Publications of Thomas Bläsius

This document lists all peer-reviewed publications of Thomas Bläsius, Chair for Algorithm Engineering, Hasso Plattner Institute, Potsdam, Germany. This listing was automatically generated on October 15, 2020. An up-to-date version is available online at hpi.de/friedrich/docs/publist/blaeusi.pdf.

Journal articles


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.


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. 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.


We study the problem of computing straight-line drawings of non-planar graphs with few crossings. We assume that a crossing-minimization algorithm is applied first, yielding a planarization, i.e., a planar graph with a dummy vertex for each crossing, that fixes the topology of the resulting drawing. We present and evaluate two different approaches for drawing a planarization in such a way that the edges of the input graph are as straight as possible. The first approach is based on the planarity-preserving force-directed algorithm IMPRED [Simonetto et al. Computer Graphics Forum 2011], the second approach, which we call Geometric Planarization Drawing, iteratively moves vertices to their locally optimal positions in the given initial drawing. Our evaluation shows that both approaches significantly improve the initial drawing and that our geometric approach outperforms the force-directed approach. To the best of our knowledge, this is the first paper concerned with the generation of a straight-line drawing that respects an arbitrary planarization.


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.


A simultaneous embedding (with fixed edges) of two graphs $G_1$ and $G_2$ with common graph $G = G_1 \cap G_2$ is a pair of planar drawings of $G_1$ and $G_2$ that coincide on $G$. It is an open question whether there is a polynomial-time algorithm that decides whether two graphs admit a simultaneous embedding (problem SEFE). In this paper, we present two results. First, a set of three linear-time preprocessing algorithms that remove certain substructures from a given SEFE instance, producing a set of equivalent SEFE instances without such substructures. The structures we can remove are (1) cutvertices of the union graph $G^1 = G^1 \cup G^2$, (2) most separating pairs of $G^i$, and (3) connected components of $G$ that are biconnected but not a cycle. Second, we give an $O(n^3)$-time algorithm solving SEFE for instances with the following restriction. Let $u$ be a pole of a P-node $\mu$ in the SPQR-tree of a block of $G_1$ or $G_2$. Then at most three virtual edges of $\mu$ may contain common edges incident to $u$. All algorithms extend to the sunflower case, i.e., to the case of more than two graphs pairwise intersecting in the same common graph.
is the smallest integer d such that H is the intersection graph of axis-aligned boxes in \( \mathbb{R}^d \). These intersection representations can be interpreted as covering representations of the complement \( \overline{H} \) of H with co-interval graphs, that is, complements of interval graphs. We follow the recent framework of global, local and folded covering numbers (Knauer and Ueckert, 2016) to define two new parameters: the local boxicity \( box_l(H) \) and the union boxicity \( box_u(H) \) of H. The union boxicity of H is the smallest d such that \( \overline{H} \) can be covered with d vertex-disjoint unions of co-interval graphs, while the local boxicity of H is the smallest d such that \( \overline{H} \) can be covered with d vertex-disjoint unions of co-interval graphs, at most d at every vertex. We show that for every graph H we have \( box_u(H) \leq box_l(H) \leq box(H) \) and that each of these inequalities can be arbitrarily far apart. Moreover, we show that local and union boxicity are also characterized by intersection representations of appropriately partitioned vertices in \( \mathbb{R}^d \). We demonstrate with a few striking examples, that in a sense, the local boxicity is a better indication for the complexity of a graph, than the classical boxicity.

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 Krioukov et al. and has shown theoretically and empirically to fulfill all typical properties of real world networks, including power-law degree distribution and high clustering. We study cliques 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 \( \beta = 3 \). More precisely, for \( \beta \in (2, 3) \) we prove \( E[K_1] = n^{\Theta(1/3-\beta/2)}/\Theta(k^{-\beta}) \) and \( \omega(G) = \Theta(n^{(3-\beta)/2}) \), while for \( \beta \geq 3 \) we prove \( E[K_1] = n/\Theta(k)^{2-\beta} \) 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 \( O(n^{\log(n)}) \). If the underlying geometry is known, cliques can be found with worst-case runtime \( O(n^{\log(n)}) \) for all values of \( \beta \).

