

---

# FLOWGEN: Fast and slow graph generation

---

Aman Madaan<sup>1</sup> Yiming Yang<sup>1</sup>

## Abstract

We present FLOWGEN, a graph-generation model inspired by the dual-process theory of mind that generates large graphs incrementally. Depending on the difficulty of completing the graph at the current step, graph generation is routed to either a FAST (weaker) or a SLOW (stronger) model. FAST and SLOW models have identical architectures, but vary in the number of parameters and consequently the strength. Experiments on real-world graphs show that FLOWGEN can successfully generate graphs similar to those generated by a single large model in a fraction of time.

## 1. Introduction

Why should the auto-regressive factorization be parameterized by a single model?

Graphs provide a rich abstraction for a wide range of tasks including molecular design (De Cao & Kipf, 2018; Samanta et al., 2019; Lim et al., 2020), chemical compound design (Lim et al., 2020), online user interaction modeling (Zhou et al., 2020), and map layout design (Mi et al., 2021). Developing generative models of graphs is is therefore an important classical problem, which has seen renewed interest with the success of deep learning models. Specifically, *implicit* generative models are a popular choice for graph generative modeling. Unlike explicit models, implicit generative models do not explicitly model the distribution of graphs but instead allow sampling graphs. A popular example of such implicit models are GANs, and have recently shown state of the art results for generative modeling of graphs (Bojchevski et al., 2018).

Like typical machine learning models, generative models of graphs currently use identical model complexity and computational strength while generating graphs. However, since these models are constructive by design (i.e., they build a

graph piece-by-piece), it is natural to expect that generating different parts of a graph requires different levels of reasoning. For example, generating a 2-hop neighborhood frequently seen during training might be *easier* than generating a novel 4-hop neighborhood. Notably, humans typically use different reasoning power depending on the problem difficulty (e.g., solving  $2 * 2 = ?$  vs.  $19 * 3 = ?$ ), an idea known as the dual-process theory of mind (Posner & Snyder, 1975; Shiffrin & Schneider, 1977; Evans, 1984; Stanovich, 2000; Kahneman, 2003; Frankish, 2010). Specifically, Stanovich (2000) propose to divide mental processing as being done by two metaphorical systems referred by them as *System 1* (instinctive, used for  $2 * 2$ ) and *System 2* (analytical, planner, used for  $19 * 3$ ). The terms FAST and SLOW for Systems 1 and 2 were subsequently popularized by Kahneman (2011). There is now a growing interest in utilizing a combination of fast and slow reasoning systems in diverse areas of Machine Learning (Anthony et al., 2017; Mujika et al., 2017; Schwarzschild et al., 2021b).

This paper introduces FLOWGEN, a generative graph model that is inspired by the dual-process theory of mind. FLOWGEN decomposes the problem of generating a graph into the problem of learning to generate walks. Generating walks provides a setting where identifying the easy and challenging portions is easier: starting from a given node, the model begins by generating walks seen during the training in known neighborhoods. The difficulty of generating such walks then gradually increases for two reasons. First, conditioning on increasingly longer contexts is required for generating longer walks. Second, as the length of the walks exceeds the length seen during training, a model is forced to create neighborhoods not seen during the training: a task that requires more robust generalization capabilities. FLOWGEN leverages this mismatch in problem difficulty by dynamically switching from a small (FAST) model to a large (SLOW) model for efficient graph generation. Figure 1 provides an overview of our approach. FLOWGEN method achieves the same results as using the SLOW method alone on three different graphs, while taking up to 50% less time.

## 2. FLOWGEN

In this section, we describe our novel graph generation method. First, we describe how auto-regressive models can

---

<sup>1</sup>Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA. Correspondence to: Aman Madaan <amadaan@cs.cmu.edu>.

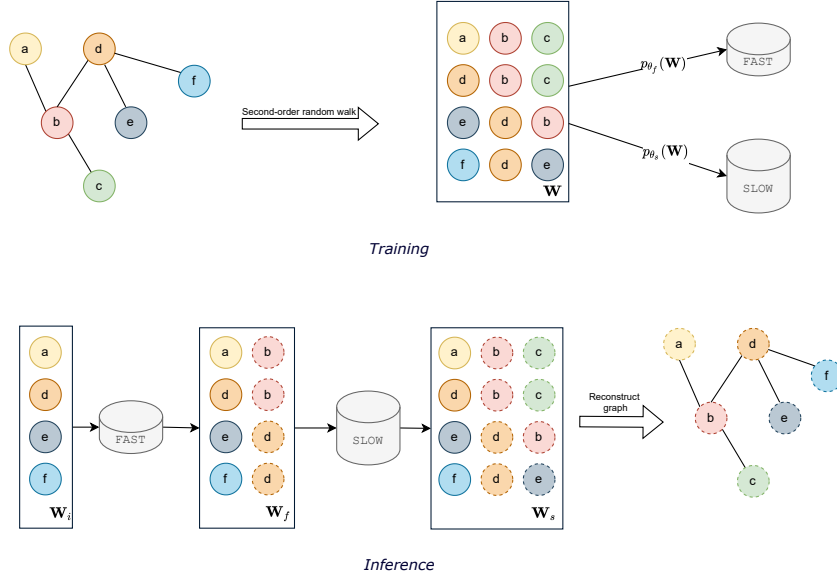


Figure 1. An overview of FLOWGEN: During training (top, Section 2.1), two auto-regressive models (FAST and SLOW) are trained on a corpus of random walks. The two models have the same architecture, but differ in size (number of parameters). During inference (below, Section 2.2), the two models are used in tandem for generating a graph. The FAST model generates the simpler, initial parts of the walk, and the SLOW model takes over for generating the latter, more challenging parts.

be used for graph generation. Next, we describe how we use two of these models for dynamically for efficient graph generation.

