Publications of Sarel Cohen

Hasso Plattner Institut
Digital Engineering · Universität Potsdam

This document lists all peer-reviewed publications of Sarel Cohen, Chair for Algorithm Engineering, Hasso Plattner Institute, Potsdam, Germany. This listing was automatically generated on December 9, 2022. An up-to-date version is available online at hpi.de/friedrich/docs/publist/cohen.pdf.

Journal articles

[1] Azar, Y., Cohen, S., An improved algorithm for online machine minimization. In: *Operations Research Letters*, pp. 128–133, 2018.

The online machine minimization problem seeks to design a preemptive scheduling algorithm on multiple machines — each job j arrives at its release time r_j, has to be processed for p_j time units, and must be completed by its deadline d_j. The goal is to minimize the number of machines the algorithm uses. We improve the $\mathcal{O}(\log m)$-competitive algorithm by Chen, Megow and Schewior (SODA 2016) and provide an $\mathcal{O}(\frac{\log m}{\log \log m})$-competitive algorithm.

[2] Cohen, S., Fiat, A., Hershcovitch, M., Kaplan, H., Minimal indices for predecessor search. In: *Information and Computation*, pp. 12–30, 2015.

We give a new predecessor data structure which improves upon the index size of the Pătraşcu–Thorup data structures, reducing the index size from $\mathcal{O}(nw^{4/5})$ bits to $\mathcal{O}(n \log w)$ bits, with optimal probe complexity. Alternatively, our new data structure can be viewed as matching the space complexity of the (probe-suboptimal) z-fast trie of Belazzougui et al. Thus, we get the best of both approaches with respect to both probe count and index size. The penalty we pay is an extra $\mathcal{O}(\log w)$ inter-register operations. Our data structure can also be used to solve the weak prefix search problem, the index size of $\mathcal{O}(n \log w)$ bits is known to be optimal for any such data structure. The technical contributions include highly efficient single word indices, with out-degree $w \log w$ (compared to $w^{1/5}$ of a fusion tree node). To construct these indices we device highly efficient bit selectors which, we believe, are of independent interest.

Conference papers

[3] Wood, A., Hershcovitch, M., Ennmouri, I., Zong, W., Chennuri, S., Cohen, S., Sundararaman, S., Waddington, D., Chin, P., Towards Fast Crash-Consistent Cluster Checkpointing. In: *High Performance Extreme Computing Conference (HPEC)*, 2022.

Machine Learning models are expensive to train: they require expensive high-compute hardware and have long training times. Therefore, models are extra sensitive to program faults or unexpected system crashes, which can erase hours if not days worth of work. While there are plenty of strategies designed to mitigate the risk of unexpected system downtime, the most popular strategy in machine learning is called checkpointing: periodically saving the state of the model to persistent storage. Checkpointing is an effective strategy, however, it requires carefully balancing two operations: how often a checkpoint is made (the checkpointing schedule), and the cost of creating a checkpoint itself. In this paper, we leverage Python Memory Manager (PyMM), which provides Python support for Persistent Memory and emerging Persistent Memory technology (Optane DC) to accelerate the checkpointing operation while maintaining crash consistency. We first show that when checkpointing models, PyMM with persistent memory can save from minutes to days of checkpointing runtime. We then further optimize the checkpointing operation with PyMM and demonstrate our approach with the KMeans and Gaussian Mixture Model algorithms on two real-world datasets: MNIST and MusicNet. Through evaluation, we show that these two algorithms achieve a checkpointing speedup of a factor between 10 and 75x for KMeans and over 3x for GMM against the current state-of-the-art checkpointing approaches. We also verify that our solution recovers from crashes, while traditional approaches cannot.

[4] Bilò, D., Casel, K., Choudhary, K., Cohen, S., Friedrich, T., Lagodzinski, J. G., Schirneck, M., Wietheger, S., Fixed-Parameter Sensitivity Oracles. In: *Innovations in Theoretical Computer Science (ITCS)*, pp. 23:1–23:18, 2022.

