Publications of Tobias Friedrich

This document lists all peer-reviewed publications of Tobias Friedrich, Chair for Algorithm Engineering, Hasso Plattner Institute, Potsdam, Germany. This listing was automatically generated on November 21, 2022. An up-to-date version is available online at hpi.de/friedrich/docs/publist/friedrich.pdf.

Journal articles


   Schelling’s classical segregation model gives a coherent explanation for the wide-spread phenomenon of residential segregation. We introduce an agent-based saturated open-city variant, the Flip Schelling Process (FSP), in which agents, placed on a graph, have one out of two types and, based on the predominant type in their neighborhood, decide whether to change their types; similar to a new agent arriving as soon as another agent leaves the vertex. We investigate the probability that an edge \( u \) connects with monochrome and \( v \) have the same type in the FSP, and we provide a general framework for analyzing the influence of the underlying graph topology on residential segregation. In particular, for two adjacent vertices, we show that a highly decisive common neighborhood, i.e., a common neighborhood where the absolute value of the difference between the number of vertices with different \( \alpha \)-types is high, supports segregation and, moreover, that large common neighborhoods are more decisive. As an application, we study the expected behavior of the FSP on two common random graph models with and without geometry: (1) For random geometric graphs, we show that the existence of an edge \( u \) connects with monochrome neighborhood for \( u \) and \( v \) more likely. Based on this, we prove the existence of a constant \( c > 0 \) such that the expected fraction of monochrome edges after the FSP is at least \( 1/2 + c \).


   The hard-sphere model is one of the most extensively studied models in statistical physics. It describes the continuous distribution of spherical particles, governed by hard-core interactions. An important quantity of this model is the normalizing factor of this distribution, called the partition function. We propose a Markov chain Monte Carlo algorithm for approximating the grand canonical partition function of the hard-sphere model in \( d \) dimensions. Up to a fugacity of \( \lambda < 1/2 d \), the runtime of our algorithm is polynomial in the volume of the system. Key to our approach is to define a discretization that closely approximates the partition function of the continuous model. This results in a discrete hard-core instance that is exponential in the size of the initial hard-sphere model. Our approximation bound follows directly from the correlation decay threshold of an infinite regular tree with degree equal to the maximum degree of our discretization. To cope with the exponential blow-up of the discrete instance, we use block dynamics, a Markov chain that generalizes the more frequently studied Glauber dynamics by grouping the vertices of the graph into blocks and updating an entire block instead of a single vertex in each step. We prove rapid mixing of block dynamics, based on disjoint cliques as blocks, up to the tree threshold of the univariate hard-core model. This is achieved by adapting the spectral expansion method, which was recently used for bounding the mixing time of Glauber dynamics within the same parameter regime.


   We present fully polynomial time approximation schemes for a broad class of Holant problems with complex edge weights, which we call Holant polynomials. We transform these problems into partition functions of abstract combinatorial structures known as polymers in statistical physics. Our method involves establishing zero-free regions for the partition functions of polymer models and using the most significant terms of the cluster expansion to approximate them. Results of our technique include new approximation and sampling algorithms for a diverse class of Holant polynomials in the low-temperature regime (i.e. small external field) and approximation algorithms for general Holant problems with small signature weights. Additionally, we give randomised approximation and sampling algorithms with faster running times for more restrictive classes. Finally, we improve the known zero-free regions for a perfect matching polynomial.


   A standard approach to accelerating shortest path algorithms on networks is the bidirectional search, which explores the graph from the start and the destination, simultaneously. In practice this strategy performs particularly well on scale-free real-world networks. Such networks typically have a heterogeneous degree distribution (e.g., a power-law distribution) and high clustering (i.e., vertices with a common neighbor are likely to be connected themselves). These two properties can be obtained by assuming an underlying hyperbolic geometry. To explain the observed behavior of the bidirectional search, we analyze its running time on hyperbolic random graphs and prove that it is \( O(n^{2-1/\alpha}+r^{1/(2\alpha)}+\delta_{\text{max}}) \) with high probability, where \( \alpha \) controls the power-law exponent of the degree distribution, and \( \delta_{\text{max}} \) is the maximum degree. This bound is sublinear, improving the obvious worst-case linear bound. Although our analysis depends on the underlying geometry, the algorithm itself is oblivious to it.

This vision paper outlines the main building blocks of what we term AI Compliance, an effort to bridge two complementary research areas: computer science and the law. Such research has the goal to model, measure, and affect the quality of AI artifacts, such as data, models and applications, to then facilitate adherence to legal standards.


Multi-column dependencies in relational databases come associated with two different computational tasks. The detection problem is to decide whether a dependency of a certain type and size holds in a given database, the discovery problem asks to enumerate all valid dependencies of that type. We settle the complexity of both of these problems for unique column combinations (UCCs), functional dependencies (FDs), and inclusion dependencies (INDs). We show that the detection of UCCs and FDs is \(W[2]\)-complete when parameterized by the solution size. The discovery of inclusion-wise minimal UCCs is proven to be equivalent under parsimonious reductions to the transversal hypergraph problem of enumerating the minimal hitting sets of a hypergraph. The discovery of FDs is equivalent to the simultaneous enumeration of the hitting sets of multiple input hypergraphs. We further identify the detection of INDs as one of the first natural \(W[3]\)-complete problems. The discovery of maximal INDs is shown to be equivalent to enumerating the maximal satisfying assignments of antimonotone, 3-normalized Boolean formulas.


The transversal hypergraph problem asks to enumerate the minimal hitting sets of a hypergraph. If the solutions have bounded size, Eiter and Gottlob [SICOMP’95] gave an algorithm running in output-polynomial time, but whose space requirement also scales with the output. We improve this to polynomial delay and space. Central to our approach is the extension problem, deciding for a set \(X\) of vertices whether it is contained in any minimal hitting set. We show that this is one of the first natural problems to be \(W[3]\)-complete.

We give an algorithm for the extension problem running in time \(O(m |X|+1 n)\) and prove a SETH-lower bound showing that this is close to optimal. We apply our enumeration method to the discovery problem of minimal unique column combinations from data profiling. Our empirical evaluation suggests that the algorithm outperforms its worst-case guarantees on hypergraphs stemming from real-world databases.


The computational complexity of the VertexCover problem has been studied extensively. Most notably, it is \(NP\)-complete to find an optimal solution and typically \(NP\)-hard to find an approximation with reasonable factors. In contrast, recent experiments suggest that on many real-world networks the run time to solve VertexCover is way smaller than even the best known \(FPT\)-approaches can explain. We link these observations to two properties that are observed in many real-world networks, namely a heterogeneous degree distribution and high clustering. To formalize these properties and explain the observed behavior, we analyze how a branch-and-reduce algorithm performs on hyperbolic random graphs, which have become increasingly popular for modeling real-world networks. In fact, we are able to show that the VertexCover problem on hyperbolic random graphs can be solved in polynomial time, with high probability. The proof relies on interesting structural properties of hyperbolic random graphs. Since these predictions of the model are interesting in their own right, we conducted experiments on real-world networks showing that these properties are also observed in practice.


Unique column combinations (UCCs) are a fundamental concept in relational databases. They identify entities in the data and support various data management activities. Still, UCCs are usually not explicitly defined and need to be discovered. State-of-the-art data profiling algorithms are able to efficiently discover UCCs in moderately sized datasets, but they tend to fail on large and, in particular, on wide datasets due to run time and memory limitations. In this paper, we introduce HPIValid, a novel UCC discovery algorithm that implements a faster and more resource-saving search strategy. HPIValid models the metadata discovery as a hitting set enumeration problem in hypergraphs. In this way, it combines efficient discovery techniques from data profiling research with the most recent theoretical insights into enumeration algorithms. Our evaluation shows that HPIValid is not only orders of magnitude faster than related work, it also has a much smaller memory footprint.


Large real-world networks typically follow a power-law degree distribution. To study such networks, numerous random graph models have been proposed. However, real-world networks are not drawn at random. Therefore, Brach, Cygan, Lacki, and Sankowski [SODA 2016] introduced two natural deterministic conditions: (1) a power-law upper bound on the degree distribution (PLB-U) and (2) power-law neighborhoods, that is, the degree distribution of neighbors of each vertex is also upper bounded by a power law (PLB-N). They showed that many real-world networks satisfy both deterministic properties and exploit them to design faster algorithms for a number of classical graph problems. We complement the work of Brach et al. by showing that some well-studied random graph models exhibit both the mentioned PLB properties and additionally also a power-law lower bound on the degree distribution (PLB-L). All three properties hold with high probability for Chung–Lu Random Graphs and Geometric Inhomogeneous Random Graphs and almost surely for Hyperbolic Random Graphs. As a consequence, all results of Brach et al. also hold with high probability or almost surely for those random graph classes. In the second part of this work we study three classical NP-hard combinatorial optimization problems on PLB networks. It is known that on general graphs with maximum degree \(\Delta\), a greedy algorithm, which chooses nodes in the order of their degree, only achieves a \(\Omega(\log \Delta)\)-approximation for Minimum Vertex Cover and Minimum Dominating Set, and a \(\Omega(\Delta)\)-approximation for Maximum Independent Set. We prove that the PLB-L property suffices for the greedy approach to achieve a constant-factor approximation for all three problems. We also show that all three combinatorial optimization problems are \(APX\)-complete even if all PLB-properties holds hence, PTAS cannot be expected unless \(P=NP\).

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 success in real-world networks, and we try to minimize the lengths of the resulting greedy paths. We evaluate our algorithm on multiple generated and real-world 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. In addition to measuring the stretch, we empirically evaluate our algorithms regarding the size of the coordinates of the resulting embeddings and observe how it impacts the success rate when coordinates are not represented with very high precision. Since numerical difficulties are a major issue when performing computations in the hyperbolic plane, we consider variations of our algorithm that improve the success rate when using coordinates with lower precision.


In the article "Reoptimization Time Analysis of Evolutionary Algorithms on Linear Functions Under Dynamic Uniform Constraints", we claimed a worst-case runtime of \(O(nD \log D)\) and \(O(nD)\) for the Multi-Objective Evolutionary Algorithm and the Multi-Objective \((\mu + \lambda)\)-Genetic Algorithm, respectively, on linear profit functions under dynamic uniform constraint, where \(D = |B - B^*|\) denotes the difference between the original constraint bound \(B\) and the new one \(B^*\). The technique used to prove these results contained an error. We correct this mistake and show a weaker bound of \(O(nD^2)\) for both algorithms instead.


Linear functions have gained great attention in the run time analysis of evolutionary computation methods. The corresponding investigations have provided many effective tools for analyzing more complex problems. So far, the runtime analysis of evolutionary algorithms has mainly focused on unconstrained problems, but problems occurring in applications frequently involve constraints. Therefore, there is a strong need to extend the methods for analyzing unconstrained problems to a setting involving constraints. In this paper, we consider the behavior of the classical \((1+1)\) evolutionary algorithm on linear functions under linear constraint. We show tight bounds in the case where the constraint is given by the OneMax function and the objective function is given by either the OneMax or the BinVal function. For the general case we present upper and lower bounds.


Evolutionary algorithms have been widely used in the area of creativity. Recently, evolutionary processes have been used to create artistic image transition processes using random walks. In this paper, we explore the use of quasi-random walks for evolutionary image transition and animation. Quasi-random walks show similar features as standard random walks, but with much less randomness. We utilize this extended model from discrete mathematics and show how agents carrying out quasi-random walks can be used for evolutionary image transition and animation. The key idea is to generalize the notion of quasi-random walks and let a set of autonomous agents perform quasi-random walks painting an image. Each agent has one particular target image that they paint when following a sequence of directions for their quasi-random walk. The sequence can easily be chosen by a user and allows them to produce a wide range of different transition patterns and animations.


A significant percentage of urban traffic is caused by the search for parking spots. One possible approach to improve this situation is to guide drivers along routes which are likely to have free parking spots. The task of finding such a route can be modeled as a probabilistic graph problem which is NP-complete. Thus, we propose heuristic approaches for solving this problem and evaluate them experimentally. For this, we use probabilities of finding a parking spot, which are based on publicly available empirical data from TomTom International B.V. An algorithmic heuristic that relies exclusively on conventional road attributes. Our experiments show that this algorithm comes close to the baseline by a factor of 1.3 in our cost measure. Last, we complement our experiments with results from a field study, comparing the success rates of our algorithms against real human drivers.


In the context of black-box optimization, black-box complexity is used for understanding the inherent difficulty of a given optimization problem. Central to our understanding of nature-inspired search heuristics in this context is the notion of unbiasedness. Specialized black-box complexities have been developed in order to better understand the limitations of these heuristics – especially of (population-based) evolutionary algorithms (EAs). In contrast to this, we focus on a model for algorithms explicitly maintaining a probability distribution over the search space: so-called estimation-of-distribution algorithms (EDAs). We consider the recently introduced \(n\)-Bernoulli-\(\lambda\)-EDA framework; which subsumes, for example, the commonly known EDAs PBIL, UMDA, \(\lambda\)-MMAS\(\lambda\), and \(\lambda\)-IB. We show that an \(n\)-Bernoulli-\(\lambda\)-EDA is unbiased if and only if its probability distribution satisfies a certain invariance property under isometric automorphisms of \([0,1]^n\). By restricting how an \(n\)-Bernoulli-\(\lambda\)-EDA can perform an update, in a way common to many examples, we derive conciser characterizations, which are easy to verify. We demonstrate this by showing that our examples above are all unbiased.


Rigorous runtime analysis is a major approach towards understanding evolutionary computing techniques, and in this area linear pseudo-Boolean objective functions play a central role. Having an additional linear constraint is then equivalent to the NP-hard Knapsack problem, certain classes thereof have been studied in recent works. In this article, we present a dynamic model of optimizing linear functions under uniform constraints. Starting from an optimal solution with respect to a given constraint bound, we investigate the runtimes that different evolutionary algorithms need to recompute an optimal solution when the constraint bound changes by

3
a certain amount. The classical $(1+1)$ EA and several population-based algorithms are designed for that purpose, and are shown to recompute efficiently. Furthermore, a variant of the $(1+(\lambda,\lambda))$ GA for the dynamic optimization problem is studied, whose performance is better when the change of the constraint bound is small.


Island models in evolutionary computation solve problems by a careful interplay of independently running evolutionary algorithms on the island and an exchange of good solutions between the islands. In this work, we conduct rigorous run time analyses for such island models trying to simultaneously obtain good run times and low communication effort. We improve the existing upper bounds for both measures (i) by improving the run time bounds via a careful analysis, (ii) by balancing individual computation and communication in a more appropriate manner, and (iii) by replacing the usual communicate-with-all approach with randomized rumor spreading. In the latter, each island contacts a randomly chosen neighbor. This epidemic communication paradigm is known to lead to very fast and robust information dissemination in many applications. Our results concern island models running simple $(1+1)$ evolutionary algorithms to optimize the classic test functions OneMax and LeadingOnes. We investigate binary trees, $d$-dimensional tori, and complete graphs as communication topologies.


Random walks are frequently used in randomized algorithms. We study a derandomized variant of a random walk on graphs, called rotor-router model. In this model, instead of distributing tokens randomly, each vertex serves its fixed deterministic order. For most setups, both processes behave remarkably similar: Starting with the same initial configuration, the number of tokens in the rotor-router model deviates only slightly from the expected number of tokens on the corresponding vertex in the random walk model. The maximal difference over all vertices and all times is called single vertex discrepancy. Cooper and Spencer (2006) showed that on $Z^d$ the single vertex discrepancy is only a constant $c_d$. Other authors also determined the precise value of $c_d$ for $d = 1, 2$. All previous results, however, assume that initially all tokens are only placed on one partition of the bipartite graph $Z^d$. We show that this assumption is crucial by proving that otherwise the single vertex discrepancy can become arbitrarily large. For all dimensions $d \geq 1$ and arbitrary discrepancies $\ell \geq 0$, we construct configurations that reach a discrepancy of at least $\ell$.


Hyperbolic geometry appears to be intrinsic in many large real networks. We construct and implement a new maximum likelihood estimation algorithm that embeds scale-free graphs in the hyperbolic space. All previous approaches of similar embedding algorithms require at least a quadratic runtime. Our algorithm achieves quasilinear runtime, which makes it the first algorithm that can embed networks with hundreds of thousands of nodes in less than one hour. We demonstrate the performance of our algorithm on artificial and real networks. In all typical metrics, like Log-likelihood and greedy routing, our algorithm discovers embeddings that are very close to the ground truth.


Large real-world networks are typically scale-free. Recent research has shown that such graphs are described best in a geometric space. More precisely, the internet can be mapped to a hyperbolic space such that geometric greedy routing is close to optimal (Boguñá, Papadopoulos, and Krioukov. Nature Communications, 1:62, 2010). This observation has pushed the interest in hyperbolic networks as a natural model for scale-free networks. Hyperbolic random graphs follow a power law degree distribution with controllable exponent $\beta$ and show high clustering (Gugelmann, Panagiotou, and Peter. ICALP, pp. 573–585, 2012). For understanding the structure of the resulting graphs and for analyzing the behavior of network algorithms, the next question is bounding the size of the diameter. The only known explicit bound is $O((log n)^{2(1-(\beta - 3)/3\beta)}})$ (Kiwi and Mitsche. ANALCO, pp. 26–39, 2015). We present two much simpler proofs for an improved upper bound of $O((log n)^{1/(3-\beta)})$ and a lower bound of $O((log n)^{\beta})$. If $\beta > 3$, we show that the latter bound is tight by proving an upper bound of $O((log n)^{\beta})$ for the diameter.


There are hundreds of online social networks with altogether billions of users. Many such networks publicly release structural information, with all personal information removed. Empirical studies have shown, however, that this provides a false sense of privacy - it is possible to identify almost all users that appear in two such anonymized network as long as a few initial mappings are seed nodes (for an arbitrarily small $\epsilon$) and runs in quasilinear time. This improves previous theoretical results which need $n^2$ seed nodes and have runtimes of order $n^{1+\Omega(1)}$. Additionally, the applicability of our algorithm is studied experimentally on different networks.


Population diversity is essential for the effective use of any crossover operator. We compare seven commonly used diversity mechanisms and prove rigorous run time bounds for the $(\mu+1)$ GA using uniform crossover on the fitness function Jump. All previous results in this context only hold for unrealistically low crossover probability $p_c=O(1/n)$, while we give analyses for the setting of constant $p_c \leq 1$ in all but one case. Our bounds show a dependence on the problem size $n$, the jump length $k$, the population size $\mu$, and each constant case $p_c$. For crossover probability $p_c$, we can compare the resulting expected optimisation times for different diversity mechanisms assuming an optimal choice of $\mu$. $O(n^{2(1-\alpha)})$ for duplicate elimination/minimisation, $O(n^2 \log n)$ for maximising the convex hull, $O(n \log n)$ for det. crowding (assuming $p_c = k/n$), $O(n \log n)$ for maximising the Hamming distance, $O(n \log n)$ for fitness sharing, $O(n \log n)$ for the single-receiver island model. This proves a sizeable advantage of all variants of the $(\mu+1)$ GA compared to the $(1+1)$ EA, which requires $O(n^2)$. In a short empirical study we confirm that the asymptotic differences can also be observed experimentally.

Most complex real world networks display scale-free features. This characteristic motivated the study of numerous random graph models with a power-law degree distribution. There is, however, no established and simple model which also has a high clustering of vertices as typically observed in real data. Hyperbolic random graphs bridge this gap. This natural model has recently been introduced by Krzaklewski et al. and has shown theoretically to fulfill all typical properties of real world networks, including power-law degree distribution and high clustering. We study cliques in hyperbolic random graphs and present new results on the expected number of $k$-cliques $E[K_k]$ and the size of the largest clique $\omega(G)$. We observe that there is a phase transition at power-law exponent $\beta = 3$. More precisely, for $\beta \in (2, 3)$ we prove $E[K_k] = n^{(3-\beta)k}/k^{k-1}$ and $\omega(G) = \Theta(n^{(3-\beta)/2})$, while for $\beta \geq 3$ we prove $E[K_k] = n^{k}\Theta(k)^{-k}$ and $\omega(G) = \Theta(\log(n)/\log \log n)$. Furthermore, we show that for $\beta \geq 3$, cliques in hyperbolic random graphs can be computed in time $O(n)$. If the underlying geometry is known, cliques can be found with worst-case runtime $O(mn^{3/2})$ for all values of $\beta$.


In this paper we initiate the study of job scheduling on related and unrelated machines so as to minimize the maximum flow time or the maximum weighted flow time (when each job has an associated weight). Previous work for these metrics considered only the setting of parallel machines, while previous work on unrelated machines only considered $L_p, p < \infty$ norms. Our main results are: (1) We give an $O(\epsilon^{-1})$-competitive algorithm to minimize maximum weight regret on related machines where we assume that the machines of the online algorithm can process $1 + \epsilon$ units of a job in 1 time-unit ($\epsilon$ speed augmentation). (2) For the objective of minimizing maximum flow time on unrelated machines we give a simple $2/\epsilon$-competitive algorithm when we augment the speed by $\epsilon$. For $m$ machines we show a lower bound of $O(m)$ on the competitive ratio if speed augmentation is not permitted. Our algorithm does not assign jobs to machines as soon as they arrive. To justify this "drawback" we show a lower bound of $\Omega(log m)$ on the competitive ratio of any online algorithm which is permitted to use $s = O(1)$ speed machines. In our lower bound construction, job $j$ has a processing time of $\mu_j$ on a subset of $m$ machines and infinity on others, and we have a weight $1/\mu_j$. Hence this lower bound applies to the subset parallel setting for the special case of minimizing maximum stretch.


Practical optimization problems frequently include uncertainty about the quality measure, for example due to noisy evaluations. Thus, they do not allow for a straightforward application of traditional optimization techniques. In these settings, randomized search heuristics such as evolutionary algorithms are a popular choice because they are often assumed to exhibit some kind of robustness to noise. Empirical evidence suggests that some algorithms, such as estimation of distribution algorithms (EDAs) are robust against a scaling of the noise intensity, even without resorting to explicit noise-handling techniques such as resampling. In this paper, we want to support such claims with mathematical rigor. We introduce the concept of graceful scaling in which the run time of an algorithm scales polynomially with noise intensity. We study a monotone fitness function over binary strings with additive noise taken from a Gaussian distribution. We show that myopic heuristics cannot efficiently optimize the function under arbitrarily intense noise without any explicit noise-handling. Furthermore, we prove that using a population does not help. Finally we prove that a simple EDA called the compact Genetic Algorithm can overcome the shortsightedness of mutation-only heuristics to scale gracefully with noise. We conjecture that recombative genetic algorithms also have this property.


Recently Ant Colony Optimization (ACO) algorithms have been proven to be efficient in uncertain environments, such as noisy or dynamically changing fitness functions. Most of these analyses focus on combinatorial problems, such as path finding. We analyze an ACO algorithm in a setting where we try to optimize the simple OneMax test function, but with additive posterior noise sampled from a Gaussian distribution. Without noise the classical $(\mu+1)$-EA outperforms any ACO algorithm, with smaller $\mu$ being better. However, with large noise, the $(\mu+1)$-EA fails, even for high values of $\mu$ (which are known to help against small noise). In this paper, we show that ACO is able to deal with arbitrarily large noise in a graceful manner, that is, as long as the evaporation factor $\rho$ is small enough dependent on the parameter $\sigma^2$ of the noise and the dimension $D$ of the search space ($p = o(D/n + D \delta n \log D)$), optimization will be successful.