## 2.1. Graph generation using auto-regressive models

**Notation** We denote a graph by  $\mathcal{G}$ . A random walk  $w$  is a sequence of  $k$  nodes  $v_1, v_2, \dots, v_k$  obtained by traversing the  $\mathcal{G}$  for  $k$  steps starting from  $v_1$ . A random walk matrix of  $m$  such walks is denoted by  $\mathbf{W} \in \mathbb{R}^{m \times k}$ . An element  $v_{ij} \in \mathbf{W}$  denotes the  $j^{\text{th}}$  node in the  $i^{\text{th}}$  random walk. For a single random walk  $w$ ,  $v_i$  denotes the  $i^{\text{th}}$  node in  $w$ . The nodes connected to  $v_i$  are denoted by  $\text{Adj}(v_i)$ . The graph-generation model is trained in four steps.

1. **Generating walks:** Given a graph  $\mathcal{G}$ , we generate  $m$  walks, each of length  $k$  using second-order traversal (Grover & Leskovec, 2016). This yields a random walk matrix  $\mathbf{W} \in \mathbb{R}^{m \times k}$  (we include more details on sampling in Appendix C).
2. **Training:** We next train an auto-regressive language model  $p_\theta$  on  $\mathbf{W}$ . Specifically, we treat  $\mathbf{W}$  as a corpus of  $m$  random walks  $[w_1, w_2, \dots, w_k]$  from  $\mathcal{G}$ . The model is trained to generate the  $i^{\text{th}}$  node in the walk, conditioned on the preceding ( $< i$ ) nodes. We model the probability  $p(\mathbf{W})$  of a random walk as a series of conditional next token distributions:  $p(\mathbf{W}) = \prod_{i=1}^m \prod_{j=1}^k p_\theta(v_{ij} | v_{i, < j})$ . We parameterize  $p_\theta$  using a decoder-only language model based on the architecture used by GPT-

2 (Radford et al., 2019). The number of self-attention layers (or *depth*) of the language model decides the number of parameters  $\theta$ , and, consequently, the strength of the model. This setup is analogous to language model training, with each random walk acting as a sentence, and each node a token (Perozzi et al., 2014).

3. **Sampling walks:** During inference, an approximate random walk matrix  $\mathbf{W}'$  is obtained by randomly sampling from  $p(\mathbf{W})$ . To sample a random walk of length  $l$ , we first generate a random node  $v_1 \in \mathcal{G}$ . The generation process begins by  $v_2 \sim p_\theta(v | v_1)$ . The next token is then drawn by sampling  $v_3 \sim p_\theta(v | v_1, v_2)$ . The process is repeated for  $l-1$  steps to generate a random walk of size  $l$ . We generate  $n$ , and stack them to create a generated random walks matrix  $\mathbf{W}'$ .
4. **Reconstructing graph:**  $\mathbf{W}'$  is used to reconstruct the graph  $\mathcal{G}'$ . Downstream tasks are used to evaluate the quality of  $\mathcal{G}'$  vs.  $\mathcal{G}$ . We follow the two-step procedure used by Bojchevski et al. (2018) to assemble the generated graph  $\mathcal{G}'$  from generated random walks  $\mathbf{W}'$ . First,  $\mathbf{W}'$  is converted to a count matrix  $\mathbf{S}$ , where  $S_{ij}$  is the number of times the nodes  $v_i$  and  $v_j$  appeared consecutively (indicating an edge between  $v_i$  and  $v_j$ ). Next, an edge is added between  $v_i$  and  $v_j$  in the generated graph  $\mathcal{G}'$  with probability  $p_{ij} = \frac{S_{ij}}{\sum_{v \in \text{Adj}(i)} S_{iv}}$ .

**A note on evaluation** Note that a large model may simply remember a small graph. However, our goal is not such

memorization, but rather generalization. To evaluate this,  $\sim 20\%$  of the edges from  $\mathcal{G}$  are hidden during training.  $\mathcal{G}'$  is then evaluated for presence of these edges.

## 2.2. Fast and slow graph generation

Let  $v$  be a random walk to be generated. Consider a breakdown of the walk into two parts using chain rule. We posit that there is a  $k$  such that the first half and the second half of the model can generate wildly different parts

We train two different generation models (i.e., two different  $p_\theta$ ) using procedure outlined in Section 2.1: FAST and SLOW. Both these models have the same architecture type (transformers), but differ in the number of parameters: FAST is a 1-4 layered transformer whereas SLOW has 6 or more layers (depending on the graph). A speed vs. performance trade-off is expected for the FAST and SLOW models: FAST will struggle with generating new walks, whereas SLOW will generate these at the cost of slower inference.

