Probability

The distinct elements problem is one of the fundamental problems in streaming algorithms — given a stream of integers in the range $\{1, \ldots, n\}$, we wish to provide a $(1 + \varepsilon)$ approximation of the number of distinct elements in the input. After a long line of research optimal solution for this problem with constant probability of success, using $\mathcal{O}(\frac{1}{\varepsilon^2} + \log n)$ bits of space, was given by Kane, Nelson and Woodruff in 2010. The standard approach used in order to achieve low failure probability δ, is to take a median of $\log \delta^{-1}$ parallel repetitions of the original algorithm. We show that such a multiplicative space blow-up is unnecessary: we provide an optimal algorithm using $\mathcal{O}(\frac{\log \delta^{-1}}{\varepsilon^2} + \log n)$ bits of space — matching known lower bounds for this problem. That is, the $\log \delta^{-1}$ factor does not multiply the $\log n$ term. This settles completely the space complexity of the distinct elements problem with respect to all standard parameters. We consider also *strong tracking* (or *continuous monitoring*) variant of the distinct elements problem, where we want an algorithm which provides an approximation of the number of distinct elements seen so far, at all times of the stream. We show that this variant can be solved using $\mathcal{O}(\frac{\log \log n + \log \delta^{-1}}{\varepsilon^2} + \log n)$ bits of space, which we show to be optimal.

Jaroslaw Blasiok
Harvard Unibersity
jblasiok@g.harvard.edu

CP32

Efficient $\widetilde{O}(n/\epsilon)$ Spectral Sketches for the Laplacian and Its Pseudoinverse

In this paper we consider the problem of efficiently computing ϵ-sketches for the Laplacian and its pseudoinverse. For a given matrix A and an error tolerance ϵ, we seek to construct a function f such that for any vector x (chosen obliviously from f), with high probability $(1-\epsilon)x^\top Ax \leq f(x) \leq (1+\epsilon)x^\top Ax$. Our goal is to construct such a sketch f efficiently and to store it in the least space possible.
We provide nearly-linear time algorithms that, when given a Laplacian matrix $\mathcal{L} \in \mathbb{R}^{n \times n}$ and an error tolerance ϵ, produces $\tilde{O}(n/\epsilon)$-size sketches of both $\mathcal{L}$ and its pseudoinverse's quadratic form, where the $\tilde{O}()$ notation hides polylogarithmic factors in n and ϵ. Our algorithms improve upon the previous best sketch size of $\tilde{O}(n/\epsilon^{1.6})$ for sketching Laplacians by Andoni et al [ITCS 2016] and the previous best sketch size of $O(n/\epsilon^2)$ for sketching Laplacian pseudoinverses by Batson, Spielman, and Srivastava [STOC 2009].
Furthermore, we show how to compute all-pairs effective resistances from our $\tilde{O}(n/\epsilon)$-size sketch in $\tilde{O}(n^2/\epsilon)$ time. This improves upon the previous best running time of $\tilde{O}(n^2/\epsilon^2)$ by Spielman and Teng, [STOC 2004].

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

CP32

A Nearly Instance Optimal Algorithm for Top-K Ranking under the Multinomial Logit Model

We study the active learning problem of top-k ranking from multi-wise comparisons under the popular multinomial logit model. Our goal is to identify the top-k items with high probability by adaptively querying sets for comparisons and observing the noisy output of the most preferred item from each comparison. To achieve this goal, we design a new active ranking algorithm without using any information about the underlying items' preference scores. We also establish a matching lower bound on the sample complexity even when the set of preference scores is given to the algorithm. These two results together show that the proposed algorithm is nearly instance optimal (similar to instance optimal [Ronald Fagin, Amnon Lotem, and Moni Naor, Optimal aggregation algorithms for middleware, J. Comput. Syst. Sci., 66(4):614-656, 2003], but up to polylog factors). Our work extends the existing literature on rank aggregation in three directions. First, instead of studying a static problem with fixed data, we investigate the top-k ranking problem in an active learning setting. Second, we show our algorithm is nearly instance optimal, which is a much stronger theoretical guarantee. Finally, we extend the pairwise comparison to the multi-wise comparison, which has not been fully explored in ranking literature.