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 Π on a graph G with parameter k preprocesses G in time $O(g(f,k)poly(n))$. When queried with a set F of at most f edges of G, the oracle reports the answer to the Π-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 $O(f \min\{f, \log(f)+k\})$. 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 $O\big(\big(\frac{f+k}{f}\big)^f \big(\frac{f+k}{k}\big)^k fk \cdot \log n\big)$ at the expense of increasing the query time to $O\big(\big(\frac{f+k}{f}\big)^f \big(\frac{f+k}{k}\big)^k f \min\{f,k\} \cdot \log n\big)$. 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 $O(3^{f+k})$ space and has $O(2^f)$ query time, the second one has a size of $O(2^{f+k^2+k})$ and a query time of $O(f+k^2)$; finally, the third one takes $O(fk+k^2)$ space and can be queried in time $O(1.2738^k + fk^2)$. All our oracles are computable in time (at most) proportional to their size and the time needed to detect a k-path or k-vertex cover, respectively. We also provide an interesting connection between k-Vertex Cover and the fault-tolerant shortest path problem, by giving a DSO of size $O(poly(f,k) \cdot n)$ with query time in $O(poly(f,k))$, where k is the size of a vertex cover. Following our line of research connecting fault-tolerant FPT and shortest paths problems, we introduce parameterization to the computation of distance preservers. Given a graph with a fixed source s and parameters f,k, we study the problem of constructing polynomial-sized oracles that reports efficiently, for any target vertex v and set F of at most f edge failures, whether the distance from s to v increases

at most by an additive term of k in $G - F$. The oracle size is $O(2^k k^2 \cdot n)$, while the time needed to answer a query is $O(2^k f^\omega k^\omega)$, where $\omega < 2.373$ is the matrix multiplication exponent. The second problem we study is about the construction of bounded-stretch fault-tolerant preservers. We construct a subgraph with $O(2^{fk+f+k} k \cdot n)$ edges that preserves those s-v-distances that do not increase by more than k upon failure of F. This improves significantly over the $\tilde{O}(fn^{2-\frac{1}{2^f}})$ bound in the unparameterized case by Bodwin et al. [ICALP 2017].

[5] Bilò, D., Choudhary, K., Cohen, S., Friedrich, T., Schirneck, M., Deterministic Sensitivity Oracles for Diameter, Eccentricities and All Pairs Distances. In: *International Colloquium on Automata, Languages and Programming (ICALP)*, pp. 68:1–68:19, 2022.

We construct data structures for extremal and pairwise distances in directed graphs in the presence of transient edge failures. Henzinger et al. (ITCS 2017) initiated the study of fault-tolerant (sensitivity) oracles for the diameter and vertex eccentricities. We extend this with a special focus on space efficiency. We present several new data structures, among them the first fault-tolerant eccentricity oracle for dual failures in subcubic space. We further prove lower bounds that show limits to approximation vs. space and diameter vs. space trade-offs for fault-tolerant oracles. They highlight key differences between data structures for undirected and directed graphs. Initially, our oracles are randomized leaning on a sampling technique frequently used in sensitivity analysis. Building on the work of Alon, Chechik, and Cohen (ICALP 2019) as well as Karthik and Parter (SODA 2021), we develop a hierarchical framework to derandomize fault-tolerant data structures. We first apply it to our own diameter/eccentricity oracles and then show its usefulness by derandomizing algorithms from the literature: the distance sensitivity oracle of Ren (JCSS 2022) and the Single-Source Replacement Path algorithm of Chechik and Magen (ICALP 2020).

[6] Böther, M., Kißig, O., Taraz, M., Cohen, S., Seidel, K., Friedrich, T., What's Wrong with Deep Learning in Tree Search for Combinatorial Optimization. In: *International Conference on Learning Representations (ICLR)*, 2022.

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 often are 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.

[7] 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.

[8] Hildebrandt, P., Schulze, M., Cohen, S., Doskoč, V., Saabni, R., Friedrich, T., Optical Character Recognition Guided Image Super Resolution. In: *Symposium on Document Engineering (DocEng)*, pp. 1–4, 2022.

