Publications of Philipp Fischbeck

This document lists all peer-reviewed publications of Philipp Fischbeck, Chair for Algorithm Engineering, Hasso Plattner Institute, Potsdam, Germany. This listing was automatically generated on March 12, 2024. An up-to-date version is available online at hpi.de/friedrich/docs/publist/fischbeck.pdf.

Journal articles

[1] Bläsius, T., Fischbeck, P., On the External Validity of Average-Case Analyses of Graph Algorithms. In: *ACM Transactions on Algorithms* 20, 2024.

The number one criticism of average-case analysis is that we do not actually know the probability distribution of real-world inputs. Thus, analyzing an algorithm on some random model has no implications for practical performance. At its core, this criticism doubts the existence of external validity, i.e., it assumes that algorithmic behavior on the somewhat simple and clean models does not translate beyond the models to practical performance real-world input. With this paper, we provide a first step towards studying the question of external validity systematically. To this end, we evaluate the performance of six graph algorithms on a collection of 2740 sparse real-world networks depending on two properties; the heterogeneity (variance in the degree distribution) and locality (tendency of edges to connect vertices that are already close). We compare this with the performance on generated networks with varying locality and heterogeneity. We find that the performance in the idealized setting of network models translates surprisingly well to real-world networks. Moreover, heterogeneity and locality appear to be the core properties impacting the performance of many graph algorithms.

[2] Böther, M., Schiller, L., Fischbeck, P., Molitor, L., Krejca, M. S., Friedrich, T., Evolutionary Minimization of Traffic Congestion. In: *IEEE Transactions on Evolutionary Computation* 27, pp. 1809–1821, 2023.

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 destination and 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 Berlin, Germany. We find that a combination of all operators yields the best result, reducing 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.

[3] Bläsius, T., Fischbeck, P., Friedrich, T., Katzmann, M., Solving Vertex Cover in Polynomial Time on Hyperbolic Random Graphs. In: *Theory of Computing Systems* 67, pp. 28–51, 2023.

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.

[4] Casel, K., Fischbeck, P., Friedrich, T., Göbel, A., Lagodzinski, J. A. G., Zeros and approximations of Holant polynomials on the complex plane. In: *Computational Complexity* 31, pp. 11, 2022.

We present fully polynomial time approximation schemesfor 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 mostsignificant 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.

[5] Doerr, B., Fischbeck, P., Frahnow, C., Friedrich, T., Kötzing, T., Schirneck, M., Island Models Meet Rumor Spreading. In: *Algorithmica* 81, pp. 886–915, 2019.

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.

Conference papers

[6] Cohen, S., Fischbeck, P., Friedrich, T., Krejca, M. S., The Common-Neighbors Metric is Noise-Robust and Reveals Substructures of Real-World Networks. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pp. 67–79, 2023.

Real-world networks typically display a complex structure that is hard to explain by a single model. A common approach is to partition the edges of the network into disjoint simpler structures. An important property in this context is locality—incident vertices usually have many common neighbors. This allows to classify edges into two groups, based on the number of the common neighbors of their incident vertices. Formally, this is captured by the common-neighbors (CN) metric, which forms the basis of many metrics for detecting outlier edges. Such outliers can be interpreted as noise or as a substructure. We aim to understand how useful the metric is, and empirically analyze several scenarios. We randomly insert outlier edges into real-world and generated graphs with high locality, and measure the metric accuracy for partitioning the combined edges. In addition, we use the metric to decompose real-world networks, and measure properties of the partitions. Our results show that the CN metric is a very good classifier that can reliably detect noise up to extreme levels (83% noisy edges). We also provide mathematically rigorous analyses on special random-graph models. Last, we find the CN metric consistently decomposes real-world networks into two graphs with very different structures.

[7] Khomutovskiy, I., Dunker, R., Dierking, J., Egbert, J., Helms, C., Schöllkopf, F., Casel, K., Fischbeck, P., Friedrich, T., Isaac, D., Krogmann, S., Lenzner, P., Applying Skeletons to Speed Up the Arc-Flags Routing Algorithm. In: *SIAM Symposium on Algorithm Engineering and Experiments (ALENEX)*, pp. 110–122, 2023.

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.

[8] Bläsius, T., Fischbeck, P., On the External Validity of Average-Case Analyses of Graph Algorithms. In: *European Symposium on Algorithms (ESA)*, pp. 21:1–21:14, 2022.

