Schedule

Dagstuhl Seminar 23061

Scheduling 2023

Monday

09:00 – 09:45	Welcome and Round of Introduction
09:45 – 10:15	**Thomas Kesselheim**: Online Load Balancing Beyond l_p Norms
10:15 – 10:45	Coffee break
10:45 – 11:45	**Siddhartha Banerjee**: We Need to Talk About How we Talk About Online Decision-Making
12:15 – 13:30	Lunch
13:30 – 15:30	Collaboration Time
15:30 – 16:00	Coffee & Cake
16:00 – 16:30	**Daniel Freund:** Efficient Decentralized Multi-Agent Learning in Symmetric Queuing Systems
16:30 – 17:00	**Etienne Bamas**: Better Trees for Santa Claus
17:00 – 18:00	**Claire Mathieu:** Stable matching in Practice
18:00 – 19:30	Dinner

Tuesday

09:00 – 10:00	**Sergei Vassilvitskii**: Five(ish) Years of Algorithms with Predictions
10:00 – 10:30	Coffee break
10:30 – 11:00	**Thomas Lavastida**: Algorithms with Prediction Portfolios
11:00 – 11:30	**Amit Kumar**: Online Algorithms with Multiple Predictions
11:30 – 12:00	**Open problems** & proposals for joint work at Dagstuhl
12:15 – 13:30	Lunch
13:30 – 15:30	Collaboration Time
15:30 – 16:00	Coffee & Cake
16:00 – 16:30	Short Talks: **Yossi Azar, Jose Verschae, Jiri Sgall**
16:40 – 17:10	**Martin Skutella**: Hiking through the complexity landscape with Gerhard
17:10 – 18:00	**Open Problems** that Gerhard would have liked (and other stories)
18:00 – 19:30	Dinner

Wednesday

09:00 – 10:00	**Silvio Lattanzi**: Clustering with Advice
10:00 – 10:30	Coffee break
10:30 – 11:00	**Franziska Eberle**: Stochastic Configuration Balancing
11:00 – 11:30	**Chaitanya Swamy**: Stochastic Minimum Norm Combinatorial Optimization
11:30 – 12:00	**Tjark Vredeveld**: Bayesian Scheduling: Analysis of Simple Policies
12:15 – 13:30	Lunch
meet 13:30	Hike tour – free afternoon
15:30 – 16:00	Coffee & Cake
18:00 – 19:30	Dinner

Thursday

09:00 – 10:00	**Debmalya Panigrahi**: Santa Claus with Predictions
10:00 – 10:30	Coffee break
10:30 – 11:00	**Ali Vakilian**: Faster Fundamental Graph Algorithms via Learned Predictions
11:00 – 11:30	**Yossi Azar**: Flow Time Scheduling with Uncertain Processing Time
11:30 – 12:00	**Seffi Naor**: Online Rounding of Bipartite Matching
12:15 – 13:30	Lunch
13:30 – 15:30	Collaboration Time
15:30 – 16:00	Coffee & Cake
16:00 – 17:00	**Sami Davies:** Scheduling with Communication Delays
17:00 – 17:45	Short talks: **Alberto Marchetti-Spaccamela, Antonios Antoniadis, Kevin Schewior, Alexander Lindermayr**
18:00 – 19:30	Dinner

Friday

09:00 – 09:30	**Vineet Goyal**: Online Matching with Reusable Capacities
09:30 – 10:00	**Rudy Zhou**: Minimizing Completion Times for Stochastic Jobs via Batched Free Times
10:00 – 10:30	Coffee break
10:30 – 11:00	**Jens Schlöter**: Learning-Augmented Query Policies for Minimum Spanning Tree with Uncertainty
11:00 – 11:30	**Cliff Stein:** Queueing Safely and Efficiently for Elevators
12:15 – 13:30	Lunch