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 enlarge and enhance text images for better readability, as well as recognize text automatically afterwards. We achieve improved peak signal-to-noise ratio and text recognition accuracy scores over a state-of-the-art text super-resolution model TBSRN 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.

[9] Cohen, S., Hershcovitch, M., Taraz, M., Kißig, O., Wood, A., Waddington, D., Chin, P., Friedrich, T., Drug Repurposing using Link Prediction on Knowledge Graphs with Applications to Non-Volatile Memory. In: *Complex Networks and their Applications (ComplexNetworks)*, pp. 742–753, 2021.

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.

[10] Berger, J. Fine-Grained Localization, Classification and Segmentation of Lungs with Various Diseases. In: *CVPR Workshop on Fine-Grained Visual Categorization (FGVC@CVPR)*, 2021.

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. We do so using transfer learning to fine-tune a UNet model on the Montgomery and Shenzhen datasets. In our experiments, we compare standard augmentations (like crop, pad, rotate, warp, zoom, brightness, and contrast variations) to more complex ones (for 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.

[11] Bilò, D., Cohen, S., Friedrich, T., Schirneck, M., Near-Optimal Deterministic Single-Source Distance Sensitivity Oracles. In: *European Symposium on Algorithms (ESA)*, pp. 18:1–18:17, 2021.

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 compress 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 $O(M^{1/2}n^{3/2})$ and query time $\widetilde{O}(1)$. 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 $\widetilde{O}(m\sqrt{n}+n^2)$ 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 $\widetilde{O}(m\sqrt{n}+n^2)$ preprocessing time, $O(n^{3/2})$ space, and $\widetilde{O}(1)$ 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 $\sqrt{n}$-factor compared to previous results. Grandoni and Vassilevska Williams [FOCS 2012, TALG 2020] gave an algebraic randomized $\widetilde{O}(Mn^{\omega})$ time SSRP algorithm for (undirected and directed) graphs with integer edge weights in the range $[1,M]$, where $\omega < 2.373$ is the matrix multiplication exponent. We derandomize their algorithm for undirected graphs and apply our compression to obtain an algebraic Single-Source DSO with $\widetilde{O}(Mn^{\omega})$ preprocessing time, $O(M^{1/2}n^{3/2})$ space, and $\widetilde{O}(1)$ 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 $O(M^{1/3}n^{5/3})$ and that all our oracles can be made path-reporting. On sparse graphs with $m = O(\frac{n^{5/4-\varepsilon}}{M^{7/4}})$ edges, for any constant $\varepsilon > 0$, we reduce the preprocessing to randomized $\widetilde{O}(M^{7/8}m^{1/2}n^{11/8}) = O(n^{2-\varepsilon/2})$ time. To the best of our knowledge, this is the first truly subquadratic time algorithm for building Single-Source DSOs on sparse graphs.

[12] Wood, A., Hershcovitch, M., Waddington, D., Cohen, S., Chin, P., Non-Volatile Memory Accelerated Posterior Estimation. In: *High Performance and Embedded Computing (HPEC)*, 2021.

Bayesian inference allows machine learning models to express uncertainty. Current machine learning models use only a single learnable parameter combination when making predictions, and as a result are highly overconfident when their predictions are wrong. To use more learnable parameter combinations efficiently, these samples must be drawn from the posterior distribution. Unfortunately computing the posterior directly is infeasible, so often researchers approximate it with a well known distribution such as a Gaussian. In this paper, we show that through the use of high-capacity persistent storage, models whose posterior distribution was too big to approximate are now feasible, leading to improved predictions in downstream tasks.

[13] Kißig, O., Taraz, M., Cohen, S., Doskoč, V., Friedrich, T., Drug Repurposing for Multiple COVID Strains using Collaborative Filtering. In: *ICLR Workshop on Machine Learning for Preventing and Combating Pandemics (MLPCP@ICLR)*, 2021.

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 repurposing 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.

