

## References

- Ralph Abboud, İsmail İlkan Ceylan, Martin Grohe, and Thomas Lukasiewicz. The surprising power of graph neural networks with random node initialization. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 2112–2118. ijcai.org, 2021. doi: 10.24963/ijcai.2021/291. URL <https://doi.org/10.24963/ijcai.2021/291>.
- Abien Fred Agarap. Deep learning using rectified linear units (relu). *CoRR*, abs/1803.08375, 2018. URL <http://arxiv.org/abs/1803.08375>.
- Sungsoo Ahn, Younggyo Seo, and Jinwoo Shin. Learning what to defer for maximum independent sets. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 134–144. PMLR, 2020. URL <http://proceedings.mlr.press/v119/ahn20a.html>.
- Maria Chiara Angelini and Federico Ricci-Tersenghi. Modern graph neural networks do worse than classical greedy algorithms in solving combinatorial optimization problems like maximum independent set. *Nat. Mac. Intell.*, 5(1):29–31, 2023. doi: 10.1038/s42256-022-00589-y. URL <https://doi.org/10.1038/s42256-022-00589-y>.
- Lei Jimmy Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization. *CoRR*, abs/1607.06450, 2016. URL <http://arxiv.org/abs/1607.06450>.
- Jean Barbier, Florent Krzakala, Lenka Zdeborová, and Pan Zhang. The hard-core model on random graphs revisited. *Journal of Physics: Conference Series*, 473(1):012021, dec 2013. doi: 10.1088/1742-6596/473/1/012021. URL <https://dx.doi.org/10.1088/1742-6596/473/1/012021>.
- Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinícius Flores Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Çağlar Gülçehre, H. Francis Song, Andrew J. Ballard, Justin Gilmer, George E. Dahl, Ashish Vaswani, Kelsey R. Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matthew M. Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks. *CoRR*, abs/1806.01261, 2018. URL <http://arxiv.org/abs/1806.01261>.
- Irwan Bello, Hieu Pham, Quoc V. Le, Mohammad Norouzi, and Samy Bengio. Neural combinatorial optimization with reinforcement learning. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Workshop Track Proceedings*. OpenReview.net, 2017. URL <https://openreview.net/forum?id=Bk9mx1SFx>.
- Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML ’09*, page 41–48, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605585161. doi: 10.1145/1553374.1553380. URL <https://doi.org/10.1145/1553374.1553380>.
- Yoshua Bengio, Andrea Lodi, and Antoine Prouvost. Machine learning for combinatorial optimization: A methodological tour d’horizon. *Eur. J. Oper. Res.*, 290(2):405–421, 2021. doi: 10.1016/j.ejor.2020.07.063. URL <https://doi.org/10.1016/j.ejor.2020.07.063>.
- Maximilian Böther, Otto Kißig, Martin Taraz, Sarel Cohen, Karen Seidel, and Tobias Friedrich. What’s wrong with deep learning in tree search for combinatorial optimization. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL <https://openreview.net/forum?id=mk0HzdqY7i1>.
- Quentin Cappart, Didier Chételat, Elias B. Khalil, Andrea Lodi, Christopher Morris, and Petar Velickovic. Combinatorial optimization and reasoning with graph neural networks. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 4348–4355. ijcai.org, 2021. doi: 10.24963/ijcai.2021/595. URL <https://doi.org/10.24963/ijcai.2021/595>.
- Corinna Cortes, Vitaly Kuznetsov, Mehryar Mohri, and Umar Syed. Structural maxent models. In Francis R. Bach and David M. Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pages 391–399. JMLR.org, 2015. URL <http://proceedings.mlr.press/v37/cortes15.html>.
- Miroslav Dudík, Steven J. Phillips, and Robert E. Schapire. Maximum entropy density estimation with generalized regularization and an application to species distribution modeling. *J. Mach. Learn. Res.*, 8: 1217–1260, 2007. doi: 10.5555/1314498.1314540. URL <https://dl.acm.org/doi/10.5555/1314498.1314540>.

- Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2023. URL <https://www.gurobi.com>.
- Mohamed Hibat-Allah, Estelle M. Inack, Roeland Wiersema, Roger G. Melko, and Juan Carrasquilla. Variational neural annealing. *Nat. Mach. Intell.*, 3(11):952–961, 2021. doi: 10.1038/s42256-021-00401-3. URL <https://doi.org/10.1038/s42256-021-00401-3>.
- Tommi S. Jaakkola and Michael I. Jordan. Improving the mean field approximation via the use of mixture distributions. In Michael I. Jordan, editor, *Learning in Graphical Models*, volume 89 of *NATO ASI Series*, pages 163–173. Springer Netherlands, 1998. doi: 10.1007/978-94-011-5014-9\_6. URL [https://doi.org/10.1007/978-94-011-5014-9\\_6](https://doi.org/10.1007/978-94-011-5014-9_6).
- E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, May 1957. doi: 10.1103/PhysRev.106.620. URL <https://link.aps.org/doi/10.1103/PhysRev.106.620>.
- Nikolaos Karalias and Andreas Loukas. Erdos goes neural: an unsupervised learning framework for combinatorial optimization on graphs. In Hugo Larochelle, Marc’Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL <https://proceedings.neurips.cc/paper/2020/hash/49f85a9ed090b20c8bed85a5923c669f-Abstract.html>.
- Nikolaos Karalias, Joshua Robinson, Andreas Loukas, and Stefanie Jegelka. Neural set function extensions: Learning with discrete functions in high dimensions. In *NeurIPS*, 2022. URL [http://papers.nips.cc/paper\\_files/paper/2022/hash/6294a235c0b80f0a2b224375c546c750-Abstract-Conference.html](http://papers.nips.cc/paper_files/paper/2022/hash/6294a235c0b80f0a2b224375c546c750-Abstract-Conference.html).
- Jun’ichi Kazama and Jun’ichi Tsujii. Evaluation and extension of maximum entropy models with inequality constraints. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP 2003, Sapporo, Japan, July 11-12, 2003*, 2003. URL <https://aclanthology.org/W03-1018/>.
- Elias B. Khalil, Hanjun Dai, Yuyu Zhang, Bistra Dilikina, and Le Song. Learning combinatorial optimization algorithms over graphs. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 6348–6358, 2017. URL <https://proceedings.neurips.cc/paper/2017/hash/d9896106ca98d3d05b8cbdf4fd8b13a1-Abstract.html>.
- Shoummo Ahsan Khandoker, Jawaril Munshad Abedin, and Mohamed Hibat-Allah. Supplementing recurrent neural networks with annealing to solve combinatorial optimization problems. *Mach. Learn. Sci. Technol.*, 4(1): 15026, 2023. doi: 10.1088/2632-2153/acb895. URL <https://doi.org/10.1088/2632-2153/acb895>.
- Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL <https://openreview.net/forum?id=SJU4ayYgl>.
- S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598): 671–680, 1983. doi: 10.1126/science.220.4598.671. URL <https://www.science.org/doi/abs/10.1126/science.220.4598.671>.
- Wouter Kool, Herke van Hoof, and Max Welling. Attention, learn to solve routing problems! In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019. URL <https://openreview.net/forum?id=ByxBFsRqYm>.
- Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data>, June 2014.
- Fenglin Liu, Xuancheng Ren, Zhiyuan Zhang, Xu Sun, and Yuexian Zou. Rethinking skip connection with layer normalization in transformers and resnets. *CoRR*, abs/2105.07205, 2021. URL <https://arxiv.org/abs/2105.07205>.
- Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. In *International Conference on Learning Representations*, 2017. URL <https://openreview.net/forum?id=Skq89Scxx>.
- Andrew Lucas. Ising formulations of many np problems. *Frontiers in Physics*, 2, 2014. ISSN 2296-424X. doi: 10.3389/fphy.2014.00005. URL <https://www.frontiersin.org/articles/10.3389/fphy.2014.00005>.
- David J. C. MacKay. *Information theory, inference, and learning algorithms*. Cambridge University Press, 2003. ISBN 978-0-521-64298-9.

475 Nina Mazyavkina, Sergey Sviridov, Sergei Ivanov, and Evgeny Burnaev. Reinforcement learning for combinatorial optimization: A survey. *Comput. Oper. Res.*, 134:105400, 2021. doi: 10.1016/j.cor.2021.105400. URL <https://doi.org/10.1016/j.cor.2021.105400>.

478 Song Mei, Theodor Misiakiewicz, and Andrea Montanari. Mean-field theory of two-layers neural networks: dimension-free bounds and kernel limit. In Alina Beygelzimer and Daniel Hsu, editors, *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, volume 99 of *Proceedings of Machine Learning Research*, pages 2388–2464. PMLR, 2019. URL <http://proceedings.mlr.press/v99/mei19a.html>.

482 Yimeng Min, Frederik Wenkel, Michael Perlmuter, and Guy Wolf. Can hybrid geometric scattering networks help solve the maximum clique problem? In *NeurIPS*, 2022. URL [http://papers.nips.cc/paper\\_files/paper/2022/hash/8ec88961d36d9a87ac24baf45402744f-Abstract-Conference.html](http://papers.nips.cc/paper_files/paper/2022/hash/8ec88961d36d9a87ac24baf45402744f-Abstract-Conference.html).

485 Christopher Morris, Nils M. Kriege, Franka Bause, Kristian Kersting, Petra Mutzel, and Marion Neumann. Tudataset: A collection of benchmark datasets for learning with graphs. *CoRR*, abs/2007.08663, 2020. URL <https://arxiv.org/abs/2007.08663>.