Xi Chen
New York Unviersity

xchen3@stern.nyu.edu

Yuanzhi Li, Jieming Mao
Princeton University
yuanzhil@cs.princeton.edu, jiemingm@cs.princeton.edu

CP32

Estimating Graph Parameters from Random Order Streams

We develop a new algorithmic technique that allows to transfer some constant time approximation algorithms for general graphs into random order streaming algorithms. We illustrate our technique by proving that in random order streams with probability at least 2/3,

- the number of connected components of G can be approximated up to an additive error of εn using $(\frac{1}{\varepsilon})^{O(1/\varepsilon^3)}$ space,
- the weight of a minimum spanning tree of an input graph with integer edges weights from $\{1,\ldots,W\}$ can be approximated within a multiplicative factor of $1+\varepsilon$ using $\left(\frac{1}{\varepsilon}\right)^{\tilde{O}(W^3/\varepsilon^3)}$ space,
- the size of a maximum independent set in planar graphs can be approximated within a multiplicative factor of $1+\varepsilon$ using space $2^{(1/\varepsilon)^{(1/\varepsilon)^{\log^{O(1)}(1/\varepsilon)}}}$.

Pan Peng
Faculty of Computer Science, University of Vienna
pan.peng@univie.ac.at

Christian Sohler
Department of Computer Science
Technische Universitat Dortmund
christian.sohler@tu-dortmund.de

CP32

Set Cover in Sub-Linear Time

We study the classic set cover problem from the perspective of sub-linear algorithms. Given access to a collection of m sets over n elements in the query model, we show that sub-linear algorithms derived from existing techniques have almost tight query complexities. On one hand, first we show an adaptation of the streaming algorithm of Har-Peled et al. to the sub-linear query model, that returns an α-approximate cover using $\tilde{O}(m(n/k)^{1/(\alpha-1)}+nk)$ queries to the input, where k denotes the value of a minimum set cover. We then complement this upper bound by proving that for lower values of k, the required number of queries is $\tilde{\Omega}(m(n/k)^{1/(2\alpha)})$, even for estimating the optimal cover size. Moreover, we prove that even checking whether a given collection of sets covers all the elements would require $\Omega(nk)$ queries. These two lower bounds provide strong evidence that the upper bound is almost tight for certain values of the parameter k. On the other hand, we show that this bound is not optimal for larger values of the parameter k, as there exists a $(1+\varepsilon)$-approximation algorithm with $\tilde{O}(mn/k\varepsilon^2)$ queries. We show that this bound is essentially tight for sufficiently small constant ε, by establishing a lower bound of $\tilde{\Omega}(mn/k)$ query complexity.

Piotr Indyk
Massachusetts Institute of Technology
indyk@mit.edu

Sepideh Mahabadi
MIT
mahabadi@mit.edu

Ronitt Rubinfeld
CSAIL, MIT and Tel Aviv University
ronitt@csail.mit.edu

Ali Vakilian
MIT
CSAIL
vakilian@mit.edu

Anak Yodpinyanee
MIT
anak@mit.edu

CP33

Envy-Free Chore Division for An Arbitrary Number of Agents

Chore division, introduced by Gardner in 1970s, is the problem of fairly dividing a chore among n different agents. In particular, in an envy-free chore division, we would like to divide a negatively valued heterogeneous object among a number of agents who have different valuations for different parts of the object, such that no agent envies another agent. It is the dual variant of the celebrated cake cutting problem, in which we would like to divide a desirable object among agents. In this paper, we provide the first discrete and bounded envy-free protocol for chore division for an arbitrary number of agents. We produce major and powerful tools for designing protocols for the fair division of negatively valued objects. These tools are based on structural results and important observations. In general, we believe these structures and techniques may be useful not only in chore division but also in other fairness problems.

Hadi Yami, Sina Dehghani, Alireza Farhadi,
MohammadTaghi Hajiaghayi
University of Maryland
hadi.yami93@gmail.com, sina.dehghani@gmail.com,
farhadi94@gmail.com, hajiaghayi@gmail.com

CP33

Almost Envy-Freeness with General Valuations

The goal of fair division is to distribute resources among competing players in a "fair" way. Envy-freeness is the most extensively studied fairness notion in fair division. Envy-free allocations do not always exist with indivisible goods, motivating the study of relaxed versions of envy-freeness. We study the *envy-freeness up to any good* (EFX) property, which states that no player prefers the bundle of another player following the removal of any single good, and prove the first general results about this property. We use the leximin solution to show existence of EFX allocations in several contexts, sometimes in conjunction with Pareto optimality. For two players with valuations obeying a mild assumption, one of these results provides stronger guarantees than the currently deployed algorithm on Spliddit, a popular fair division website. Unfortunately, finding the leximin solution can require exponential time. We show that this is necessary by proving an exponential lower bound on the number of value queries needed to identify an EFX allocation, even for two players with identical valuations. We consider both additive and more general valuations, and our work suggests that there is a rich landscape of problems to explore in the fair division of indivisible

goods with different classes of player valuations.