Next, we note that *it is easier to generate the first few nodes of a random walk*: the first node of the walk is given (the starting point), and generating the second node requires an understanding of a second-order random walk. Generating subsequent random walks require models to pay attention to the walk seen so far and gets progressively more difficult as the walk length increases. Further, generating walks longer than  $k$  (random walk length used for training) requires a model with better generalization capabilities.

Our method, FLOWGEN, relies on these key intuitions to pair a fast and slow process together. We start by generating walks using a FAST model and then switch to a SLOW model to explore novel neighborhoods. Since generation is autoregressive, such a formulation is natural: subsequent walks can be conditioned on the walks seen so far without any changes to the two models.

**Switching from FAST to SLOW** While generating a walk of length  $l$ , the switch from FAST to SLOW can happen at any point  $v_j$ , where  $j \in (0, l)$ . The choice of  $v_j$  is important due to the speed vs. accuracy trade-off: a large  $j$  implies that the walk will be generated quickly but mainly using the FAST model. On the other hand, a smaller  $j$  will shift most of the responsibility to the SLOW model, for better accuracy but slower inference.

- **Neighborhood  $\mathcal{N}$ :** a sequence of  $p$  nodes that appear in a random walk. For instance, given a random walk  $(v_1, v_2, v_3, v_4, v_5)$ , and  $p = 4$ , the two neighborhoods are  $(v_1, v_2, v_3, v_4)$  and  $(v_2, v_3, v_4, v_5)$ .
- **Exploration and exploitation:** a random walk  $w$  to be in a state of *exploration* if it is in a neighborhood where it is discovering new neighborhoods not present in the training data. Otherwise, the walk is said to be

in the *exploitation* phase. As mentioned earlier, a random walk starts from a given node, and thus is expected to be in exploitation mode in the beginning (known neighborhoods), before switching to exploration mode (new neighborhoods). Both exploration and exploitation phases are essential: exploration helps the model generalize to new edges, whereas exploitation helps the model recreate the structure.

Given these definitions, a sweet spot for the handover point  $v_j$  will be the step where the random walk exits the exploration mode and enters the exploitation mode. To perform this check efficiently, we create a *bloom filter* (Bloom, 1970) of all the neighborhoods seen in the training data.

**Bloom filter** A bloom filter  $\mathcal{B}$  created over a set  $\mathbb{S}$  provides an efficient way to check if a key  $x$  does not exist in  $\mathbb{S}$ . Bloom filters are particularly useful in data-intensive applications, where an application might want to be sure about a query’s existence before checking an offline database (Broder & Mitzenmacher, 2004; Kleppmann, 2017). Given a search key  $x$ , if the search over  $\mathcal{B}$  is unsuccessful, it is guaranteed that  $x \notin \mathbb{S}$ . Otherwise,  $x$  may be present with a probability  $1 - P$ , where  $P$  is the false positive rate. Crucially, while creating the bloom filter incurs a one-time cost of  $\mathcal{O}(|\mathbb{S}|h)$ , the lookup can be done in  $\mathcal{O}(h)$  time where  $h$  is the number of hash functions (a detailed analysis and relevant algorithms in Appendix B). We generate 100M (second-order) random walks of length 16 for each graph. We re-use these walks to create a bloom filter  $\mathcal{B}$ . For each walk, we use a sliding window of length  $p = 4$  and inserted the neighborhood in  $\mathcal{B}$ . Note that this is a one-time procedure. Using a false-positive rate of  $P = 0.01$ , the  $\mathcal{B}$  is approximately  $130\times$  smaller than saving the neighborhoods in a hashmap on average.

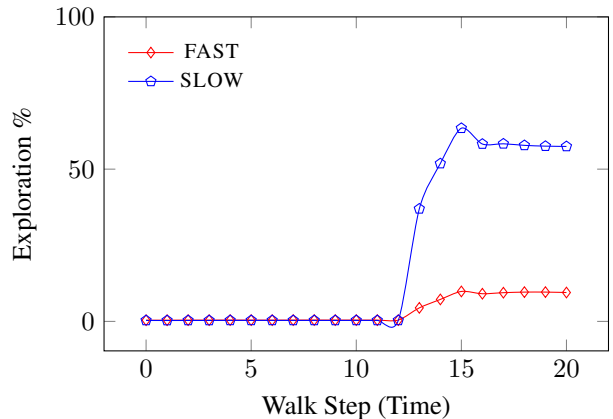


Figure 2. Exploration % (y-axis) vs. random walk step (x-axis) for CORAML (additional results in Figure 5). The larger SLOW model explores once the walk exceeds a certain threshold, whereas the lighter FAST model repeats the training data.