Evolutionary computation and other nature-inspired search heuristics are known and applied on a daily basis for 50 years. They remain an important tool in situations where difficult problems need to be solved and no good problem-specific solution is available. Thanks to continuous efforts directed at a theoretical foundation of this broad and complex set of heuristics we have a much improved understanding of their properties, strengths and limitations. This special issue contains a collection of theoretical analyses of quite different nature-inspired heuristics, all presented in preliminary form either at the field’s largest conference, the Genetic and Evolutionary Computation Conference (GECCO 2013), or at a small and specialized workshop on Theory of Randomized Search Heuristics (ThRaSH 2013) that provides an opportunity for theoreticians in the field to meet and discuss their work since 2007. All improved understanding of their properties, strengths and limitations. This special issue contains a collection of theoretical analyses of quite different nature-inspired heuristics, all presented in preliminary form either at the field’s largest conference, the Genetic and Evolutionary Computation Conference (GECCO 2013), or at a small and specialized workshop on Theory of Randomized Search Heuristics (ThRaSH 2013) that provides an opportunity for theoreticians in the field to meet and discuss their work since 2007. All articles presented here have been reworked and significantly expanded beyond their initial presentation and they all witness that theory is now developed well beyond the understanding of very simple evolutionary algorithms on simple example functions.


We present a new randomized diffusion-based algorithm for balancing indivisible tasks (tokens) on a network. Our aim is to minimize the discrepancy between the maximum and minimum load. The algorithm works as follows. Every vertex distributes its tokens as evenly as possible among its neighbors and itself. If this is not possible without splitting some tokens, the vertex redistributes its excess tokens among all its neighbors randomly (without replacement). In this paper we prove several upper bounds on the load discrepancy for general networks. These bounds depend on some expansion properties of the network, that is, the second largest eigenvalue, and a novel measure which we refer to as refined local divergence. We then apply these general bounds to obtain results for some specific networks. For constant-degree expanders and torus graphs, these yield exponential improvements on the discrepancy bounds. For hypercubes we compared to the algorithm of Rabani, Sinclair, and Wanka. In contrast to previous papers, our algorithm is vertex-based and not edge-based. This means excess tokens are assigned to vertices instead to
edges, and the reallocate all of its excess tokens by itself. This approach avoids nodes having "negative loads", but causes additional dependencies for the analysis.


Many combinatorial optimization problems have underlying goal functions that are submodular. The classical goal is to find a good solution for a given submodular function \( f \) under a given set of constraints. In this paper, we investigate the runtime of a simple single-objective evolutionary algorithm called (1+1) E and a multi-objective evolutionary algorithm called GSEMO until they have obtained a good approximation for submodular functions. For the case of monotone submodular functions and uniform cardinality constraints we show that the GSEMO achieves a \((1 - \frac{1}{e})\)-approximation in expected polynomial time. For the case of monotone functions where the constraints are given by the intersection of \( k \geq 2 \) matroids, we show that the (1+1) EA achieves a \((1 + \frac{k}{\delta})\)-approximation in expected polynomial time for any constant \( \delta > 0 \). Turning to non-monotone symmetric submodular functions with \( k \geq 1 \) matroid intersection constraints, we show that the GSEMO achieves a \( \frac{1}{\log(k+\delta)}\)-approximation in expected time \( O(k^{\frac{k}{\delta}} \log(n))/\epsilon \).


The theory of population genetics and evolutionary computation have been evolving separately for nearly 30 years. Many results have been independently obtained in both fields and many others are unique to its respective field. We aim to bridge this gap by developing a unifying framework for evolutionary processes that allows both evolutionary algorithms and population genetics models to be cast in the same formal framework. The framework we present here decomposes the evolutionary process into its several components in order to facilitate the identification of similarities between different models. In particular, we propose a classification of evolutionary operators based on the defining properties of the different components. We cast several commonly used operators from both fields into this common framework. Using this, we map different evolutionary and genetic algorithms to different evolutionary regimes and identify candidates with the most potential for the translation of results between the fields. This provides a unified description of evolutionary processes and represents a stepping stone towards new tools and results to both fields.


The k-Clique problem is a fundamental combinatorial problem that plays a prominent role in classical as well as in parameterized complexity theory. It is among the most well-known NP-complete and W[1]-complete problems. Moreover, its average-case complexity analysis has created a long thread of research already since the 1970s. Here, we continue this line of research by studying the dependence of the average-case complexity of the k-Clique problem on the parameter \( k \). To this end, we define two natural parameterized analogs of efficient average-case algorithms. We then show that k-Clique admits both analogues for Erdős-Rényi random graphs of arbitrary density. We also show that k-Clique is unlikely to admit either of these analogs for some specific computable input distribution.


Covering all edges of a graph by a minimum number of cliques is a well known NP-hard problem. For the parameter \( k \) being the maximal number of cliques to be used, the problem becomes fixed parameter tractable. However, assuming the Exponential Time Hypothesis, there is no kernel of subexponential size in the worst-case. We study the average kernel size for random intersection graphs of arbitrary density. We also show that k-Clique is unlikely to admit either of these analogs for some specific computable input distribution.


Most experimental studies initialize the population of evolutionary algorithms with random genotypes. In practice, however, optimizers are typically seeded with good candidate solutions either previously known or created according to some problem-specific method. This seeding has been studied extensively for single-objective problems. For multi-objective problems, however, very little literature is available on the approaches to seeding and their individual benefits and disadvantages. In this article, we are trying to narrow this gap via a comprehensive computational study on common real-valued test functions. We investigate the effect of two seeding techniques for five algorithms on 48 optimization problems with 2, 3, 4, 6, and 8 objectives. We observe that some functions (e.g., DTLZ4 and the LZ family) benefit significantly from seeding, while others (e.g., WFG) profit less. The advantage of seeding


Many optimization problems arising in applications have to consider several objective functions at the same time. Evolutionary algorithms seem to be a very natural choice for dealing with multi-objective problems as the population of such an algorithm can be used to represent the trade-offs with respect to the given objective functions. In this paper, we contribute to the theoretical understanding of evolutionary algorithms for multi-objective problems. We consider indicator-based algorithms whose goal is to maximize the hypervolume for a given set of points on the Pareto front by distributing \( \mu \) points on the Pareto front. For the class of linear fronts and a class of convex fronts, we prove that maximizing the hypervolume gives the best possible approximation ratio when assuming that the extreme points have to be included in both distributions of the points on the Pareto front. Furthermore, we investigate the choice of the reference point on the approximation behavior of hypervolume-based approaches and examine Pareto fronts of different shapes by numerical calculations.

Finding cliques in graphs is a classical problem which is in general NP-hard and parameterized intractable. In typical applications like social networks or biological networks, however, the considered graphs are scale-free, i.e., their degree sequence follows a power law. Their specific structure can be algorithmically exploited and makes it possible to solve clique much more efficiently. We prove that on inhomogeneous random graphs with $n$ nodes and power law exponent $\beta$, cliques of size $k$ can be found in time $O(n^3 \beta)$ for $\beta \geq 3$ and in time $O(n^k \beta)$ for $2 < \beta < 3$.


Multi-objective optimization problems arise frequently in applications, but can only often be solved approximately by heuristic approaches. Evolutionary optimization algorithms have been widely used to tackle multi-objective problems. These algorithms use different measures to ensure diversity in the objective space but are not guided by a formal notion of approximation. We present a framework for evolutionary multi-objective optimization that allows to work with a formal notion of approximation. This approximation-guided evolutionary algorithm (AGE) has a worst-case runtime linear in the number of objectives and works with an archive that is an approximation of the non-dominated objective vectors seen during the run of the algorithm. Our experimental results show that AGE finds competitive or better solutions not only regarding the achieved approximation, but also regarding the total hypervolume. For all considered test problems, even for many (i.e., more than ten) dimensions, AGE discovers a good approximation of the Pareto front. This is not the case for established algorithms such as NSGA-II, SPEA2, and SMS-EMOA. In this paper we compare AGE with two additional algorithms that use very fast hypervolume-approximations to guide their search. This significantly speeds up the runtime of the hypervolume-based algorithms, which now allows a comparison of the underlying selection schemes.


We propose and analyze a quasirandom analogue of the classical push model for disseminating information in networks ("randomized rumor spreading"). In the classical model, in each round, each informed vertex chooses a neighbor at random and informs it, if it was not informed before. It is known that this simple protocol succeeds in spreading a rumor from one vertex to all others within $O(\log n)$ rounds on complete graphs, hypercubes, random regular graphs, Erdős-Rényi random graphs, and Ramanujan graphs with probability $1-o(1)$. In the quasirandom model, we assume that each vertex has a (cyclic) list of its neighbors. Once informed, it starts at a random position on the list, but from then on informs its neighbors in the order of the list. Surprisingly, irrespective of the orders of the lists, the above-mentioned bounds still hold. In some cases, even better bounds than for the classical model can be shown.


Multiobjective evolutionary algorithms typically maintain a set of solutions. A crucial part of these algorithms is the archiving, which decides what solutions to keep. A $(\mu+\lambda)$-archiving algorithm defines how to choose in each generation $\mu$ children from $\mu$ parents and $\lambda$ offspring together. We study mathematically the convergence behavior of hypervolume-based archiving algorithms.

We distinguish two cases for the offspring generation. A best-case view leads to a study of the effectiveness of archiving algorithms. This measures how much smaller hypervolumes are achieved due to the use of an archive. On the other hand, a worstcase view on the offspring generation leads to a study of the competitive ratio of archiving algorithms. This measures how much smaller hypervolumes are achieved due to not knowing the future offspring in advance. We present upper and lower bounds on the competitive advantage of different archiving algorithms and present an archiving algorithm, which is the first known computationally efficient archiving algorithm with constant competitive ratio.


We give an algorithm that computes the final state of certain growth models without computing all intermediate states. Our technique is based on a "least action principle" which characterizes the odometry function of the growth process. Starting from an approximation for the odometry, we successively correct under- and overestimates and provably arrive at the correct final state. The degree of speedup depends on the accuracy of the initial guess. Determining the size of the boundary fluctuations in growth models like internal diffusion-limited aggregation (IDLA) is a long-standing open problem in statistical physics. As an application of our method, we calculate the size of fluctuations over two orders of magnitude beyond previous simulations.


A random geometric graph (RGG) is defined by placing $n$ points uniformly at random in $[0, n^{1/4}]^d$, and joining two points by an edge whenever their Euclidean distance is at most some fixed $r$. We assume that $r$ is larger than the critical value for the emergence of a connected component with $\Omega(n)$ nodes. We show that, with high probability (w.h.p.), for any two connected nodes with a Euclidean distance of $\omega(\log n/r^{d-1})$, their graph distance is only a constant factor larger than their Euclidean distance. This implies that the diameter of the largest connected component is $\Theta(n^{1/d}/r)$ w.h.p. We also prove that the condition on the Euclidean distance above is essentially tight. We also analyze the following randomized broadcast algorithm on RGGs. At the beginning, only one node from the largest connected component of the RGG is informed. Then, in each round, each informed node chooses a neighbor independently and uniformly at random and informs it. We prove that w.h.p. this algorithm informs every node in the largest connected component of an RGG within $\Theta(n^{1/d}/r + \log n)$ rounds.

Wind energy plays an increasing role in the supply of energy worldwide. The energy output of a wind farm is highly dependent on the weather conditions present at its site. If the output can be predicted more accurately, energy suppliers can coordinate the collaborative production of different energy sources more efficiently to avoid costly overproduction. In this paper, we take a computer science perspective on energy prediction based on weather data and analyze the important parameters as well as their correlation on the energy output. To deal with the interaction of the different parameters, we use symbolic regression based on the genetic programming tool DataModeler. Our studies are carried out on publicly available weather and energy data for a wind farm in Australia. We report on the correlation of the different variables for the energy output. The model obtained for energy prediction gives a very reliable prediction of the energy output for newly supplied weather data.


In order to allow a comparison of (otherwise incomparable) sets, many evolutionary multi-objective optimizers use indicator functions to guide the search and to evaluate the performance of search algorithms. The most widely used indicator is the hypervolume indicator. It measures the volume of the dominated portion of the objective space bounded from below by a reference point. Though the hypervolume indicator is very popular, it has not been shown that maximizing the hypervolume indicator of sets of bounded size is indeed equivalent to the overall objective of finding a good approximation of the Pareto front. To address this question, we compare the optimal approximation ratio with the approximation ratio achieved by two-dimensional sets maximizing the hypervolume indicator. We bound the optimal multiplicative approximation ratio of \( n \) points by \( 1 + \Theta(1/n) \) for arbitrary Pareto fronts. Furthermore, we prove that the same asymptotic approximation ratio is achieved by sets of \( n \) points that maximize the hypervolume indicator. However, there is a provable gap between the two approximation ratios which is even exponential in the ratio between the largest and the smallest value of the front. We also examine the additive approximation ratio of the hypervolume indicator in two dimensions and prove that it achieves the optimal additive approximation ratio apart from a small ratio.


Many state-of-the-art evolutionary vector optimization algorithms compute the contributing hypervolume for ranking candidate solutions. However, with an increasing number of objectives, calculating the volumes becomes intractable. Therefore, although hypervolume-based algorithms are often the method of choice for bi-criteria optimization, they are regarded as not suitable for many-objective optimization. Recently, Monte Carlo methods have been derived and analyzed for approximating the contributing hypervolume. Turning theory into practice, we employ these results in the ranking procedure of the multi-objective covariance matrix adaptation evolution strategy (MO-CMA-ES) as an example of a state-of-the-art method for vector optimization. It is empirically shown that the approximation does not impair the quality of the obtained solutions given a budget of objective function evaluations, while considerably reducing the computation time in the case of multiple objectives. These results are obtained on common benchmark functions as well as on two design optimization tasks. Thus, employing Monte Carlo approximations makes hypervolume-based algorithms applicable to many-objective optimization.


Evolutionary algorithms have been widely used to tackle multi-objective optimization problems. Incorporating preference information into the search of evolutionary algorithms for multi-objective optimization is of great importance as it allows one to focus on interesting regions in the objective space. Zitzler et al. have shown how to use a weight distribution function on the objective space to incorporate preference information into hypervolume-based algorithms. We show that this weighted information can easily be used in other popular EMO algorithms as well. Our results for NSGA-II and SPEA2 show that this yields similar results to the hypervolume approach and requires less computational effort.


We examine the complexity of constraint satisfaction problems that consist of a set of AllDiff constraints. Such CSPs naturally model a wide range of real-world and combinatorial problems, like scheduling, frequency allocations, and graph coloring problems. As this problem is known to be NP-complete, we investigate under which further assumptions it becomes tractable. We observe that a crucial property seems to be the convexity of the variable domains and constraints. Our main contribution is an extensive study of the complexity of Multiple AllDiff CSPs for a set of natural parameters, like maximum size and maximum size of the constraint scopes. We show that, depending on the parameter, convexity can make the problem tractable even though it is provably intractable in general. Interestingly, the convexity of constraints is the key property in achieving fixed parameter tractability, while the convexity of domains does not usually make the problem easier.


Systems biology has emerged over the last decade. Driven by the advances in sophisticated measurement technology the research community generated huge molecular biology data sets. This comprises rather static data on the interplay of biological entities, for instance protein-protein interaction network data, as well as quite dynamic data collected for studying the behavior of individual cells or tissues in accordance to changing environmental conditions, such as DNA microarrays or RNA sequencing. Here we bring the two different data types together in order to gain higher level knowledge. We introduce a significantly improved version of the KeyPathwayMiner software framework. Given a biological network modelled as graph and a set of expression studies, KeyPathwayMiner efficiently finds and visualizes connected sub-networks where most components are expressed in most cases. It finds all maximal connected sub-networks where all nodes but \( k \) exceptions are expressed in all experimental studies but at least \( l \) exceptions. We demonstrate the power of the new approach by comparing it to similar approaches with gene expression data previously used to study Huntington’s disease. In addition, we demonstrate KeyPathwayMiner’s flexibility and applicability to non-array data by analyzing genome-scale DNA methylation profiles from colorectal tumor cancer patients. KeyPathwayMiner release 2 is available as a Cytoscape plugin and online at http://keypathwayminer.mpib-mpg.de.

Understanding structural and algorithmic properties of complex networks is an important task, not least because of the huge impact of the internet. Our focus is to analyze how news spreads in social networks. We simulate a simple information spreading process in different network topologies and demonstrate that news spreads much faster in existing social network topologies. We support this finding by analyzing information spreading in the mathematically defined preferential attachment network topology, which is a common model for real-world networks. We prove that here a sublogarithmic time suffices to spread a news to all nodes of the network. All previously studied network topologies need at least a logarithmic time. Surprisingly, we observe that nodes with few neighbors are crucial for the fast dissemination. Social networks like Facebook and Twitter are reshaping the way people take collective actions. They have played a crucial role in the recent uprisings of the ‘Arab Spring’ and the ‘London riots’. It has been argued that the ‘instantaneous nature’ of these networks influenced the speed at which the events were handled. It is quite remarkable that social networks spread news so fast. Both the structure of social networks and the process that distributes the news are not designed with this purpose in mind. On the contrary, they are not designed at all, but have evolved in a random and decentralized manner.

So is our view correct that social networks ease the spread of information ("rumors"). And if so, what particular properties of social networks are the reason for this? To answer these questions, we simulate a simple rumor spreading process on several graphs having the structure of existing large social networks. We see, for example, that a rumor started at a random node of the Twitter network in average reaches 45.6 million of the total of 51.2 million members within only eight rounds of communication. We also analyze this process on an abstract model of social networks, the so-called preferential attachment graphs introduced by Barabási and Albert [5]. In [17], we obtain a mathematical proof that rumors in such networks spread much faster than in many other network topologies, which are not as random. They are built faster than networks having a communication link between any two nodes (complete graphs). As an explanation, we observe that nodes of small degree build a short-cut between those having large degree (hubs), which due to their large number of possible communication partners less often talk to each other directly.


The hypervolume indicator is an increasingly popular set measure to compare the quality of two Pareto sets. The basic ingredient of most hypervolume indicator based optimization algorithms is the calculation of the hypervolume contribution of single solutions regarding a Pareto set. We show in exact calculation of the hypervolume contribution is NP-hard. The same holds for the calculation of the minimal contribution. We also prove that it is NP-hard to decide whether a solution has the least hypervolume contribution. Even deciding whether the contribution of a solution is at most $(1 + \epsilon)$ times the minimal contribution is NP-hard. This implies that it is neither possible to efficiently find the least contributing solution (unless $P = NP$) nor to approximate it (unless $NP = BPP$). Nevertheless, in the second part of the paper we present a very fast approximation algorithm for this problem. We prove that for arbitrarily given $\delta > 0$ it calculates a solution with contribution at most $(1 + \epsilon)$ times the minimal contribution with probability at least $(1 - \delta)$. Though it cannot run in polynomial time for all instances, it performs extremely fast on various benchmark datasets. The algorithm solves very large problem instances which are intractable for exact algorithms (e.g., 10000 solutions in 100 dimensions) within a few seconds.


We propose a simple distributed algorithm for balancing indivisible tokens on graphs. The algorithm is completely deterministic, theoretically to imitate (and excellently) a randomized algorithm while keeping the error as small as possible. Our new algorithm, surprisingly, closely approximates the idealized process (where the tokens are divisible) on important network topologies. On $d$-dimensional torus graphs with $n$ nodes it deviates from the idealized process only by an additive constant. In contrast, the randomized algorithm of Bringmann and Friedrich and Sauerwald [Proceedings of the 41st Annual ACM Symposium on Theory of Computing, 2009, pp. 121-130] can deviate up to $\Omega(\text{polylog}(n))$, and the deterministic algorithm of Rabani, Sinclair, and Wanka [Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science, 1998, pp. 694-705] has a deviation of $\Omega(n^{1/4})$. This makes our quasirandom algorithm the first known algorithm for this setting, which is optimal both in time and achieved smoothness. We further show that on the hypercube as well, our algorithm has a smaller deviation from the idealized process than the previous algorithms. To prove these results, we derive several combinatorial and probabilistic results that we believe to be of independent interest. In particular, we show that first-passage probabilities of a random walk on a path with arbitrary weights can be expressed as a convolution of independent geometric probability distributions.


Multi-objective optimization deals with the task of computing a set of solutions that represents possible trade-offs with respect to a given set of objective functions. Set-based approaches such as evolutionary algorithms are very popular for solving multi-objective optimization problems. Convergence of set-based approaches for multi-objective optimization is essential for their success. We take an order-theoretic view on the convergence of set-based multi-objective optimization and examine how the use of indicator functions can help to direct the search towards Pareto optimal sets. In doing so, we point out that set-based multi-objective optimization working on the dominance relation of search points has to deal with a cyclic behavior that may lead to worsening with respect to the Pareto-dominance relation defined on sets. Later on, we show in which situations well-known binary and unary indicators can help to avoid this cyclic behavior and therefore guarantee convergence of the algorithm. We also study the impact of deteriorative cycles on the runtime behavior and give an example in which they provably slow down the optimization process.


We empirically analyze two versions of the well-known "randomized rumor spreading" protocol to disseminate a piece of information in networks. In the classical model, in each round each informed node informs a random neighbor. At SODA 2008, three of the authors proposed a quasirandom variant. Here, each node has a (cyclic) list of its neighbors. Once informed, it starts at a random position of the list, but from then on informs its neighbors in the order of the list. While for sparse random graphs a better performance of the quasirandom model could be proven, all other results show that, independent of the structure of the lists, the same asymptotic performance guarantees hold as for the classical model. In this work, we compare the two models experimentally. This not only shows that the quasirandom model generally is faster (which was expected, though maybe not this extent), but also that the runtime is more concentrated around the mean value (which is surprising given that much fewer random bits are used in the quasirandom process). These advantages are also observed in a lossy communication model, where each transmission does not reach its target with a certain probability, and in an asynchronous model, where nodes send at random times drawn from an exponential distribution. We also show that the particular structure of the lists has little influence on the efficiency. In particular, there is no problem if all nodes use an identical order to inform their neighbors.
Friedrich, T., Horoba, C., Neumann, F., Illustration of Fairness in Evolutionary Multi-Objective Optimization. In: Theoretical Computer Science 412, pp. 1546–1556, 2011. It is widely assumed that evolutionary algorithms for multi-objective optimization problems should use certain mechanisms to achieve a good spread over the Pareto front. In this paper, we examine such mechanisms from a theoretical point of view and analyze simple algorithms incorporating the concept of fairness. This mechanism tries to balance the number of offspring of all individuals in the current population. We rigorously analyze the runtime behavior of different fairness mechanisms and present showcase examples to point out situations, where the right mechanism can speed up the optimization process significantly. We also indicate drawbacks for the use of fairness by presenting instances, where the optimization process is slowed down drastically.

Friedrich, T., Sauerwald, T., Vilenchik, D., Smoothed analysis of balancing networks. In: Random Structures and Algorithms 39, pp. 115–138, 2011. In a balancing network each processor has an initial collection of unit-size jobs (tokens) and in each round, pairs of processors connected by balancers split their load as evenly as possible. An excess token (if any) is placed according to some predefined rule. As it turns out, this rule crucially affects the performance of the network. In this work we propose a model that studies this effect. We suggest a model bridging the uniformly-random assignment rule, and the arbitrary one (in the spirit of smoothed-analysis). We start with an arbitrary assignment of balancer directions and then flip each assignment with probability α independently. For a large class of balancing networks our result implies that after O(log n) rounds the discrepancy is O((1/2 − α) log n + log log n) with high probability. This matches and generalizes known upper bounds for α = 0 and α = 1/2. We also show that a natural network matches the upper bound for any α.

Friedrich, T., Hebbinghaus, N., Average update times for fully-dynamic all-pairs shortest paths. In: Discrete Applied Mathematics 159, pp. 1751–1758, 2011. We study the fully-dynamic all pairs shortest path problem for graphs with arbitrary non-negative edge weights. It is known for digraphs that an update of the distance matrix costs $O(n^2\log^2 n)$ worst-case time. In this paper we present an FPRAS for the fully-dynamic all-pairs shortest paths problem. We also show that a natural network matches the upper bound for any α.