Benjamin Plaut
Stanford University
bplaut@stanford.edu

Tim Roughgarden
Stanford University
Computer Science Department
tim@cs.stanford.edu

CP33

The Price of Information in Combinatorial Optimization

Consider a network design application where we wish to lay down a minimum-cost spanning tree in a given graph; however, we only have stochastic information about the edge costs. To learn the precise cost of any edge, we have to conduct a study that incurs a price. Our goal is to find a spanning tree while minimizing the disutility, which is the sum of the tree cost and the total price that we spend on the studies. Situations such as the above often arise in practice when we wish to find a good solution to an optimization problem, but we start with only some partial knowledge about the input parameters. The missing information can be found only after paying a probing price, which we call the price of information. What strategy should we adopt to optimize our expected utility/disutility? A classical example of the above setting is Weitzman's "Pandora's box' problem where we are given probability distributions on values of n independent random variables. The goal is to choose a single variable with a large value, but we find the actual outcomes only after paying a price. Our work is a generalization of this model to other combinatorial optimization problems such as matching, set cover, facility location, and prize-collecting Steiner tree. We give a technique that reduces such problems to their non-price counterparts, and use it to design exact/approximation algorithms to optimize our utility/disutility.

Sahil Singla
CMU
ssingla@cmu.edu

CP33

The Value of Information Concealment

We consider a revenue optimizing seller selling a single item to a buyer, on whose private value the seller has a noisy signal. We show that, when the signal is kept private, arbitrarily more revenue could potentially be extracted than if the signal is leaked or revealed. We then show that, if the seller is not allowed to make payments to the buyer, the gap between the two is bounded by a multiplicative factor of 3, subject to fairly mild conditions on the joint distribution of the value and signal. Our examples show that both conditions are necessary for a constant bound to hold. We connect this scenario to multi-bidder single-item auctions where bidders' values are correlated. Results similar to the above are shown for the gap between the revenue of a Bayesian incentive compatible, ex post individually rational auction and that of a dominant strategy incentive compatible auction.

Zhihao Gavin Tang
The University of Hong Kong
zhtang@cs.hku.hk

Hu Fu, Chris Liaw
University of British Columbia
fu.hu.thu@gmail.com, cvliaw@cs.ubc.ca

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

CP33

Targeting and Signaling in Ad Auctions

Modern ad auctions allow advertisers to target more specific segments of the user population. Unfortunately, this is not always in the best interest of the ad platform – partially hiding some information could improve the platform's revenue. In this paper, we examine how to optimally reveal information about the ad opportunity to the advertisers in order to maximize revenue in a second-price ad auction. We consider a model in which bidders' valuations depend on a random state of the ad opportunity drawn from an extremely large support, and focus on developing algorithms whose running time is independent of the number of ad opportunity realizations. When the auctioneer is restricted to send a public signal to all bidders, we study a well-motivated Bayesian valuation setting in which the auctioneer and bidders both have private information, and present two results: (1) an exponential lower bound on the minimum number of signals required in any constant-approximate signaling scheme; (2) a "simple" signaling scheme that serves as a constant approximation under mild assumptions. We also initiate an exploration on the power of being able to send private signals to different bidders. In a basic setting where the auctioneer knows bidders' valuations, we exhibit a polynomial-time private scheme that extracts almost full surplus even in the worst Bayes Nash equilibrium. This illustrates the surprising power of private signaling schemes in extracting revenue.

Ashwinkumar Badanidiyuru, Kshipra Bhawalkar
Google Research
ashwinkumarbv@google.com, kshipra@google.com

Haifeng Xu
University of Southern California
haifengx@usc.edu

CP34

The Complexity of Distributed Edge Coloring with Small Palettes

The complexity of distributed edge coloring depends heavily on the palette size as a function of the maximum degree Δ. In this paper we explore the complexity of edge coloring in the LOCAL model in different palette size regimes.