The clustered planarity problem (c-planarity) asks whether a hierarchically clustered graph admits a planar drawing such that the clusters can be nicely represented by regions. We introduce the cd-tree data structure and give a new characterization of c-planarity. It leads to efficient algorithms for c-planarity testing in the following cases. (i) Every cluster and every co-cluster (complement of a cluster) has at most two connected components. (ii) Every cluster has at most five outgoing edges. Moreover, the cd-tree reveals interesting connections between c-planarity and planarity with constraints on the order of edges around vertices. On one hand, this gives rise to a bunch of new open problems related to c-planarity, on the other hand it provides a new perspective on previous results.

Traditionally, the quality of orthogonal planar drawings is quantified by the total number of bends or the maximum number of bends per edge. However, this neglects that, in typical applications, edges have varying importance. We consider the problem OptimalFlexDraw that is defined as follows. Given a planar graph \( G \) on \( n \) vertices with maximum degree 4 (4-planar graph) and for each edge \( e \) a cost function \( cost_e : \mathbb{N}_0 \rightarrow \mathbb{R} \) defining costs depending on the number of bends \( e \) has, compute a planar orthogonal drawing of \( G \) of minimum cost. In this generality OptimalFlexDraw is NP-hard. We show that it can be solved efficiently if (1) the cost function of each edge is convex and (2) the first bend on each edge does not cause any cost. Our algorithm takes time \( O(n \cdot T_{\text{flow}}(n)) \) for biconnected and connected graphs, respectively, where \( T_{\text{flow}}(n) \) denotes the time to compute a minimum-cost flow in a planar network with multiple sources and sinks. Our result is the first polynomial-time bend-optimization algorithm for general 4-planar graphs optimizing over all embeddings. Previous work considers restricted graph classes and unit costs.

We consider the problem of creating plane orthogonal drawings of 4-planar graphs (planar graphs with maximum degree 4) with constraints on the number of bends per edge. More precisely, we have a flexibility function assigning to each edge \( e \) a natural number \( flex(e) \), its flexibility. The problem FlexDraw asks whether there exists an orthogonal drawing such that each edge \( e \) has at most \( flex(e) \) bends. It is known that FlexDraw is NP-hard if \( flex(e) = 0 \) for every edge \( e \). On the other hand, FlexDraw can be solved efficiently if \( flex(e) \geq 1 \) and is trivial if \( flex(e) \geq 2 \) for every edge \( e \). To close the gap between the NP-hardness for \( flex(e) = 0 \) and the efficient algorithm for \( flex(e) \geq 1 \), we investigate the computational complexity of FlexDraw in case only few edges are inflexible (i.e., have flexibility 0). We show that for any > 0 FlexDraw is NP-complete for instances with \( O(n^\epsilon) \) inflexible edges with pairwise distance \( O(n^{\epsilon-1}) \) (including the case where they induce a matching). On the other hand, we give an FPT-algorithm with running time \( O(2^{\Delta^2}n \cdot T_{\text{flow}}(n)) \) where \( T_{\text{flow}}(n) \) is the time necessary to compute a maximum flow in a planar flow network with multiple sources and sinks, and \( k \) is the number of inflexible edges having at least one endpoint of degree 4.