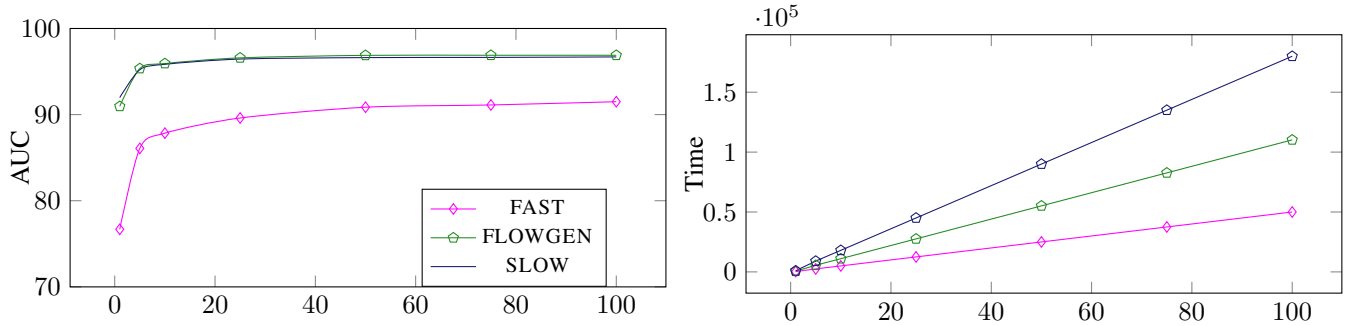


Figure 3. AUC and total time taken in seconds (y-axis) for the three models for CORAML, as the number of random walks sampled (x-axis) increases from 500k to 100M. FLOWGEN closely matches the performance of SLOW across scale, while taking a fraction of time.

**Calculating handover point** We calculate the handover point (the step where we switch from FAST to SLOW) for each graph separately. We create a bloom filter  $\mathcal{B}$  using all the four-node neighborhoods in the training data. For each graph, we generate 10,000 random walks of length  $l = 24$  using both FAST and SLOW models. Then, the handover point is calculated by finding the *knee* of the exploration % curve, and we use Satopaa et al. (2011) to find such points.<sup>1</sup> We plot the % of neighborhoods not found in  $\mathcal{B}$  (or exploration %) in Figure 2 for CORAML. We also experiment with using entropy for deciding the switch, but found it to be not effective for determining exploration vs. exploitation Appendix (D.3).

### 3. Experiments

We experiment with three graphs formed by citation networks (CITSEER (Sen et al., 2008), CORAML (Mccallum, 2008)) and political blogs (POLBLOGS (Adamic & Glance, 2005)). Dataset statistics are provided in Table 2. For the link prediction experiments, we use the train/test/val splits provided by Bojchevski et al. (2018).

We base FLOWGEN on a decoder-only transformer architecture. The SLOW models use six layers, and the FAST model uses a single layer. We do not perform any hyper-parameter tuning: all the models use the same hyperparameters, and use a single Nvidia 2080-ti for all experiments.

We use a modern open-source implementation of scalable bloom filters (Almeida et al., 2007).<sup>2</sup> The handover point was selected to be 13 for all the graphs. Note that the model was trained for 16 steps (i.e.,  $k = 16$ ). Instead of using a fixed handover point, we can also switch dynamically at each step. However, we found that constantly switching between models incurs a cost as the model has to perform a

forward pass on all the tokens seen so far. This is required, as the auto-regressive attention at each step depends on the hidden layer representations for all layers and previous steps. A static handover point avoids constant switching and does not degrade the performance.

Here, we show that our FLOWGEN generalizes as well as the SLOW model by reporting AUC on the link prediction task. Detailed results, including performance on structured metrics and additional analysis are included in Appendix D. While our primary goal is not establishing a state-of-the-art graph generation method, we find that FLOWGEN closely matches or outperforms existing graph-generation methods (Section D.1) for the three graphs.

	FAST	SLOW	FLOWGEN
CORAML	91.5 (5.0e-4)	96.7 (1.80e-3)	<b>96.9</b> (1.10e-3)
CITSEER	96.1 (6.2e-4)	<b>96.8</b> (1.72e-3)	96.5 (1.37e-3)
POLBLOGS	66.2 (4.8e-4)	<b>93.8</b> (1.56e-3)	<b>93.8</b> (1.08e-3)

Table 1. **Main results:** AUC for the link prediction task using FAST, SLOW, and FLOWGEN. The time (seconds) taken by for generating each walk is in parentheses. For each graph, 100M random walks of length 24 are sampled.

**Results** FLOWGEN closely matches or outperforms the larger model SLOW while taking a fraction of time. The size of the underlying graph also plays a role in how significant the gains are from our approach: FLOWGEN outperforms the SLOW model for the large POLBLOGS graphs. In contrast, the FAST model is competitive for a smaller graph like CITSEER. Further, Figure 3 shows that FLOWGEN continue to be effective across the scale of sampled walks.

### 4. Related Work

**Graph generation** Our work relies on using random walks for learning generative models of graph, similar to (Bojchevski et al., 2018) and (You et al., 2018). (You

<sup>1</sup><https://pypi.org/project/kneed/>

<sup>2</sup><https://github.com/joseph-fox/python-bloomfilter>

et al., 2018) learn a generative model of molecules, where each inference step generates the complete graph. Their setup also leverages graph-aware specialized decoding procedures, and scales for their setup since molecular graphs are typically small. In contrast, our random walk based method allows learning generative models of large graphs that cannot be generated in a single inference step. Additionally, in contrast with (Bojchevski et al., 2018) that use GAN-based training, we leverage relatively simple graph generation model. The idea of modeling random walks as sequence of nodes is identical to DeepWalk (Perozzi et al., 2014). However, different from DeepWalk, our main goal is generative graph modeling, and not learning node representations. Further, our underlying architecture (transformers) is also different than the one used by DeepWalk (MLP).

**Fast and slow machine learning** There are several works that use the fast-slow metaphor. For instance, Mujika et al. (2017) present a hierarchical RNN architecture, where the lower (or fast) layer contains one RNN cell for each time-step. The higher layer in contrast connects several different neurons together. Hill et al. (2020) focus on language reasoning tasks, where slow and fast denote the two phases of learning: slow supervised training, and a fast k-shot adaptation.