- We simplify the round elimination technique of Brandt et al. and prove that $(2\Delta-2)$-edge coloring requires $\Omega(\log_\Delta \log n)$ time w.h.p. and $\Omega(\log_\Delta n)$ time deterministically, even on trees.
- We give a randomized edge coloring algorithm using palette sizes as small as $\Delta+\tilde{O}(\sqrt{\Delta})$, which is a natural barrier for randomized approaches. The running time of the algorithm is at most $O(\log\Delta \cdot T_{LLL})$, where T_{LLL} is the complexity of a permissive version of the constructive Lovasz local lemma.
- We develop a new distributed Lovasz local lemma algorithm for tree-structured dependency graphs, which leads to a $(1+\epsilon)\Delta$-edge coloring algorithm for trees running in $O(\log\log n)$ time.

- A natural approach to computing $(\Delta+1)$-edge colorings (Vizing's theorem) is to extend partial colorings by re-coloring parts of the graph. We prove that this approach may be viable, but in the worst case requires recoloring subgraphs of diameter $\Omega(\Delta \log n)$. This is in contrast to distributed algorithms for Brooks' theorem, which exploit the existence of $O(\log_\Delta n)$-length augmenting paths.

Yi-Jun Chang
University of Michigan
cyijun@umich.edu

Qizheng He, Wenzheng Li
IIIS, Tsinghua University
hqz14@mails.tsinghua.edu.cn,
wz-li14@mails.tsinghua.edu.cn

Seth Pettie
University of Michigan
seth@pettie.net

Jara Uitto
ETH Zurich
jara.uitto@inf.ethz.ch

CP34

Labeling Schemes for Nearest Common Ancestors Through Minor-Universal Trees

A labeling scheme for nearest common ancestors assigns a distinct binary string, called the label, to every node of a tree. Given the labels of two nodes (and no further information about the topology of the tree), we can compute the label of their nearest common ancestor. The goal is to make the labels short. Alstrup, Gavoille, Kaplan, and Rauhe [Theor. Comput. Syst. 37(3):441-456 2004] showed that $O(\log n)$-bit labels are enough. More recently, Alstrup, Halvorsen, and Larsen [SODA 2014] refined this to only $2.772 \log n$, and provided a lower bound of $1.008 \log n$. We connect the question of designing such a scheme to the existence of a tree, called a minor-universal tree, that contains every tree on n nodes as a topological minor. Even though it is not clear if a labeling scheme must be based on such a notion, we argue that all existing schemes can be reformulated as such. We show that this notion allows us to easily obtain good bounds on the length of the labels. As the main upper bound, we show that $2.318 \log n$-bit labels are enough. On the lower bound side, we show that any minor-universal tree for trees on n nodes must contain $\Omega(n^{2.174})$ nodes. We complement the existential results with a generic transformation that allows us, for any labeling scheme for nearest common ancestors based on a minor-universal tree, to decrease the query time to constant, while increasing the length of the labels by lower order terms.

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

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

Jakub Lopuszanski
University of Wroclaw
qbolec@gmail.com

Konstantinos Panagiotou
University of Munich
kpanagio@math.lmu.de

Pascal Su
ETH Zurich
sup@inf.ethz.ch

CP34

Mst in O(1) Rounds of Congested Clique

We present a distributed randomized algorithm finding Minimum Spanning Tree (MST) of a given graph in $O(1)$ rounds, with high probability, in the congested clique model. The input graph in the congested clique model is a graph of n nodes, where each node initially knows only its incident edges. The communication graph is a clique with limited edge bandwidth: each two nodes (not necessarily neighbours in the input graph) can exchange $O(\log n)$ bits. As in previous works, the key part of the MST algorithm is an efficient Connected Components (CC) algorithm. However, unlike the former approaches, we do not aim at simulating the standard Boruvka's algorithm, at least at initial stages of the CC algorithm. Instead, we develop a new technique which combines connected components of sample sparse subgraphs of the input graph in order to accelerate the process of uncovering connected components of the original input graph. Our result addresses a problem proposed by Lotker et al. [SPAA 2003; SICOMP 2005] and improves over previous $O(\log^* n)$ algorithm of Ghaffari et al. [PODC 2016], and $O(\log \log \log n)$ algorithm of Hegeman et al. [PODC 2015]. It also determines $\Theta(1)$ round complexity in the congested clique for MST, as well as other graph problems, including bipartiteness, cut verification, s-t connectivity, and cycle containment.