488 Marc Mézard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, 01 2009. ISBN 9780198570837. doi: 10.1093/acprof:oso/9780198570837.001.0001. URL <https://doi.org/10.1093/acprof:oso/9780198570837.001.0001>.

491 Christos H Papadimitriou and Kenneth Steiglitz. *Combinatorial optimization: algorithms and complexity*. Courier Corporation, 1998.

493 G Parisi. Statistical field theory, vol. 66 in. *Frontiers in Physics*, page 343, 1988.

494 Ruizhong Qiu, Zhiqing Sun, and Yiming Yang. DIMES: A differentiable meta solver for combinatorial optimization problems. In *NeurIPS*, 2022. URL [http://papers.nips.cc/paper\\_files/paper/2022/hash/a3a7387e49f4de290c23beea2dfcdc75-Abstract-Conference.html](http://papers.nips.cc/paper_files/paper/2022/hash/a3a7387e49f4de290c23beea2dfcdc75-Abstract-Conference.html).

497 Prabhakar Raghavan. Probabilistic construction of deterministic algorithms: approximating packing integer programs. *Journal of Computer and System Sciences*, 37(2):130–143, 1988.

499 John Schulman, Philipp Moritz, Sergey Levine, Michael I. Jordan, and Pieter Abbeel. High-dimensional continuous control using generalized advantage estimation. In Yoshua Bengio and Yann LeCun, editors, *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016. URL <http://arxiv.org/abs/1506.02438>.

503 John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *CoRR*, abs/1707.06347, 2017. URL <http://arxiv.org/abs/1707.06347>.

505 Haoran Sun, Etash Kumar Guha, and Hanjun Dai. Annealed training for combinatorial optimization on graphs. In *OPT 2022: Optimization for Machine Learning (NeurIPS 2022 Workshop)*, 2022. URL <https://openreview.net/forum?id=fo3b0XjTkU>.

508 Jan Tönshoff, Martin Ritzert, Hinrikus Wolf, and Martin Grohe. Graph neural networks for maximum constraint satisfaction. *Frontiers Artif. Intell.*, 3:580607, 2020. doi: 10.3389/frai.2020.580607. URL <https://doi.org/10.3389/frai.2020.580607>.

511 Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.*, 1(1-2):1–305, 2008. doi: 10.1561/22000000001. URL <https://doi.org/10.1561/22000000001>.

514 Haoyu Wang, Nan Wu, Hang Yang, Cong Hao, and Pan Li. Unsupervised learning for combinatorial optimization with principled objective relaxation. In *NeurIPS*, 2022. URL [http://papers.nips.cc/paper\\_files/paper/2022/hash/cbc1ad2066f0afebbcea930c5688fc1f-Abstract-Conference.html](http://papers.nips.cc/paper_files/paper/2022/hash/cbc1ad2066f0afebbcea930c5688fc1f-Abstract-Conference.html).

517 Haoyu Peter Wang and Pan Li. Unsupervised learning for combinatorial optimization needs meta learning. In *The Eleventh International Conference on Learning Representations*, 2023. URL <https://openreview.net/forum?id=-ENYHCE8zBp>.

520 Nicholas C Wormald. Differential equations for random processes and random graphs. *The annals of applied probability*, pages 1217–1235, 1995.

522 Dian Wu, Lei Wang, and Pan Zhang. Solving statistical mechanics using variational autoregressive networks. *Phys. Rev. Lett.*, 122:080602, Feb 2019. doi: 10.1103/PhysRevLett.122.080602. URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.080602>.

- 525 Xiaoxia Wu, Ethan Dyer, and Behnam Neyshabur. When do curricula work? In *9th International Conference on*  
526 *Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL  
527 <https://openreview.net/forum?id=tW4QEInpni>.
- 528 Ke Xu, Frédéric Boussemart, Fred Hemery, and Christophe Lecoutre. A simple model to generate hard satisfiable  
529 instances. In Leslie Pack Kaelbling and Alessandro Saffiotti, editors, *IJCAI-05, Proceedings of the Nineteenth*  
530 *International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30 - August 5, 2005*,  
531 pages 337–342. Professional Book Center, 2005. URL [http://ijcai.org/Proceedings/05/Papers/](http://ijcai.org/Proceedings/05/Papers/0989.pdf)  
532 [0989.pdf](http://ijcai.org/Proceedings/05/Papers/0989.pdf).
- 533 Gal Yehuda, Moshe Gabel, and Assaf Schuster. It’s not what machines can learn, it’s what we cannot teach.  
534 In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020,*  
535 *Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 10831–10841. PMLR, 2020.  
536 URL <http://proceedings.mlr.press/v119/yehuda20a.html>.