Ajwani, D., Friedrich, T., Average-case analysis of incremental topological ordering. In: Discrete Applied Mathematics 158, pp. 240–250, 2010. Many applications like pointer analysis and incremental compilation require maintaining a topological ordering of the nodes of a directed acyclic graph (DAG) under dynamic updates. All known algorithms for this problem are either only analyzed for worst-case insertion sequences or only evaluated experimentally on random DAGs. We present the first average-case analysis of incremental topological ordering algorithms. We prove an expected runtime of $O(n^2\text{polylog}(n))$ under insertion of the edges of a complete DAG in a random order for the algorithms of Alpern et al. (1990) [4], Katriel and Bodlaender (2006) [18], and Pearce and Kelly (2006) [23].

Friedrich, T., Hebbinghaus, N., Neumann, F., Plateaus can be harder in Multi-Objective Optimization. In: Theoretical Computer Science 411, pp. 854–864, 2010. In recent years a lot of progress has been made in understanding the behavior of evolutionary computation methods for single- and multi-objective problems. Our aim is to analyze the diversity mechanisms that are implicitly used in evolutionary algorithms for multi-objective problems by rigorous runtime analyses. We show that, even if the population size is small, the runtime can be exponential where corresponding single-objective problems are optimized within polynomial time. To illustrate this behavior we analyze a simple plateau function in a first step and extend our result to a class of instances of the well-known SetCover problem.

Bringmann, K., Friedrich, T., Approximating the volume of unions and intersections of high-dimensional geometric objects. In: Computational Geometry 43, pp. 601–610, 2010. We consider the computation of the volume of the union of the volume of the union of high-dimensional geometric objects. While showing that this problem is #P-hard already for very simple bodies (i.e., axis-parallel boxes), we give a fast FPRAS for all objects where one can: (1) test whether a given point lies inside the object, (2) sample a point uniformly, (3) calculate the volume of the object in polynomial time. All three oracles can be weak, that is, just approximate. This implies that Klee’s measure problem and the hypervolume indicator can be approximated efficiently even though they are #P-hard and hence cannot be solved exactly in time polynomial in the number of dimensions unless P = NP. Our algorithm also allows to approximate efficiently the volume of the union of convex bodies given by weak membership oracles. For the analogous problem of the intersection of high-dimensional geometric objects we prove #P-hardness for boxes and show that there is no multiplicative-polynomial-time $2^{\epsilon n} \text{-approximation}$ for certain boxes unless NP = BPP, but give a simple additive-polynomial-time $\epsilon$-approximation.

Friedrich, T., Sauerwald, T., The Cover Time of Deterministic Random Walks. In: Electronic Journal of Combinatorics 17, 2010. The rotor router model is a popular deterministic analogue of a random walk on a graph. Instead of moving to a random neighbor, the neighbors are served in a fixed order. We examine how fast this "deterministic random walk” covers all vertices (or all edges). We present general techniques to derive upper bounds for the vertex and edge cover time and derive matching lower bounds for several important graph classes. Depending on the topology, the deterministic random walk can be asymptotically faster, slower or equally fast compared to the classical random walk.

Jim Propp’s rotor router model is a deterministic analogue of a random walk on a graph. Instead of distributing chips randomly, each vertex serves its neighbors in a fixed order. Cooper and Spencer [Comb. Probab. Comput. (2006)] show a remarkable similarity of both models. If an (almost) arbitrary population of chips is placed on the vertices of a grid $Z^d$ and does a simultaneous walk in the Propp model, then at all times, the number of chips deviates from the expected number the random walk would have gotten there, by at most a constant. This constant is independent of the starting configuration and the order in which each vertex serves its neighbors. This result raises the question if all graphs do have this property. With quite some effort, we are now able to answer this question negatively. For the graph being an infinite $k$-ary tree ($k \geq 3$), we show that for any deviation $D$ there is an initial configuration of chips such that after running the Propp model for a certain time there is a vertex with at least $D$ more chips than expected in the random walk model. However, to achieve a deviation of $D$ it is necessary that at least $4\phi(D^2)$ vertices contribute by being occupied by a number of chips not divisible by $k$ in a certain time interval.


The hypervolume indicator serves as a sorting criterion in many recent multi-objective evolutionary algorithms (MOEAs). Typical algorithms remove the solution with the smallest loss with respect to the dominated hypervolume from the population. We present a new algorithm which determines for a population of size $n$ with $d$ objectives, a solution with minimal hypervolume contribution in time $n^{d/2}$ for $d = 2$. This improves all previously known algorithms by a factor of $\sqrt{n}$ for $d = 3$. We also analyze hypervolume indicator based optimization algorithms which remove $\lambda > 1$ solutions from a population of size $n = \mu + \lambda$. We show that there are populations such that the hypervolume contribution of iteratively chosen $\lambda$ solutions is much larger than the hypervolume contribution of an optimal set of $\lambda$ solutions. Selecting the optimal set of $\lambda$ solutions implies calculating $\binom{n}{\lambda}$ conventional hypervolume contributions, which is considered to be computationally too expensive. We present the first hypervolume algorithm which directly calculates the contribution of every set of $\lambda$ solutions. This gives an additive term of $\binom{n}{\lambda}$ in the runtime of the calculation instead of a multiplicative factor of $\binom{n}{\lambda}$. More precisely, for a population of size $n$ with $d$ objectives, our algorithm can calculate a set of $\lambda$ solutions with minimal hypervolume contribution in time $O(n^{d/2} \log n + n^d)$ for $d > 2$. This improves all previously published algorithms by a factor of $n^{d/2} / \lambda^d$ for $d > 3$ and by a factor of $n$ for $d = 3$.


The main aim of randomized search heuristics is to produce good approximations of optimal solutions within a small amount of time. In contrast to numerous experimental results, there are only a few theoretical explorations on this subject. We consider the approximation ability of randomized search heuristics for the class of covering problems and compare single-objective and multi-objective models for such problems. For the VertexCover problem, we point out situations where the multi-objective model leads to a fast construction of optimal solutions while in the single-objective case, no good approximation can be achieved within the expected polynomial time. Examining the more general SetCover problem, we show that optimal solutions can be approximated within a logarithmic factor of the size of the ground set, using the multi-objective approach, while the approximation quality obtainable by the single-objective approach in expected polynomial time may be arbitrarily bad.


Deterministic and randomized balancing schemes are used to distribute workload evenly in networks. In this paper, we compare two very general ones: the random walk and the (deterministic) Propp machine. Roughly speaking, we show that on the two-dimensional grid, the Propp machine always has the same number of tokens on a node as does the random walk in expectation, apart from an additive error of less than eight. This constant is independent of the total number of tokens and the runtime of the two processes. However, we also show that it makes a difference whether the Propp machine serves the neighbors in a circular or non-circular order.


In this paper, we examine how adding objectives to a given optimization problem affects the computational effort required to generate the set of Pareto-optimal solutions. Experimental studies show that additional objectives may change the running time behavior of an algorithm drastically. Often it is assumed that more objectives make a problem harder as the number of different tradeoffs may increase with the problem dimension. We show that additional objectives, however, may be both beneficial and obstructive depending on the chosen objective. Our results are obtained by rigorous running time analyses that show the different effects of adding objectives to a well-known plateau function. Additional experiments show that the theoretically shown behavior can be observed for problems with more than one objective.


Hybrid methods are very popular for solving problems from combinatorial optimization. In contrast, the theoretical understanding of the interplay of different optimization methods is rare. In this paper, we make a first step towards the analysis of such combinations for combinatorial optimization problems. The subject of our analyses is the vertex cover problem for which several approximation algorithms have been proposed. We point out specific instances where solutions can (or cannot) be improved by the search process of a simple evolutionary algorithm in expected polynomial time.


It is widely assumed and observed in experiments that the use of diversity mechanisms in evolutionary algorithms may have a great impact on its running time. Up to now there is no rigorous analysis pointing out how different diversity mechanisms influence the runtime behavior. We consider evolutionary algorithms that differ from each other in the way they ensure diversity and point out situations where the right mechanism is crucial for the success of the algorithm. The considered evolutionary algorithms either diversify the population with respect to the search points or with respect to function values. Investigating simple plateau functions, we show that using the “right” diversity strategy makes the difference between an exponential and a polynomial runtime. Later on, we examine how the drawback of the “wrong” diversity mechanism can be compensated by increasing the population size.
Conference papers


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


We study the problem of multicommodity flow and multicut in treewidth-2 graphs and prove bounds on the multicommodity-multicut gap. In particular, we give a primal-dual algorithm for computing multicommodity flow and multicut in treewidth-2 graphs and prove the following approximate max-flow min-cut theorem: given a treewidth-2 graph, there exists a multicommodity flow of value $f$ with congestion 4, and a multicut of capacity $c$ such that $c \leq 20f$. This implies a multicommodity-multicut gap of 80 and improves upon the previous best known bounds for such graphs. Our algorithm runs in polynomial time when all the edges have capacity one. Our algorithm is completely combinatorial and builds upon the primal-dual algorithm of Garg, Vazirani and Yannakakis for multicut in trees and the augmenting paths framework of Ford and Fulkerson.


In the house allocation problem with lower and upper quotas, we are given a set of applicants and a set of projects. Each applicant has a strictly ordered preference list over the projects, while the projects are equipped with a lower and an upper quota. A feasible matching assigns the applicants to the projects in such a way that a project is either matched to no applicant or to a number of applicants between its lower and upper quota. A feasible matching is Pareto optimal if there is no other feasible matching that makes some applicant happier and leaves all others at least as happy. Furthermore, a feasible matching is popular if there is no other feasible matching that makes more applicants happy and leaves all others at least as happy. We study the complexity of finding such a matching.


Combinatorial optimization problems are a prominent application area of evolutionary algorithms, where the $(1+1)$ EA is one of the most investigated. We extend this algorithm by introducing some problem knowledge with a specialized mutation operator which works under the assumption that the number of Is of a solution is critical, as frequently happens in combinatorial optimization. This slight modification increases the chance to correct wrongly placed bits while preserving the simplicity and problem independence of the $(1+1)$ EA. As an application of our algorithm we examine the vertex cover problem on certain instances, where we show that it leads to theoretically better runtime behaviour on directed acyclic graphs (DAG) with $n$ nodes, under an online edge insertion sequence, in $O(n^{2.75})$ time, independent of the number $m$ of edges inserted. For dense DAGs, this is an improvement over the previous best result of $O(\min(m^{5/2} \log n, n^{5/2} + n^2 \log n))$ by Katriel and Bodlaender 2006. We also provide an empirical comparison of our algorithm with other algorithms for incremental topological sorting.
reduces this to $\Theta(n^3)$, and present experimental evidence that such a configuration is reached. Concerning the complete bipartite graph our modification finds the optimum in polynomial time with probability $1 - 1/2^{\Omega(n^5)}$ for every positive constant $\xi \leq 1$, which improves the known probability of $1 - 1/poly(n)$ for the $(1+1)$ EA.


Human lives are increasingly influenced by algorithms, which therefore need to meet higher standards not only in accuracy but also with respect to explainability. This is especially true for high-stakes areas such as real estate valuation. Unfortunately, the methods applied there often exhibit a trade-off between accuracy and explainability. One explainable approach is case-based reasoning (CBR), where each decision is supported by specific previous cases. However, such methods can be wanting in accuracy. The unexplainable machine learning approaches are often observed to provide higher accuracy but are not tractable in their decision-making. In this paper, we apply evolutionary algorithms (EAs) to CBR predictors in order to improve their performance. In particular, we deploy EAs to the similarity functions (used in CBR to find comparable cases), which are fitted to the data set at hand. As a consequence, we achieve higher accuracy than state-of-the-art deep neural networks (DNNs), while keeping interpretability and explainability. These results stem from our empirical evaluation on a large data set of real estate offers where we compare known similarity functions, their EA-improved counterparts, and DNNs. Surprisingly, DNNs are only on par with standard CBR techniques. However, using EA-learned similarity functions does yield an improved performance.


In order to understand better how and why crossover can benefit optimization, we consider pseudo-Boolean functions with an upper bound $B$ on the number of 1s allowed in the bit string (cardinality constraint). We consider the natural translation of the OneMax test function, a linear function where $f(k)$, the best-known FPT algorithm on $G$ with parameter $k$ preprocessed $G$ in time $O(f(k)\cdot poly(n))$. When queried with a set $F$ of at most $f$ edges of $G$, the oracle reports the answer to the $B$-with-the-same parameter $k$-on the graph $G - F$, i.e., $G$ deprived of $F$. The oracle should answer queries in a time that is significantly faster than merely running the best-known FPT algorithm on $G - F$ from scratch. We design sensitivity oracles for the $k$-Path and the $k$-Vertex Cover problem. Our first oracle for $k$-Path has size $O(k^{(f+1)}$ size and query time $O(\min\{f, log(f) + k\})$. We use a technique inspired by the work of Weimann and Yuster [FOCS 2010, TALG 2013] on distance sensitivity problems to reduce the space to $O(\log n)$, where $f$ is a parameter that gives $O(n \cdot \log n)$ space and has $O(f \cdot \log f)$ query time, the second one has a size of $O(n)$ and query time $O\left(f, k \cdot \log k \right)$. Both oracles can be modified to handle vertex-failures, but we need to replace $k$ with $2k$ in all the claimed bounds. Regarding $k$-Vertex Cover, we design three oracles offering different trade-offs between the size and the query time. The first oracle takes $O(2^{f} + k)$ space and has $O(2^{f} + k)$ query time, the second one has a size of $O(2^{f+k} + k)$ and a query time of $O(f + k^{2})$, finally, the third one takes $O(f + k^{2})$ space and can be queried in time $O(2^{2f+2} + k^{2})$.


We combine ideas from distance sensitivity oracles (DSOs) and fixed-parameter tractability (FPT) to design sensitivity oracles for FPT graph problems. An oracle with sensitivity $f$ for an FPT problem $\Pi$ on a graph $G$ with parameter $k$ preprocessed $G$ in time $O(f(k)\cdot poly(n))$. When queried with a set $F$ of at most $f$ edges of $G$, the oracle reports the answer to the $\Pi$-with-the-same parameter $k$-on the graph $G - F$, i.e., $G$ deprived of $F$. The oracle should answer queries in a time that is significantly faster than merely running the best-known FPT algorithm on $G - F$ from scratch. We design sensitivity oracles for the $\Pi$-Path and the $\Pi$-Vertex Cover problem. Our first oracle for $\Pi$-Path has size $O(k^{(f+1)}$ size and query time $O(\min\{f, log(f) + k\})$. We use a technique inspired by the work of Weimann and Yuster [FOCS 2010, TALG 2013] on distance sensitivity problems to reduce the space to $O(\log n)$, where $f$ is a parameter that gives $O(n \cdot \log n)$ space and has $O(f \cdot \log f)$ query time, the second one has a size of $O(n)$ and query time $O\left(f, k \cdot \log k \right)$. Both oracles can be modified to handle vertex-failures, but we need to replace $k$ with $2k$ in all the claimed bounds. Regarding $\Pi$-Vertex Cover, we design three oracles offering different trade-offs between the size and the query time. The first oracle takes $O(2^{f} + k)$ space and has $O(2^{f} + k)$ query time, the second one has a size of $O(2^{f+k} + k)$ and a query time of $O(f + k^{2})$, finally, the third one takes $O(f + k^{2})$ space and can be queried in time $O(2^{2f+2} + k^{2})$.


During a pandemic people have to find a trade-off between meeting others and staying safely at home. While meeting others is socially much worse stable states.

We study the algorithmic applications of a natural discretization for the hard-sphere model and the Widom–Rowlinson model in a region of $d$-dimensional Euclidean space $V \subset \mathbb{R}^d$. These continuous models are frequently used in statistical physics to describe mixtures of one or multiple particle types subjected to hard-core interactions. For each type, particles are distributed according to a Poisson point process with a type-specific activity parameter, called fugacity. The Gibbs distribution over all possible system states is characterized by the mixture of these point processes conditioned that no two particles are closer than some type-dependent distance threshold. A key part in better understanding the Gibbs distribution is its normalizing constant, called partition function. Our main algorithmic result is the deterministic approximative algorithm for the partition function of the hard-sphere model and the Widom–Rowlinson model in box-shaped regions of Euclidean space. Our algorithms have quasi-polynomial running time in the volume of the region $v(V)$ if the fugacity is below a certain threshold. For the $d$-dimensional hard-sphere model with particles of unit volume, this threshold is $\epsilon/2^d$. As the number of dimensions $d$ increases, this bound asymptotically matches the best known results for randomized approximation of the hard-sphere partition function. We prove similar bounds for the Widom–Rowlinson model.


Combinatorial optimization lies at the core of many real-world problems. Especially since the rise of graph neural networks (GNNs), the deep learning community has been developing solvers that derive solutions to NP-hard problems by learning the problem-specific solution structure. However, reproducing the results of these publications proves to be difficult. We make three contributions. First, we present an open-source benchmark suite for the NP-hard Maximum Independent Set problem, in both its weighted and unweighted variants. The suite offers a unified interface to various state-of-the-art traditional and machine learning-based solvers. Second, using our benchmark suite, we conduct an in-depth analysis of the popular guided tree search algorithm by Li et al. [NeurIPS 2018], testing various configurations on small and large synthetic and real-world graphs. By re-implementing their algorithm with a focus on code quality and extensibility, we show that the graph convolution network used in the tree search does not learn a meaningful representation of the solution structure, and can in fact be replaced by random values. Instead, the tree search relies on algorithmic techniques like graph kernelization to find good solutions. Thus, the results from the original publication are not reproducible. Third, we extend the analysis to compare the tree search implementations to other solvers, showing that the classical algorithmic solvers are often faster, while providing solutions of similar quality. Additionally, we analyze a recent solver based on reinforcement learning and observe that for this solver, the GNN is responsible for the competitive solution quality.


Estimation of distribution algorithms (EDAs) provide a distribution-based approach for optimization which adapts its probability distribution during the run of the algorithm. We contribute to the theoretical understanding of EDAs and point out that their distribution approach makes them more suitable to deal with rugged fitness landscapes than classical local search algorithms. Concretely, we make the OneMax function rugged by adding noise to each fitness value. The cGA can nevertheless find solutions with $n(1 - \epsilon)$ many 1s, even for high variance of noise. In contrast to this, RLS and the $(1+1)$ EA, with high probability, only find solutions with $n(1/2 + o(1))$ many 1s, even for noise with small variance.


Local search is the most basic strategy in optimization settings when no specific problem knowledge is employed. While this strategy finds good solutions for certain optimization problems, it generally suffers from getting stuck in local optima. This stagnation can be avoided if local search is modified. Depending on the optimization landscape, different modifications vary in their success. We discuss several features of optimization landscapes and give analyses as examples for how they affect the performance of modifications of local search. We consider modifying random local search by restarting it and by considering larger search radii. The landscape features we analyze include the number of local optima, the distance between different optima, as well as the local landscape around a local optimum. For each feature, we show which modifications of local search handle them well and which do not.


Bootstrap percolation is a classical model for the spread of information in a network. In the round-based version, nodes of an undirected graph become active once at least $r$ neighbors were active in the previous round. We propose the perturbed percolation process: a superposition of two percolation processes on the same node set. One process acts on a local graph with activation
threshold 1, the other acts on a global graph with threshold $r$ – representing local and global edges, respectively. We consider grid-like local graphs and expanders as global graphs on $n$ nodes. For the extreme case $r = 1$, all nodes are active after $O(\log n)$ rounds, while the process spreads only polynomially fast for the other extreme case $r \geq n$. For a range of suitable values of $r$, we prove that the process exhibits both phases of the above extremes. It starts with a polynomial growth and eventually transitions from at most $cn$ to $n$ active nodes, for some constant $c \in (0, 1)$, in $O(\log n)$ rounds. We observe this behavior also empirically, considering additional global-graph models.


Recognizing disturbed text in real-life images is a difficult problem, as information that is missing due to low resolution or out-of-focus text has to be recreated. Combining text super-resolution and optical character recognition deep learning models can be a valuable tool to enhance and enhance text images for better readability, as well as recognize text automatically afterwards. We achieve improved peak-signal-to-noise ratio test text recognition accuracy scores over a state-of-the-art model. Our algorithm is based on the real-world low-resolution dataset TextZoom while having a smaller theoretical model size due to the usage of quantization techniques. In addition, we show how different training strategies influence the performance of the resulting model.


Several large-scale machine learning tasks, such as data summarization, can be approached by maximizing functions that satisfy submodularity. These optimization problems often involve complex side constraints, imposed by the underlying application. In this paper, we develop an algorithm with poly-logarithmic adaptivity for non-monotone submodular maximization under general side constraints. The adaptive complexity of a problem is the minimal rounds of the algorithm needed to achieve a $(1 - \frac{1}{e})$-approximation. Our algorithm is suitable to maximize a non-monotone submodular function under a $p$-system side constraint, and it achieves a $(p + O(\sqrt{p}))$-approximation for this problem, after only poly-logarithmic adaptive rounds and polynomial queries to the valuation oracle function. Furthermore, our algorithm achieves a $p + O(1)$-approximation when the given side constraint is a $p$-extendible system. This algorithm yields an exponential speed-up, with respect to the adaptivity, over any other known constant-factor approximation algorithm for this problem. It also competes with previous known results in terms of approximation complexity. We perform various experiments on various real-world applications. We find that, in comparison with commonly used heuristics, our algorithm performs better on these instances.


The active global SARS-CoV-2 pandemic caused more than 167 million cases and 3.4 million deaths worldwide. The development of completely new drugs for such a novel disease is a challenging, time intensive process and despite researchers around the world working on this task, no effective treatments have been developed yet. This emphasizes the importance of drug repurposing, where treatments are found among existing drugs that are meant for different diseases. A common approach to this is based on knowledge graphs, that condense relationships between entities like drugs, diseases and genes. Graph neural networks (GNNs) can then be used for the task at hand by predicting links in such knowledge graphs. Expanding on state-of-the-art GNN research, Doshi et al. recently developed the Dr-COVID model. We further extend their work using additional output interpretation strategies. The best aggregation strategy derives a top-100 ranking of candidate drugs, 32 of which currently being in COVID-19-related clinical trials. Moreover, we present an alternative application for the model, the generation of additional candidates based on a given pre-selection of drug candidates using collaborative filtering. In addition, we improved the implementation of the Dr-COVID model by significantly shortening the inference and pre-processing time by exploiting data-parallelism. As drug repurposing is a task that requires high computation and memory resources, we further accelerate the post-processing phase using a new emerging hardware — we propose a new approach to leverage the use of high-capacity Non-Volatile Memory for aggregate drug ranking.


Understanding real-world networks has been a core research endeavor throughout the last two decades. Network Creation Games are a promising approach for this from a game-theoretic perspective. In these games, selfish agents corresponding to nodes in a network strategically decide which links to form to optimize their centrality. Many versions have been introduced and analyzed, but none of them fits to modeling the evolution of social networks. In real-world social networks, connections are often established by recommendations from common acquaintances or by a chain of such recommendations. Thus establishing and maintaining a contact is proportional to the distance of the endpoints before establishing the connection. We provide results for generic cost functions, which essentially only must be convex functions in the distance of the endpoints without the respective edge. For this formal class of cost functions, we provide many structural properties of equilibrium networks and prove (almost) tight bounds on the diameter, the Price of Anarchy and the Price of Stability. Moreover, as a proof-of-concept we show via experiments that the created equilibrium networks of our model indeed closely mimics real-world social networks. We observe degree distributions that seem to follow a power-law, high clustering, and low diameters. This can be seen as a promising first step towards game-theoretic network creation models that predict networks featuring all core real-world properties.