End

Abstracts

Monday

Thomas Kesselheim: Online Load Balancing Beyond l_p Norms

The classic problem of online makespan minimization can be understood as minimizing the l_infinity norm of the vector of machine loads. It was extended to l_p norms already more than 25 years ago. We study the problem beyond l_p norms. We show that general norms admit good algorithms as long as the norm can be approximated by a function that is “gradient-stable”, a notion that we introduce. Roughly it says that the gradient of the function should not drastically decrease in any component as we increase the input vector. In particular, we give the first O(log^2 m)-competitive algorithm for online load balancing with respect to an arbitrary monotone symmetric norm. Our techniques extend to applications beyond symmetric norms as well, e.g., to Online Vector Scheduling and to Online Generalized Assignment with Convex Costs.
The set of techniques can also be applied in stochastic settings, e.g., in which the sequence of jobs arrives in random order. Given that they can be understood as based on duality, it would be interesting to see if they also can be used with error-prone predictions.

Based on joint work with Marco Molinaro and Sahil Singla in SODA’23.

Siddhartha Banerjee: We Need to Talk About How we Talk About Online Decision-Making

Scheduling, and other problems involving online allocation of resources, are topics of great interest across many academic communities. However, the huge diversity in underlying models and methodologies means that existing assumptions/algorithms/guarantees are difficult to understand and compare (and often not very useful...).
I will try to present the stochastic control viewpoint on these problems, and discuss how there seems to be a fundamental divide between the view of online algorithms in CS and in controls. I will then present a sample-path coupling technique, which provides a simple way of reasoning about online algorithms, and regret guarantees against any chosen benchmark. I will describe how this framework gives new algorithms and insights for a variety of problems, including (time permitting):
1. *Constant regret* algorithms (i.e., having additive loss compared to the hindsight optimal solution which is *independent of the horizon and budget*) for several widely-studied settings including online packing, load balancing, dynamic pricing, assortment optimization, and online bin packing.
2. Incorporating side information and historical data in these settings (and achieve constant regret with as little as a single data trace).
3. Fundamental tradeoffs in multi-objective settings (in particular, for fairness in online allocation).

Daniel Freund (MIT): Efficient decentralized multi-agent learning in asymmetric queuing systems

Learning in multi-agent systems often poses significant challenges due to interference between agents. In particular, unlike classical stochastic systems, the performance of an agent's action is not drawn i.i.d. from some distribution but is directly affected by the (unobserved) actions of the other agents. This is the reason why most collaborative multi-agent learning approaches aim to globally coordinate all agents' actions to evade this interference.
In this talk, we focus on agents in a decentralized bipartite queuing system, where N agents request service from K servers. Prior decentralized approaches aim to globally identify a coordinated schedule or do not take advantage of the bipartite structure, which leads to significant shortcomings: performance that degrades exponentially in the number of servers, requirement of shared randomness and unique identifiers, and computationally demanding algorithms. In contrast, we provide a low-complexity algorithm that is run decentrally by each agent, avoids the shortcomings of "global coordination" and leads the

queuing system to have efficient performance in asymmetric bipartite queuing systems while also having additional robustness properties.

Etienne Bamas: Better Trees for Santa Claus

A notorious open problem in approximation algorithms is whether there exists a constant factor approximation for MaxMin Fair Allocation of indivisible items (also known as the Santa Claus problem). Bateni, Charikar, and Guruswami [STOC'09] introduced the MaxMin Arborescence problem as an important special case: Given a directed graph with sources and sinks we have to find vertex disjoint arborescences rooted in the sources such that at each non-sink vertex of an arborescence the out-degree is at least k, where k is to be maximized. This problem is of particular interest, since it appears to capture much of the difficulty of the general Santa Claus problem. Indeed, the progress made by Bateni et al. was quickly generalized by Chakrabarty, Chuzhoy, and Khanna [FOCS'09] to the general case. These two results remain the state-of-the-art for both problems, and they yield a polylogarithmic approximation in quasi-polynomial time.

In this talk, I will present the main ideas behind an exponential improvement to this, a poly(loglog n)-approximation in quasi-polynomial time for the MaxMin Arborescence problem.