Our work is closest in spirit to the remarkable recent work by Schwarzschild et al. (2021b;a), who focus on three different generalization tasks. They observe increasing the number of test iterations (which corresponds to the network depth in their setting) helps the models in generalizing better to the difficult problem. Our study replicates this general finding, by showing that FAST (small) and SLOW (larger) models can be combined for efficient graph generation. Our method can be seen as an extension of their method for graph generation, with the following novel additions. First, instead of varying the depth of the network, we actually leverage two different transformer networks (FAST and SLOW), and the output of FAST is used by SLOW. Second, instead of hardcoding the switching point, we dynamically determine it using bloom filters. Schwarzschild et al. (2021b) note that the confidence of the model was a good proxy for correctness in their setting. We find that not to be the case, and also propose a method for finding a switching point for the network.

## 5. Conclusion

Future machine learning applications will potentially have API-level access to several models of varying strengths and costs of usage. In such scenarios, building systems that can adapt to the difficulty of the sample will be critical for scale and efficiency. FLOWGEN presents a real-world use case for such FAST-SLOW systems. The success of

transformers-based auto-regressive language model on such a corpus finds a parallel in existing language modeling research and presents novel insights that might help guide the development of efficient language generation models.



## References

- Adamic, L. A. and Glance, N. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pp. 36–43, 2005.
- Almeida, P. S., Baquero, C., Preguiça, N., and Hutchison, D. Scalable bloom filters. *Information Processing Letters*, 101(6):255–261, 2007.
- Anthony, T., Tian, Z., and Barber, D. Thinking Fast and Slow with Deep Learning and Tree Search. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL <https://proceedings.neurips.cc/paper/2017/hash/d8e1344e27a5b08cdfd5d027d9b8d6de-Abstract.html>.
- Bloom, B. H. Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 13(7): 422–426, 1970.
- Bojchevski, A., Shchur, O., Zügner, D., and Günnemann, S. Netgan: Generating graphs via random walks. In *International Conference on Machine Learning*, pp. 610–619. PMLR, 2018.
- Broder, A. and Mitzenmacher, M. Network applications of bloom filters: A survey. *Internet mathematics*, 1(4): 485–509, 2004.
- Chang, F., Feng, W.-c., and Li, K. Approximate caches for packet classification. In *IEEE INFOCOM 2004*, volume 4, pp. 2196–2207. IEEE, 2004.
- Christensen, K., Roginsky, A., and Jimeno, M. A new analysis of the false positive rate of a bloom filter. *Information Processing Letters*, 110(21):944–949, 2010.
- De Cao, N. and Kipf, T. MolGAN: An implicit generative model for small molecular graphs. *arXiv:1805.11973 [cs, stat]*, May 2018. URL <http://arxiv.org/abs/1805.11973>. arXiv: 1805.11973.
- Evans, J. S. B. T. Heuristic and analytic processes in reasoning\*. *British Journal of Psychology*, 75(4):451–468, 1984. ISSN 2044-8295. doi: 10.1111/j.2044-8295.1984.tb01915.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2044-8295.1984.tb01915.x>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2044-8295.1984.tb01915.x>.
- Frankish, K. Dual-Process and Dual-System Theories of Reasoning. *Philosophy Compass*, 5(10):914–926, 2010. ISSN 1747-9991. doi: 10.1111/j.1747-9991.2010.00330.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1747-9991.2010.00330.x>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1747-9991.2010.00330.x>.
- Grover, A. and Leskovec, J. node2vec: Scalable Feature Learning for Networks. *arXiv:1607.00653 [cs, stat]*, July 2016. URL <http://arxiv.org/abs/1607.00653>. arXiv: 1607.00653.
- Hill, F., Tieleman, O., von Glehn, T., Wong, N., Merzic, H., and Clark, S. Grounded Language Learning Fast and Slow. *arXiv:2009.01719 [cs]*, October 2020. URL <http://arxiv.org/abs/2009.01719>. arXiv: 2009.01719.
- Kahneman, D. Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review*, 93(5):1449–1475, 2003. ISSN 0002-8282. URL <https://www.jstor.org/stable/3132137>. Publisher: American Economic Association.
- Kahneman, D. *Thinking, fast and slow*. Macmillan, 2011.
- Kleppmann, M. *Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems*. ” O’Reilly Media, Inc.”, 2017.
- Lim, J., Hwang, S.-Y., Moon, S., Kim, S., and Youn Kim, W. Scaffold-based molecular design with a graph generative model. *Chemical Science*, 11(4):1153–1164, 2020. doi: 10.1039/C9SC04503A. URL <https://pubs.rsc.org/en/content/articlelanding/2020/sc/c9sc04503a>. Publisher: Royal Society of Chemistry.
- Mccallum, A. K. Automating the Construction of Internet Portals with Machine Learning. pp. 37, 2008.
- Mi, L., Zhao, H., Nash, C., Jin, X., Gao, J., Sun, C., Schmid, C., Shavit, N., Chai, Y., and Anguelov, D. HDMapGen: A Hierarchical Graph Generative Model of High Definition Maps. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4225–4234, Nashville, TN, USA, June 2021. IEEE. ISBN 978-1-66544-509-2. doi: 10.1109/CVPR46437.2021.00421. URL <https://ieeexplore.ieee.org/document/9577586/>.
- Mujika, A., Meier, F., and Steger, A. Fast-Slow Recurrent Neural Networks. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL <https://proceedings.neurips.cc/paper/2017/hash/e4a93f0332b2519177ed55741ea4e5e7-Abstract.html>.