Tomasz Jurdzinski
University of Wroclaw
tju@cs.uni.wroc.pl

Krzysztof Nowicki
University of Wroclaw
University of Wroclaw
knowicki@cs.uni.wroc.pl

CP34

Fast Space Optimal Leader Election in Population Protocols

The model of *population protocols* refers to the growing in popularity theoretical framework suitable for studying *pairwise interactions* within a large collection of simple indistinguishable entities, frequently called *agents*. In this paper the emphasis is on the space complexity in fast *leader election* via population protocols governed by the *random scheduler*, which uniformly at random selects pairwise interactions from the population of n agents. The main result of this paper is a new fast and space optimal *leader election protocol*. The new protocol operates in *parallel time* $O(\log^2 n)$ equivalent to $O(n \log^2 n)$ sequential *pairwise interactions*, in which each agent utilises $O(\log \log n)$ states. This double logarithmic space utilisation matches asymptotically the existing lower bound $\frac{1}{2} \log \log n$ on the number of states utilised by agents in any leader election algorithm with the running time $o(\frac{n}{\text{polylog } n})$. Our solution takes an advantage of the concept of phase clocks, a fundamental synchronisation and coordination tool in Distributed Computing. We propose a new fast and robust

population protocol for initialisation of phase clocks to be run simultaneously in multiple modes and intertwined with the leader election process. We also provide the reader with the relevant formal argumentation indicating that our solution is always correct and fast with high probability.

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

Grzegorz Stachowiak
Uniwersytet Wroclawski
gst@cs.uni.wroc.pl

CP34

Ergodic Effects in Token Circulation

We consider a dynamical process in a network which distributes all particles (tokens) located at a node among its neighbors, in a round-robin manner. We show that in the recurrent state of this dynamics (i.e., disregarding a polynomially-long initialization phase of the system), the number of particles located on a given edge, averaged over an interval of time, is tightly concentrated around the average particle density in the system. Formally, for a system of k particles in a graph of m edges, during any interval of length T, this time-averaged value is $k/m \pm \widetilde{O}(1/T)$, whenever $gcd(m,k) = \widetilde{O}(1)$ (and so, e.g., whenever m is a prime number). To achieve these bounds, we link the behavior of the studied dynamics to ergodic properties of traversals based on Eulerian circuits on a symmetric directed graph. Our results are proved through sum set methods. As a corollary, we obtain bounds on the idleness of the studied dynamics, i.e., on the longest possible time between two consecutive appearances of a token on an edge, taken over all edges. Minimizing idleness is fundamental to the study of the patrolling problem in networks. Our results immediately imply a bound of $\widetilde{O}(m/k)$ on the idleness of the studied process, showing that it is a distributed $\widetilde{O}(1)$-competitive solution to the patrolling task, for all of the covered cases. Our work provides further insights which may be interesting in load-balancing applications.

Adrian Kosowski
Inria Paris
adrian.kosowski@inria.fr

Przemyslaw Uznanski
ETH Zurich
przemyslaw.uznanski@inf.ethz.ch

CP35

Non Interactive Simulation of Correlated Distributions Is Decidable

A basic problem in information theory is the following: Let $\mathbf{P} = (\mathbf{X}, \mathbf{Y})$ be an arbitrary distribution where the marginals $\mathbf{X}$ and $\mathbf{Y}$ are (potentially) correlated. Let Alice and Bob be two players where Alice gets samples $\{x_i\}_{i\geq 1}$ and Bob gets samples $\{y_i\}_{i\geq 1}$ and for all i, $(x_i, y_i) \sim \mathbf{P}$. What joint distributions $\mathbf{Q}$ can be simulated by Alice and Bob without any interaction? Classical works in information theory by Gács-Körner and Wyner answer this question when at least one of $\mathbf{P}$ or $\mathbf{Q}$ is the distribution Eq (Eq is defined as uniform over the points $(0,0)$ and $(1,1)$). However, other than this special case, the answer to this question is understood in very few cases. Recently, Ghazi, Kamath and Sudan showed that this problem is decidable for $\mathbf{Q}$ supported on $\{0,1\} \times \{0,1\}$. We extend their result to $\mathbf{Q}$ supported on any finite alphabet. Moreover, we show that If can be simulated, our algorithm also provides a (non-interactive) simulation protocol. We rely on recent results in Gaussian geometry (by the authors) as well as a new *smoothing argument* inspired by the method of *boosting* from learning theory and potential function arguments from complexity theory and additive combinatorics.