Claire Mathieu: Stable matching in Practice

Stable matching methods, based on the algorithm designed by Gale and Shapley, are used around the world in many applications.the college assigned to the applicant in their preference list; robustness; running time; etc. After a brief theoretical review, we present issues arising in practice, (1) in the contexte of college admissions in France since 2018, and (2) in the context of upcoming medical studies specialization in France starting in 2024.

Tuesday

Sergei Vassilvitskii: Five(ish) Years of Algorithms with Predictions

Abstract: Worst-case analysis has proven invaluable for understanding aspects of both the complexity and practicality of algorithms. In some cases, however, we do not face worst-case scenarios, and the question arises of how we can tune our algorithms to work even better on the kinds of instances we are likely to see, while keeping a rigorous formal framework similar to what we have developed through worst-case analysis.

We give an overview of a recent trend that develops algorithms parameterized by additional parameters which capture "the kinds of instances we are likely to see," and obtains a finer grained analysis of algorithms' performance. We will give examples of re-analyzing classical algorithms through this lens, as well as developing new algorithms that expose new structural insights about the problems.

Thomas Lavastida: Algorithms with Prediction Portfolios

The research area of algorithms with predictions has seen recent success showing how to incorporate machine learning into algorithm design to improve performance when the predictions are correct, while retaining worst-case guarantees when they are not. Most previous work has assumed that the algorithm has access to a single predictor. However, in practice, there are many machine learning methods available, often with incomparable generalization guarantees, making it hard to pick a best method a priori. In this work we consider scenarios where multiple predictors are available to the algorithm and the question is how to best utilize them.

Ideally, we would like the algorithm’s performance to depend on the quality of the best predictor. However, utilizing more predictions comes with a cost, since we now have to identify which prediction is the best. We study the use of multiple predictors for a number of fundamental problems, including matching, load balancing, and non-clairvoyant scheduling, which have been well-studied in the single predictor setting. For each of these problems we introduce new algorithms that take advantage of multiple predictors, and prove bounds on the resulting performance. (Joint work with Michael Dinitz, Sungjin Im, Benjamin Moseley, Sergei Vassilvitskii)

Amit Kumar: Online Algorithms with Multiple Predictions

This paper studies online algorithms augmented with multiple machine-learned predictions. While online algorithms augmented with a single prediction have been extensively studied in recent years, the literature for the multiple predictions setting is sparse. In this paper, we give a generic algorithmic framework for online covering problems with multiple predictions that obtains an online solution that is competitive against the performance of the best predictor. Our algorithm incorporates the use of predictions in the classic potential-based analysis of online algorithms. We apply our algorithmic framework to solve classical problems such as online set cover, (weighted) caching, and online facility location in the multiple predictions setting. Our algorithm can also be robustified, i.e., the algorithm can be simultaneously made competitive against the best prediction and the performance of the best online algorithm (without prediction).

Wednesday

Silvio Lattanzi: Clustering with Advice

In this talk, starting from practical questions, we motivate different models of machine learning advice and present new algorithms that leverage the additional information to obtain stronger guarantees. In particular, we start by describing the semi-supervised active clustering framework and how one can recover convex and non-convex clusters in this setting. Then we describe how partial knowledge about input instances can be leveraged to obtain better guarantees in online correlation clustering and the classic k-means problem.

Franziska Eberle: Stochastic Configuration Balancing

The configuration balancing problem with stochastic requests generalizes many well-studied resource allocation problems such as load balancing and virtual circuit routing. In it, we have m resources and n requests. Each request has multiple possible configurations, each of which increases the load of each resource by some amount. The goal is to select one configuration for each request to minimize the makespan: the load of the most-loaded resource. In our work, we focus on a stochastic setting, where we only know the distribution for how each configuration increases the resource loads, learning the realized value only after a configuration is chosen.