[14] Kißig, O., Taraz, M., Cohen, S., Friedrich, T., Drug Repurposing Using Link Prediction on Knowledge Graphs. In: *ICML Workshop on Computational Biology (WCB@ICML)*, pp. 742–753, 2021.

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.

[15] Bilò, D., Cohen, S., Friedrich, T., Schirneck, M., Space-Efficient Fault-Tolerant Diameter Oracles. In: *Mathematical Foundations of Computer Science (MFCS)*, pp. 18:1–18:16, 2021.

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 $|F| \leq f$, returns the diameter of $G - F$. An f-FDO has stretch $\sigma \geq 1$ if the returned value $\widehat{D}$ satisfies $\operatorname{diam}(G-F) \leq \widehat{D} \leq \sigma \operatorname{diam}(G-F)$. For the case of a single edge failure ($f = 1$) in an unweighted directed 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 $\widetilde{O}(mn + n^{1.5}\sqrt{Dm/\varepsilon})$, 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 $\widetilde{O}(mn + n^2/\varepsilon)$, 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 $\widetilde{O}(n^{2.5794} + n^2/\varepsilon)$. 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 $0 < \varepsilon < 3/2$, 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 $O(f^2 \log^2 n)$, $\widetilde{O}(fn)$ space, and preprocessing time $\widetilde{O}(fm)$. 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 $mn^{1+o(1)}$, query time $n^{o(1)}$, and space $n^{2+o(1)}$. With fast matrix multiplication, the preprocessing time can be improved to $n^{\omega+o(1)}$, where $\omega < 2.373$ is the matrix multiplication exponent.

[16] Fogel, S., Averbuch-Elor, H., Cohen, S., Mazor, S., Litman, R., ScrabbleGAN: Semi-Supervised Varying Length Handwritten Text Generation. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4323–4332, 2020.

Optical character recognition (OCR) systems performance have improved significantly in the deep learning era. This is especially true for handwritten text recognition (HTR), where each author has a unique style, unlike printed text, where the variation is smaller by design. That said, deep learning based HTR is limited, as in every other task, by the number of training examples. Gathering data is a challenging and costly task, and even more so, the labeling task that follows, of which we focus here. One possible approach to reduce the burden of data annotation is semi-supervised learning. Semi supervised methods use, in addition to labeled data, some unlabeled samples to improve performance, compared to fully supervised ones. Consequently, such methods may adapt to unseen images during test time. We present ScrabbleGAN, a semi-supervised approach to synthesize handwritten text images that are versatile both in style and lexicon. ScrabbleGAN relies on a novel generative model which can generate images of words with an arbitrary length. We show how to operate our approach in a semi-supervised manner, enjoying the aforementioned benefits such as performance boost over state of the art supervised HTR. Furthermore, our generator can manipulate the resulting text style. This allows us to change, for instance, whether the text is cursive, or how thin is the pen stroke.

[17] Chechik, S., Cohen, S., Distance sensitivity oracles with subcubic preprocessing time and fast query time. In: *Symposium Theory of Computing (STOC)*, pp. 1375–1388, 2020.

