

# Publications of Karen Seidel

This document lists all peer-reviewed publications of [Karen Seidel](#), Chair for Algorithm Engineering, Hasso Plattner Institute, Potsdam, Germany. This listing was automatically generated on September 25, 2022. An up-to-date version is available online at [hpi.de/friedrich/docs/publist/seidel.pdf](https://hpi.de/friedrich/docs/publist/seidel.pdf).

## Journal articles

- [1] Göbel, A., Lagodzinski, J. A. G., Seidel, K., [Counting Homomorphisms to Trees Modulo a Prime](#). In: *ACM Transactions on Computation Theory* 13, 2021.

Many important graph theoretic notions can be encoded as counting graph homomorphism problems, such as partition functions in statistical physics, in particular independent sets and colourings. In this article we study the complexity of  $\#_p\text{HomsTo}H$ , the problem of counting graph homomorphisms from an input graph to a graph  $H$  modulo a prime number  $p$ . Dyer and Greenhill proved a dichotomy stating that the tractability of non-modular counting graph homomorphisms depends on the structure of the target graph. Many intractable cases in non-modular counting become tractable in modular counting due to the common phenomenon of cancellation. In subsequent studies on counting modulo 2, however, the influence of the structure of  $H$  on the tractability was shown to persist, which yields similar dichotomies. Our main result states that for every tree  $H$  and every prime  $p$  the problem  $\#_p\text{HomsTo}H$  is either polynomial time computable or  $\#_p\text{P}$ -complete. This relates to the conjecture of Faben and Jerrum stating that this dichotomy holds for every graph  $H$  when counting modulo 2. In contrast to previous results on modular counting, the tractable cases of  $\#_p\text{HomsTo}H$  are essentially the same for all values of the modulo when  $H$  is a tree. To prove this result, we study the structural properties of a homomorphism. As an important interim result, our study yields a dichotomy for the problem of counting weighted independent sets in a bipartite graph modulo some prime  $p$ . These results are the first suggesting that such dichotomies hold not only for the one-bit functions of the modulo 2 case but also for the modular counting functions of all primes  $p$ .

## Conference papers

- [2] Böther, M., Kifig, 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.

- [3] Khazraei, A., Kötzing, T., Seidel, K., [Towards a Map for Incremental Learning in the Limit from Positive and Negative Information](#). In: *Computability in Europe (CiE)*, pp. 273–284, 2021.

In order to model an efficient learning paradigm, iterative learning algorithms access data one by one, updating the current hypothesis without regress to past data. Prior research investigating the impact of additional requirements on iterative learners left many questions open, especially in learning from informant, where the input is binary labeled.

- [4] Kötzing, T., Seidel, K., [Learning Languages in the Limit from Positive Information with Finitely Many Memory Changes](#). In: *Computability in Europe (CiE)*, pp. 318–329, 2021.

We investigate learning collections of languages from texts by an inductive inference machine with access to the current datum and a bounded memory in form of states. Such a bounded memory states (BMS) learner is considered successful in case it eventually settles on a correct hypothesis while exploiting only finitely many different states.

- [5] Gao, Z., Jain, S., Khossainov, B., Li, W., Melnikov, A., Seidel, K., Stephan, F., [Random Subgroups of Rationals](#). In: *Mathematical Foundations of Computer Science (MFCS)*, pp. 25:1–25:14, 2019.

This paper introduces and studies a notion of algorithmic randomness for subgroups of rationals. Given a randomly generated additive subgroup  $(G, +)$  of rationals, two main questions are addressed: first, what are the model-theoretic and recursion-theoretic properties of  $(G, +)$ ; second, what learnability properties can one extract from  $G$  and its subclass of finitely generated subgroups? For the first question, it is shown that the theory of  $(G, +)$  coincides with that of the additive group of integers and is therefore decidable; furthermore, while the word problem for  $G$  with respect to any generating sequence for  $G$  is not even semi-decidable, one can build a generating sequence  $\beta$  such that the word problem for  $G$  with respect to  $\beta$  is co-recursively enumerable (assuming that the set of generators of  $G$  is limit-recursive). In regard to the second question, it is proven that there is a generating sequence  $\beta$  for  $G$  such that every non-trivial finitely generated subgroup of  $G$  is recursively enumerable and the class of all such subgroups of  $G$  is behaviourally correctly learnable, that is, every non-trivial finitely generated subgroup can be semantically identified in the limit (again assuming that the set of generators of  $G$  is limit-recursive). On the other hand, the class of non-trivial finitely generated subgroups of  $G$  cannot be syntactically identified in the limit with respect to any generating sequence for  $G$ . The present work thus contributes to a recent line of research studying algorithmically random infinite structures and uncovers an interesting connection between the arithmetical complexity of the set of generators of a randomly generated subgroup of rationals and the learnability of its finitely generated subgroups.