Anindya De
Northwestern University
anindya.de1@northwestern.edu

Elchanan Mossel
MIT
n/a

Joe Neeman
UT Austin
na/

CP35

Which Distribution Distances Are Sublinearly Testable?

Given samples from an unknown distribution p and a description of a distribution q, are p and q close or far? This question of "identity testing" has received significant attention in the case of testing whether p and q are equal or far in total variation distance. However, in recent work, the following questions have been been critical to solving problems at the frontiers of distribution testing: -Alternative Distances: Can we test whether p and q are far in other distances? -Robustness: Can we test when p and q are close, rather than equal? And if so, close in which distances? Motivated by these questions, we characterize the complexity of distribution testing under a variety of distances, including L1, L2, Hellinger, KL, and χ^2. For each pair of distances d1 and d2, we study the complexity of testing if p and q are close in d1 versus far in d2, with a focus on identifying which problems allow strongly sublinear testers. We provide matching upper and lower bounds for each case. We also study these questions in the case where we only have samples from q, showing qualitative differences from identity testing in terms of when robustness can be achieved. Our algorithms fall into the classical paradigm of chi-squared statistics, but require crucial changes to handle the challenges introduced by each distance we consider. Finally, we survey other recent results in an attempt to serve as a reference for the complexity of various distribution testing problems.

Constantinos Daskalakis, Gautam Kamath, John Wright
MIT
costis@csail.mit.edu, g@csail.mit.edu, jswright@mit.edu

CP35

Robustly Learning a Gaussian: Getting Optimal Error, Efficiently

We study the fundamental problem of learning the parameters of a high-dimensional Gaussian in the presence of noise – where an ε-fraction of our samples were chosen by an adversary. We give robust estimators that achieve estimation error $O(\varepsilon)$ in the total variation distance, which is optimal up to a universal constant that is independent of the dimension. In the case where just the mean is unknown, our robustness guarantee is optimal up to a factor of $\sqrt{2}$ and the running time is polynomial in d and $1/\varepsilon$. When both

the mean and covariance are unknown, the running time is polynomial in d and quasipolynomial in $1/\varepsilon$. Moreover all of our algorithms require only a polynomial number of samples. Our work shows that the same sorts of error guarantees that were established over fifty years ago in the one-dimensional setting can also be achieved by efficient algorithms in high-dimensional settings.

Ilias Diakonikolas
USC
diakonik@usc.edu

Gautam Kamath
MIT
g@csail.mit.edu

Daniel Kane
UC San Diego
dakane@cs.ucsd.edu

Jerry Li, Ankur Moitra
MIT
jerryzli@mit.edu, moitra@mit.edu

Alistair Stewart
USC
alistais@usc.edu

CP35

Cycles in Adversarial Regularized Learning

Regularized learning is a fundamental technique in online optimization, machine learning and many other fields of computer science. A natural question that arises in these settings is how regularized learning algorithms behave when faced against each other. We study a natural formulation of this problem by coupling regularized learning dynamics in zero-sum games. We show that the system's behavior is Poincaré recurrent, implying that almost every trajectory revisits any (arbitrarily small) neighborhood of its starting point infinitely often. This cycling behavior is robust to the agents' choice of regularization mechanism (each agent could be using a different regularizer), to positive-affine transformations of the agents' utilities, and it also persists in the case of networked competition, i.e., for zero-sum polymatrix games.

Panayotis Mertikopoulos
Univ. Grenoble Alpes, CNRS
panayotis.mertikopoulos@imag.fr

Christos Papadimitriou
Columbia University
christos@cs.columbia.edu

Georgios Piliouras
Singapore University of Technology and Design
georgios.piliouras@gmail.com