We develop both offline and online algorithms for configuration balancing with stochastic requests. When the requests are known offline, we give a non-adaptive policy for configuration balancing with stochastic requests that $O(\log m / \log \log m)$-approximates the optimal adaptive policy. In particular, this closes the adaptivity gap for this problem as there is an asymptotically matching lower bound even for the very special case of load balancing on identical machines. When requests arrive online in a list, we give a non-adaptive policy that is $O(\log m)$ competitive. Again, this result is asymptotically tight due to information-theoretic lower bounds for very special cases (e.g., for load balancing on unrelated machines). Finally, we show how to leverage adaptivity in the special case of load balancing on related machines to obtain a constant-factor approximation offline and an $O(\log \log m)$-approximation online. A crucial technical ingredient in all of our results is a new structural characterization of the optimal adaptive policy that allows us to limit the correlations between its decisions.

This is joint work with Anupam Gupta, Nicole Megow, Ben Moseley, and Rudy Zhou and recently got accepted for presentation at IPCO 2023.

Chaitanya Swamy: Stochastic Minimum Norm Combinatorial Optimization

We develop a framework for designing approximation algorithms for a wide class of (1-stage) stochastic-optimization problems with norm-based objective functions. We introduce the model of stochastic minimum-norm combinatorial optimization, wherein the costs involved are random variables with given distributions, and we are given a monotone, symmetric norm f. Each feasible solution induces a random multidimensional cost vector whose entries are independent random variables, and the goal is to find a solution that minimizes the expected f-norm of the induced cost vector. This is a very rich class of objectives, containing all l_p norms, as also Top-l norms (sum of l largest coordinates in absolute value), which enjoys various closure properties.

Our chief contribution is a framework for designing approximation algorithms for stochastic minimum-norm optimization, which has two key components:
(i) A reduction showing that one can control the expected f-norm by simultaneously controlling a (small) collection of expected Top-l norms; and
(ii) Showing how to tackle the minimization of a single expected Top-l-norm by leveraging techniques used to deal with minimizing the expected maximum, circumventing the difficulties posed by the non-separable nature of Top-l norms.
We apply our framework to obtain strong approximation guarantees for two concrete problem settings: (1) stochastic load balancing, wherein jobs have random processing times and the induced cost vector is the machine-load vector; and (2) stochastic spanning tree, where edges have stochastic weights and the cost-vector coordinates are the weights of the edges in the spanning tree returned.

This is joint work with Sharat Ibrahimpur.

Tjark Vredeveld: Bayesian Scheduling: Analysis of Simple Policies

Bayesian scheduling is an extension of stochastic scheduling in which there is uncertainty about the system parameters. By processing jobs, we can learn about the true values of these parameters. We consider the basic scheduling problem of non-preemptively scheduling jobs on a single machine so as to minimize the expected total completion time. We consider a setting in which there are m classes of (a finite number of) jobs and the processing times of jobs of the same class are independent and identically exponentially distributed with an unknown parameter. The initial beliefs on this parameter are modelled as a prior distribution and after processing a job of the class, the posterior distribution models the updated beliefs on the parameter. For this setting, optimal policies based on Gittins-indices exist (see, e.g., Hamada and Glazebrook, 1993). However, computing these indices may be computationally challenging.

In this talk, we consider two simple policies, based on SEPT (Shortest Expected Processing Time), the optimal policy for the variant in which the parameters of the exponential distribution are known. We consider one version of SEPT, where the expected processing time are based on the initial belief and one version that updates the expected value each time a job has been processed. We can show that the first version of SEPT is at most a factor of m away from the optimal policy and that this is tight. Moreover, for the second variant we can show that it is not worse than the first and also a lower bound of $1 + \sqrt{m-1}$ on the approximation factor.

This is joint work with Sebastian Marban and Cyriel Rutten.