We present the first distance sensitivity oracle (DSO) with subcubic preprocessing time and poly-logarithmic query time for directed graphs with integer weights in the range $[-M, M]$. Weimann and Yuster [FOCS 10] presented a distance sensitivity oracle for a single vertex/edge failure with subcubic preprocessing time of $\mathcal{O}(Mn^{\omega+1-\alpha})$ and subquadratic query time of $\tilde{\mathcal{O}}(n^{1+\alpha})$, where α is any parameter in $[0, 1]$, n is the number of vertices, m is the number of edges, the $\tilde{\mathcal{O}}()$ notation hides poly-logarithmic factors in n and $\omega < 2.373$ is the matrix multiplication exponent.Later, Grandoni and Vassilevska Williams [FOCS 12] substantially improved the query time to sublinear in n. In particular, they presented a distance sensitivity oracle for a single vertex/edge failure with $\tilde{\mathcal{O}}(Mn^{\omega+1/2} + Mn^{\omega+\alpha(4-\omega)})$ preprocessing time and $\tilde{\mathcal{O}}(n^{1-\alpha})$ query time. Despite the substantial improvement in the query time, it still remains polynomial in the size of the graph, which may be undesirable in many settings where the graph is of large scale. A natural question is whether one can hope for a distance sensitivity oracle with subcubic preprocessing time and very fast query time (of poly-logarithmic in n). In this paper we answer this question affirmatively by presenting a distance sensitive oracle supporting a single vertex/edge failure in subcubic $\tilde{\mathcal{O}}(Mn^{2.873})$ preprocessing time for $\omega = 2.373$, $\tilde{\mathcal{O}}(n^{2.5})$ space and near optimal query time of $\tilde{\mathcal{O}}(1)$. For comparison, with the same $\tilde{\mathcal{O}}(Mn^{2.873})$ preprocessing time the DSO of Grandoni and Vassilevska Williams has $\tilde{\mathcal{O}}(n^{0.693})$ query time. In fact, the best query time their algorithm can obtain is $\tilde{\mathcal{O}}(Mn^{0.385})$ (with $\tilde{\mathcal{O}}(Mn^3)$ preprocessing time).

[18] Alon, N., Chechik, S., Cohen, S., Deterministic Combinatorial Replacement Paths and Distance Sensitivity Oracles. In: *International Colloquium on Automata, Languages and Programming (ICALP)*, pp. 12:1–12:14, 2019.

In this work we derandomize two central results in graph algorithms, replacement paths and distance sensitivity oracles (DSOs) matching in both cases the running time of the randomized algorithms. For the replacement paths problem, let $G = (V, E)$ be a directed unweighted graph with n vertices and m edges and let P be a shortest path from s to t in G. The replacement paths problem is to find for every edge e in P the shortest path from s to t avoiding e. Roditty and Zwick [ICALP 2005] obtained a randomized algorithm with running time of $\mathcal{O}(m\sqrt{n})$. Here we provide the first deterministic algorithm for this problem, with the same $\mathcal{O}(m\sqrt{n})$ time. Due to matching conditional lower bounds of Williams et al. [FOCS 2010], our deterministic combinatorial algorithm for the replacement paths problem is optimal up to polylogarithmic factors (unless the long standing bound of $\mathcal{O}(mn)$ for the combinatorial boolean matrix multiplication can be improved). This also implies a deterministic algorithm for the second simple shortest path problem in $\mathcal{O}(m\sqrt{n})$ time, and a deterministic algorithm for the k-simple shortest paths problem in $\mathcal{O}(kmsqrtn)$ time (for any integer constant $k > 0$). For the problem of distance sensitivity oracles, let $G = (V, E)$ be a directed graph with real-edge weights. An f-Sensitivity Distance Oracle (f-DSO) gets as input the graph $G = (V, E)$ and a parameter f, preprocesses it into a data-structure, such that given a query (s, t, F) with $s, t \in V$ and $F \subseteq E \cup V$, $|F| <= f$ being a set of at most f edges or vertices (failures), the query algorithm efficiently computes the distance from s to t in the graph $G \setminus F$ (i.e., the distance from s to t in the graph G after removing from it the failing edges and vertices F). For weighted graphs with real edge weights, Weimann and Yuster [FOCS 2010] presented several randomized f-DSOs. In particular, they presented a combinatorial f-DSO with $\mathcal{O}(mn^{4-\alpha})$ preprocessing time and subquadratic $\mathcal{O}(n^{2-2(1-\alpha)/f})$ query time, giving a tradeoff between preprocessing and query time for every value of $0 < \alpha < 1$. We derandomize this result and present a combinatorial deterministic f-DSO with the same asymptotic preprocessing and query time.

[19] Chechik, S., Cohen, S., Near Optimal Algorithms For The Single Source Replacement Paths Problem. In: *Symposium on Discrete Algorithms (SODA)*, pp. 2090–2109, 2019.