The fine-grained localization and classification of various lung abnormalities is a challenging yet important task for combating diseases and, also, pandemics. In this paper, we present one way to detect and classify abnormalities within chest X-ray scans. In particular, we investigate the use of binary image classification (to distinguish between healthy and infected chests) and the weighted box fusion (which constructs a detection box using the proposed boxes within range). We observe that both methods increase the performance of a base model significantly. Furthermore, we improve state of the art on lung segmentation, even in the presence of abnormalities.
example, block masking and diffused noise augmentations). This way, we obtain a state-of-the-art model with a dice score of 97.9%. In particular, we show that simple augmentations outperform complex ones in our setting.


We study the maximum-flow/minimum-cut problem on scale-free networks, i.e., graphs whose degree distribution follows a power-law. We propose a simple algorithm that capitalizes on the fact that often only a small fraction of such a network is relevant for the flow. At its core, our algorithm augments Dinitz’s algorithm with a balanced bidirectional search. Our experiments on a scale-free random network model indicate sublinear run time. On scale-free real-world networks, we outperform the commonly used highest-label Push-Relabel implementation by up to two orders of magnitude. Compared to Dinitz’s original algorithm, our modifications reduce the search space, e.g., by a factor of 275 on an autonomous systems graph. Beyond these good run times, our algorithm has an additional advantage compared to Push-Relabel. The latter computes a preflow, which makes the extraction of a minimum cut potentially more difficult. This is relevant, for example, for the computation of Gomory-Hu trees. On a social network with 7000 nodes, our algorithm computes the Gomory-Hu tree in 3 seconds compared to 12 hours when using Push-Relabel.


We introduce the balanced crown decomposition that captures the structure imposed on graphs by their connected induced subgraphs of a given size. Such subgraphs are a popular modeling tool in various application areas, where the non-local nature of the connectivity condition usually results in very challenging algorithmic tasks. The balanced crown decomposition is a combination of a crown decomposition and a balanced partition which makes it applicable to graph editing as well as graph packing and partitioning problems. We illustrate this by deriving improved approximation algorithms and kernelization for a variety of such problems. In particular, through this structure, we obtain the first constant-factor approximation for the Balanced Connected Partition (BCP) problem, where the task is to partition a vertex-weighted graph into $k$ connected components of approximately equal weight. We derive a $3$-approximation for the two most commonly used objectives of maximizing the weight of the lightest component or minimizing the weight of the heaviest component.


Given a graph with a distinguished source vertex $s$, the Single Source Replacement Paths (SSRP) problem is to compute and output, for any target vertex $t$ and edge $e$, the length $d(s, t, e)$ of a shortest path from $s$ to $t$ that avoids a failing edge $e$. A Single-Source Distance Sensitivity Oracle (Single-Source DSO) is a compact data structure that answers queries of the form $(t, e)$ by returning the distance $d(s, t, e)$. We show how to compute the output of the SSRP problem on $n$-vertex, $m$-edge graphs with integer edge weights in the range $[1, M]$ into a deterministic Single-Source DSO that has size $O(M^{1/3}n^{2/3})$ and query time $O(1)$. We prove that the space requirement is optimal (up to the word size). Our techniques can also handle vertex failures within the same bounds. Chechik and Cohen [SODA 2019] presented a combinatorial randomized $\tilde{O}(mn^{1/2} + n^2)$ time SSRP algorithm for undirected and unweighted graphs. We derandomize their algorithm with the same asymptotic running time and apply our compression to obtain a deterministic Single-Source DSO with $O(mn^{1/2} \log n + n^2)$ preprocessing time, $O(\sqrt{n} \log n)$ space, and $O(1)$ query time. Our combinatorial Single-Source DSO has near-optimal space, preprocessing and query time for dense unweighted graphs, improving the preprocessing time by a $1/\sqrt{n}$-factor compared to previous results. Grandoni and Vassilevska Williams [FOCS 2012, TALG 2020] gave an algebraic randomized $\tilde{O}(Mn\sqrt{n} + n^2)$ time SSRP algorithm for (undirected and directed) graphs with integer edge weights in the range $[1, M]$, where $\omega \leq 2.373$ is the matrix multiplication exponent. We derandomize their algorithm for undirected graphs and apply our compression to obtain an algebraic Single-Source DSO with $O(Mn^2)$ preprocessing time, $O(M^{1/3}n^{2/3})$ space, and $O(1)$ query time. This improves the preprocessing time of algebraic Single-Source DSOs by polynomial factors compared to previous results. We also present further improvements of our Single-Source DSOs. We show that the query time can be reduced to a constant at the cost of increasing the size of the oracle to $O(M^{1/3}n^{2/3})$ and that all our oracles can be made path-reporting. On sparse graphs with $m = O(\sqrt{n}/\sqrt{k})$ edges, for any constant $k > 0$, we reduce the preprocessing to randomized $\tilde{O}(M^{7/8}m^{1/2}n^{1/8}) = O(n^{9/8-k/2})$ time. To the best of our knowledge, this is the first truly subquadratic time algorithm for building Single-Source DSOs on sparse graphs.


We generalize the tree doubling and Christofides algorithm to parameterized approximations for ATSP. The parameters we consider for the respective generalizations are upper bounded by the number of asymmetric distances, which yields algorithms to efficiently compute good approximations also for moderately asymmetric TSP instances. As generalization of the Christofides algorithm, we derive a parameterized $2$-approximation, where the parameter is the size of a vertex cover for the subgraph induced by the asymmetric distances. Our generalization of the tree doubling algorithm gives a parameterized $3$-approximation, where the parameter is the minimum number of asymmetric distances in a minimum spanning arborescence. Further, we combine these with a notion of symmetry relaxation which allows to trade approximation guarantee for runtime. Since the two parameters we consider are theoretically incomparable, we present experimental results which show that generalized tree doubling frequently outperforms parameterized Christofides with respect to parameter size.


Traffic congestion is a major issue that can be solved by suggesting drivers alternative routes they are willing to take. This concept has been formalized as a strategic routing problem in which a single alternative route is suggested to an existing one. We extend this formalization and introduce the Multiple-Routes problem, which is given a start and a destination and then aims at finding up to $n$ different routes that the drivers strategically disperse over, minimizing the overall travel time of the system. Due to the NP-hard nature of the problem, we introduce the Multiple-Routes evolutionary algorithm (MREA) as a heuristic solver. We study several mutation and crossover operators and evaluate them on real-world data of the city of Berlin, Germany. We find that a combination of all operators yields the best result, improving the overall travel time by a factor between 1.8 and 3, in the median, compared to all drivers taking the fastest route. For the base case $n = 2$, we compare our MREA to the highly tailored optimal solver by Bläsius et al. [ATMOS 2020] and show that, in the median, our approach finds solutions of quality at least 99.69% of an optimal solution while only requiring 40% of the time.
The ongoing COVID-19 pandemic demands for a swift discovery of suitable treatments. The development of completely new compounds for such a novel disease is a challenging, time intensive process. This amplifies the relevance of drug repurposing, a technique where existing drugs are used to treat other diseases. A common bioinformatical approach to this is based on knowledge graphs, which compile relationships between drugs, diseases, genes and other biomedical entities. Then, graph neural networks (GNNs) are used for the drug repursuing task as they provide a good link prediction performance on such knowledge graphs. Building on state-of-the-art GNN research, Doshi & Chepuri (2020) construct the remarkable model DR-COVID. We re-implement their model and extend the approach to perform significantly better. We propose and evaluate several strategies for the aggregation of link predictions into drug recommendation rankings. With the help of clustering of similar target diseases we improve the model by a substantial margin, compiling a top-100 ranking of candidates including 32 currently being in COVID-19-related clinical trials. Regarding the re-implementation, we offer more flexibility in the selection of the graph neighborhood sizes fed into the model and reduce the training time significantly by making use of data parallelism.


The active global SARS-CoV-2 pandemic caused more than 167 million cases and 3.4 million deaths worldwide. The development of completely new drugs for such a novel disease is a challenging, time intensive process and despite researchers around the world working on this task, no effective treatments have been developed yet. This emphasizes the importance of drug repurposing, where treatments are found among existing drugs that are meant for different diseases. A common approach to this is based on knowledge graphs, that condense relationships between entities like drugs, diseases and genes. Graph neural networks (GNNs) can then be used for the task at hand by pre-training on graphs. Exploit links in such knowledge graphs. Kišig et al. recently developed the Dr-COVID model. We further extend their work using additional output interpretation strategies. The best aggregation strategy derives a top-100 ranking of candidate drugs, 32 of which currently being in COVID-19-related clinical trials. Moreover, we present an alternative application for the model, the generation of additional candidates based on a given pre-selection of drug candidates using collaborative filtering. In addition, we improved the implementation of the Dr-COVID model by significantly shortening the inference and pre-processing time by exploiting data-parallelism. As drug repurposing is a task that requires high computation and memory resources, we further accelerate the post-processing phase using a new emerging hardware—we propose a new approach to leverage the use of high-capacity Non-Volatile Memory for aggregate drug ranking.


We study the problem of counting the number of homomorphisms from an input graph $G$ to a fixed (quantum) graph $H$ in any finite field of prime order $\mathbb{F}_p$. The subproblem with graph $H$ was introduced by Faben and Jerrum [ToC’15] and its complexity is still uncharacterised despite active research, e.g., the very recent work of Focke, Goldberg, Roth, and Zivný [SODA’21]. Our contribution is threefold. First, we introduce the study of quantum graphs to the study of modular counting homomorphisms. We show that the complexity for a quantum graph $H$ collapses to the complexity criteria found at dimension 1: graphs. Second, in order to prove cases of intractability we establish a further reduction to the study of bipartite graphs. Lastly, we establish a dichotomy for all bipartite $(K_{3,3} \text{-free})$-graphs by a thorough structural study incorporating both local and global arguments. This result subsumes all results on bipartite graphs known for all prime moduli and extends them significantly. Even for the subproblem with $p = 2$ this establishes new results.


The hard-sphere model is one of the most extensively studied models in statistical physics. It describes the continuous distribution of spherical particles, governed by hard-core interactions. An important quantity of this model is the normalizing factor of the continuous model. This results in a discrete hard-core instance that is exponential in the size of the initial hard-sphere model. Our approximation bound follows directly from the correlation decay threshold of an infinite regular tree with degree equal to the maximum degree of our discretization. To cope with the exponential blow-up of the discrete instance we use clique dynamics, a Markov chain that was recently introduced in the setting of abstract polymer models. We prove rapid mixing of clique dynamics up to the tree threshold of the univariate hard-core model. This is achieved by adapting clique dynamics to block dynamics and adapting the spectral expansion method, which was recently used to bound the mixing time of Glauber dynamics within the same parameter regime.


Schelling’s classical segregation model gives a coherent explanation for the wide-spread phenomenon of residential segregation. We consider an agent-based saturated open-city variant, the Flip-Schelling-Process (FSP), in which agents, placed on a graph, have one out of two types and, based on the predominant type in their neighborhood, decide whether to change their types; similar to a new agent arriving as soon as another agent leaves the vertex. We investigate the probability that an edge $uv$ is monochrome, i.e., that both vertices $u$ and $v$ have the same type in the FSP, and we provide a general framework for analyzing the influence of the underlying graph topology on residential segregation. In particular, for two adjacent vertices, we show that a highly decisive common neighborhood, i.e., a common neighborhood where the absolute value of the difference between the number of vertices with different types is high, supports segregation and moreover, that large common neighborhoods are more decisive. As an application, we study the expected behavior of the FSP on two common random graph models behavior of the FSP: [1] For random geometric graphs, we show that the existence of an edge $uv$ makes a highly decisive common neighborhood for $u$ and $v$ more likely. Based on this, we prove the existence of a constant $c > 0$ such that the expected fraction of monochrome edges after the FSP is at least $1/2 + c$. This proves the model DR-COVID. We re-implement their model and extend the approach to perform significantly better. We propose and evaluate several strategies for the aggregation of link predictions into drug recommendation rankings. With the help of clustering of similar target diseases we improve the model by a substantial margin, compiling a top-100 ranking of candidates including 32 currently being in COVID-19-related clinical trials. Regarding the re-implementation, we offer more flexibility in the selection of the graph neighborhood sizes fed into the model and reduce the training time significantly by making use of data parallelism.
c. (2) For Erdős–Rényi graphs we show that large common neighborhoods are unlikely and that the expected fraction of monochrome edges after the FSP is at most $1/2 + o(1)$. Our results indicate that the cluster structure of the underlying graph has a significant impact on the obtained segregation strength.


Chromatic Correlation Clustering (CCC) models clustering of objects with categorical pairwise relationships. The model can be viewed as clustering the vertices of a graph with edge-labels (colors). Bonchi et al. [KDD 2015] introduced it as a natural generalization of the well studied problem Correlation Clustering (CC), motivated by real-world applications from data-mining, social networks and bioinformatics. We give theoretical as well as practical contributions to the study of CCC. Our main theoretical contribution is an alternative analysis of the famous Pivot algorithm for CCC. We show that, when simply run color-blind, Pivot is also a linear time 3-approximation for CCC. The previous best theoretical results for CCC were a 4-approximation with a high-degree polynomial runtime and a linear time 11-approximation for a variation, both by Anava et al. [WWW 2015]. With Theoretical result justifies Pivot as a baseline comparison for other heuristics, its blunt color-blindness performs poorly in practice. We develop a color-sensitive, practical heuristic we call Greedy Expansion that empirically outperforms all heuristics proposed for CCC so far, both on real-world and synthetic instances. Further, we propose a novel generalization of CCC allowing for multi-labelled edges. We argue that it is more suitable for many of the real-world applications and extend our results to this model.


We design $f$-edge fault-tolerant diameter oracles ($f$-FDO, or simply FDO if $f = 1$). For a given directed or undirected and possibly edge-weighted graph $G$ with $n$ vertices and $m$ edges and a positive integer $f$, we preprocess the graph and construct a data structure that, when queried with a set $F$ of edges, where $|F| < f$, returns the diameter of $G - F$. An $f$-FDO has stretch $\sigma \geq 1$ if the returned value $D$ satisfies diam$(G - F) \leq D \leq \sigma$ diam$(G - F)$. For the case of a single edge failure ($f = 1$) in an undirected weighted graph, there exists an approximate FDO by Henzinger et al. [ITCS 2017] with stretch $(1 + \varepsilon)$, constant query time, space $O(m)$, and a combinatorial preprocessing time of $O(mn + n^{1.5} \sqrt{Dm}/\varepsilon)$, where $D$ is the diameter. We present a near-optimal FDO with the same stretch, query time, and space. It has a preprocessing time of $O(mn + n^2/\varepsilon)$, which is better for any constant $\varepsilon > 0$. The preprocessing time nearly matches a conditional lower bound for combinatorial algorithms, also by Henzinger et al. When using fast matrix multiplication instead, we achieve a preprocessing time of $O(n^{2.794} + n^2/\varepsilon)$. We further prove an information-theoretic lower bound showing that any FDO with stretch better than $3/2$ requires $\Omega(m)$ bits of space. Thus, for constant $0 < \varepsilon < \frac{3}{2}$, our combinatorial $(1 + \varepsilon)$-approximate FDO is near-optimal in all the parameters. In the case of multiple edge failures ($f > 1$) in undirected graphs with non-negative edge weights, we give an $f$-FDO with stretch $(f + 2)$, query time $O(f^2 \log n)$, $O(fn)$ space, and preprocessing time $O(fm)$. We complement this with a lower bound excluding any finite stretch in $o(fn)$ space. Many real-world networks have polylogarithmic diameter. We show that for those graphs and up to $f = o((\log n)/\log \log n)$ failures one can swap approximation for query time and space. We present an exact combinatorial $f$-FDO with preprocessing time $O(n\log f)$, query time $n^{\omega(1)}$, and space $n^{\omega(1)}$. With fast matrix multiplication, the preprocessing time can be improved to $n^{\omega(1)}$, where $\omega < 2.373$ is the matrix multiplication exponent.


Force-directed drawing algorithms are the most commonly used approach to visualize networks. While they are usually very robust, the performance of Euclidean spring embedders decreases if the graph exhibits the high level of heterogeneity that typically occurs in real-world applications from data-mining, social networks and bioinformatics. We give theoretical as well as practical contributions to the study of CCC. Our main theoretical contribution is an alternative analysis of the famous Pivot algorithm for CCC. We show that, when simply run color-blind, Pivot is also a linear time 3-approximation for CCC. The previous best theoretical results for CCC were a 4-approximation with a high-degree polynomial runtime and a linear time 11-approximation for a variation, both by Anava et al. [WWW 2015]. With Theoretical result justifies Pivot as a baseline comparison for other heuristics, its blunt color-blindness performs poorly in practice. We develop a color-sensitive, practical heuristic we call Greedy Expansion that empirically outperforms all heuristics proposed for CCC so far, both on real-world and synthetic instances. Further, we propose a novel generalization of CCC allowing for multi-labelled edges. We argue that it is more suitable for many of the real-world applications and extend our results to this model.


Network Design problems typically ask for a minimum cost sub-network from a given host network. This classical point-of-view assumes a central authority enforcing the optimum solution. But how should networks be designed to cope with selfish agents that own parts of the network? In this setting, minimum cost networks may be very unstable in that agents will deviate from a proposed solution if this decreases their individual cost. Hence, designed networks should be both efficient in terms of total cost and stable in terms of the agents’ willingness to accept the network. We study this novel type of Network Design problem by investigating the creation of $(\beta, \gamma)$-networks, that are in $\beta$-approximate Nash equilibrium and have a total cost of at most $\gamma$ times the optimal cost, for the recently proposed Euclidean Generalized Network Creation Game by Bilò et al.SPAA2019. There, $n$ agents corresponding to points in Euclidean space create costly edges among themselves to optimize their centrality in the created network. Our main result is a simple $O(n^2)$-time algorithm that computes a $(\beta, \beta)$-network with low $\beta$ for any given set of points. Moreover, on integer grid point sets or random point sets our algorithm achieves a low constant $\beta$. Besides these results for the Euclidean model, we discuss a generalization of our algorithm to instances with arbitrary, even non-metric, edge lengths. Moreover, in contrast to these algorithmic results, we show that no such positive results are possible when focusing on either optimal networks, i.e., $(\beta, 1)$-networks, or perfectly stable networks, i.e., $(1, \gamma)$-networks, as in both cases NP-hard problems arise, there exist instances with very unstable optimal networks, and there are instances for perfectly stable networks with high total cost. Along the way, we significantly improve several results from the existing literature et al. and we asymptotically resolve their conjecture about the Price of Anarchy by providing a tight bound.


Recently, there has been an interest in studying non-uniform random $k$-satisfiability ($k$-SAT) models in order to address the non-uniformity of formulas arising from real-world applications. While uniform random $k$-SAT has been extensively studied from both a theoretical and experimental perspective, understanding the algorithmic complexity of heterogeneous distributions is still an open challenge. When a sufficiently dense formula is guaranteed to be satisfiable by conditioning or a planted assignment, it is well-known
that uniform random $k$-SAT is easy on average. We generalize this result to the broad class of non-uniform random $k$-SAT models that are characterized only by an ensemble of distributions over variables with a mild balancing condition. This balancing condition rules out extremely skewed distributions in which nearly half the variables occur less frequently than a small constant fraction of the most frequent variables, but generally recent studies non-uniform $k$-SAT distributions such as power-law and geometric formulas. We show that for all formulas generated from this model of at least logarithmic densities, a simple greedy algorithm can find a solution with high probability. As a side result we show that the total variation distance between planted and filtered (conditioned on satisfiability) models is $o(1)$ once the planted model produces formulas with a unique solution with probability $1 - o(1)$. This holds for all random $k$-SAT models where the signs of variables are drawn uniformly and independently at random.


Traditional navigation services find the fastest route for a single driver. Though always using the fastest route seems desirable for every individual, selfish behavior can have undesirable effects such as higher energy consumption and avoidable congestion, even leading to higher overall and individual travel times. In contrast, strategic routing aims at optimizing the traffic for all agents regarding a global optimization goal. We describe a framework to formalize real-world strategic routing scenarios as algorithmic problems and study one of them, which we call Single Alternative Path (SAP), in detail. There, we are given an original route between a single origin-destination pair. The goal is to suggest an alternative route to all agents that optimizes the overall travel time under the assumption that the agents distribute among both routes according to a psychological model, for which we introduce the concept of Pareto-conformity. We show that the SAP problem is NP-complete, even for such models. Nonetheless, assuming Pareto-conformity, we give multiple algorithms for different variants of SAP, using multi-criteria shortest path algorithms as subroutines. Moreover, we prove that several natural models are in fact Pareto-conform. The implementation and evaluation of our algorithms serve as a proof of concept, showing that SAP can be solved in reasonable time even though the algorithms have exponential running time in the worst case.


We study the problem of maximizing a non-monotone submodular function under multiple knapsack constraints. We propose a simple discrete greedy algorithm to approach this problem, and prove that it yields strong approximation guarantees for functions with bounded curvature. In contrast to other heuristics, this does not require problem relaxation to continuous domains and it maintains a constant-factor approximation guarantee in the problem size. In the case of a single knapsack, our analysis suggests that the standard greedy can be used in non-monotone settings. Additionally, we study this problem in a dynamic setting, in which knapsacks change during the optimization process. We modify our greedy algorithm to avoid a complete restart at each constraint update. This modification retains the approximation guarantee of the static case. We evaluate our results experimentally on a video summarization and sensor placement task. We show that our proposed algorithm competes with the state-of-the-art in static settings. Furthermore, we show that in dynamic settings with tight computational time budget, our modified greedy yields significant improvements over starting the greedy from scratch, in terms of the solution quality achieved.


We investigate the maximum-entropy model $B_{n,m,p}$ for random $n$-vertex, $m$-edge hypergraphs with expected edge size $pn$. We show that the expected size of the minimization $\min(B_{n,m,p})$, i.e., the number of inclusion-wise minimal edges of $B_{n,m,p}$, undergoes a phase transition with respect to $m$. If $m$ is at most $1/((1-p)^{1-p}m)$, then $\E[\min(B_{n,m,p})]$ is of order $\Theta(n)$, while for $m \geq 1/((1-p)^{1-p}n)$ for any $\varepsilon > 0$, it is $\Theta(2^{\Omega((1-\varepsilon)\log n)})/\sqrt{n}$. Here, $\H$ denotes the binary entropy function and $\alpha = (\log_2 m)/n$. The result implies that the maximum expected number of minimal edges over all $m$ is $\Theta((1+p)^n)/\sqrt{n}$. Our structural findings have algorithmic implications for minimizing an input hypergraph. This has applications in the profiling of relational databases as well as for the Orthogonal Vectors problem studied in fine-grained complexity. We make several technical contributions that are of independent interest in probability. First, we improve the Chernoff–Hoeffding theorem on the tail of the binomial distribution. In detail, for a binomial variable $X \sim \text{Bin}(n,p)$ and any $0 < x < p$, it holds that $P[X \leq x] = \Theta(2^{-\Omega(xp)}\sqrt{n})$, where $D$ is the binary Kullback–Leibler divergence between Bernoulli distributions. We give explicit upper and lower bounds on the constants hidden in the big-O notation that hold for all $n$. Secondly, we establish the fact that the probability of a set of cardinality $\alpha n$ being minimal after $m$ i.i.d. maximum-entropy trials exhibits a sharp threshold behavior at $\alpha^* = n + \log_2 m$.