In this article, we define and study the new problem of SIMULTANEOUS PQ-ORDERING. Its input consists of a set of PQ-trees, which represent series of circular orders of their leaves, together with a set of child-parent relations between these PQ-trees, such that the leaves of the child form a subset of the leaves of the parent. SIMULTANEOUS PQ-ORDERING asks whether orders of the leaves of each of the trees can be chosen simultaneously; that is, for every child-parent relation, the order chosen for the parent is an extension of the order chosen for the child. We show that SIMULTANEOUS PQ-ORDERING is NP-complete in general, and we identify a family of instances that can be solved efficiently, the 2-fixed instances. We show that this result serves as a framework for several other problems that can be formulated as instances of SIMULTANEOUS PQ-ORDERING. In particular, we give linear-time algorithms for recognizing simultaneous interval graphs and extending partial interval representations. Moreover, we obtain a linear-time algorithm for PARTIALLY PQ-CONSTRAINED PLANARITY for biconnected graphs, which asks for a planar embedding in the presence of 16 PQ-trees that restrict the possible orderings of edges around vertices, and a quadratic-time algorithm for SIMULTANEOUS EMBEDDING WITH FIXED EDGES for biconnected graphs with a connected intersection. Both results can be extended to the case where the input graphs are not necessarily biconnected but have the property that each cutvertex is contained in at most two nontrivial blocks. This includes, for example, the case where both graphs have a maximum degree of 5.


For two planar graphs \( G_1 = (V_1, E_1) \) and \( G_2 = (V_2, E_2) \) sharing a common subgraph \( G = G_1 \cap G_2 \) the problem Simultaneous Embedding with Fixed Edges (SEFE) asks whether they admit planar drawings such that the common graph is drawn the same. Previous algorithms only work for cases where \( G \) is connected, and hence do not need to handle relative positions of connected components. We consider the problem where \( G, G_1, \) and \( G_2 \) are not necessarily connected. First, we show that a general instance of SEFE can be reduced in linear time to an equivalent instance where \( V_1 = V_2 \) and \( G_1 \) and \( G_2 \) are connected. Second, for the case where \( G \) consists of disjoint cycles, we introduce the CC-tree which represents all embeddings of \( G \) that extend to planar embeddings of \( G \). We show that CC-trees can be computed in linear time, and that their intersection is again a CC-tree. This yields a linear-time algorithm for SEFE if all \( k \) input graphs (possibly \( k \geq 2 \)) pairwise share the same set of disjoint cycles. These results, including the CC-tree, extend to the case where \( G \) consists of arbitrary connected components, each with a fixed planar embedding on the sphere. Then the running time is \( O(n^2) \).


We introduce and study the problem Mutual Planar Duality, which asks for planar graphs \( G_1 \) and \( G_2 \) whether \( G_1 \) can be embedded such that its dual is isomorphic to \( G_2 \). We show NP-completeness for general graphs and give a linear-time algorithm for biconnected graphs. We consider the common dual relation, where \( G_1 \) and \( G_2 \) if and only if they admit embeddings that result in the same dual graph. We show that this is an equivalence relation on the set of biconnected graphs and devise a succinct, SPQR-tree-like representation of its equivalence classes. To solve Mutual Planar Duality for biconnected graphs, we show how to do isomorphism testing for two such representations in linear time. A special case of Mutual Planar Duality is testing whether a graph is self-dual. Our algorithm can handle the case of biconnected graphs in linear time and our NP-hardness proof extends to self-duality and also to map self-duality testing (which additionally requires to preserve the embedding).


Traditionally, the quality of orthogonal planar drawings is quantified by either the total number of bends, or the maximum number of bends per edge. However, this neglects that in typical applications, edges have varying importance. In this work, we investigate an approach that allows to specify the maximum number of bends for each edge individually, depending on its importance. We consider constraints. In: Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS), 2020.
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 introduce 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 devise algorithms for different variants of SAP, using multi-criteria shortest paths for algorithms as subroutines. Moreover, we prove that several natural models are in fact Pareto-conformal. 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 investigate the maximum-entropy model $B_{n, \alpha, p}$ for random $n$-vertex, $m$-edge hypergraphs with expected edge size $pn$. We show that the expected size of the minimization $\min(B_{n, \alpha, p})$, i.e., the number of inclusion-wise minimal edges of $B_{n, \alpha, p}$, undergoes a phase transition with respect to $m$. If $m$ is at most $1/(1-(1-p)^{-\alpha}n^n)$, then $\mathbb{E} \lceil \min(B_{n, \alpha, p}) \rceil$ is of order $\Theta (m)$, while for $m \geq 1/(1-p)^{1-p(1+\varepsilon)n}$ for any $\varepsilon > 0$, it is $\Theta (2 \ln (1-p) / \sqrt{m})$. Here, $\mathbb{H}$ denotes the binary entropy function and $\alpha = - \left( \log_2 -1 \right) / m$.