- Perozzi, B., Al-Rfou, R., and Skiena, S. Deepwalk: On-line learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 701–710, 2014.
- Posner, M. I. and Snyder, C. R. Attention and cognitive control. 1975.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. Language models are unsupervised multitask learners. OpenAI blog, 1(8):9, 2019.
- Samanta, B., De, A., Jana, G., Chattaraj, P. K., Ganguly, N., and Rodriguez, M. G. NeVAE: A Deep Generative Model for Molecular Graphs. Proceedings of the AAAI Conference on Artificial Intelligence, 33:1110–1117, July 2019. ISSN 2374-3468, 2159-5399. doi: 10.1609/aaai.v33i01.33011110. URL <https://aaai.org/ojs/index.php/AAAI/article/view/3903>.
- Satopaa, V., Albrecht, J., Irwin, D., and Raghavan, B. Finding a”kneedle” in a haystack: Detecting knee points in system behavior. In 2011 31st international conference on distributed computing systems workshops, pp. 166–171. IEEE, 2011.
- Schwarzschild, A., Borgnia, E., Gupta, A., Bansal, A., Emam, Z., Huang, F., Goldblum, M., and Goldstein, T. Datasets for Studying Generalization from Easy to Hard Examples. arXiv:2108.06011 [cs], September 2021a. URL <http://arxiv.org/abs/2108.06011>. arXiv: 2108.06011.
- Schwarzschild, A., Borgnia, E., Gupta, A., Huang, F., Vishkin, U., Goldblum, M., and Goldstein, T. Can You Learn an Algorithm? Generalizing from Easy to Hard Problems with Recurrent Networks. In Advances in Neural Information Processing Systems, volume 34, pp. 6695–6706. Curran Associates, Inc., 2021b. URL <https://proceedings.neurips.cc/paper/2021/hash/3501672ebc68a5524629080e3ef60aef-Abstract.html>.
- Sen, P., Namata, G., Bilgic, M., Getoor, L., Galligher, B., and Eliassi-Rad, T. Collective classification in network data. AI magazine, 29(3):93–93, 2008.
- Shiffrin, R. M. and Schneider, W. Controlled and automatic human information processing: II. perceptual learning, automatic attending and a general theory. Psychological Review, 84:127–190, 1977.
- Stanovich, K. E. Individual differences in reasoning: Implications for the rationality debate? BEHAVIORAL AND BRAIN SCIENCES, pp. 82, 2000.
- You, J., Ying, R., Ren, X., Hamilton, W., and Leskovec, J. Graphrnn: Generating realistic graphs with deep auto-regressive models. In International Conference on Machine Learning, pp. 5708–5717. PMLR, 2018.
- Zhou, D., Zheng, L., Han, J., and He, J. A Data-Driven Graph Generative Model for Temporal Interaction Networks. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 401–411, Virtual Event CA USA, August 2020. ACM. ISBN 978-1-4503-7998-4. doi: 10.1145/3394486.3403082. URL <https://dl.acm.org/doi/10.1145/3394486.3403082>.

**Algorithm 1:** Creating a bloom filter with  $M$  bits and  $h$  hash functions  $\mathbf{H}$  over a set  $\mathbb{S}$ . Each hash function takes  $\mathcal{O}(1)$ , and thus creating a bloom filter incurs a one time cost  $\mathcal{O}(h|\mathbb{S}|)$ .

---

**Given:**  $\mathcal{B}, \mathbf{H}, \mathbb{S}$   
**Init:**  $\mathcal{B}(i) \leftarrow 0; i \in [1, M]$   
**for**  $q \in \mathbb{S}$  **do** //  $\mathcal{O}(|\mathbb{S}|)$   
    **for**  $i \leftarrow 1, 2, \dots, h$  **do** //  $\mathcal{O}(|\mathbf{H}|) = \mathcal{O}(k)$   
         $\mathcal{B}(\mathbf{H}_i(q)) \leftarrow 1$   
    **end**  
**end**

---

**Algorithm 2:** Querying a bloom filter. The cost is a fixed constant  $\mathcal{O}(h)$ .

---

**Given:**  $\mathcal{B}, \mathbf{H}$   
**for**  $i \leftarrow 1, 2, \dots, h$  **do** //  $\mathcal{O}(h)$   
    **if**  $\mathcal{B}(\mathbf{H}_i(q)) = 0$  **then** // certainly absent  
        **return** *False*  
    **end**  
**end**  
 /\* Maybe present with a false  
    positive rate  $p$ . \*/  
**return** *True*

---

## A. Dataset details

Table 2 lists the details of the three graphs used in our experiments.

	$N_{LCC}$	$E_{LCC}$
CORAML	2,810	7,981
CITeseer	2,110	3,757
POLBLOGS	1,222	16,714

Table 2. Graphs statistics. The  $N_{LCC}$  and  $E_{LCC}$  refer to the number of nodes and edges in the largest connected component.