CP35

Improved Coresets for Kernel Density Estimates

We study the construction of coresets for kernel density estimates. That is we show how to approximate the kernel density estimate described by a large point set with another kernel density estimate with a much smaller point set. For characteristic kernels (including Gaussian and Laplace kernels), our approximation preserves the L_∞ error between kernel density estimates within error ϵ, with coreset size $2/\epsilon^2$, but no other aspects of the data, including the dimension, the diameter of the point set, or the bandwidth of the kernel common to other approximations. When the dimension is unrestricted, we show this bound is tight for these kernels as well as a much broader set. This work provides a careful analysis of the iterative Frank-Wolfe algorithm adapted to this context, an algorithm called kernel herding. This analysis unites a broad line of work that spans statistics, machine learning, and geometry. When the dimension d is constant, we demonstrate much tighter bounds on the size of the coreset specifically for Gaussian kernels, showing that it is bounded by the size of the coreset for axis-aligned rectangles. Currently the best known constructive bound is $O(\frac{1}{\epsilon}\log^d\frac{1}{\epsilon})$, and non-constructively, this can be improved by $\sqrt{\log\frac{1}{\epsilon}}$. This improves the best constant dimension bounds polynomially for $d \geq 3$.

Wai Ming Tai, Jeff Phillips
University of Utah
taiwaiming2003@gmail.com, jeffp@cs.utah.edu

CP36

Covering a Tree with Rooted Subtrees – Parameterized and Approximation Algorithms

We consider the multiple traveling salesman problem on a weighted tree. In this problem there are m salesmen located at the root initially. Each of them will visit a subset of vertices and return to the root. The goal is to assign a tour to every salesman such that every vertex is visited and longest tour among all salesmen is minimized. The problem is equivalent to the subtree cover problem, in which we cover a tree with rooted subtrees such that the weight of the maximum weighted subtree is minimized. The classical machine scheduling problem can be viewed as a special case of our problem when the given tree is a star. We provide approximation and parameterized algorithms for this problem. We first present a PTAS (Polynomial Time Approximation Scheme). We then observe that, the problem remains NP-hard even if tree height and edge weight are constant, and present an FPT algorithm for this problem parameterized by the largest tour length. To achieve the FPT algorithm, we show a more general result. We prove that, integer linear programming that has a tree-fold structure is in FPT, which extends the FPT result for the n-fold integer programming by Hemmecke, Onn and Romanchuk (Math. Programming, 2013).

Lin Chen
Zhejiang University
chenlin198662@gmail.com

Dániel Marx
Hungarian Academy of Sciences (MTA SZTAKI),
Budapest,Hungary
dmarx@cs.bme.hu

CP36

Fully Polynomial Fpt Algorithms for Some Classes of Bounded Clique-Width Graphs

Recently, hardness results for problems in P were achieved using reasonable complexity theoretic assumptions such as the Strong Exponential Time Hypothesis. According to these assumptions, many graph theoretic problems do not admit truly subquadratic algorithms. A central technique used to tackle the difficulty of the above mentioned prob-

lems is fixed-parameter algorithms with *polynomial dependency* in the fixed parameter (P-FPT). Applying this technique to *clique-width*, an important graph parameter, remained to be done. In this paper we study several graph theoretic problems for which hardness results exist such as *cycle problems*, *distance problems* and *maximum matching*. We give hardness results and P-FPT algorithms, using clique-width and some of its upper-bounds as parameters. We believe that our most important result is an $\mathcal{O}(k^4 \cdot n + m)$-time algorithm for computing a maximum matching where k is either the modular-width or the P_4-sparseness. The latter generalizes many algorithms that have been introduced so far for specific subclasses such as cographs. Our algorithms are based on preprocessing methods using modular decomposition and split decomposition. Thus they can also be generalized to some graph classes with unbounded clique-width.

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

Guillaume Ducoffe
ICI National Institute for Research and Development in Informatics
guillaume.ducoffe@ici.ro

Alexandru Popa
ICI National Institute for Research and Development in Informatics, Romania
alexandru.popa@ici.ro

CP36

Parameterized Algorithms for Survivable Network Design with Uniform Demands

In the Survivable Network Design Problem (SNDP), the input is an edge-weighted (di)graph G and an integer r_{uv} for every pair of vertices $u, v \in V(G)$. The objective is to construct a subgraph H of minimum weight which contains r_{uv} edge-disjoint (or node-disjoint) u-v paths. An important restriction of this problem is one where the connectivity demands are the same for every pair of vertices. In this paper, we first consider the edge-connectivity version of this problem which we call λ-Edge Connected Subgraph (λ-ECS). In this problem, the input is a λ-edge connected (di)graph G and an integer k and the objective is to check whether G contains a spanning subgraph H that is also λ-edge connected and H excludes *at least* k edges of G. If we replace λ-edge connectivity with λ-vertex connectivity we get the λ-Veretx Connected Subgraph (λ-VCS) problem. We show that λ-ECS is fixed-parameter tractable (FPT) for both graphs and digraphs even if the (di)graph has non-negative real weights on the edges and the objective is to exclude from H, some edges of G whose total weight exceeds a prescribed value. We also show that λ-VCS is FPT on digraphs; however the problem on undirected graphs remains open.