Thursday

Debmalya Panigrahi: Learning-augmented Assignment: Santa Claus does Load Balancing

Assignment problems are among the most well-studied in online algorithms. In these problems, a sequence of items arriving online must be assigned among a set of agents so as to optimize a given objective. This encompasses scheduling problems for minimizing makespan, p-norms, and other objectives, as well as fair division problems such as the Santa Claus problem and Nash welfare maximization. One common feature is that many of these problems are characterized by strong worst-case lower bounds in the online setting. To circumvent these impossibility results, recent research has focused on using additional (learned) information about the problem instance and this has led to dramatic improvements in the competitive ratio over the worst case. In this talk, I will first survey some of this literature (Lattanzi et al., SODA 20; Li and Xian, ICML 21; Banerjee et al., SODA 22; Barman et al., AAAI 22) that addresses specific problems in this domain. I will then proceed to describe recent work with Ilan Cohen that brings these problems under one umbrella: we give a single algorithmic framework for learning-augmented online assignment for a large class of maximization and minimization objectives.

Ali Vakilian: Faster Fundamental Graph Algorithms via Learned Predictions

We consider the question of speeding up classic graph algorithms with machine-learned predictions. In this model, algorithms are furnished with extra advice learned from past or similar instances. Given the additional information, we aim to improve upon the traditional worst-case run-time guarantees. Our contributions are the following:

(i) We give a faster algorithm for minimum-weight bipartite matching via learned duals, improving the recent result by Dinitz, Im, Lavastida, Moseley and Vassilvitskii (NeurIPS, 2021);

(ii) We extend the learned dual approach to the single-source shortest path problem (with negative edge lengths), achieving an almost linear runtime given sufficiently accurate predictions which improves upon the classic fastest algorithm due to Goldberg (SIAM J. Comput., 1995);

(iii) We provide a general reduction-based framework for learning-based graph algorithms, leading to new algorithms for degree-constrained subgraph and minimum-cost 0-1 flow, based on reductions to bipartite matching and the shortest path problem.

Finally, we give a set of general learnability theorems, showing that the predictions required by our algorithms can be efficiently learned in a PAC fashion.

Yossi Azar: Flow Time Scheduling with Uncertain Processing Time

We consider the problem of online scheduling on a single machine to minimize unweighted and weighted flow time. The existing algorithms for these problems require exact knowledge of the processing time of each job. This assumption is crucial, as even a slight perturbation of the processing time would lead to polynomial competitive ratio. However, this assumption very rarely holds in real-life scenarios. We present a competitive algorithm (the competitive ratio is a function of the distortion) for unweighted flow time that does not require knowledge of the distortion in advance. For the weighted flow time we present competitive algorithms but, in this case, we need to know (an upper bound on) the distortion in advance.
This is joint work with Stefano Leonardi and Noam Touitou based on papers that appeared in STOC 21 and SODA 2022.

Seffi Naor: Online Rounding of Bipartite Matchings

Two complementary facets of the online bipartite matching problem are discussed.
(1) For numerous online bipartite matching problems, such as edge-weighted matching and matching under two-sided vertex arrivals, state-of-the-art fractional algorithms outperform their randomized integral counterparts. Thus, a natural question is whether we can achieve lossless online rounding of fractional solutions in this setting. Even though lossless online rounding is impossible in general, randomized algorithms do induce fractional algorithms of the same competitive ratio, which by definition are losslessly roundable online. This motivates the addition of constraints that decrease the ``online integrality gap'', thus allowing for lossless online rounding. We characterize a set of non-convex constraints which allow for such lossless online rounding and allow for better competitive ratios than yielded by deterministic algorithms.
(2) In a different vein, we study the problem of rounding fractional bipartite matchings in online settings. We assume that a fractional solution is already generated for us online by a black box (via a fractional algorithm, or some machine-learned advice) and provided as part of the input, which we then wish to round. We provide improved bounds on the rounding ratio and discuss several applications.

Based on joint papers with Niv Buchbinder, Aravind Srinivasan, and David Wajc.