The VertexCover problem is proven to be computationally hard in different ways: It is NP-complete to find an optimal solution and even 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. 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 $O(m^k + n^2)$ 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 parameterised by the cardinality of the set which is to be included. 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.


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 underlying 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. Our resulting implementation is the fastest sequential HRG generator, despite the fact that we support non-zero temperatures. Though non-zero temperatures are crucial for many applications, most existing generators are restricted to $T = 0$. We also support parallelization, although this is not the focus of this paper. Moreover, we note that our generators draw from the correct probability distribution, i.e., they involve no approximation. Besides the generators themselves, we also provide an efficient algorithm to find the exponential non-trivial time of the random generator. The results pave the way for the analysis of the runtime of the HRG model. This makes it possible to specify $d$ as input and obtain a graph with expected average degree $d$. Moreover, we investigate the differences between HRGs and GIRGs, shedding new light on the nature of the relation between the two models. Although HRGs

4
represent, in a certain sense, a special case of the GIRG model, we find that a straightforward inclusion does not hold in practice. However, the difference is negligible for most use cases.


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 complete SAT instances, this phenomenon is inverted: backtracking solvers can tackle large industrial problems, while 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.


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, which is 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 length 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.


A common way to accelerate shortest path algorithms on graphs is to use 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 $O(n^{2−1/\alpha+n^{1/(2\alpha)}+\delta_{max}})$ with high probability, where $\alpha \in (0, 1)$ controls the power-law exponent of the degree distribution, and $\delta_{max}$ 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.


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 $A^*$.\[27\]


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 study the problem of computing straight-line drawings of non-planar graphs with few crossings. We assume that a crossing-minimization algorithm is applied first, yielding a planarization, i.e., a planar graph with a dummy vertex for each crossing, that fixes the topology of the resulting drawing. We present and evaluate two different approaches for drawing a planarization in such a way that the edges of the input graph are as straight as possible. The first approach is based on the planarity-preserving force-directed algorithm ImPrEd, the second approach, which we call Geometric Planarization Drawing, iteratively moves vertices to their locally optimal positions in the given initial drawing.


We present TrussFab, an integrated end-to-end system that allows users to fabricate large scale structures that are sturdy enough to carry human weight. TrussFab achieves the large scale by complementing 3D print with plastic bottles. It does not use these bottles as "bricks" though, but as beams that form structurally sound node-link structures, also known as trusses, allowing it to handle the forces resulting from scale and load. TrussFab embodies the required engineering knowledge, allowing non-engineers to design such structures and to validate their design using integrated structural analysis. We have used TrussFab to design and fabricate tables and chairs, a 2.5 m long bridge strong enough to carry a human, a functional boat that seats two, and a 5 m diameter dome.


Isocontours in road networks represent the area that is reachable from a source within a given resource limit. We study the problem of computing accurate isocontours in realistic, large-scale networks. We propose isocontours represented by polygons with minimum number of segments that separate reachable and unreachable components of the network. Since the resulting problem is not known to be solvable in polynomial time, we introduce several heuristics that run in (almost) linear time and are simple enough to be implemented in practice. A key ingredient is a new practical linear-time algorithm for minimum-link paths in simple polygons. Experiments in a challenging realistic setting show excellent performance of our algorithms in practice, computing near-optimal solutions in a few milliseconds on average, even for long ranges.


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 $O(\sqrt{n^{(3-\beta)}})$, $O(\log n)$, 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 whp a treewidth of $O(\sqrt{n^{(3-\beta)}})$, $O(\log^2 n)$, and $O(\log n)$, respectively. For $2 < \beta < 3$, this matches a known lower bound.


