Transformers and Graph Tasks

May 21, 2026

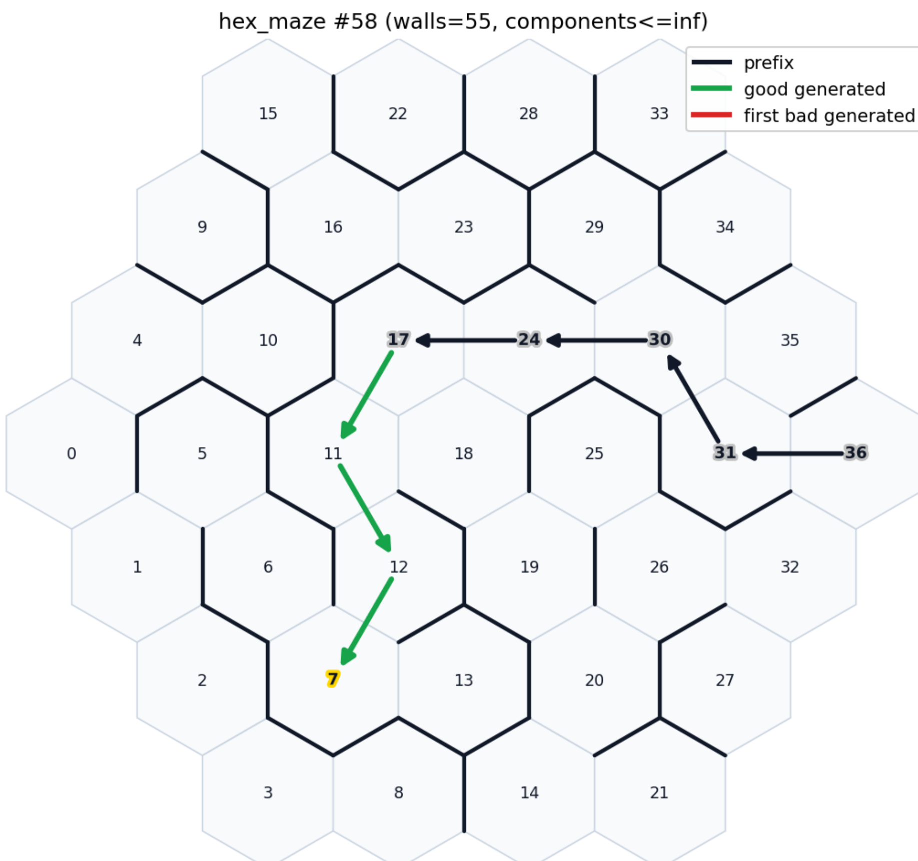

Figure 1: A test case in the qualifying problems.

Since their inception [Vaswani et al., 2017], transformer models have come to dominate first the field of natural language processing, and then the field of computer vision [Dosovitskiy et al., 2021]. Later on, they were finetuned to be helpful virtual assistants [Ouyang et al., 2022], and by now they can perform assignments as diverse as emotional support, ideas brainstorming, mathematical problem solving, and pair programming.

Besides their astounding robustness in training, and emergent capabilities upon scaling up [Brown et al., 2020], these models can also have unexpected difficulties in learning certain simple algorithms such as addition [Zhou et al., 2024]. Here, we mean learning in the machine learning sense of generalization: to produce correct output on previously unseen input. In sequence models such as transformers, length generalization is of special importance: to produce correct output on inputs longer than those seen in training.

In this project, we aim to study how transformers learn graph tasks. Graph tasks, besides of great intrinsic interest to Mathematicians, are very versatile abstractions of various real world tasks such as reasoning [Khona et al., 2024] or logistical optimization [Kool et al., 2019]. Moreover, we can study transformer capabilities on graph tasks using small models, thus eschewing the very high computational power one needs when working with Large Language Models. In particular, you will be able to participate perfectly well on any modern laptop, without access to a powerful GPU.

Prerequisites

Strong command of the `Python` numerical library `numpy`.

Qualifying problems

See the contents of the attached `zip` file. `qualifying_problems.py` contains functions to be written, which you can test with `test.py`. To aid you debug, or just bask in the glory of your working function, you can use `visualize.py`.

I very much welcome partial solutions! The functions to be written are ordered in increasing difficulty. Please try to avoid explicit Python loops. Instead, use `numpy` functions that operate on the entire arrays at once.

You can hand in multiple versions. I will evaluate the latest submission. A valid submission must arrive to my email address by the deadline written in the general RES course description on the BSM webpage. You can expect me to answer a question before this time if it arrived to my email address at least 24 hours before this deadline.

In your email, please also write me the following:

1. Your Mathematics and Computer Science background.
2. Your Mathematics and Computer Science interests.
3. What do you find especially interesting in this project?

Have fun with the problems and I hope to see you in the group!

Using Virtual Assistants

Please do not use any AI tool (ChatGPT, Claude, Gemini, etc., or derivatives such as Codex or Copilot) to solve the qualifying problems. The objective of the qualifying problems is to gauge your own programming capabilities. This is important even in the age of AI as you need to be able to verify the work of the virtual assistant.

This being said, in the group, I will highly welcome using virtual assistants as they can accelerate work greatly.

Contact

Pál Zsámboki, `zsamboki@renyi.hu`, HUN-REN Alfréd Rényi Institute of Mathematics

References

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf`.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil

Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.

Yongchao Zhou, Uri Alon, Xinyun Chen, Xuezhi Wang, Rishabh Agarwal, and Denny Zhou. Transformers can achieve length generalization but not robustly. In *ICLR 2024 Workshop on Mathematical and Empirical Understanding of Foundation Models*, 2024. URL https://openreview.net/forum?id=DWkWIh3vFJ.

Mikail Khona, Maya Okawa, Jan Hula, Rahul Ramesh, Kento Nishi, Robert P. Dick, Ekdeep Singh Lubana, and Hidenori Tanaka. Towards an understanding of stepwise inference in transformers: A synthetic graph navigation model. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=8VEGkphQaK.

Wouter Kool, Herke van Hoof, and Max Welling. Attention, learn to solve routing problems! In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=ByxBFsRqYm.