The Single Source Replacement Paths (SSRP) problem is as follows; Given a graph $G = (V, E)$, a source vertex s and a shortest paths tree T_s rooted in s, output for every vertex $t \in V$ and for every edge e in T_s the length of the shortest path from s to t avoiding e. We present near optimal upper bounds, by providing $\tilde{\mathcal{O}}(m\sqrt{n} + n^2)$ time randomized combinatorial algorithm for unweighted undirected graphs, and matching conditional lower bounds for the SSRP problem.

[20] Arar, M., Chechik, S., Cohen, S., Stein, C., Wajc, D., Dynamic Matching: Reducing Integral

Algorithms to Approximately-Maximal Fractional Algorithms. In: *International Colloquium on Automata, Languages and Programming (ICALP)*, pp. 7:1–7:16, 2018.

We present a simple randomized reduction from fully-dynamic integral matching algorithms to fully-dynamic "approximately-maximal" fractional matching algorithms. Applying this reduction to the recent fractional matching algorithm of Bhattacharya, Henzinger, and Nanongkai (SODA 2017), we obtain a novel result for the integral problem. Specifically, our main result is a randomized fully-dynamic $(2+\varepsilon)$-approximate integral matching algorithm with small polylog worst-case update time. For the $(2+\varepsilon)$-approximation regime only a fractional fully-dynamic $(2+\varepsilon)$-matching algorithm with worst-case polylog update time was previously known, due to Bhattacharya et al. (SODA 2017). Our algorithm is the first algorithm that maintains approximate matchings with worst-case update time better than polynomial, for any constant approximation ratio. As a consequence, we also obtain the first constant-approximate worst-case polylogarithmic update time maximum weight matching algorithm.

[21] Chechik, S., Cohen, S., Fiat, A., Kaplan, H., $(1+\varepsilon)$-Approximate f-Sensitive Distance Oracles. In: *Symposium on Discrete Algorithms (SODA)*, pp. 1479–1496, 2017.

An f-Sensitive Distance Oracle with stretch a preprocesses a graph $G(V, E)$ and produces a small data structure that is used to answer subsequent queries. A query is a triple consisting of a set $F \subset E$ of at most f edges, and vertices s and t. The oracle answers a query $(F, s., t)$ by returning a value d which is equal to the length of some path between s and t in the graph $G \setminus F$ (the graph obtained from G by discarding all edges in F). Moreover, d is at most a times the length of the shortest path between s and t in $G \setminus F$. The oracle can also construct a path between s and t in $G \setminus F$ of length d. To the best of our knowledge we give the first nontrivial f-sensitive distance oracle with fast query time and small stretch capable of handling multiple edge failures. Specifically, for any $\mathcal{O}(\frac{\log n}{\log\log n})$ and a fixed $\varepsilon > 0$ our oracle answers queries (F, s, t) in time $\mathcal{O}(l)$ with $(1+\varepsilon)$ stretch using a data structure of size $n^{2+\mathcal{O}(1)}$ For comparison, the naive alternative requires $m^f n^2$ space for sublinear query time.

[22] Cohen, S., Fiat, A., Hershcovitch, M., Kaplan, H., Minimal Indices for Successor Search - (Extended Abstract). In: *Mathematical Foundations of Computer Science (MFCS)*, pp. 278–289, 2013.

We give a new successor data structure which improves upon the index size of the Pătraşcu-Thorup data structures, reducing the index size from $\mathcal{O}(nw^{4/5}$ bits to $\mathcal{O}(n \log w)$ bits, with optimal probe complexity. Alternatively, our new data structure can be viewed as matching the space complexity of the (probe-suboptimal) z-fast trie of Belazzougui et al. Thus, we get the best of both approaches with respect to both probe count and index size. The penalty we pay is an extra $\mathcal{O}(n \log w)$ inter-register operations. Our data structure can also be used to solve the weak prefix search problem, the index size of $\mathcal{O}(n \log w)$ bits is known to be optimal for any such data structure. The technical contributions include highly efficient single word indices, with out-degree $w/\log w$ (compared to the $w^{1/5}$ out-degree of fusion tree based indices). To construct such high efficiency single word indices we device highly efficient bit selectors which, we believe, are of independent interest.