Joergen Bang-Jensen
Department of Mathematics and Computer Science
University of Southern Denmark
jbj@imada.sdu.dk

Manu Basavaraju
NIT Suratkal
iammanu@gmail.com

Kristine Vittin Klinkby
UiB+SDU
kristine.vkk@gmail.com

Pranabendu Misra
UiB
pranabendu.m@gmail.com

Ramanujan M. S.
University of Warwick
r.maadapuzhi-sridharan@warwick.ac.uk

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

Meirav Zehavi
Ben-Gurion University
zehavimeirav@gmail.com

CP36

An Fpt Algorithm Beating 2-Approximation for K-Cut

In the k-cut problem, we are given an edge-weighted graph G and an integer k, and have to remove a set of edges with minimum total weight so that G has at least k connected components. Prior work on this problem gives, for all $h \in [2, k]$, a $(2 - h/k)$-approximation algorithm for k-cut that runs in time $n^{O(h)}$. Hence to get a $(2 - \epsilon)$-approximation algorithm for some absolute constant ϵ, the best runtime using prior techniques is $n^{O(k\epsilon)}$. Moreover, it was recently shown that getting a $(2-\epsilon)$-approximation for general k is NP-hard, assuming the Small Set Expansion Hypothesis. If we use the size of the cut as the parameter, an FPT algorithm to find the exact k-cut is known, but solving the k-cut problem exactly is $W[1]$-hard if we parameterize only by the natural parameter of k. An immediate question is: *can we approximate k-cut better in FPT-time, using k as the parameter?* We answer this question positively. We show that for some absolute constant $\epsilon > 0$, there exists a $(2 - \epsilon)$-approximation algorithm that runs in time $2^{O(k^6)} \cdot \widetilde{O}(n^4)$. This is the first FPT algorithm that is parameterized only by k and strictly improves the 2-approximation.

Jason M. Li, Anupam Gupta, Euiwoong Lee
Carnegie Mellon University
jmli@cs.cmu.edu, anupamg@cs.cmu.edu, euiwoonl@cs.cmu.edu

CP36

Covering Small Independent Sets and Separators with Applications to Parameterized Algorithms

We present two new combinatorial tools for the design of parameterized algorithms. The first is a simple linear time randomized algorithm that given as input a d-degenerate graph G and an integer k, outputs an independent set Y, such that for every independent set X in G of size at most k, the probability that X is a subset of Y is at least $\left(\binom{(d+1)k}{k} \cdot k(d+1)\right)^{-1}$. The second is a new (deterministic) polynomial time graph sparsification procedure that given a graph G, a set $T = \{\{s_1, t_1\}, \{s_2, t_2\}, \ldots, \{s_\ell, t_\ell\}\}$ of terminal pairs and an integer k, returns an induced subgraph $G^\star$ of G that maintains *all* the inclusion minimal multicuts of G of size at most k, and does not contain any $(k+2)$-vertex connected set of size $2^{O(k)}$. In particular, $G^\star$

excludes a clique of size $2^{O(k)}$ as a topological minor. Put together, our new tools yield new randomized fixed parameter tractable (FPT) algorithms for Stable s-t Separator, Stable Odd Cycle Transversal and Stable Multicut on general graphs, and for Stable Directed Feedback Vertex Set on d-degenerate graphs, resolving two problems left open by Marx et al. [ACM Transactions on Algorithms, 2013]. All of our algorithms can be derandomized at the cost of a small overhead in the running time.

Daniel Lokshtanov
University of Bergen
daniello@ii.uib.no

Fahad Panolan
University of Bergen, Norway
fahad.panolan@ii.uib.no

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

Roohani Sharma
Institute of Mathematical Sciences, HBNI, Chennai, India
roohani@imsc.res.in

Meirav Zehavi
Ben-Gurion University
zehavimeirav@gmail.com