How do rational agents self-organize when trying to connect to a common target? We study this question with a simple tree formation game which is related to the well-known fair single-source connection game by Anshelevich et al. (FOCS’04) and selfish spanning tree games by Gourvès and Monnot (WINE’08). In our game agents correspond to nodes in a network that activate a single outgoing edge to connect to the common target node (possibly via other nodes). Agents pay for their path to the common target, and edge costs are shared fairly among all agents using an edge. The main novelty of our model is dynamic edge costs that depend on the in-degree of the respective endpoint. This reflects that connecting to popular nodes that have increased internal coordination costs is more expensive since they can charge higher prices for their routing service. In contrast to related models, we show that equilibria are not guaranteed to exist, but we prove the existence for infinitely many numbers of agents. Moreover, we analyze the structure of equilibrium trees and employ probabilistic insights to prove a constant upper bound on the height of such a tree as well as non-trivial lower bounds on both the Price of Anarchy and the Price of Stability. We also show that in comparison with the social optimum tree the overall cost of an equilibrium tree is more fairly shared among the agents. Thus, we prove that self-organization of rational agents yields on average only slightly higher cost per agent compared to the centralized optimum, and at the same time, it induces a more fair cost distribution. Moreover, equilibrium trees achieve a beneficial trade-off between a low height and low maximum degree, and hence these trees might be of independent interest from a combinatorics point-of-view. We conclude with a discussion of promising extensions of our model.


Time series are sequences of data indexed by time. Such data are collected in various domains, often in massive amounts, such that storing them proves challenging. Thus, time series are commonly stored in a compressed format. An important compression approach is piecewise linear approximation (PLA), which only keeps a small set of time points and interpolates the remainder linearly. Picking a subset of time points such that the PLA minimizes the mean squared error to the original time series is a challenging task, naturally lending itself to heuristic approaches. We propose the piecewise linear approximation genetic algorithm (PLA-GA) for compressing time series by PLA. The PLA-GA is a memetic (μ + λ) GA that makes use of two different operators tailored to time series compression. First, we add special individuals to the initial population that are derived using established PLA heuristics. Second, we propose a novel local search operator that greedily improves a compressed time series. We compare the PLA-GA empirically with existing evolutionary approaches and with a deterministic PLA algorithm, known as Bellman’s algorithm, that is optimal for the restricted setting of sampling. In both cases, the PLA-GA approximates the original time series better and quicker. Further, it drastically outperforms Bellman’s algorithm with increasing instance size with respect to run time until finding a solution of equal or better quality — we observe speed-up factors between 7 and 100 for instances of 90,000 to 100,000 data points.


Network Creation Games (NCGs) model the creation of decentralized communication networks like the Internet. In such games strategic agents corresponding to network nodes selfishly decide with whom to connect to optimize some objective function. Past research intensively analyzed models where the agents strive for a central position in the network. This models agents optimizing the network for low-latency applications like VoIP. However, with today’s abundance of streaming services it is important to ensure that the created network can satisfy the increased bandwidth demand. To the best of our knowledge, this natural problem of the decentralized strategic creation of networks with sufficient bandwidth has not yet been studied. We introduce Flow-Based NCGs where the selfish agents focus on bandwidth instead of latency. In essence, budget-constrained agents create network links to maximize their own, or average network flow value to all other network nodes. Equivalently, this can also be understood as agents who create links to increase their connectivity and thus also the robustness of the network. For this novel type of NCG we prove that pure Nash equilibria exist, we give a simple algorithm for computing optimal networks, we show that the Price of Stability is 1 and we prove an (almost) tight bound of 2 on the Price of Anarchy. Last but not least, we show that our models do not admit a potential function.


The VertexCover problem is proven to be computationally hard in different ways: It is NP-complete to find an optimal solution and even APX-hard to find an approximation with reasonable factors. In contrast, recent theoretical results suggest that many real-world networks the run time to solve VertexCover is way smaller than even the best known FPT-approaches can explain. Similarly, greedy algorithms deliver very good approximations to the optimal solution in practice. We link these observations to two properties that are observed in many real-world networks, namely a heterogeneous degree distribution and high clustering. To formalize these properties and explain the observed behavior, we analyze how a branch-and-reduce algorithm performs on hyperbolic random graphs, which have become increasingly popular for modeling real-world networks. In fact, we are able to show that the VertexCover problem on hyperbolic random graphs can be solved in polynomial time, with high probability. The proof relies on interesting structural properties of hyperbolic random graphs. Since these predictions of the model are interesting in their own right, we conducted experiments on real-world networks showing that these properties are also observed in practice. When utilizing the same structural properties in an adaptive greedy algorithm, further experiments suggest that, on real instances, this leads to better approximations than the standard greedy approach within reasonable time. We link these observations to two properties that are observed in many real-world networks, namely a heterogeneous degree distribution and high clustering. To formalize these properties and explain the observed behavior, we analyze how a branch-and-reduce algorithm performs on hyperbolic random graphs, which have become increasingly popular for modeling real-world networks. In fact, we are able to show that the VertexCover problem on hyperbolic random graphs can be solved in polynomial time, with high probability. The proof relies on interesting structural properties of hyperbolic random graphs. Since these predictions of the model are interesting in their own right, we conducted experiments on real-world networks showing that these properties are also observed in practice. When utilizing the same structural properties in an adaptive greedy algorithm, further experiments suggest that this leads to even better approximations than the standard greedy approach on real instances.


We devise an enumeration method for inclusion-wise minimal hitting sets in hypergraphs. It has delay $O(m^{k+1} \cdot n^2)$ and uses linear space. Hereby, $n$ is the number of vertices, $m$ the number of hyperedges, and $k$ the rank of the transversal hypergraph. In particular, on classes of hypergraphs for which the cardinality $k$ of the largest minimal hitting set is bounded, the delay is polynomial. The algorithm solves the extension problem for minimal hitting sets as a subroutine. We show that the extension problem is W[3]-complete when parametrised by the cardinality of the set which is to be extended. For the subroutine, we give an algorithm that is optimal under the exponential time hypothesis. Despite these lower bounds, we provide empirical evidence showing that the enumeration outperforms the theoretical worst-case guarantee on hypergraphs arising in the profiling of relational databases, namely, in the detection of unique column combinations.


In this paper, we consider subset selection problems for functions $f$ with constraints where the constraint bound $B$ changes over time. We point out that adaptive variants of greedy approaches commonly used in the area of submodular optimization are not able to maintain their approximation quality. Investigating the recently introduced POMC Pareto optimization approach, we show that this algorithm efficiently computes $\phi = (\alpha_f/2)(1 - \frac{1}{\alpha_f})$-approximation, where $\alpha_f$ is the submodularity ratio, for each possible constraint bound $b \leq B$. Furthermore, we show that POMC is able to adapt its set of solutions quickly in the case that $B$ increases. Our experimental investigations for the influence maximization in social networks show the advantage of POMC over generalized greedy algorithms.
We investigate the performance of a deterministic GREEDY algorithm for the problem of maximizing functions under a partition matroid constraint. We consider non-monotone submodular functions and monotone subadditive functions. Even though constrained maximization problems of monotone submodular functions have been extensively studied, little is known about greedy maximization of non-monotone submodular functions or monotone subadditive functions. We give approximation guarantees for GREEDY on these problems, in terms of the curvature. We find that this simple heuristic yields a strong approximation guarantee on a broad class of functions. We discuss the applicability of our results to three real-world problems: maximizing the determinant function of a positive semidefinite matrix, and related problems such as the maximum entropy sampling problem, the constrained maximum cut problem on directed graphs, and combinatorial auction games. We conclude that GREEDY is well-suited to approach these problems. Overall, we present evidence to support the idea that, when dealing with constrained maximization problems with bounded curvature, one needs not search for (approximate) monotonicity to get good approximate solutions.


Hyperbolic random graphs (HRG) and geometric inhomogeneous random graphs (GIRG) are two similar generative network models that were designed to resemble complex real world networks. In particular, they have a power-law degree distribution with controllable exponent $\beta$, and high clustering that can be controlled via the temperature $T$. We present the first implementation of an efficient GIRG generator running in expected linear time. Besides varying temperatures, it also supports underlining geometries of higher dimensions. It is capable of generating graphs with ten million edges in under a second on commodity hardware. The algorithm can be adapted to HRGs. To illustrate our implementation, we show a statistical analysis of SAT and has triggered the development of sophisticated rigorous and non-rigorous techniques for analyzing random structures. Despite a long line of research and substantial progress, most theoretical work on random SAT assumes a uniform distribution on the variables. In contrast, real-world instances often exhibit large fluctuations in variable occurrence. This can be modeled by a non-uniform distribution of the variables, which can result in distributions closer to industrial SAT instances. We study satisfiability thresholds of non-uniform random 2-SAT with $n$ variables and $m$ clauses and with an arbitrary probability distribution $(p_1,\ldots,p_n)$ with $p_1 \geq p_2 \geq \cdots \geq p_n > 0$ over the $n$ variables. We show that $p_2^*=\Theta(\sum_{i=2}^{n-1} p_i^2)$ that the asymptotic satisfiability threshold is at $m = \Theta(1 - (\sum_{i=1}^{n} p_i^2)/(\sum_{i=2}^{n} p_i^2)^{1/2})$ and that it is coarse. For $p_2^* = \Theta(\sum_{i=2}^{n-1} p_i^2)$ we show that there is a sharp satisfiability threshold at $m = \Theta(\sum_{i=2}^{n-1} p_i^2)^{-1}$. This result generalizes the seminal works by Chvatal and Reed [FOCS 1992] and by Goerdt [JCSS 1996].


Assigning staff to engagements according to hard constraints while optimizing several objectives is a task encountered by many companies on a regular basis. Simplified versions of such assignment problems are NP-hard. Despite this, a typical approach to solving them consists of formulating them as mixed integer programming (MIP) problems and using a state-of-the-art solver to get solutions that closely approximate the optimum. In this paper, we consider a complex real-world staff assignment problem encountered by the professional service company KPMG, with the goal of finding an algorithm that solves it faster and with a better solution than a commercial MIP solver. We follow the evolutionary algorithm (EA) metaheuristic and design a search heuristic which iteratively improves a solution using domain-specific mutation operators. Furthermore, we use a flow algorithm to optimally solve a subproblem, which tremendously reduces the search space for the EA. For our real-world instance of the assignment problem, given the same total time budget of 100 hours, a parallel EA approach finds a solution that is only 1.7% away from an upper bound for the (unknown) optimum within under five hours, while the MIP solver Gurobi still has a gap of 10.5%.


We study non-uniform random $k$-SAT on $n$ variables with an arbitrary probability distribution $p$ on the variable occurrences. The number $t = t(n)$ of randomly drawn clauses at which random formulas go from asymptotically almost surely (a.a.s.) satisfiable to a.a.s. unsatisfiable is called the satisfiability threshold. Such a threshold is called sharp if it approaches a step function as $n$ increases. We show that a threshold $t(n)$ for random $k$-SAT with an ensemble $(p_n)_{n \in \mathbb{N}}$ of arbitrary probability distributions on the variable occurrences is sharp if $\|p_2\|_2^2 = o_n(2^{-1/k})$ and $\|p_2\|_\infty = o_n(\log^{(k-1)/(2k-1)}(t))$. This result generalizes Friedgut’s sharpness result from uniform to non-uniform random $k$-SAT and implies sharpness for thresholds of a wide range of random $k$-SAT models with heterogeneous probability distributions, for example such models where the variable probabilities follow a power-law distribution.

Network Creation Games are a well-known approach for explaining and analyzing the structure, quality and dynamics of real-world networks like the Internet and other infrastructure networks which evolved via the interaction of selfish agents without a central authority. In these games selfish agents which correspond to nodes in a network strategically buy incident edges to improve their centralization. However, past research has shown that these games have only considered the creation of networks with unit-weight edges. In practice, e.g. when constructing a fiber-optic network, the choice of which nodes to connect and also the induced price for a link crucially depends on the distance between the involved nodes and such settings can be modeled via edge-weighted graphs. We incorporate arbitrary edge weights by generalizing the well-known model by Fabrikant et al. [PODC’03] to edge-weighted host graphs and focus on the geometric setting where the weights are induced by the distances in some metric space. In stark contrast to the state-of-the-art for the unit-weight version, where the Price of Anarchy is conjectured to be constant and where resolving this is a major open problem, we prove a tight non-constant bound on the Price of Anarchy for the metric version and a slightly weaker upper bound for the non-metric case. Moreover, we analyze the existence of equilibria, the computational hardness and the game dynamics for several natural metrics. The model we propose can be seen as the game-theoretic analogue of a variant of the classical Network Design Problem. Thus, low-cost equilibria of our game correspond to decentralized and stable approximations of the optimum network design.

### References


Network science is driven by the question which properties large real-world networks have and how we can exploit them algorithmically. In the past few years, hyperbolic graphs have emerged as a very promising model for scale-free networks. The connection between hyperbolic geometry and complex networks gives insights in both directions: (1) Hyperbolic geometry forms the basis of a natural and explanatory model for real-world networks. Hyperbolic random graphs are obtained by choosing random points in the hyperbolic plane and connecting pairs of points that are geometrically close. The resulting networks share many structural properties for example with online social networks like Facebook or Twitter. They are thus well suited for algorithmic analyses in a more realistic setting. (2) Starting with a real-world network, hyperbolic geometry is well-suited for metric embeddings. The vertices of a network can be mapped to points in this geometry, such that geometric distances are similar to graph distances. Such embeddings have a variety of algorithmic applications ranging from approximations based on efficient geometric algorithms to greedy routing solely using hyperbolic coordinates for navigation decisions.


The hardness of formulas at the solubility phase transition of random propositional satisfiability (SAT) has been intensely studied for decades both empirically and theoretically. Solvers based on stochastic local search (SLS) appear to scale very well at the critical threshold, while complete backtracking solvers exhibit exponential scaling. On industrial SAT instances, this phenomenon is inverted: backtracking solvers can tackle large industrial problems, where SLS-based solvers appear to stall. Industrial instances exhibit sharply different structure than uniform random instances. Among many other properties, they are often heterogeneous in the sense that some variables appear in many while others appear in only few clauses. We conjecture that the heterogeneity of SAT formulas alone already contributes to the trade-off in performance between SLS solvers and complete backtracking solvers. We empirically determine how the run time of SLS vs. backtracking solvers depends on the heterogeneity of the input, which is controlled by drawing variables according to a scale-free distribution. Our experiments reveal that the efficiency of complete solvers at the phase transition is strongly related to the heterogeneity of the degree distribution. We report results that suggest the depth of satisfying assignments in complete search trees is influenced by the level of heterogeneity as measured by a power-law exponent. We also find that incomplete SLS solvers, which scale well on uniform instances, are not affected by heterogeneity. The main contribution of this paper utilizes the scale-free random 3-SAT model to isolate heterogeneity as an important factor in the scaling discrepancy between complete and SLS solvers at the uniform phase transition found in previous works.


The phenomenon of residential segregation was captured by Schelling’s famous segregation model where two types of agents are placed on a grid and an agent is content with her location if the fraction of her neighbors which have the same type as her is at least \(r\). However, past research has only considered the creation of networks with unit-weight edges. In practice, e.g. when constructing a fiber-optic network, the choice of which nodes to connect and also the induced price for a link crucially depends on the distance between the involved nodes and such settings can be modeled via edge-weighted graphs. We incorporate arbitrary edge weights by generalizing the well-known model by Fabrikant et al. [PODC’03] to edge-weighted host graphs and focus on the geometric setting where the weights are induced by the distances in some metric space. In stark contrast to the state-of-the-art for the unit-weight version, where the Price of Anarchy is conjectured to be constant and where resolving this is a major open problem, we prove a tight non-constant bound on the Price of Anarchy for the metric version and a slightly weaker upper bound for the non-metric case. Moreover, we analyze the existence of equilibria, the computational hardness and the game dynamics for several natural metrics. The model we propose can be seen as the game-theoretic analogue of a variant of the classical Network Design Problem. Thus, low-cost equilibria of our game correspond to decentralized and stable approximations of the optimum network design.


Given a public transportation network of stations and connections, we want to find a minimum subset of stations such that each connection runs through a selected station. Although this problem is NP-hard in general, real-world instances are regularly solved almost completely by a set of simple reduction rules. To explain this behavior, we view transportation networks as hitting set instances and identify two characteristic properties, locality and heterogeneity. We then devise a randomized model to generate hitting set instances with adjustable properties. While the heterogeneity does influence the effectiveness of the reduction rules, the generated instances show that locality is the significant factor. Beyond that, we prove that the effectiveness of the reduction rules is independent of the underlying graph structure. Finally, we show that high locality is also prevalent in instances from other domains, facilitating a fast computation of minimum hitting sets.

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 world 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.


Greedy algorithms provide a fast and often also effective solution to many combinatorial optimization problems. However, it is well known that they sometimes lead to low quality solutions on certain instances. In this paper, we explore the use of randomness in greedy algorithms for the minimum vertex cover and dominating set problem and compare the resulting performance against their deterministic counterpart. Our algorithms are based on a parameter $\gamma$ which allows to explore the spectrum between uniform and deterministic greedy selection in the steps of the algorithm and our theoretical and experimental investigations point out the benefits of incorporating randomness into greedy algorithms for the two considered combinatorial optimization problems.


In the context of black box optimization, the most common way to handle deceptive attractors is to periodically restart the algorithm. In this paper, we explore the benefits of combining the simple $(1+1)$ Evolutionary Algorithm (EA) with the Luby Universal Strategy - the $(1+1)EA_{U}$, a meta-heuristic that does not require parameter tuning. We first consider two artificial pseudo-Boolean landscapes, on which the $(1+1)EA$ exhibits exponential run time. We prove that the $(1+1)EA_{U}$ has polynomial run time on both instances. We then consider the Minimum Vertex Cover on two classes of graphs. Again, the $(1+1)EA$ yields exponential run time on those instances, and the $(1+1)EA_{U}$ finds the global optimum in polynomial time. We conclude by studying the Makespan Scheduling. We consider an instance on which the $(1+1)EA$ does not find a $(4/3-\epsilon)$-approximation in polynomial time, and we show that the $(1+1)EA_{U}$ reaches a $(4/3-\epsilon)$-approximation in polynomial time. We then prove that the $(1+1)EA_{U}$ serves as an Efficient Polynomial-time Approximation Scheme (EPTAS) for the Partition Problem, for a $(1+\epsilon)$-approximation with $\epsilon > 4/n$.


In many Evolutionary Algorithms (EAs), a parameter that needs to be tuned is that of the mutation rate, which determines the probability for each decision variable to be mutated. Typically, this rate is set to $1/n$ for the duration of the optimization, where $n$ is the number of decision variables. This setting has the appeal that the expected number of mutated variables per iteration is one. In a recent theoretical study, it was proposed to sample the number of mutated variables from a power-law distribution. This results into a significantly higher probability on larger numbers of mutations, so that escaping local optima becomes more probable. In this paper, we propose another class of non-uniform mutation rates. We study the benefits of this operator in terms of average-case black-box complexity analysis and experimental comparison. We consider both pseudo-Boolean artificial landscapes and combinatorial problems (the Minimum Vertex Cover and the Maximum Cut). We observe that our non-uniform mutation rates significantly outperform the standard choices, when dealing with landscapes that exhibit large deceptive basins of attraction.


A common way to accelerate shortest path algorithms on graphs is the use of a bidirectional search, which simultaneously explores the graph from the start and the destination. It has been observed recently that this strategy performs particularly well on scale-free real-world networks. Such networks typically have a heterogeneous degree distribution (e.g., a power-law distribution) and high clustering (i.e., vertices with a common neighbor are likely to be connected themselves). These two properties can be obtained by assuming an underlying hyperbolic geometry. To explain the observed behavior of the bidirectional search, we analyze its running time on hyperbolic random graphs and prove that it is $O(n^{\frac{1}{2}-\frac{1}{\gamma}}+\sqrt{\log n})$ with high probability, where $\gamma \geq (0.5, 1)$ controls the power-law exponent of the degree distribution, and $\Delta_{\text{max}}$ is the maximum degree. This bound is sublinear, improving the obvious worst-case linear bound. Although our analysis depends on the underlying geometry, the algorithm itself is oblivious to it.


A core feature of evolutionary algorithms is their mutation operator. Recently, much attention has been devoted to the study of mutation operators with dynamic and non-uniform mutation rates. Following up on this line of work, we propose a new mutation operator and analyze its performance on the $(1+1)$ Evolutionary Algorithm (EA). Our analyses show that this mutation operator competes with pre-existing ones, when used by the $(1+1)EA$ on classes of problems for which results on the other mutation operators are available. We present a jump function for which the performance of the $(1+1)EA$ using any static uniform mutation and any restart strategy can be worse than the performance of the $(1+1)EA$ using our mutation operator with no restarts. We show that the $(1+1)EA$ using our mutation operator finds a $(1/3)$-approximation ratio on any non-negative submodular function in polynomial time. This performance matches that of combinatorial local search algorithms specifically designed to solve this problem. Finally, we evaluate experimentally the performance of the $(1+1)EA$ using our operator, on real-world graphs of different origins with up to $\sim 37000$ vertices and $\sim 1.6$ million edges. In comparison with uniform mutation and a recently proposed dynamic scheme our operator comes out on top on these instances.

Modern routing algorithms reduce query time by depending heavily on preprocessed data. The recently developed Navigation Data Standard (NDS) enforces a separation between algorithms and map data, rendering preprocessing inapplicable. Furthermore, map data is partitioned into tiles with respect to their geographic coordinates. With the limited memory found in portable devices, the number of tiles loaded becomes the major factor for run time. We study routing under these restrictions and present new algorithms as well as empirical evaluations. Our results show that, on average, the most efficient algorithm presented uses more than 20 times fewer tile loads than a normal A*.


We study non-uniform random k-SAT on n variables with an arbitrary probability distribution p on the variable occurrences. The number t = t(n) of randomly drawn clauses at which random formulas go from asymptotically almost surely (a.a.s.) satisfiable to a.a.s. unsatisfiable is called the satisfiability threshold. Such a threshold is called sharp if it approaches a step function as n increases. We show that a threshold t(n) for random k-SAT with an ensemble (p_α)_α∈A of arbitrary probability distributions on the variable occurrences is sharp if ∥p∥_2 = O((t_γ)^{-\epsilon}) and ∥p∥_∞ = o((t_γ)^{-1/2}log^2((t_γ)^{-1})) for any t_γ ≥ 2. This result generalizes Friedgut's sharpness result from uniform to non-uniform random k-SAT and implies sharpness for thresholds of a wide range of random k-SAT models with heterogeneous probability distributions, for example such models where the variable probabilities follow a power-law distribution.


Generative graph models play an important role in network science. Unlike real-world networks, they are accessible for mathematical analysis and the number of available networks is not limited. The explanatory power of results on generative models, however, heavily depends on how realistic they are. We present a framework that allows for a systematic evaluation of generative network models. It is based on the question whether real-world networks can be distinguished from generated graphs with respect to certain graph parameters. As a proof of concept, we apply our framework to four popular random graph models (Erdős-Rényi, Barabási-Albert, Chung-Lu, and hyperbolic random graphs). Our experiments for example show that all four models are bad representations for Facebook’s social networks, while Chung-Lu and hyperbolic random graphs are good representations for other networks, with different strengths and weaknesses.


We investigate how well-performing local search algorithms for small or medium size instances can be scaled up to perform well for large inputs. We introduce a parallel kernelization technique that is motivated by the assumption that graphs in medium to large scale are composed of components which are on their own easy for state-of-the-art solvers but when hidden in large graphs are hard to solve. To show the effectiveness of our kernelization technique, we consider the well-known minimum vertex cover problem and two state-of-the-art solvers called NuMVC and FastVC. Our kernelization approach reduces an existing large problem instance significantly and produces better quality results on a wide range of benchmark instances and real world graphs.


A common strategy for improving optimization algorithms is to restart the algorithm when it is believed to be trapped in an inferior part of the search space. However, while specific restart strategies have been developed for specific problems (and specific algorithms), restarts are typically not regarded as a general tool to speed up an optimization algorithm. In fact, many optimization algorithms do not employ restarts at all. Recently, "bet-and-run" was introduced in the context of mixed-integer programming, where first a number of short runs with randomized initial conditions is made, and then the most promising run of these is continued. In this article, we consider two classical NP-complete combinatorial optimization problems, traveling salesperson and minimum vertex cover, and study the effectiveness of different bet-and-run strategies. In particular, our restart strategies do not take any problem knowledge into account, nor are tailored to the optimization algorithm. Therefore, they can be used off-the-shelf. We observe that state-of-the-art solvers for these problems can benefit significantly from restarts on standard benchmark instances.