The number one criticism of average-case analysis is that we do not actually know the probability distribution of real-world inputs. Thus, analyzing an algorithm on some random model has no implications for practical performance. At its core, this criticism doubts the existence of external validity, i.e., it assumes that algorithmic behavior on the somewhat simple and clean models does not translate beyond the models to practical performance real-world input. With this paper, we provide a first step towards studying the question of external validity systematically. To this end, we evaluate the performance of six graph algorithms on a collection of 2751 sparse real-world networks depending on two properties; the heterogeneity (variance in the degree distribution) and locality (tendency of edges to connect vertices that are already close). We compare this with the performance on generated networks with varying locality and heterogeneity. We find that the performance in the idealized setting of network models translates surprisingly well to real-world networks. Moreover, heterogeneity and locality appear to be the core properties impacting the performance of many graph algorithms.

[9] Cohen, S., Fischbeck, P., Friedrich, T., Krejca, M. S., Sauerwald, T., Accelerated Information Dissemination on Networks with Local and Global Edges. In: *Structural Information and Communication Complexity (SIROCCO)*, pp. 79–97, 2022.

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.

[10] Bläsius, T., Fischbeck, P., Gottesbüren, L., Hamann, M., Heuer, T., Spinner, J., Weyand, C., Wilhelm, M., A Branch-and-Bound Algorithm for Cluster Editing. In: *Symposium on Experimental Algorithms (SEA)*, pp. 13:1–13:19, 2022.

The editing problem asks to transform a given graph into a disjoint union of cliques by inserting and deleting as few edges as possible. We describe and evaluate an exact branch-and-bound algorithm for cluster editing. For this, we introduce new reduction rules and adapt existing ones. Moreover, we generalize a known packing technique to obtain lower bounds and experimentally show that it contributes significantly to the performance of the solver. Our experiments further evaluate the effectiveness of the different reduction rules and examine the effects of structural properties of the input graph on solver performance. Our solver won the exact track of the 2021 PACE challenge.

[11] Böther, M., Schiller, L., Fischbeck, P., Molitor, L., Krejca, M. S., Friedrich, T., Evolutionary Minimization of Traffic Congestion. In: *Genetic and Evolutionary Computation Conference (GECCO)*, pp. 937–945, 2021. Best-Paper Award (RWA Track).

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 etal. [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.

[12] Bläsius, T., Fischbeck, P., Gottesbüren, L., Hamann, M., Heuer, T., Spinner, J., Weyand, C., Wilhelm, M., PACE Solver Description: The KaPoCE Exact Cluster Editing Algorithm. In: *International Symposium on Parameterized and Exact Computation (IPEC)*, pp. 27:1–27:3, 2021.

The cluster editing problem is to transform an input graph into a cluster graph by performing a minimum number of edge editing operations. A cluster graph is a graph where each connected component is a clique. An edit operation can be either adding a new edge or removing an existing edge. In this write-up we outline the core techniques used in the exact cluster editing algorithm of the KaPoCE framework (contains also a heuristic solver), submitted to the exact track of the 2021 PACE challenge.

[13] Bläsius, T., Fischbeck, P., Gottesbüren, L., Hamann, M., Heuer, T., Spinner, J., Weyand, C., Wilhelm, M., PACE Solver Description: KaPoCE: A Heuristic Cluster Editing Algorithm. In: *International Symposium on Parameterized and Exact Computation (IPEC)*, pp. 31:1–31:4, 2021.

The cluster editing problem is to transform an input graph into a cluster graph by performing a minimum number of edge editing operations. A cluster graph is a graph where each connected component is a clique. An edit operation can be either adding a new edge or removing an existing edge. In this write-up we outline the core techniques used in the heuristic cluster editing algorithm of the Karlsruhe and Potsdam Cluster Editing (KaPoCE) framework, submitted to the heuristic track of the 2021 PACE challenge.

[14] Bläsius, T., Böther, M., Fischbeck, P., Friedrich, T., Gries, A., Hüffner, F., Kißig, O., Lenzner, P., Molitor, L., Schiller, L., Wells, A., Wietheger, S., A Strategic Routing Framework and Algorithms for Computing Alternative Paths. In: *Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS)*, pp. 10:1–10:14, 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 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.

[15] Bläsius, T., Fischbeck, P., Friedrich, T., Katzmann, M., Solving Vertex Cover in Polynomial Time on Hyperbolic Random Graphs. In: *Symposium on the Theoretical Aspects of Computer Science (STACS)*, pp. 25:1–25:14, 2020.

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.

[16] Bläsius, T., Fischbeck, P., Friedrich, T., Schirneck, M., Understanding the Effectiveness of Data Reduction in Public Transportation Networks. In: *Workshop on Algorithms and Models for the Web Graph (WAW)*, pp. 87–101, 2019.

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.

[17] Doerr, B., Fischbeck, P., Frahnow, C., Friedrich, T., Kötzing, T., Schirneck, M., Island Models Meet Rumor Spreading. In: *Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1359–1366, 2017.

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.