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 $O(n^2)$. 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.


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 unique 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.

We study contact representations for graphs, which we call pixel representations in 2D and voxel representations in 3D. Our representations are based on the unit square grid whose cells we call pixels in 2D and voxels in 3D. Two pixels are adjacent if they share an edge, two voxels if they share a face. We call a connected set of pixels or voxels a blob. Given a graph, we represent its vertices by disjoint blobs such that two blobs contain adjacent pixels or voxels if and only if the corresponding vertices are adjacent. We are interested in the size of a representation, which is the number of pixels or voxels it consists of. We first show that finding minimum-size representations is NP-complete. Then, we bound representation sizes needed for certain graph classes. In 2D, we show that, for k-unipolar graphs with n vertices, Θ(n) pixels are always sufficient and sometimes necessary. In particular, outerplanar graphs can be represented with a linear number of pixels, whereas general planar graphs sometimes need a quadratic number. In 3D, Θ(nT(n, 1)) voxels are always sufficient and sometimes necessary for any n-vertex graph. We improve this bound to Θ(n) for graphs of treewidth r and to O((g + 1)^2n log^2 n) for graphs of genus g. In particular, planar graphs admit representations with O(n log^2 n) voxels.


We consider the problem of creating plane orthogonal drawings of 4-planar graphs (planar graphs with maximum degree 4) with constraints on the number of bends per edge. More precisely, we have a flexibility function assigning to each edge e a natural number flex(e), its flexibility. The problem FlexDraw asks whether there exists an orthogonal drawing such that each edge e has at most flex(e) bends. It is known that FlexDraw is NP-hard if flex(e) = 0 for every edge e [7]. On the other hand, FlexDraw can be solved efficiently if flex(e) ≥ 2 [1] for every edge e. To close the gap between the NP-hardness for flex(e) = 0 and the efficient algorithm for flex(e) ≥ 1, we investigate the computational complexity of FlexDraw in case only few edges are inflexible (i.e., have flexibility 0). We show that for any ε > 0 FlexDraw is NP-complete for instances with O(n^ε) inflexible edges with pairwise distance Ω(n^{1−ε}) (including the case where they induce a matching). On the other hand, we give an FPT-algorithm with running time O(2^k n · T_{flow}(n)), where T_{flow}(n) is the time necessary to compute a maximum flow in a planar flow network with multiple sources and sinks, and k is the number of inflexible edges having at least one endpoint of degree 4.


We show that finding orthogonal grid drawings of plane graphs (planar with fixed combinatorial embedding) with the minimum number of bends in the so-called Kandinsky model (which allows vertices of degree 4) is NP-complete, thus solving a long-standing open problem. On the positive side, we give an efficient algorithm for several restricted variants, such as graphs of bounded branch width and a subexponential exact algorithm for general plane graphs.


The clustered planarity problem (c-planarity) asks whether a hierarchically clustered graph admits a planar drawing such that the clusters can be nicely represented by regions. We introduce the cd-tree data structure and give a new characterization of c-planarity. It leads to efficient algorithms for c-planarity testing in the following cases. (i) Every cluster and every co-cluster has at most two connected components. (ii) Every cluster has at most five outgoing edges. Moreover, the latter reveals interesting connections between c-planarity and planarity with constraints on the order of edges around vertices. On one hand, this gives rise to a bunch of new open problems related to c-planarity, on the other hand it provides a new perspective on previous results.


We present a simple and versatile formulation of grid-based graph representation problems as an integer linear program (ILP) and a corresponding SAT instance. In a grid-based representation vertices and edges correspond to axis-parallel boxes on an underlying integer grid; boxes can be further constrained in their shapes and interactions by additional problem-specific constraints. We describe a general d-dimensional model for grid representation problems. This model can be used to solve a variety of NP-hard graph problems, including pathwidth, bandwidth, optimum st-orientation, area-minimal (bar-k) visibility representation, boxicity-k graphs and others. We implemented SAT-models for all of the above problems and evaluated them on the Rome graphs collection. The experiments show that our model successfully solves NP-hard problems within few minutes on small to medium-size Rome graphs.