We provide a brief overview on some hot topics in the area of evolutionary computation. Our main focus is on recent developments in the areas of combinatorial optimization and real-world applications. Furthermore, we highlight recent progress on the theoretical understanding of evolutionary computing methods.


Recently, a number of non-uniform random satisfiability models have been proposed that are closer to practical satisfiability problems in some characteristics. In contrast to uniform random Boolean formulas, scale-free formulas have a variable occurrence distribution that follows a power law. It has been conjectured that such a distribution is a more accurate model for some industrial instances than the uniform random model. Though it seems that there is already an awareness of a threshold phenomenon in such models, there is still a complete picture lacking. In contrast to the uniform model, the critical density threshold does not lie at a single point, but instead exhibits a functional dependency on the power-law exponent. For scale-free formulas with clauses of length k ≥ 2, we give a lower bound on the phase transition threshold as a function of the scaling parameter. We also perform computational studies that suggest our bound is tight and investigate the critical density for formulas with higher clause lengths. Similar to the uniform model, on formulas with k ≥ 3, we find that the phase transition regime corresponds to a set of formulas that are difficult to solve by backtracking search.
[137] Wagner, M., Friedrich, T., Lindauer, M., Improving local search in a minimum vertex cover solver for classes of networks. In: Congress on Evolutionary Computation (CEC), pp. 1704–1711, 2017. For the minimum vertex cover problem, a wide range of solvers has been proposed over the years. Most classical exact approaches are encountering run time issues on massive graphs that are considered nowadays. A straightforward alternative approach is then to use heuristics, which make assumptions about the structure of the studied graphs. These assumptions are typically hard-coded and are hoped to work well for a wide range of networks - which is in conflict with the nature of broad benchmark sets. With this article, we investigate in two ways. First, we identify a component in an existing solver that influences its performance depending on the class of graphs, and then customize instances of this solver for different classes of graphs. Second, we create the first algorithm portfolio for the minimum vertex cover to further improve the performance of a single integrated approach to the minimum vertex cover problem.

[138] Friedrich, T., Krohmer, A., Rothenberger, R., Sauerwald, T., Sutton, A. M., Bounds on the Satisfiability Threshold for Power Law Distributed Random SAT. In: European Symposium on Algorithms (ESA), pp. 37:1–37:15, 2017. Propositional satisfiability (SAT) is one of the most fundamental problems in computer science. The worst-case hardness of SAT lies at the core of computational complexity theory. The average-case analysis of SAT has triggered the development of sophisticated rigorous and non-rigorous techniques for analyzing random structures. Despite a long line of research and substantial progress, nearly all theoretical work on random SAT assumes a uniform distribution on the variables. In contrast, real-world instances often exhibit large fluctuations in variable occurrence. This can be modeled by a scale-free distribution of the variables, which results in distributions closer to industrial SAT instances. We study random k-SAT on n variables, m = Θ(n) clauses, and a power law distribution on the variable occurrences with exponent β. We observe a satisfiability threshold at β = (2k−1)/(k−1). This threshold is tight in the sense that instances with $\beta < 2(k-1)/(k-1) - \varepsilon$ for any constant $\varepsilon > 0$ are unsatisfiable with high probability (w.h.p.). For $\beta > (2k-1)/(k-1) + \varepsilon$, the picture is reminiscent of the uniform case: instances are satisfiable w.h.p. for sufficiently small constant clause-variable ratios $m/n$: they are unsatisfiable above a ratio $m/n$ that depends on $\beta$.

[139] Friedrich, T., Kötzing, T., Quinzan, F., Sutton, A. M., Resampling vs Recombination: a Statistical Run Time Estimation. In: Foundations of Genetic Algorithms (FOGA), pp. 25–35, 2017. Noise is pervasive in real-world optimization, but there is still little understanding of the interplay between the operators of randomized search heuristics and explicit noise-handling techniques, such as statistical resampling. In this paper, we report on several statistically motivated models and theoretical results that help to clarify this reciprocal relationship for a collection of randomized search heuristics on noisy functions. We consider the optimization of pseudo-Boolean functions under additive posterior Gaussian noise and explore the trade-off between noise reduction and the computational cost of resampling. We first perform experiments to find the optimal parameters at a given noise intensity for a mutation-only evolutionary algorithm, a genetic algorithm employing recombination, an estimation of distribution algorithm (EDA), and an ant colony optimization algorithm. We then observe how the optimal parameter depends on the noise intensity for the different algorithms. Finally, we locate the point where statistical resampling costs more than it is worth in terms of run time. We find that the EA requires the highest number of resamples to obtain the best speed-up, whereas crossovers both the run time and the number of resamples required. Most surprisingly, we find that EDA-like algorithms require no resampling, and can handle noise implicitly.

[140] Friedrich, T., Kötzing, T., Lagodzinski, J. A. G., Neumann, F., Schirneck, M., Analysis of the (1+1) EA on Subclasses of Linear Functions under Uniform and Linear Constraints. In: Foundations of Genetic Algorithms (FOGA), pp. 45–54, 2017. Linear functions have gained a lot of attention in the area of run time analysis of evolutionary computation methods and the corresponding analyses have provided many effective tools for analyzing more complex problems. In this paper, we consider the behavior of the classical (1+1) Evolutionary Algorithm for linear functions under linear constraint. We show tight bounds in the case where both the objective function and the constraint is given by the OneMax function and present upper bounds as well as lower bounds for the general case. Furthermore, we also consider the LeadingOnes fitness function.

[141] Pourhassan, M., Friedrich, T., Neumann, F., On the Use of the Dual Formulation for Minimum Weighted Vertex Cover in Evolutionary Algorithms. In: Foundations of Genetic Algorithms (FOGA), pp. 37–44, 2017. We consider the weighted minimum vertex cover problem and investigate how its dual formulation can be exploited to design evolutionary algorithms that provably obtain a 2-approximation. Investigating multi-valued representations, we show that variants of randomized local search and the (1+1) EA achieve this goal in expected pseudo-polynomial time. In order to speed up the process, we consider the use of step size adaptation in both algorithms and show that RLS obtains a 2-approximation in expected polynomial time while the (1+1) EA still encounters a pseudo-polynomial lower bound.

[142] Friedrich, T., Kötzing, T., Melnichenko, A., Analyzing Search Heuristics with Differential Equations. In: Genetic and Evolutionary Computation Conference (GECCO), pp. 313–314, 2017. Drift Theory is currently the most common technique for the analysis of randomized search heuristics because of its broad applicability and the resulting tight first hitting time bounds. The biggest problem when applying a drift theorem is to find a suitable potential function which maps a complex space into a single number, capturing the essence of the state of the search in just one value. We discuss another method for the analysis of randomized search heuristics based on the Theory of Differential Equations. This method considers the deterministic counterpart of the randomized process by replacing probabilistic outcomes by their expectation, and then bounding the error with good probability. We illustrate this by analyzing an Ant Colony Optimization algorithm (ACO) for the Minimum Spanning Tree problem (MST).

[143] Chauhan, A., Friedrich, T., Quinzan, F., Approximating Optimization Problems using EAs on Scale-Free Networks. In: Genetic and Evolutionary Computation Conference (GECCO), pp. 235–242, 2017. It has been experimentally observed that real-world networks follow certain topological properties, such as small-world, power-law etc. To study these networks, many random graph models, such as Preferential Attachment, have been proposed. In this paper, we consider the deterministic counterparts which capture power-law degree distribution and degeneracy. Networks with these properties are known as scale-free networks in the literature. Many interesting problems remain NP-hard on scale-free networks. We study the relationship between scale-free properties and the approximation-ratio of some commonly used evolutionary algorithms. For the
Vertex Cover, we observe experimentally that the \((1 + 1)\) EA always gives the better result than a greedy local search, even when it runs for only \(O(n \log(n))\) steps. We give the construction of a scale-free network in which a multi-objective algorithm and a greedy algorithm obtain optimal solutions, while the \((1 + 1)\) EA obtains the worst possible solution with constant probability. We prove that for the Dominating Set, Vertex Cover, Connected Dominating Set and Independent Set, the \((1 + 1)\) EA obtains constant-factor approximation in expected run time \(O(n \log(n))\) and \(O(n^2)\) respectively. Whereas, GSEMO gives even better approximation than \((1 + 1)\) EA in expected run time \(O(n^3)\) for Dominating Set, Vertex Cover and Connected Dominating Set on such networks.


The investigation of linear pseudo-Boolean functions play a central role in the area of runtime analysis of evolutionary computing techniques. Having an additional linear constraint on a linear function is equivalent to the NP-hard knapsack problem and special problem classes thereof have been investigated in recent works. In this paper, we extend these studies to problems with dynamic constraints and investigate the runtime of different evolutionary algorithms to recompute an optimal solution when the constraint bound changes by a certain amount. We study the classical \((1 + 1)\) EA and population-based algorithms and show that they recompute an optimal solution very efficiently. Furthermore, we show that a variant of the \((1 + (\lambda, \lambda))\) GA can recomputed the optimal solution more efficiently in some cases.


Island models in evolutionary computation solve problems by a careful interplay of independently running evolutionary algorithms on the island and an exchange of good solutions between the islands. In this work, we conduct rigorous run time analyses for such island models trying to simultaneously obtain good run times and low communication effort. We improve the existing upper bounds for the communication effort (i) by improving the run time bounds via a careful analysis, (ii) by setting the balance between individual computation and communication in a more appropriate manner, and (iii) by replacing the usual communicate-with-all-neighbors approach with randomized rumor spreading, where each island contacts a randomly chosen neighbor. This epidemic communication paradigm is known to lead to very fast and robust information dissemination in many applications. Our results concern islands running simple \((1 + 1)\) evolutionary algorithms, we regard d-dimensional tori and complete graphs as communication topologies, and optimize the classic test functions OneMax and LeadingOnes.


Inspired by real world examples, e.g. the Internet, researchers have introduced an abundance of strategic games to study natural phenomena in networks. Unfortunately, almost all of these games have the conceptual drawback of being computationally intractable, i.e. computing a best response strategy or checking if an equilibrium is reached is NP-hard. Thus, a main challenge in the field is to find tractable realistic network formation models. We address this challenge by investigating a very recently introduced model by Goyal et al. [WINE'16] which focuses on robust networks in the presence of a strong adversary who attacks (and kills) nodes in the network and lets this attack spread virus-like to neighboring nodes and their neighbors. Our main result is to establish that this natural model is one of the few exceptions which are both realistic and computationally tractable. In particular, we answer an open question of Goyal et al. by providing an efficient algorithm for computing a best response strategy, which implies that deciding whether the game has reached a Nash equilibrium can be done efficiently as well. Our algorithm essentially solves the problem of computing a minimal connection to a network which maximizes the reachability while hedging against severe attacks on the network infrastructure and may thus be of independent interest.


Inspired by real world examples, e.g. the Internet, researchers have introduced an abundance of strategic games to study natural phenomena in networks. Unfortunately, almost all of these games have the conceptual drawback of being computationally intractable, i.e. computing a best response strategy or checking if an equilibrium is reached is NP-hard. Thus, a main challenge in the field is to find tractable realistic network formation models. We address this challenge by establishing that the recently introduced model by Goyal et al. [WINE'16], which focuses on robust networks in the presence of a strong adversary, is a rare exception which is both realistic and computationally tractable. In particular, we sketch an efficient algorithm for computing a best response strategy, which implies that deciding whether the game has reached a Nash equilibrium can be done efficiently as well. Our algorithm essentially solves the problem of computing a minimal connection to a network which maximizes the reachability while hedging against severe attacks on the network infrastructure.


An estimated 30% of urban traffic is caused by search for parking spots. Traffic could be reduced by suggesting effective routes leading along potential parking spots. In this paper, we formalize parking search as a probabilistic problem on a road graph and show that it is NP-complete. We explore heuristics that optimize for the driving duration and the walking distance to the destination. Routes are constrained to reach a certain probability threshold of finding a spot. Empirically estimated probabilities of successful parking attempts are provided by TomTom on a per-street basis. We release these probabilities as a dataset of about 80,000 roads covering the Berlin area. This allows to evaluate parking search algorithms on a real road network with realistic probabilities for the first time. However, for many other areas, parking probabilities are not openly available. Because they are effortful to collect, we propose an algorithm that relies on conventional road attributes only. Our experiments show that this algorithm comes close to the baseline by a factor of 1.3 in our cost measure. This leads to the conclusion that conventional road attributes may be sufficient to compute reasonably good parking search routes.

When designing and analyzing algorithms, one can obtain better and more realistic results for practical instances by assuming a certain probability distribution on the input. The worst-case run-time is then replaced by the expected run-time or by bounds that hold with high probability (whp), i.e., with probability $1 - O(1/n)$, on the random input. Hyperbolic random graphs can be used to model complex real-world networks as they share many important properties such as a small diameter, a large clustering coefficient, and a power-law degree distribution. Divide and conquer is an important algorithmic design principle that works particularly well if the instance admits small separators. We show that hyperbolic random graphs in fact have comparatively small separators. More precisely, we show that a hyperbolic random graph can be expected to have a balanced separator hierarchy with separators of size $O(\sqrt{n \log n})$, $O(\log n)$, and $O(1)$ if $2 < \beta < 3$, $\beta = 3$ and $3 < \beta$, respectively ($\beta$ is the power-law exponent). We infer that these graphs have a treewidth of $O(\log n)$, and $O(1)$, respectively. For $2 < \beta < 3$, this matches a known lower bound. For the more realistic (but harder to analyze) binomial model, we still prove a sublinear bound on the treewidth. To demonstrate the usefulness of our results, we apply them to obtain fast matching algorithms and an approximation scheme for Independent Set.


Hyperbolic geometry appears to be intrinsic in many large real networks. We construct and implement a new maximum likelihood estimation algorithm that embeds scale-free graphs in the hyperbolic space. All previous approaches of similar embedding algorithms require a runtime of $\Omega(n^2)$. Our algorithm achieves quasilinear runtime, which makes it the first algorithm that can embed networks with hundreds of thousands of nodes in less than one hour. We demonstrate the performance of our algorithm on artificial and real networks. In all typical metrics like Log-likelihood and greedy routing our algorithm discovers embeddings that are very close to the ground truth.

Chauhan, A., Friedrich, T., Rothenberger, R., Greed is Good for Deterministic Scale-Free Networks. In: Foundations of Software Technology and Theoretical Computer Science (FSTTCS), pp. 33:1–33:15, 2016. Large real-world networks typically follow a power-law degree distribution. To study such networks, numerous random graph models have been proposed. However, real-world networks are not drawn at random. Therefore, Brach, Cygan, Lacki, and Sankowski [SODA 2016] introduced two natural deterministic conditions: (1) a power-law upper bound on the degree distribution (PLB-U) and (2) power-law neighborhoods, that is, the degree distribution of neighbors of each vertex is also upper bounded by a power law (PLB-N). They showed that many real-world networks satisfy both deterministic properties and exploit them to design faster algorithms for a number of classical graph problems. We complement the work of Brach et al. by showing that some well-studied random graph models exhibit both the mentioned PLB properties and additionally also a power-law lower bound on the degree distribution (PLB-L). All three properties hold with high probability for Chung-Lu Random Graphs and Geometric Inhomogeneous Random Graphs and almost surely for Hyperbolic Random Graphs. As a consequence, all results of Brach et al. also hold with high probability or almost surely for those random graph classes. In the second part of this work we study three classical NP-hard combinatorial optimization problems on PLB networks. We show that on general graphs with maximum degree $\Delta$, a greedy algorithm, which chooses nodes in the order of their degree, only achieves a $\Omega(\ln \Delta)$-approximation for Minimum Vertex Cover and Minimum Dominating Set, and a $(\Delta/2)$-approximation for Maximum Independent Set. We prove that the PLB-L property suffices for the greedy approach to achieve a constant-factor approximation for all three problems. We also show that all three combinatorial optimization problems are APX-complete even if all PLB-properties holds hence, PTAS cannot be expected unless $P = \text{NP}$.

Friedrich, T., Kotzing, T., Krejca, M. S., Sutton, A. M., The Benefit of Recombination in Noisy Evolutionary Search. In: Genetic and Evolutionary Computation Conference (GECCO), pp. 161–162, 2016. Practical optimization problems frequently include uncertainty about the quality measure, for example due to noisy evaluations. Thus, they do not allow for a straightforward application of traditional optimization techniques. In these settings, randomized search heuristics such as evolutionary algorithms are a popular choice because they are often assumed to exhibit some kind of resistance to noise. Empirical evidence suggests that some algorithms, such as estimation of distribution algorithms (EDAs) are robust against a scaling of the noise intensity, even without resorting to explicit noise-handling techniques. In this work, we introduce the concept of graceful scaling, in which the run time of an algorithm scales polynomially with noise intensity. We study a monotone fitness function over binary strings with additive noise taken from a Gaussian distribution. We show that myopic heuristics cannot efficiently optimize the function under arbitrarily intense noise without a judiciously designed explicit noise-handling technique. Furthermore, we prove that the common practice of reducing the population size to combat the noise does not help. Finally we show that a simple EDA called the Compact Genetic Algorithm can overcome the shortsightedness of mutation-only heuristics to scale gracefully with noise. We conjecture that recombative genetic algorithms also have this property.

Friedrich, T., Kotzing, T., Quinzan, F., Sutton, A. M., Ant Colony Optimization Beats Resampling on Noisy Functions. In: Genetic and Evolutionary Computation Conference (GECCO), pp. 3–4, 2016. Despite the pervasiveness of noise in real-world optimization, there is little understanding of the interplay between the operators of randomized search heuristics and explicit noise-handling techniques such as statistical resampling. Ant Colony Optimization (ACO) algorithms are claimed to be particularly well-suited to dynamic and noisy problems, even without explicit noise-handling techniques. In this work, we empirically investigate the trade-offs between resampling and the noise-handling abilities of ACO algorithms. Our main focus is to locate the point where resampling costs more than it is worth.

Dang, D.-C., Friedrich, T., Krejca, M. S., Kotzing, T., Lehre, P. K., Oliveto, P. S., Sudholt, D., Sutton, A. M., Escaping Local Optima with Diversity Mechanisms and Crossover. In: Genetic and Evolutionary Computation Conference (GECCO), pp. 645–652, 2016. Population diversity is essential for the effective use of any crossover operator. We compare seven commonly used diversity mechanisms and prove rigorous run time bounds for the $(\mu + 1)$ GA using uniform crossover on the fitness function $Jump_k$. All previous results in this context only hold for unrealistically low crossover probability $p_c = O(k/n)$, while we give analyses for the setting of constant $p_c < 1$ in all but one case. Our bounds show a dependence on the problem size $n$, the jump length $k$, the population size $\mu$, and the crossover probability $p_c$. For the typical case of constant $k > 2$ and constant $p_c$, we can compare the resulting expected optimisation times for different diversity mechanisms. Furthermore, we present a heuristic for computing the optimal choice $p_c$ for maximising the convex hull, $O(n \log n)$ for the det. crowding (assuming $p_c = k/n$), $O(n \log n)$ for maximising the Hamming distance, $O(\log n)$ for fitness sharing, $O(n \log n)$ for the single-receiver island model. This proves a sizeable advantage of all variants of the
Different works have shown how crossover can help with building block assembly. Typically, crossover might get lucky to select good building blocks from each parent, but these lucky choices are usually rare. In this work we consider a crossover operator that works on three parent individuals. In each component, the offspring inherits the values present in the majority of the parents; thus, we call this crossover operator majority vote. We show that, if good components are sufficiently prevalent in the individuals, majority vote creates an optimal individual with high probability. Furthermore, we show that this process can be amplified: as long as components are good independently and with probability at least $\frac{1}{2} + \delta$, we require only $O(\log(1/\delta + \log \log n))$ successive stages of majority vote to create an optimal individual with high probability! We show how this applies in two scenarios. The first scenario is the Jump test function. With sufficient diversity, we get an optimization time of $O(n \log n)$ even for jump sizes as large as $O(n^{11/28} + \epsilon)$. Our second scenario is a family of vertex cover instances. Majority vote optimizes this family efficiently, while local searches fail and only highly specialized two-parent crossovers are successful.


Estimation of Distribution Algorithms (EDAs) work by iteratively updating a distribution over the search space with the help of samples from each iteration. Up to now, theoretical analyses of EDAs are scarce and present run time results for specific EDAs. We propose a new framework for EDAs that captures the idea of several known optimizers, including PBI, UMDA, λ-MMASIB, (1+λ)-EA. Our focus is on analyzing two core features of EDAs: a balanced EDA is sensitive to signals in the fitness; a stable EDA remains uncommitted under a biasless fitness function. We prove that no EDA can be both balanced and stable. The LeadingOnes function is a prime example where, at the beginning of the optimization, the fitness function shows no bias for many bits. Since many well-known EDAs are balanced and thus not stable, they are not well-suited to optimize LeadingOnes. We give a stable EDA which optimizes LeadingOnes within a time of $O(n \log n)$.


We study the parameterized complexity of classical problems that arise in the profiling of relational data. Namely, we characterize the complexity of detecting unique column combinations (candidate keys), functional dependencies, and inclusion dependencies with the solution size as parameter. While the discovery of uniqueness and functional dependencies, respectively, turns out to be W[2]-complete, the detection of inclusion dependencies is one of the first natural problems proven to be complete for the class W[3]. As a side effect, our reductions give insights into the complexity of enumerating all minimal unique column combinations or functional dependencies.


Population diversity is essential for avoiding premature convergence in Genetic Algorithms (GAs) and for the effective use of crossover. Yet the dynamics of how diversity emerges in populations are not well understood. We use rigorous runtime analysis to gain insight into the interplay of the size of the largest collection of identical genotypes we show that the interplay of crossover followed by mutation may serve as a catalyst leading to a sudden burst of diversity. This leads to improvements of the expected optimisation time of order $\Omega(n/\log n)$ compared to mutation-only algorithms like the (1 + 1) EA.


Most theoretical work that studies the benefit of recombination focuses on the ability of crossover to speed up optimization time on specific search problems. In this paper, we take a slightly different perspective and investigate recombination in the context of evolving solutions that exhibit mutational robustness, i.e., they display insensitivity to small perturbations. Various models in population genetics have demonstrated that increasing the effective recombination rate promotes the evolution of robustness. We show this result also holds in the context of evolutionary computation by proving crossover promotes the evolution of robust solutions in the standard $(\mu+1)$ GA. Surprisingly, our results show that the effect is present even when robust solutions are at a selective disadvantage due to lower fitness values.


Recently, different evolutionary algorithms (EAs) have been analyzed in noisy environments. The most frequently used noise model for this was additive posterior noise (noise added after the fitness evaluation) taken from a Gaussian distribution. In particular, for this setting it was shown that the $(\mu+1)$-EA on OneMax does not scale gracefully (higher noise cannot efficiently be compensated by higher $\mu$). In this paper we want to understand whether there is anything special about the Gaussian distribution which makes the $(\mu+1)$-EA not scale gracefully. We keep the setting of posterior noise, but we look at other distributions. We see that for exponential tails the $(\mu+1)$-EA on OneMax does also not scale gracefully, for similar reasons as in the case of Gaussian noise. On the other hand, for uniform distributions (as well as other, similar distributions) we see that the $(\mu+1)$-EA on OneMax does scale gracefully, indicating the importance of the noise model.