- [6] Göbel, A., Lagodzinski, J. A. G., Seidel, K., [Counting Homomorphisms to Trees Modulo a Prime](#). In: *Mathematical Foundations of Computer Science (MFCS)*, pp. 49:1–49:13, 2018.

Many important graph theoretic notions can be encoded as counting graph homomorphism problems, such as partition functions in statistical physics, in particular independent sets and colourings. In this article we study the complexity of  $\#_p\text{HomsTo}H$ , the problem of counting graph homomorphisms from an input graph to a graph  $H$  modulo a prime number  $p$ . Dyer and Greenhill proved a dichotomy stating that the tractability of non-modular counting graph homomorphisms depends on the structure of the target graph. Many intractable cases in non-modular counting become tractable in modular counting due to the common phenomenon of cancellation. In subsequent studies on counting modulo 2, however, the influence of the structure of  $H$  on the tractability was shown to persist, which yields similar dichotomies. Our main result states that for every tree  $H$  and every prime  $p$  the problem  $\#_p\text{HomsTo}H$  is either polynomial time computable or  $\#_p\text{P}$ -complete. This relates to the conjecture of Faben and Jerrum stating that this dichotomy holds for every graph  $H$  when counting modulo 2. In contrast to previous results on modular counting, the tractable cases of  $\#_p\text{HomsTo}H$  are essentially the same for all values of the modulo when  $H$  is a tree. To prove this result, we study the structural properties of a homomorphism. As an important interim result, our study yields a dichotomy for the problem of counting weighted independent sets in a bipartite graph modulo some prime  $p$ . These results are the first suggesting that such dichotomies hold not only for the one-bit functions of the modulo 2 case but also for the modular counting functions of all primes  $p$ .

- [7] Kötzing, T., Schirneck, M., Seidel, K., [Normal Forms in Semantic Language Identification](#). In: *Algorithmic Learning Theory (ALT)*, pp. 493–516, 2017.

We consider language learning in the limit from text where all learning restrictions are semantic, that is, where any conjecture may be replaced by a semantically equivalent conjecture. For different such learning criteria, starting with the well-known TxtGbc-learning, we consider three different normal forms: strongly locking learning, consistent learning and (partially) set-driven learning. These normal forms support and simplify proofs and give insight into what behaviors are necessary for successful learning (for example when consistency in conservative learning implies cautiousness and strong decisiveness). We show that strongly locking learning can be assumed for partially set-driven learners, even when learning restrictions apply. We give a very general proof relying only on a natural property of the learning restriction, namely, allowing for simulation on equivalent text. Furthermore, when no restrictions apply, also the converse is true: every strongly locking learner can be made partially set-driven. For several semantic learning criteria we show that learning can be done consistently. Finally, we deduce for which learning restrictions partial set-drivenness and set-drivenness coincide, including a general statement about classes of infinite languages. The latter again relies on a simulation argument.

- [8] Hölzl, R., Jain, S., Schlicht, P., Seidel, K., Stephan, F., [Automatic Learning from Repetitive Texts](#). In: *Algorithmic Learning Theory (ALT)*, pp. 129–150, 2017.

We study the connections between the learnability of automatic families of languages and the types of text used to present them to a learner. More precisely, we study how restrictions on the number of times that a correct datum appears in a text influence what classes of languages are automatically learnable. We show that an automatic family of languages is automatically learnable from fat text iff it is automatically learnable from thick text iff it is verifiable from balanced text iff it satisfies Angluin’s tell-tale condition. Furthermore, many automatic families are automatically learnable from exponential text. We also study the relationship between automatic learnability and verifiability and show that all automatic families are automatically partially verifiable from exponential text and automatically learnable from thick text.

- [9] Seidel, K. [Zu mathematischen Argumentationen eines Experten aus einer semiotischen Perspektive](#). In: *Beiträge zum Mathematikunterricht 2017*. Jahrestagung Gesellschaft für Didaktik der Mathematik 2017, pp. 897–900, 2017.

In Argumentationen unter in der Forschung tätigen Mathematikern kombiniert die erklärende Person Gesten, Sprache und Zeichen und verleiht damit unter anderem den concept images der verwendeten mathematischen Begriffe Ausdruck. Im Folgenden wird der Frage nachgegangen, inwiefern eine multimodale Analyse von erklärenden Phasen in Argumentationen von Experten mit dem Fokus auf Begriffsbildungsprozessen Aufschluss geben kann über operationale und strukturelle Vorstellungen von Begriffen (Sfard, 1991).