Sami Davies: Scheduling with communication delays

I'll discuss progress on scheduling with communication delays. In this setting, if two dependent jobs are scheduled on different machines, a delay must pass between their execution times. The question of whether constant factor approximation algorithms exist in this setting was one of the biggest open problems in scheduling theory. We effectively answered this question by (1) finding polylog approximations algorithms when the delay is uniform between dependent jobs and (2) proving super-constant hardness when the delay is non-uniform between dependent jobs. This is based on joint work with Janardhan Kulkarni, Thomas Rothvoss, Sai Sandeep, Jakub Tarnawski, and Yihao Zhang.

Friday

Vineet Goyal: Online Matching with Reusable Capacities

Online matching problems are at the heart of resource allocation given the inherent demand uncertainty. For instance, in a typical setting users arrive sequentially to a platform and the platform needs to make irrevocable matching decisions in an online manner. We consider fundamental generalizations of the classical variants in order to incorporate some of the natural stochasticity in resource usage that arises in many applications.

This talk mainly focuses on understanding the impact of reusability of resource capacities - a key aspect of resource allocation in applications such as cloud computing and sharing economies. Here allocated resources are required by users for some a priori unknown (stochastic) durations. Resources are returned after use and are available for re-allocation. We propose a new policy that achieves the best possible guarantee of (1-1/e - o(1)) under reasonable assumptions. Further, in the process of analyzing this policy we develop a novel framework of analysis that is useful more broadly in other settings with stochastic resource consumption. Based on joint work with Garud Iyengar and Rajan Udwani

Rudy Zhou: Minimizing Completion Times for Stochastic Jobs via Batched Free Times

In this talk, we consider the classic problem of minimizing the expected total completion time of jobs on m identical machines in the setting where the sizes of the jobs are stochastic. Specifically, the size of each job is a random variable whose distribution is known to the algorithm, but whose realization is revealed only after the job is scheduled. We give a O(m^(1/2) poly(log n))-approximation for stochastic jobs which have Bernoulli processing times. This is the first approximation for this problem that is both independent of the variance in the job sizes, and is sublinear in the number of machines m. Our algorithm is based on a novel reduction from minimizing the total completion time to a natural makespan-like objective, which we call the weighted free time. We hope this free time objective will be useful in further improvements to this problem, as well as other stochastic scheduling problems.

This is a joint work with Anupam Gupta and Benjamin Moseley.

Jens Schlöter: Learning-Augmented Query Policies for Minimum Spanning Tree with Uncertainty

We study how to utilize (possibly erroneous) predictions in a model for computing under uncertainty in which an algorithm can query unknown data. Our aim is to minimize the number of queries needed to solve the minimum spanning tree problem, a fundamental combinatorial optimization problem that has been central also to the research area of explorable uncertainty. For all integral $\gamma\ge 2$, we present algorithms that are γ-robust and $(1+(1/\gamma))$-consistent, meaning that they use at most γOPT queries if the predictions are arbitrarily wrong and at most $(1+1/\gamma) OPT$ queries if the predictions are correct, where OPT is the optimal number of queries for the given instance. Moreover, we show that this trade-off is best possible. Furthermore, we argue that a suitably defined hop distance is a useful measure for the amount of prediction error and design algorithms with performance guarantees that degrade smoothly with the hop distance. Our results demonstrate that untrusted predictions can circumvent the known lower bound of 2, without any degradation of the worst-case ratio. To obtain our results, we provide new structural insights for the minimum spanning tree problem that might be useful in the context of query-based algorithms regardless of predictions. In particular, we generalize the concept of witness sets---the key to lower-bounding the optimum---by proposing novel global witness set structures and completely new ways of adaptively using those. This is joint work with Thomas Erlebach, Murilo de Lima and Nicole Megow.

Cliff Stein: Queueing Safely and Efficiently for Elevators

Motivated by various COVID-19 related restrictions, we studied low tech interventions to reduce waiting time for elevators. We will describe several new interventions and discuss both mathematical and experimental analysis. We will also discuss some of the
issues involved in implementing these interventions in real buildings.

Joint work with Sai Mali Ananthanarayana, Charles Branas, Adam Elmachtoub, and Yeqing Zhou.