We consider how well-known branching approaches for the classical minimum vertex cover problem can be turned into randomized initialization strategies with provable performance guarantees and investigate them by experimental investigations. Furthermore, we show how these techniques can be built into local search components and analyze a basic local search variant that is similar to a state-of-the-art approach called NuMVC. Our experimental results for the two local search approaches show that making use of more complex branching strategies in the local search component can lead to better results on various benchmark graphs.
The majority of empirical comparisons of multi-objective evolutionary algorithms (MOEAs) are performed on synthetic benchmark functions. One of the advantages of synthetic test functions is the a-priori knowledge of the optimal Pareto front. This allows measuring the proximity to the optimal front for the solution sets returned by the different MOEAs. Such a comparison is only meaningful if the cardinality of all solution sets is bounded by some fixed $k$. In order to compare MOEAs to the theoretical optimum achievable with $k$ solutions, we determine best possible $\epsilon$-indicator values achievable with solution sets of size $k$, up to an error of $\delta$. We present a new algorithm with runtime $O(k \cdot \log^2(\delta - 1))$, which is an exponential improvement regarding the dependence on the error $\delta$ compared to all previous work. We show mathematical correctness of our algorithm and determine optimal solution sets for sets of cardinality $k \in \{2, 3, 4, 5, 10, 20, 50, 100, 1000\}$ for the well known test suits DTLZ, ZDT, WFG and LZ09 up to error $\delta = 10^{-25}$.

Recently Ant Colony Optimization (ACO) algorithms have been proven to be efficient in uncertain environments, such as noisy or dynamically changing fitness functions. Most of these analyses focus on combinatorial problems, such as path finding. We analyze an ACO algorithm in a setting where we can work with additive posterior noise sampled from a Gaussian distribution. Without noise the classical ($\mu + 1$)-EA outperforms any ACO algorithm, with smaller $\mu$ being better; however, with large noise, the ($\mu + 1$)-EA fails, even for high values of $\mu$ (which are known to help against small noise). In this paper we show that ACO is able to deal with arbitrarily large noise in a graceful manner, that is, as long as the evaporation factor $\mu$ is small enough dependent on the parameter $\delta^2$ of the noise and the dimension $n$ of the search space ($p = o(1/n(n + \delta \log n)^2 \log n)$), optimization will be successful.

Large real-world networks are typically scale-free. Recent research has shown that such graphs are described best in a geometric space. More precisely, the internet can be mapped to a hyperbolic space such that geometric greedy routing performs close to optimal (Boguna, Papadopoulos, and Krioukov. Nature Communications, 1:62, 2010). This observation pushed the interest in hyperbolic networks as a natural model for scale-free networks. Hyperbolic random graphs follow a power-law degree distribution with controllable exponent $\beta$ and show high clustering (Gugul worn, Papagiotou, and Peter. ICALP, pp. 573-585, 2012). For understanding the structure of the resulting graphs and for analyzing the behavior of network algorithms, the next question is bounding the size of the diameter. The only known explicit bound is $O((\log n)^{2/3})$ (Kiwi and Mitsche. ANALCO, pp. 26-39, 2015). We present two much simpler proofs for an improved upper bound of $O((\log n)^{2/3})$ and a lower bound of $\Omega((\log n))$. If the average degree is bounded from above by some constant, we show that the latter bound is tight by proving an upper bound of $O((\log n))$.

The performance of large distributed systems crucially depends on efficiently balancing their load. This has motivated a large amount of theoretical research how an imbalanced load vector can be smoothed with local algorithms. For technical reasons, the vast majority of previous work focuses on regular (or almost regular) graphs including symmetric topologies such as grids and hypercubes, and ignores the fact that large networks are often highly heterogeneous. We model large scale-free networks by Chung-Lu random graphs and analyze a simple local algorithm for iterative load balancing. On $n$-node graphs our distributed algorithm balances the load within $O((\log log n)^2)$ steps. It does not need to know the exponent $2 < \beta < 3$ of the power-law degree distribution or the weights $w_i$ of the graph model. To the best of our knowledge, this is the first result which shows that load-balancing can be done in double-logarithmic time on realistic graph classes.

Most complex real-world networks display scale-free features. This motivated the study of numerous random graph models with a power-law degree distribution. There is, however, no established and simple model which also has a high clustering of vertices as typically observed in real data. Hyperbolic random graphs bridge this gap. This natural model has recently been introduced in hyperbolic random graphs $G$ and present new results on the expected number of $k$-cliques $E[K_k]$ and the size of the largest clique $\omega(G)$. We observe that there is a phase transition at power-law exponent $\gamma = 3$. More precisely, for $\gamma \in (2, 3)$ we prove $E[K_k] = n^{1-\gamma/2} \Theta(\gamma)$ and $\omega(G) = \Theta(n^{1-\gamma/2})$ while for $\gamma \geq 3$ we prove $E[K_k] = \Theta(k^{-\gamma})$ and $\omega(G) = \Theta((\log n)/(\log \log n))$. We empirically compare the $\omega(G)$-predictions by hyperbolic random graphs to the results of several scale-free random graph models with real-world networks. Our experiments show that the $\omega(G)$-predictions by hyperbolic random graphs are much closer to the data than other scale-free random graph models.
Practical optimization problems frequently include uncertainty about the quality measure, for example due to noisy evaluations. Thus, they do not allow for a straightforward application of traditional optimization techniques. In these settings meta-heuristics are a popular choice for deriving good optimization algorithms, most notably evolutionary algorithms which mimic evolution in nature. While there is ample empirical evidence suggesting that genetic recombination is helpful in uncertain environments because it can stabilize a noisy fitness signal. With this paper we want to support this claim with mathematical rigor. The setting we consider is that of noisy optimization. We study a simple noisy fitness function that is derived by adding Gaussian noise to a monotone function. First, we show that a classical evolutionary algorithm that does not employ sexual recombination (the $(\mu+1)$-EA) cannot handle the noise efficiently, regardless of the population size. Then we show that an evolutionary algorithm which does employ sexual recombination (the Compact Genetic Algorithm, short: cGA) can handle the noise using a graceful scaling of the population.


Random walks are frequently used in randomized algorithms. We study a derandomized variant of a random walk on graphs, called rotor-router model. In this model, instead of distributing tokens randomly, each vertex serves its neighbors in a fixed deterministic order. For most setups, both processes behave remarkably similar: Starting with the same initial configuration, the number of tokens in the rotor-router model deviates only slightly from the expected number of tokens on the corresponding vertex in the random walk model. The maximal difference over all vertices and all times is called single vertex discrepancy. Cooper and Spencer (2006) showed that on $Z^d$ the single vertex discrepancy is only a constant $c_d$. Other authors also determined the precise value of $c_d$ for $d = 1, 2$.

All these results, however, assume that initially all tokens are only placed on one partition of the bipartite graph $Z^d$. We show that this assumption is crucial by proving that otherwise the single vertex discrepancy can become arbitrarily large. For all dimensions $d \geq 1$ and arbitrary discrepancies $\ell \geq 0$, we construct configurations that reach a discrepancy of at least $\ell$.


There are hundreds of online social networks with billions of users in total. Many such networks publicly release structural information, with all personal information removed. Empirical studies have shown, however, that this provides a false sense of privacy - it is possible to identify almost all users that appear in two such anonymized network as long as a few initial mappings are known. We analyze this problem theoretically by reconciling two versions of an artificial power-law network arising from independent subsampling of vertices and edges. We present a new algorithm that identifies most vertices and makes no wrong identifications with high probability.

The number of vertices matched is shown to be asymptotically optimal. For an $n$-vertex graph, our algorithm uses $n^\epsilon$ seed nodes (for an arbitrarily small $\epsilon$) and runs in quasilinear time. This improves previous theoretical results which need $\Theta(n)$ seed nodes and have runtimes of order $n^{1+\Omega(1)}$. Additionally, the applicability of our algorithm is studied experimentally on different networks.


The goal of bi-objective optimization is to find a small set of good compromise solutions. A common problem for bi-objective evolutionary algorithms is the following subset selection problem (SSP): Given $n$ solutions $P \subseteq R^2$ in the objective space, select $k$ solutions $P^*$ from $P$ that optimize an indicator function. In the hypervolume SSP we want to select $k$ points $P^*$ that maximize the hypervolume indicator $I_{hyp}(P^*, r)$ for some reference point $r \in R^2$. Similarly, the $\epsilon$-indicator SSP aims at selecting $k$ points $P^*$ that minimize the $\epsilon$-indicator $I_\epsilon(P^*, R)$ for some reference set $R \subseteq R^2$ of size $m$ (which can be $R = P$). We present a novel algorithm for the hypervolume SSP with runtime $O(n(k + \log n))$. Our second main result is a new algorithm for the $\epsilon$-indicator SSP with runtime $O(n \log n + m \log m)$. Both results improve the current state of the art running times by a factor of (nearly) $n$ and make the problems tractable for new applications. Preliminary experiments confirm that the theoretical results translate into substantial empirical runtime improvements.


Most bi-objective evolutionary algorithms maintain a population of fixed size $\mu$ and return the final population at termination. During the optimization process many solutions are considered, but most are discarded. We present two generic postprocessing algorithms which utilize the archive of all non-dominated solutions evaluated during the search. We choose the best $\mu$ solutions from the archive such that the hypervolume or $\epsilon$-indicator is maximized. This postprocessing costs no additional fitness function evaluations and has negligible runtime compared to most EMOAs. We experimentally examine our postprocessing for four standard algorithms (NSGA-II, SPEA2, SMS-EMOA, IBEA) on ten standard test functions (DTLZ 1–7, ZDT 1–3, WFG 3–6) and measure the average quality improvement. The median decrease of the distance to the optimal $\epsilon$-indicator is 95%, the median decrease of the distance to the optimal hypervolume value is 86%. We observe similar performance on a real-world problem (wind turbine placement).


Many combinatorial optimization problems have underlying goal functions that are submodular. The classical goal is to find a good solution for a given submodular function $f$ under a given set of constraints. In this paper, we investigate the runtime of a multi-objective evolutionary algorithm called GSEMO until it has obtained a good approximation for submodular functions. For the case of monotone submodular functions and uniform cardinality constraints we show that GSEMO achieves a $(1-1/e)$-approximation in expected time $O(n^{\omega} \log n + f)$, where $k$ is the value of the given constraint. For the case of non-monotone submodular functions with $k$ matroid intersection constraints, we show that GSEMO achieves a $(1/(k + 2 + 1/k + \epsilon))$-approximation in expected time $O(n^{k+\epsilon} \log(n)/\epsilon)$.

The Pareto front of a multi-objective optimization problem is typically very large and can only be approximated. Approximation-Guided Evolution (AGE) is a recently presented evolutionary multi-objective optimization algorithm that aims at minimizing iteratively the approximation factor, which measures how well the current population approximates the Pareto front. It outperforms state-of-the-art algorithms for problems with many objectives. However, AGE’s performance is not competitive on problems with very few objectives. We study the reason for this behavior and observe that AGE selects parents uniformly at random, which has a detrimental effect on its performance. We then investigate different algorithm-specific selection strategies for AGE. The main difficulty here is finding a computationally efficient selection scheme which does not harm AGEs linear runtime in the number of objectives. We present several improved selections schemes that are computationally efficient and substantially improve AGE on low-dimensional objective spaces, but have no negative effect in high-dimensional objective spaces.


The hypervolume indicator (HYP) is a popular measure for the quality of a set of n solutions in $\mathbb{R}^d$. We discuss its asymptotic worst-case runtimes and several lower bounds depending on different complexity-theoretic assumptions. Assuming that P ≠ NP, there is no algorithm with runtime $\text{poly}(n,d)$. Assuming the exponential time hypothesis, there is no algorithm with runtime $n^{o(d)}$.

In contrast to these worst-case lower bounds, we study the average-case complexity of HYP for points distributed i.i.d. at random on a d-dimensional simplex. We present a general framework which translates any algorithm for HYP with worst-case runtime $n^{o(d)}$ to an algorithm with worst-case runtime $n^{O(d)}$ and fixed-parameter-tractable (FPT) average-case runtime. This can be used to show that HYP can be solved in expected time $O(d^2n^2 + d^3)$, which implies that HYP is FPT on average while it is W1-hard in the worst-case. For constant dimension d this gives an algorithm for HYP with runtime $O(n^2)$ on average. This is the first result proving that HYP is asymptotically easier in the average case. It gives a theoretical explanation why most HYP algorithms perform much better on average than their theoretical worst-case runtime predicts.


In this paper we study the job scheduling problem on related and unrelated machines so as to minimize the maximum flow time or the maximum weighted flow time (when each job has an associated weight). Previous work for these metrics considered only the setting of parallel machines, while most work for scheduling on unrelated machines only considered $L_p$, $p < \infty$ norms. Our main results are: (1) we give an $O(\epsilon^{-1})$-competitive algorithm to minimize maximum weighted flow time on related machines where we assume that the machines of the online algorithm can process 1 + $\epsilon$ units of a job in 1 time-unit ($\epsilon$ speed augmentation). (2) For the objective of minimizing maximum flow time on unrelated machines we give a simple $2/\epsilon$-competitive algorithm when we augment the speed by $\epsilon$. For $m$ machines we show a lower bound of $\Omega(m)$ on the competitive ratio if speed augmentation is not permitted. Our algorithm does not assign jobs to machines as soon as they arrive. To justify this “drawback” we show a lower bound of $\Omega(\log m)$ on the competitive ratio of immediate dispatch algorithms. In both these lower bound constructions we use jobs whose processing times are in $[1, \infty]$, and hence they apply to the more restrictive subset parallel setting. (3) For the objective of minimizing maximum weighted flow time on unrelated machines we establish a lower bound of $\Omega(\log m)$ on the competitive ratio of any online algorithm whenever we need to use $O(1)$ parallel machines. In our lower bound construction, job j has a processing time of $p_j$ on a subset of machines and infinity on others and has a weight $1/p_j$. Hence this lower bound applies to the subset parallel setting for the special case of minimizing maximum stretch.


The standard algorithm for fast generation of Erdős-Rényi random graphs only works in the Real RAM model. The critical point is the generation of geometric random variates $Geo(p)$, for which there is no algorithm that is both exact and efficient in any bounded precision machine model. For a RAM model with word size $w = \Omega(\log \log (1/p))$, we show that this is possible and present an exact algorithm for sampling $Geo(p)$ in optimal expected time $O(1 + \log(1/p)/w)$. We also give an exact algorithm for sampling $min\{n, Geo(p)\}$ in optimal expected time $O(1 + \log(\min\{1, 1/(npw)\})/w)$. This yields a new exact algorithm for sampling Erdős-Rényi and Chung-Lu random graphs of $n$ vertices and $m$ (expected) edges in optimal expected runtime $O(n + m)$ on a RAM with word size $w = \Theta(\log n)$.


The integrated analysis of data of different types and with various interdependencies is one of the major challenges in computational biology. Recently, we developed KeyPathwayMiner, a method that combines biological networks modeled as graphs with disease-specific genetic expression data gained from a set of cases (patients, cell lines, tissues, etc.). We aimed for finding all maximal connected subgraphs where all nodes are sashays but at most K are expressed in all cases but at most L, i.e. accumulated over all cases and all nodes in a solution. We call this strategy GLONE (global node exceptions). Previous studies assumed the offspring generation to be best-case. We assume the initial population and the offspring generation to be worst-case and use the competitive ratio to measure how much smaller hypervolumes an archiving algorithm finds due to not knowing the future in advance. We prove that all archiving algorithms which increase the hypervolume...
in each step (if they can) are only μ-competitive. We also present a new archiving algorithm which is \((4 + 2/μ)\)-competitive. This algorithm not only achieves a constant competitive ratio, but is also efficiently computable. Both properties provably do not hold for the commonly used greedy archiving algorithms, for example those used in SBEA, SMS-EMOA, or the generational MO-CMA-ES.

Finding cliques in graphs is a classical problem which is in general NP-hard and parameterized intractable. However, in typical applications like social networks or protein-protein interaction networks, the considered graphs are scale-free, i.e., their degree sequence follows a power law. Their specific structure can be algorithmically exploited and makes it possible to solve clique much more efficiently. We prove that on inhomogeneous random graphs with \(n\) nodes and power law exponent \(γ\), cliques of size \(k\) can be found in time \(O(n^2)\) for \(γ \geq 3\) and in time \(O(n \exp(\sqrt{k}^δ))\) for \(2 < γ < 3\).

Randomized rumor spreading was recently shown to be a very efficient mechanism to spread information in preferential attachment networks. Most interesting from the algorithm design point of view was the observation that the asymptotic run-time drops when memory is used to avoid re-contacting neighbors within a small number of rounds. In this experimental investigation, we confirm that a small amount of memory indeed reduces the run-time of the protocol even for small network sizes. We observe that one memory cell per node suffices to reduce the run-time significantly; more memory helps comparably little. Aside from extremely sparse graphs, preferential attachment graphs perform faster than all other graph classes examined. This holds independent of the amount of memory, but preferential attachment graphs benefit the most from the use of memory. We also analyse the influence of the network density and the size of the memory. For the asynchronous version of the rumor spreading protocol, we observe that the theoretically predicted asymptotic advantage of preferential attachment graphs is smaller than expected. There are other topologies which benefit even more from asynchrony. We complement our findings on artificial network models by the corresponding experiments on crawls of popular online social networks, where again we observe extremely rapid information dissemination and a sizable benefit from using memory and asynchrony.

We show that the asynchronous push-pull protocol spreads rumors in preferential attachment graphs (as defined by Barabasi and Albert) in time \(O(\sqrt{\log n})\) to all but a lower order fraction of the nodes with high probability. This is significantly faster than what synchronized protocols can achieve; an obvious lower bound for these is the average distance, which is known to be \(Θ(\log n/\log \log n)\).

We give an algorithm that computes the final state of certain growth models without computing all intermediate states. Our technique is based on a "least action principle" which characterizes the odometer function of the growth process. Starting from an approximation for the odometer, we successively correct under- and overestimates and provably arrive at the correct final state. The degree of speedup depends on the accuracy of the initial guess. Determining the size of the boundary fluctuations in growth models like internal diffusion-limited aggregation (IDLA) is a long-standing open problem in statistical physics. As an application of our method, we calculate the size of fluctuations over two orders of magnitude beyond previous simulations.

Evolutionary algorithms have been widely used to tackle multi-objective optimization problems. Incorporating preference information into the search of evolutionary algorithms for multi-objective optimization is of great importance as it allows one to focus on interesting regions in the objective space. Zitzler et al. have shown how to use a weight distribution function on the objective space to incorporate preference information into hypervolume-based algorithms. We show that this weighted information can easily be used in other popular EMO algorithms as well. Our results for NSGA-II and SPEA2 show that this yields similar results to the hypervolume approach and requires less computational effort.

It was recently proven that sets of points maximizing the hypervolume indicator do not give a good multiplicative approximation of the Pareto front. We introduce a new "logarithmic hypervolume indicator" and prove that it achieves a close-to-optimal multiplicative approximation ratio. This is experimentally verified on several benchmark functions by comparing the approximation quality of the multi-objective covariance matrix evolution strategy (MO-CMA-ES) with the classic hypervolume indicator and the MO-CMA-ES with the logarithmic hypervolume indicator.

The core of hypervolume-based multi-objective evolutionary algorithms is an archiving algorithm which performs the environmental selection. A \((μ + λ)\)-archiving algorithm defines how to choose \(μ\) children from \(μ\) parents and \(λ\) offspring together. We study theoretically \((μ + λ)\)-archiving algorithms which never decrease the hypervolume from one generation to the next. Zitzler, Thiele, and Bader (IEEE Trans. Evolutionary Computation, 14:58–79, 2010) proved that all \((μ + 1)\)-archiving algorithms are ineffective, which means there is an initial population such that independent of the used reproduction rule, a set with maximum hypervolume cannot be reached. We extend this and prove that for \(λ < μ\) all archiving algorithms are ineffective. On the other hand, locally optimal algorithms, which maximize the hypervolume in each step, are effective for \(λ = μ\) and can always find a population with hypervolume at least half the optimum for \(λ \leq μ\). We also prove that there is no hypervolume-based archiving algorithm which can always find a population with hypervolume greater than \(1/(1 + 0.1338, (1/λ - 1/μ))\) times the optimum.

Multi-objective optimization problems arise frequently in applications but can often only be solved approximately by heuristic approaches. Evolutionary algorithms have been widely used to tackle multi-objective problems. These algorithms use different measures to ensure diversity in the objective space but are not guided by a formal notion of approximation. We present a new framework of an evolutionary algorithm for multi-objective optimization that allows to work with a formal notion of approximation. Our experimental results show that our approach outperforms state-of-the-art evolutionary algorithms in terms of the quality of the approximation that is obtained in particular for problems with many objectives.


We examine the complexity of constraint satisfaction problems that consist of a set of AllDiff constraints. Such CSPs naturally model a wide range of real-world and combinatorial problems, like scheduling, frequency allocations, and graph coloring problems. As this problem is known to be NP-complete, we investigate under which further assumptions it becomes tractable. We observe that a crucial property seems to be the convexity of the variable domains and constraints. Our main contribution is an extensive study of the complexity of Multiple AllDiff CSPs for a set of natural parameters, like maximum domain size and maximum size of the constraint scopes. We show that, depending on the parameter, convexity can make the problem tractable even though it is provably intractable in general. Interestingly, the convexity of constraints is the key property in achieving fixed parameter tractability, while the convexity of domains does not usually make the problem easier.


A random geometric graph (RGG) is defined by placing $n$ points uniformly at random in $[0, n^{1/d}]^d$, and joining two points by an edge whenever their Euclidean distance is at most some fixed $r$. We assume that $r$ is larger than the critical value for the emergence of a connected component with $\Omega(n)$ nodes. We show that, with high probability (w.h.p.), for any two connected nodes with a Euclidean distance of $\omega(\log n/r^{d-1})$, their graph distance is only a constant factor larger than their Euclidean distance. This implies that the diameter of the largest connected component is $\Theta(n^{1/d}/r)$ w.h.p. We also prove that the condition on the Euclidean distance above is essentially tight. We also analyze the following randomized broadcast algorithm on RGGs. At the beginning, only one node from the largest connected component of the RGG is informed. Then, in each round, each informed node chooses a neighbor independently and uniformly at random and informs it. We prove that w.h.p. this algorithm informs every node in the largest connected component of an RGG within $\Theta(n^{1/d}/r + \log n)$ rounds.


We present a new randomized diffusion-based algorithm for balancing indivisible tasks (tokens) on a network. Our aim is to minimize the discrepancy between the maximum and minimum load. The algorithm works as follows. Every vertex distributes its tokens as evenly as possible among its neighbors and itself. If this is not possible without splitting some tokens, the vertex redistributes its excess tokens among all its neighbors randomly (without replacement). If this is not possible without splitting some tokens, the vertex reallocates all of its excess tokens by itself. This approach avoids nodes having “negative loads”, but causes additional dependencies for the analysis.


With the prevalence of social networks, it has become increasingly important to understand their features and limitations. It has been observed that information spreads extremely fast in social networks. We study the performance of randomized rumor spreading protocols on graphs in the preferential attachment model. The well-known random phone call model of Karp et al. (FOCS 2000) is a push-pull strategy where in each round, each vertex chooses a random neighbor and exchanges information with it. We prove the following: 1. The push-pull strategy delivers a message to all nodes within $\Theta(\log n)$ rounds with high probability. The best known bound so far was $O(\log^2 n)$. 2. If we slightly modify the protocol so that contacts are chosen uniformly from all neighbors but the one contacted in the previous round, then this time reduces to $\Theta(\log n/\log \log n)$, which is the diameter of the graph. This is the first time that a sublogarithmic broadcast time is proven for a natural setting. Also, this is the first time that avoiding double-contacts reduces the run-time to a smaller order of magnitude.