A simultaneous embedding (with fixed edges) of two graphs G₁ and G₂ with common graph G = G₁ ∩ G₂ is a pair of planar drawings of G₁ and G₂ that coincide on G. It is an open question whether there is a polynomial-time algorithm that decides whether two graphs admit a simultaneous embedding (problem Sefe). In this paper, we present two results. First, a set of three linear-time preprocessing algorithms that remove certain substructures from a given Sefe instance, producing a set of equivalent Sefe instances without such substructures. The structures we can remove are (1) cutvertices of the union graph (2) most separating pairs of connected components. (ii) Every cluster and every co-cluster has at most two connected components. (iii) Every cluster has at most five outgoing edges. Moreover, the latter reveals interesting connections between c-planarity and planarity with constraints on the order of edges around vertices. On one hand, this gives rise to a bunch of new open problems related to c-planarity, on the other hand it provides a new perspective on previous results.


Traditionally, the quality of orthogonal planar drawings is quantified by the total number of bends or the maximum number of bends per edge. However, this neglects that, in typical applications, edges have varying importance. We consider the problem OptimalFlexDraw that is defined as follows. Given a planar graph G on n vertices with maximum degree 4 (4-planar graph) and for each edge e a cost function cost_e: N₀ → R defining costs depending on the number of bends e has, compute a planar orthogonal drawing of G of minimum cost. In this generalization of the classical FlexDraw problem we show that it can be solved efficiently if (1) the cost function of each edge is convex and (2) the first bend on each edge does not cause any cost. Our algorithm takes time O(n · T_{flow}(n)) and O(n^2 · T_{flow}(n)) for biconnected and connected graphs, respectively, where T_{flow}(n) denotes the time to compute a minimum-cost flow in a planar network with multiple sources and sinks. Our result is the first polynomial-time bend-optimization algorithm for general 4-planar graphs optimizing over all embeddings. Previous work considers restricted graph classes and unit costs.
We introduce and study the problem Mutual Planar Duality, which asks for planar graphs $G_1$ and $G_2$ whether $G_1$ can be embedded such that its dual is isomorphic to $G_2$. We show NP-completeness for general graphs and give a linear-time algorithm for biconnected graphs. We consider the common dual relation $\sim$, where $G_1 \sim G_2$ if and only they admit embeddings that result in the same dual graph. We show that is an equivalence relation on the set of biconnected graphs and devise a succinct, SPQR-tree-like representation of its equivalence classes. To solve Mutual Planar Duality for biconnected graphs, we show how to do isomorphism testing for two such representations in linear time. A special case of Mutual Planar Duality is testing whether a graph is self-dual. Our algorithm can handle the case of biconnected graphs in linear time and our NP-hardness proof extends to self-duality and also to map self-duality testing (which additionally requires to preserve the embedding).

In this article, we define and study the new problem of SIMULTANEOUS PQ-ORDERING. Its input consists of a set of PQ-trees, which represent sets of circular orders of their leaves, together with a set of child-parent relations between these PQ-trees, such that the leaves of the child form a subset of the leaves of the parent. SIMULTANEOUS PQ-ORDERING asks whether orders of the leaves of each of the trees can be chosen simultaneously; that is, for every child-parent relation, the order chosen for the parent is an extension of the order chosen for the child. We show that SIMULTANEOUS PQ-ORDERING is NP-complete in general, and we identify a family of instances that can be solved efficiently, the 2-fixed instances. We show that this result serves as a framework for several other problems that can be formulated as instances of SIMULTANEOUS PQ-ORDERING. In particular, we give linear-time algorithms for recognizing simultaneous interval graphs and extending partial interval representations. Moreover, we obtain a linear-time algorithm for PARTIALLY PQ-CONSTRAINED PLANARITY for biconnected graphs, which asks for a planar embedding in the presence of 16 PQ-trees that restrict the possible orderings of edges around vertices, and a quadratic-time algorithm for SIMULTANEOUS EMBEDDING WITH FIXED EDGES for biconnected graphs with a connected intersection. Both results can be extended to the case where the input graphs are not necessarily biconnected but have the property that each cutvertex is contained in at most two nontrivial blocks. This includes, for example, the case where both graphs have a maximum degree of 5.

For two planar graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ sharing a common subgraph $G = G_1 \cap G_2$ the problem Simultaneous Embedding with Fixed Edges (SEFE) asks whether they admit planar drawings such that the common graph is drawn the same. Previous algorithms only work for cases where $G$ is connected, and hence do not need to handle relative positions of connected components. We consider the problem where $G, G_1$ and $G_2$ are not necessarily connected. First, we show that a general instance of SEFE can be reduced in linear time to an equivalent instance where $V_1 = V_2$ and $G_1$ and $G_2$ are connected. Second, for the case where $G$ consists of disjoint cycles, we introduce the CC-tree which represents all embeddings of $G$ that extend to planar embeddings of $G^1$. We show that CC-trees can be computed in linear time, and that their intersection is again a CC-tree. This yields a linear-time algorithm for SEFE if all $k$ input graphs (possibly $k > 2$) pairwise share the same set of disjoint cycles. These results, including the CC-tree, extend to the case where $G$ consists of arbitrary connected components, each with a fixed planar embedding on the sphere. Then the running time is $O(n^2)$. 

In this article, we define and study the new problem of SIMULTANEOUS PQ-ORDERING. Its input consists of a set of PQ-trees, which represent sets of circular orders of their leaves, together with a set of child-parent relations between these PQ-trees, such that the leaves of the child form a subset of the leaves of the parent. SIMULTANEOUS PQ-ORDERING asks whether orders of the leaves of each of the trees can be chosen simultaneously; that is, for every child-parent relation, the order chosen for the parent is an extension of the order chosen for the child. We show that SIMULTANEOUS PQ-ORDERING is NP-complete in general, and we identify a family of instances that can be solved efficiently, the 2-fixed instances. We show that this result serves as a framework for several other problems that can be formulated as instances of SIMULTANEOUS PQ-ORDERING. In particular, we give linear-time algorithms for recognizing simultaneous interval graphs and extending partial interval representations. Moreover, we obtain a linear-time algorithm for PARTIALLY PQ-CONSTRAINED PLANARITY for biconnected graphs, which asks for a planar embedding in the presence of 16 PQ-trees that restrict the possible orderings of edges around vertices, and a quadratic-time algorithm for SIMULTANEOUS EMBEDDING WITH FIXED EDGES for biconnected graphs with a connected intersection. Both results can be extended to the case where the input graphs are not necessarily biconnected but have the property that each cutvertex is contained in at most two nontrivial blocks. This includes, for example, the case where both graphs have a maximum degree of 5.

For two planar graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ sharing a common subgraph $G = G_1 \cap G_2$ the problem Simultaneous Embedding with Fixed Edges (SEFE) asks whether they admit planar drawings such that the common graph is drawn the same. Previous algorithms only work for cases where $G$ is connected, and hence do not need to handle relative positions of connected components. We consider the problem where $G, G_1$ and $G_2$ are not necessarily connected. First, we show that a general instance of SEFE can be reduced in linear time to an equivalent instance where $V_1 = V_2$ and $G_1$ and $G_2$ are connected. Second, for the case where $G$ consists of disjoint cycles, we introduce the CC-tree which represents all embeddings of $G$ that extend to planar embeddings of $G^1$. We show that CC-trees can be computed in linear time, and that their intersection is again a CC-tree. This yields a linear-time algorithm for SEFE if all $k$ input graphs (possibly $k > 2$) pairwise share the same set of disjoint cycles. These results, including the CC-tree, extend to the case where $G$ consists of arbitrary connected components, each with a fixed planar embedding on the sphere. Then the running time is $O(n^2)$.