The rotor router model is a popular deterministic analogue of a random walk on a graph. Instead of moving to a random neighbor, the neighbors are served in a fixed order. We examine how fast this "deterministic random walk" covers all vertices (or all edges). We present general techniques to derive upper bounds for the vertex and edge cover time and derive matching lower bounds for several important graph classes. Depending on the topology, the deterministic random walk can be asymptotically faster, slower or equally fast compared to the classical random walk.

It is known that the performance of multi-objective evolutionary algorithms (MOEAs) in general deteriorates with increasing number of objectives. For few objectives, MOEAs relying on the contributing hypervolume as (second-level) sorting criterion are the methods of choice. However, the computational complexity of calculating the contributing hypervolume prevents the broad application of these powerful MOEAs to objective functions with many objectives. In this study, we employ the approximation within the steady-state MO-CMA-ES and the SMS-EMOA to empirically investigate whether the Monte-Carlo approximation is indeed useful in practice.


Evolutionary multi-objective optimization deals with the task of computing a minimal set of search points according to a given set of objective functions. The task has been made explicit in a recent paper by Zitzler et al. [13]. We take an order-theoretic view on this task and examine how the use of indicator functions can help to direct the search towards Pareto optimal sets. Thereby, we point out that evolutionary algorithms for multi-objective optimization working on the dominance relation of search points have to deal with a cyclic behavior that may lead to worsenings with respect to the Pareto-dominance relation defined on sets. Later on, we point out in which situations well-known binary and unary indicators can help to avoid this cyclic behavior.


In order to allow a comparison of (otherwise incomparable) sets, many evolutionary multiobjective optimizers use indicator functions to guide the search and to evaluate the performance of search algorithms. The most widely used indicator is the hypervolume indicator. It measures the volume of the dominated portion of the objective space. Though the hypervolume indicator is very popular, it has not been shown that maximizing the hypervolume indicator is indeed equivalent to the overall objective of finding a good approximation of the Pareto front. To address this question, we compare the optimal approximation factor achieved by sets maximizing the hypervolume indicator. We bound the optimal approximation factor of \( n \) points by \( 1 + O(1/n) \) for arbitrary Pareto fronts. Furthermore, we prove that the same asymptotic approximation ratio is achieved by sets of \( n \) points that maximize the hypervolume indicator. This shows that the speed of convergence of the approximation ratio achieved by maximizing the hypervolume indicator is asymptotically optimal. This implies that for large values of \( n \), sets maximizing the hypervolume indicator quickly approach the optimal approximation ratio. Moreover, our bounds show that for relatively small values of \( n \), sets maximizing the hypervolume indicator achieve a near-optimal approximation ratio.


The hypervolume indicator is widely used to guide the search and to evaluate the performance of evolutionary multi-objective optimization algorithms. It measures the volume of the dominated portion of the objective space which is considered to give a good approximation of the Pareto front. There is surprisingly little theoretically known about the quality of this approximation. We examine the multiplicative approximation ratio achieved by two-dimensional sets maximizing the hypervolume indicator and prove that it deviates significantly from the optimal approximation ratio. This provable gap is even exponential in the ratio between the largest and the smallest value of the front. We also examine the additive approximation ratio of the hypervolume indicator and prove that it achieves the optimal additive approximation ratio apart from a small factor \( \leq n/(n-2) \), where \( n \) is the size of the population. Hence the hypervolume indicator can be used to achieve a very good additive but not a good multiplicative approximation of a Pareto front.


A Random Geometric Graph (RGG) in two dimensions is constructed by distributing \( n \) nodes independently and uniformly at random in \( [0, \sqrt{n}]^2 \) and creating edges between every pair of nodes having Euclidean distance at most \( r \), for some prescribed \( r \). We analyze the following randomized broadcast algorithm on RGGs. At the beginning, only one node from the largest connected component of the RGG is informed. Then, in each round, each informed node chooses a neighbor independently and uniformly at random and informs it. We prove that with probability \( 1 - O(n^{-1}) \) this algorithm informs every node in the largest connected component of an RGG within \( O(\sqrt{n}/r + \log n) \) rounds. This holds for any value of \( r \) larger than the critical value for the emergence of a connected component with \( \Omega(n) \) nodes. In order to prove this result, we show that for any two nodes sufficiently distant from each other in \( [0, \sqrt{n}]^2 \), the length of the shortest path between them in the RGG, when such a path exists, is only a constant factor larger than the optimum. This result has independent interest and, in particular, gives that the diameter of the largest connected component of an RGG is \( \Theta(\sqrt{n}/r) \), which surprisingly has been an open problem so far.


We propose a simple distributed algorithm for balancing indivisible tokens on graphs. The algorithm is completely deterministic, though it tries to imitate (and enhance) a randomized algorithm by keeping the accumulated rounding errors as small as possible. Our new algorithm, surprisingly, closely approximates the idealized process (where the tokens are divisible) on important network topologies. On \( d \)-dimensional torus graphs with \( n \) nodes it deviates from the idealized process only by an additive constant. In contrast, the randomized rounding approach of Friedrich and Sauerwald [Proceedings of the 41st Annual ACM Symposium on Theory of Computing, 2009, pp. 121-130] can deviate up to \( \Omega(\log(n)) \), and the deterministic algorithm of Rabani, Sinclair, and Wanka [Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science, 1998, pp. 694-705] has a deviation of \( \Omega(n^{1/4}) \). This makes our quasirandom algorithm the first known algorithm for this setting, which is optimal both in time and achieved smoothness. We further show that on the hypercube as well, our algorithm has a smaller deviation from the idealized process than the previous algorithms. To prove these results, we derive several combinatorial and probabilistic results that we believe to be of independent interest. In particular, we show that first-passage probabilities of a random walk on a path with arbitrary weights can be expressed as a convolution of independent geometric probability distributions.

We empirically analyze two versions of the well-known "randomized rumor spreading" protocol to disseminate a piece of information in networks. In the classical model, in each round each informed node informs a random neighbor. At SODA 2008, three of the authors proposed a quasirandom variant. Here, each node has a (cyclic) list of its neighbors. Once informed, it starts at a random position of the list, but from then on informs its neighbors in the order of the list. While for sparse random graphs a better performance of the quasirandom model could be proven, all other results show that, independent of the structure of the lists, the same asymptotic performance guarantees hold as for the classical model. In this work, we compare the two models experimentally. This not only shows that the quasirandom model generally is faster (which was expected, though maybe not to this extent), but also that the runtime is more concentrated around the mean value (which is surprising given that much fewer random bits are used in the quasirandom process). These advantages are also observed in a lossy communication model, where each transmission does not reach its target with a certain probability, and in an asynchronous model, where nodes send at random times drawn from an exponential distribution. We also show that the particular structure of the lists has little influence on the efficiency. In particular, there is no problem if all nodes use an identical order to inform their neighbors.


The hypervolume indicator is an increasingly popular set measure to compare the quality of two Pareto sets. The basic ingredient of most hypervolume indicator based optimization algorithms is the calculation of the hypervolume contribution of single solutions regarding a Pareto set. We show that exact calculation of the hypervolume contribution is #P-hard while its approximation is NP-hard. The same holds for the calculation of the minimal contribution. We also prove that it is NP-hard to decide whether a solution has the least hypervolume contribution. Even deciding whether the contribution of a solution is at most $(1+\epsilon)$ times the minimal contribution is NP-hard. This implies that it is neither possible to efficiently find the least contributing solution (unless P = NP) nor to approximate it (unless NP = BPP). Nevertheless, in the second part of the paper we present a very fast approximation algorithm for this problem. We prove that for arbitrarily given $\delta > 0$ it calculates a set of solutions with contribution at most $(1+\epsilon)$ times the minimal contribution with probability at least $(1-\delta)$. Though it cannot run in polynomial time for all instances, it performs extremely fast on various benchmark datasets. The algorithm solves very large problem instances which are intractable for exact algorithms (e.g., 10000 solutions in 100 dimensions) within a few seconds.


Runtime analysis of evolutionary algorithms has become an important part in the theoretical analysis of randomized search heuristics. The first combinatorial problem where rigorous runtime results have been achieved is the well-known single source shortest path (SSSP) problem. Scharnow, Tinnerfeld and Wegener [PPSN 2002, J. Math. Model. Alg. 2004] proposed a multi-objective approach which solves the problem in expected polynomial time. They also suggest a related single-objective fitness function. However, it was left open whether this does solve the problem efficiently, and, in a broader context, whether multi-objective fitness functions for problems like the SSSP yield more efficient evolutionary algorithms. In this paper, we show that the single objective approach yields an efficient (1+1) EA with runtime bounds very close to those of the multi-objective approach.


Most hypervolume indicator based optimization algorithms like SIBEA [Zitzler et al. 2007], SMS-EMOA [Beume et al. 2007], or MO-CMA-ES [Igel et al. 2007] remove the solution with the smallest loss with respect to the dominated hypervolume from the population. This is usually iterated $\alpha$ times until the size of the population no longer exceeds a fixed size $\mu$. We show that there are populations such that the contributing hypervolume of the $\alpha$ solutions chosen by this greedy selection scheme can be much smaller than the contributing hypervolume of an optimal set of $\lambda$ solutions. Selecting the optimal $\lambda$-set implies calculating $\frac{\lambda}{\mu}$ conventional hypervolume contributions, which has been considered computationally too expensive. We present a multi-objective approach which calculates directly the contribution of every set of $\lambda$ solutions. This gives an additive term of $\frac{\lambda}{\mu}$ in the runtime of the calculation instead of a multiplicative factor of binomial $\frac{n^\lambda}{\mu}$. Given a population of size $n = \mu + \lambda$ our algorithm can calculate a set of $\lambda \geq 1$ solutions with minimal d-dimensional hypervolume contribution in time $O(n^{(d^2+1)/2})$ for $d > 3$. This improves all previously published algorithms by a factor of order $n^{\min((d-2)/2)}$ for $d > 3$. Therefore even if we remove the solutions one by one greedily as usual, we gain a speedup factor of $n$ for all $d > 3$.


Indicator-based algorithms have become a very popular approach to solve multi-objective optimization problems. In this paper, we contribute to the theoretical understanding of algorithms maximizing the hypervolume for a given problem by distributing $\mu$ points on the Pareto front. We examine this common approach with respect to the achieved multiplicative approximation ratio for a given multi-objective problem and relate it to a set of $\mu$ points on the Pareto front that achieves the best possible approximation ratio. For the class of linear fronts and a class of concave fronts, we prove that the hypervolume gives the best possible approximation ratio. In addition, we examine Pareto fronts of different shapes by numerical calculations and show that the approximation computed by the hypervolume may differ from the optimal approximation ratio.


Randomized rumor spreading is an efficient protocol to distribute information in networks. Recently, a quasirandom version has been proposed and proven to work equally well on many graphs and better for sparse random graphs. In this work we show three main results for the quasirandom rumor spreading model. We exhibit a natural expansion property for networks which suffices to make quasirandom rumor spreading inform all nodes of the network in logarithmic time with high probability. This expansion property is satisfied, among others, by many expander graphs, random regular graphs, and Erdős-Rényi random graphs. For all network topologies, we show that if one of the push or pull model works well, so does the other. We also show that quasirandom rumor spreading is robust against transmission failures. If each message sent out gets lost with probability $f$, then the runtime increases only by a factor of $O(1/(1 – f))$. 

35

In a balancing network each processor has an initial collection of unit-size jobs (tokens) and in each round, pairs of processors connected by balancers split their load as evenly as possible. An excess token (if any) is placed according to some predefined rule. As it turns out, this rule crucially affects the performance of the network. In this work we propose a model that studies this effect. We suggest a model bridging the uniformly-random assignment rule, and the arbitrary one (in the spirit of smoothed-analysis). We start with an arbitrary assignment of balancer directions and then flip each assignment with probability α independently. For a large class of balancing networks our result implies that after O(log n) rounds the discrepancy is O((1/2 − α)n log n + log n) with high probability. This matches and generalizes known upper bounds for α = 0 and α = 1/2. We also show that a natural network matches the upper bound for any α.


We consider and analyze a new algorithm for balancing indivisible loads on a distributed network with n processors. The aim is minimizing the discrepancy between the maximum and minimum load. In every time-step paired processors balance their load as evenly as possible. The direction of the excess token is chosen according to a randomized rounding of the participating loads. We prove that in comparison to the corresponding model of Rabani, Sinclair, and Wanka (1998) with arbitrary roundings, the randomization yields an improvement of roughly a square root of the achieved discrepancy in the same number of time-steps on all graphs. For the important case of expanders we can even achieve a constant discrepancy in O(log n log log n) rounds. This is optimal up to log log n factors while the best previous algorithms in this setting either require Ω(log2 n) time or can only achieve a logarithmic discrepancy. This result also demonstrates that with randomized rounding the difference between discrete and continuous load balancing vanishes almost completely.


Representation techniques are important issues when designing successful evolutionary algorithms. Within this field the use of neutrality plays an important role. We examine the use of bit-wise neutrality introduced by Poli and Lopez (2007) from a theoretical point of view and show that this mechanism only enhances mutation-based evolutionary algorithms if not the same number of genotypic bits for each phenotypic bit is used. Using different numbers of genotypic bits for the bits in the phenotype we point out by rigorous runtime analyses that it may reduce the optimization time significantly.


Maintaining diversity is important for the performance of evolutionary algorithms. Diversity-preserving mechanisms can enhance global exploration of the search space and enable crossover to find dissimilar individuals for recombination. We focus on the global exploration capabilities of mutation-based algorithms. Using a simple bimodal test function and rigorous runtime analyses, we compare well-known diversity-preserving mechanisms like deterministic crowding, fitness sharing, and others with a plain algorithm without diversification. We show that diversification is necessary for global exploration, but not all mechanisms succeed in finding both optima efficiently. Our theoretical results are accompanied by additional experiments for different population sizes.


We study the fully-dynamic all pairs shortest path problem for graphs with arbitrary non-negative edge weights. It is known for digraphs that an update of the distance matrix costs O(n2 + ε) worst-case time Thorup, STOC ’05 and O(n2) amortized time Demetrescu and Italiano, J.ACM ’04 where n is the number of vertices. We present the first average-case analysis of the undirected problem. For a random update we show that the expected time per update is bounded by O(nV/3 ε + ε) for all ε > 0.


We consider the computation of the volume of the union of high-dimensional geometric objects. While showing that this problem is #P-hard already for very simple bodies (i.e., axis-parallel boxes), we give a fast FPRAS for all objects where one can: (1) test whether a given point lies inside the object, (2) sample a point uniformly, and (3) calculate the volume of the object in polynomial time. All three oracles can be weak, that is, just approximate. This implies that Klee’s measure problem and the hypervolume indicator can be approximated efficiently even though they are #P-hard and hence cannot be solved exactly in time polynomial in the number of dimensions unless P=NP. Our algorithm also allows to approximate efficiently the volume of the union of convex bodies given by weak membership oracles. For the analogous problem of the intersection of high-dimensional geometric objects we prove #P-hardness for boxes and show that there is no multiplicative polynomial-time approximation for certain boxes unless NP=BPP, but give a simple additive polynomial-time approximation.


Indicator-based methods to tackle multiobjective problems have become popular recently, mainly because they allow to incorporate user preferences into the search explicitly. Multiobjective Evolutionary Algorithms (MOEAs) using the hypervolume indicator in particular showed better performance than classical MOEAs in experimental comparisons. In this paper, the use of indicator-based MOEAs is investigated for the first time from a theoretical point of view. We carry out running time analyses for an evolutionary algorithm with a (μ+1)-selection scheme based on the hypervolume indicator as it is used in most of the recently proposed MOEAs. Our analyses point out two important aspects of the search process. First, we examine how such algorithms can approach the Pareto front. Later on, we point out how they can achieve a good approximation for an exponentially large Pareto front.

It is widely assumed that evolutionary algorithms for multi-objective optimization problems should use certain mechanisms to achieve a good spread over the Pareto front. In this paper, we examine such mechanisms from a theoretical point of view and analyze simple algorithms incorporating the concept of fairness introduced by Laumanns et al. This mechanism tries to balance the number of offspring of all individuals in the current population. We rigorously analyze the runtime behavior of different fairness mechanisms and present showcase examples to point out situations where the right mechanism can speed up the optimization process significantly.


Jim Propp’s rotor router model is a deterministic analogue of a random walk on a graph. Instead of distributing chips randomly, each vertex serves its neighbors in a fixed order. Cooper and Spencer (Comb. Probab. Comput. (2006)) show a remarkable similarity of both models. If an (almost) arbitrary population of chips is placed on the vertices of a grid \( Z^d \) and does a simultaneous walk in the Propp model, then at all times and on each vertex, the number of chips deviates from the expected number the random walk would have gotten there, by at most a constant. This constant is independent of the starting configuration and the order in which each vertex serves its neighbors. This result raises the question if all graphs do have this property. With quite some effort, we are now able to answer this question negatively. For the graph being an infinite \( k \)-ary tree \( (k \geq 3) \), we show that for any deviation \( D \) there is an initial configuration of chips such that after running the Propp model for a certain time there is a vertex with at least \( D \) more chips than expected in the random walk model. However, to achieve a deviation of \( D \) it is necessary that at least \( k^{\Omega(2)} \) vertices contribute by being occupied by a number of chips not divisible by \( k \) in a certain time interval.


We propose and analyse a quasirandom analogue to the classical push model for disseminating information in networks ("randomized rumor spreading"). In the classical model, in each round each informed node chooses a neighbor at random and informs it. Results of Frieze and Grimmett (Discrete Appl. Math. 1985) show that this simple protocol succeeds in spreading a rumor from one node of a complete graph to all others within \( O(\log n) \) rounds. For the network being a hypercube or a random graph \( G(n,p) \) with \( p \geq (1 + \epsilon)(\log n)/n \), also \( O(\log n) \) rounds suffice (Feige, Peleg, Raghavan, and Upfal, Random Struct. Algorithms 1990). In the quasirandom model, we assume that each node has a (cyclic) list of its neighbors. Once informed, it starts at a random position of the list, but from then on informs its neighbors in the order of the list. Surprisingly, irrespective of the orders of the lists, the above mentioned bounds still hold. In addition, we also show a \( O(\log n) \) bound for sparsely connected random graphs \( G(n,p) \) with \( p = (\log n + f(n))/n \), where \( f(n) \to \infty \) and \( f(n) = O(\log \log n) \). Here, the classical model needs \( \Omega(\log(\log n)) \) rounds. Hence the quasirandom model achieves similar or better broadcasting times with a greatly reduced use of random bits.


In recent years a lot of progress has been made in understanding the behavior of evolutionary computation methods for single- and multi-objective problems. Our aim is to analyze the diversity mechanisms that are implicitly used in evolutionary algorithms for multi-objective problems by rigorous runtime analyses. We show that, even if the population size is small, the runtime can be exponential where corresponding single-objective problems are optimized within polynomial time. To illustrate this behavior we analyze a simple plateau function in a first step and extend our result to a class of instances of the well-known SetCover problem.


Hybrid methods are very popular for solving problems from combinatorial optimization. In contrast to this the theoretical understanding of the interplay of different optimization methods is rare. The aim of this paper is to make a first step into the rigorous analysis of such combinations for combinatorial optimization problems. The subject of our analyses is the vertex cover problem for which several approximation algorithms have been proposed. We point out specific instances where solutions can (or cannot) be improved by the search process of a simple evolutionary algorithm in expected polynomial time.


The main aim of randomized search heuristics is to produce good approximations of optimal solutions within a small amount of time. In contrast to numerous experimental results, there are only a few theoretical ones on this subject. We consider the approximation ability of randomized search heuristics for the class of covering problems and compare single-objective and multi-objective models for such problems. For the Vertex-Cover problem, we point out situations where the multi-objective model leads to a fast construction of optimal solutions while in the single-objective case even no good approximation can be achieved within expected polynomial time. Examining the more general Set-Cover problem we show that optimal solutions can be approximated within a factor of \( \log n \), where \( n \) is the problem dimension, using the multi-objective approach while the approximation quality obtainable by the single-objective approach in expected polynomial time may be arbitrarily bad.


It is widely assumed and observed in experiments that the use of diversity mechanisms in evolutionary algorithms may have a great impact on its running time. Up to now there is no rigorous analysis pointing out the use of different mechanisms with respect to the runtime behavior. We consider evolutionary algorithms that differ from each other in the way they ensure diversity and point out situations where the right mechanism is crucial for the success of the algorithm. We consider the most common diversity mechanisms considered in the past and compare them with a technique that diversifies the population with respect to the search points or with respect to function values. Investigating simple plateau functions, we show that using the "right" diversity strategy makes the difference between an exponential and a polynomial runtime.

In this paper, we examine how adding objectives to a given optimization problem affects the computation effort required to generate the set of Pareto-optimal solutions. Experimental studies show that additional objectives may change the runtime behavior of an algorithm drastically. Often it is assumed that more objectives make a problem harder as the number of different trade-offs may increase with the problem dimension. We show that additional objectives, however, may be both beneficial and obstructive depending on the chosen objective. Our results are obtained by rigorous runtime analyses that show the different effects of adding objectives to a well-known plateau-function.


Many applications like pointer analysis and incremental compilation require maintaining a topological ordering of the nodes of a directed acyclic graph (DAG) under dynamic updates. All known algorithms for this problem are either only analyzed for worst-case insertion sequences or only evaluated experimentally on random DAGs. We present the first average-case analysis of online topological ordering algorithms. We prove an expected runtime of $O(n^2\log(n))$ under insertion of the edges of a complete DAG in a random order for the algorithms of Alpern et al. (SODA, 1990), Katriel and Bodlaender (TALG, 2006), and Pearce and Kelly (JEA, 2006). This is much less than the best known worst-case bound $O(n^7.75)$ for this problem.


Deterministic and randomized balancing schemes are used to distribute workload evenly in networks. In this paper, we compare two very general ones: The random walk and the (deterministic) Propp machine. Roughly speaking, we show that on the two-dimensional grid, the Propp machine always has the same number of tokens on a node as does the random walk in expectation, apart from an additive error of less than eight. This constant is independent of the total number of tokens and the runtime of the two processes. However, we also show that it makes a difference whether the Propp machine serves the neighbors in a circular or non-circular order.


We show several ways to round a real matrix to an integer one such that the rounding errors in all rows and columns as well as the whole matrix are less than one. This is a classical problem with applications in many fields, in particular, statistics. We improve earlier solutions of different authors in two ways. For rounding matrices of size $m \times n$, we reduce the runtime from $O((mn)^2)$ to $O(mn \log(mn))$. Second, our roundings also have a rounding error of less than one in all initial intervals of rows and columns. Consequently, arbitrary intervals have an error of at most two. This is particularly useful in the statistics application of controlled rounding. The same result can be obtained via (dependent) randomized rounding. This has the additional advantage that the rounding is unbiased, that is, for all entries $y_{ij}$ of our rounding, we have $E(y_{ij}) = x_{ij}$, where $x_{ij}$ is the corresponding entry of the input matrix.


We present a simple algorithm which maintains the topological order of a directed acyclic graph with $n$ nodes under an online edge insertion sequence in $O(n^{2.75})$ time, independent of the number of edges $m$ inserted. For dense DAGs, this is an improvement over the previous best result of $O(\min(m^{7/2} \log n, m^{7/2} + n^2 \log n))$ by Katriel and Bodlaender. We also provide an empirical comparison of our algorithm with other algorithms for online topological sorting.


We show that any real matrix can be rounded to an integer matrix in such a way that the rounding errors of all row sums are less than one, and the rounding errors of all column sums as well as all sums of consecutive row entries are less than two. Such roundings can be computed in linear time. This extends and improves previous results on rounding sequences and matrices in several directions. It has particular applications in just-in-time scheduling, where balanced schedules on machines with negligible switch over costs are sought after. Here we extend existing results to multiple machines and non-constant production rates.