## B. Bloom Filters

A bloom filter  $\mathcal{B}$  over a set  $\mathbb{S}$  is a data structure for efficient set-membership queries. The time to search is independent of the number of elements in  $\mathbb{S}$ . As a trade-off, a bloom filter can generate false positives (indicate that a query  $q \in \mathbb{S}$  when it is absent). We will return to an analysis of false-positive rate after expanding on details of a bloom filter.

Given a search key  $x$ , if the search over  $\mathcal{B}$  is unsuccessful, it is guaranteed that  $x \notin \mathbb{S}$ . Otherwise,  $x$  may be present with a probability  $1 - P$ , where  $P$  is the false positive rate. Internally, a bloom filter  $\mathcal{B}$  is implemented as an array of  $M$  bits accompanied by  $h$  hash functions  $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_h$ . To add an element  $x \in \mathbb{S}$  to  $\mathcal{B}$ , each of the  $h$  hash functions

map  $x$  to  $[1, M]$ , and thus the corresponding bits are set to 1. Concretely,  $\mathcal{B}[\mathbf{H}_i(x)] = 1 \forall i \in [1, h]$ .

To check the presence of an element  $x$  in  $\mathcal{B}$ , it suffices to check if  $\exists i \in [1, h] \mathcal{B}[\mathbf{H}_i(x)] = 0$ . If so, then it is guaranteed that  $x \notin \mathbb{S}$  (otherwise, all the bits would be set to 1). Otherwise, the element *may be* present. Crucially, while creating the bloom filter incurs a one-time cost of  $\mathcal{O}(|\mathbb{S}|h)$ , the lookup can be done in  $\mathcal{O}(h)$  time. Combined with the space requirements for  $\mathcal{B}$ ,  $M \ll |\mathbb{S}|$ , a bloom filter provides an efficient way to determine if an element is absent from a set. The key elements in the design of a bloom filter are its size  $M$ ,  $h$  hash functions  $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_h$ , and the size of set  $\mathbb{S}$  over which search operations are to be performed.

Algorithms 1 and 2 show the algorithms for creating and querying a bloom filter, respectively.

One of the biggest follies of a bloom filter are its false positive rates. Chang et al. (2004) proposed bucketed bloom filters to alleviate the false positive rate. In their method, each hash function  $\mathbf{H}_i$  maps to the indices  $[(i-1) * m + 1, m]$ , where  $m = M/h$  is the number of bits in each bucket.

Let  $P$  be the rate of false positives,  $|\mathbb{S}| = n$ . Allowing each bucket of bloom filter to be 50% full, it can be shown that the number of elements  $n \sim M^{\frac{(\ln 2)^2}{|\ln P|}}$  (Almeida et al., 2007). See Christensen et al. (2010) for a comprehensive analysis of false positive rate for classical implementation of bloom filters.

We next approximate the size of bloom filter required for storing all neighborhoods of a graph  $\mathcal{G}$ . Let  $|\mathbb{V}|$  be the number of nodes in  $\mathcal{G}$ . Let  $d_{\max}$  be the max-degree of  $\mathcal{G}$ . Then, the number of neighborhoods  $\mathcal{N}$  of size  $p$  are upper-bounded by  $|\mathbb{V}| * d_{\max}^{p-1}$ . Clearly, this can be non-tractable for large, dense graphs. However, if  $d_{\max}$  is a fixed constant, then the number of neighborhoods is  $\mathcal{O}(|\mathbb{V}|)$  ( $d_{\max}^{p-1}$  is absorbed in the constant). Thus, for such graphs, bloom filter can be tractably grown. Crucially, note that our goal is not to store all the graphs. Rather, we want to only approximately answer the membership queries in the graph.

## C. Second-order sampling for generating the training data

For completeness, we now present the second order sampling method used by Grover & Leskovec (2016) that we adopt for generating the training data for our system.

Following the notation used by Grover & Leskovec (2016), let  $t$  be the previous node visited by the walk, and  $v$  be the current node (i.e., the walk just traversed  $[t, v]$ ). The distribution over the next node  $x$ ,  $p(x | t, v)$ , is given as  $p(x | t, v) = \frac{\pi(x, t)}{\sum_{y \in \text{Adj}(v)} \pi(y, t)}$ . Here,  $\pi(x, t)$  is defined as



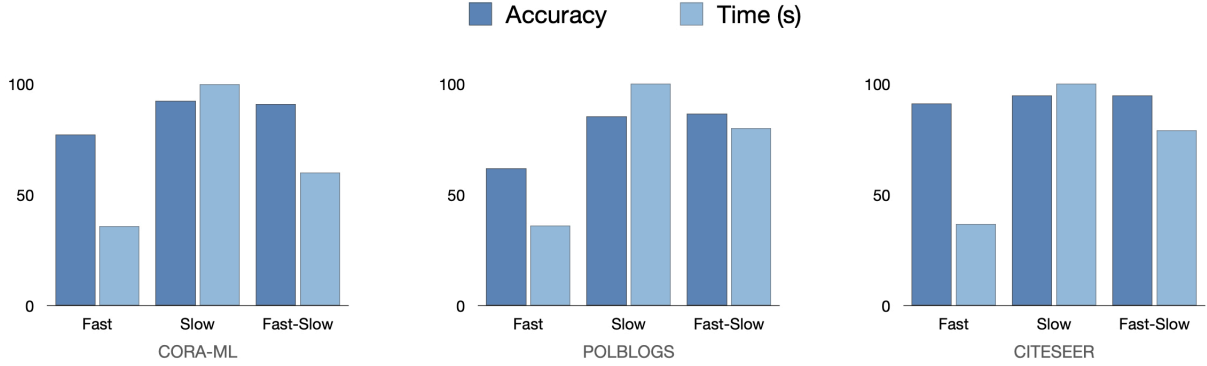


Figure 4. AP and time for three different graphs using FAST-SLOW

follows:

$$\pi(x, t) = \begin{cases} \frac{1}{p} & \text{if } d_{tx} = 0 \\ 1 & \text{if } d_{tx} = 1 \\ \frac{1}{q} & \text{if } d_{tx} = 2 \end{cases}$$

The parameter  $p$  decides the likelihood of revisiting a node. Specifically, a low  $p$  will encourage the walk to go back to the node  $t$  recently visited. Similarly,  $q$  controls the likelihood of the walk visiting new nodes. A lower value of  $q$  will encourage the walk to move towards node that are farther away from the node recently visited, allowing higher exploration. Following [Bojchevski et al. \(2018\)](#), we set  $p = q = 1$  to balance between the two properties. For more insights into the properties of second order random walk, please see Section 3.2 of ([Grover & Leskovec, 2016](#)).

## D. Additional Experiments

### D.1. Can auto-regressive language models successfully learn generative models of graphs?

In contrast with prior work on generative graph modeling, our backbone graph-generation model is a simple transformer-based language model. The simplicity of this method allows us to experiment with the fast and slow settings easily. In this section, we show that this simplicity does not come at the cost of performance.

Table 3 shows the results for comparing structural properties of generated graph for CORAML. As the Table shows, FLOWGEN outperforms state-of-the-art graph generation models (Additional results from structural generation are in the Appendix).

Table 4 shows the performance of FLOWGEN on link prediction tasks. Similar to link prediction, FLOWGEN outperforms strong state-of-the-art methods for all the graphs.

### D.2. Performance of FLOWGEN with scale

How does the performance of FLOWGEN change as the scale of data increases? To test this, we vary the number of random walks  $n$  generated during inference to recreate the graph. The results are shown in Figure 3. FLOWGEN matches or outperforms SLOW, while being consistently faster across the number of walks. Table 7 shows the results for two different scales.

### D.3. Using entropy for deciding the switch

Recall that we are using an auto-regressive language model for generating the walks. Thus, at each step  $i$ , the model generates a distribution over the next node,  $p(v_i | v_1, v_2, \dots, v_{i-1})$ . Thus, for a well calibrated model, in the exploitation phase, when the model is still generating walks from the training set, the entropy of this distribution will be fairly low (the model will be confident about the next node), and that the entropy will increase further in the walk. If that was the case, the entropy of the distribution can be a useful indicator of the switching point. We investigate the same in this section.

Specifically, we generate a walk of length 32, and for each step  $i$ , we calculate the entropy of the distribution  $p(v_i | v_1, v_2, \dots, v_{i-1})$ . The average entropy at each step is calculated, and the knee ([Satopaa et al., 2011](#)) of the entropy plot is used as the switching point. The results are shown in Figures 7 and 8. As the Figures show, the knee point is detected early on for all the cases (within 4 steps), and then fluctuates around a mean value.

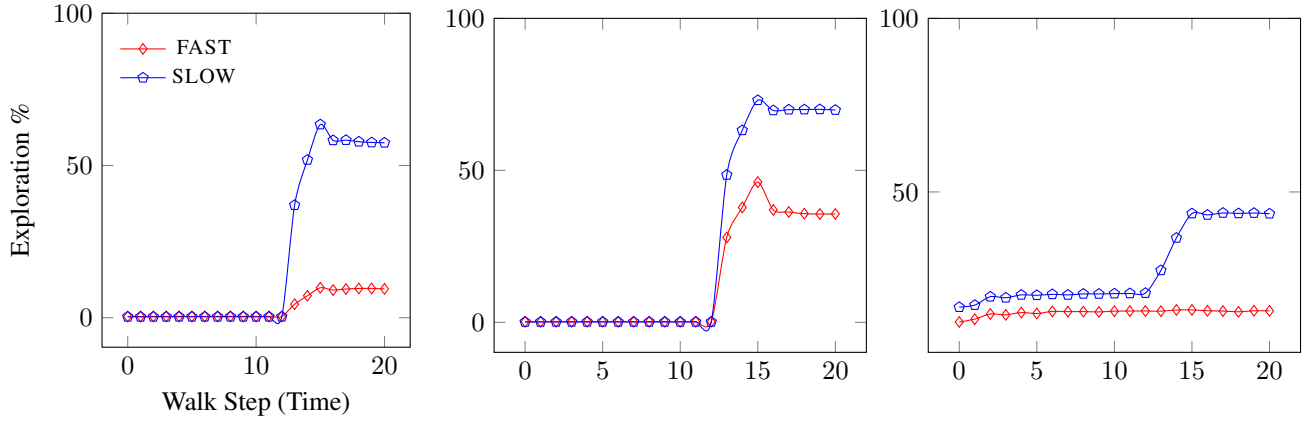


Figure 5. Exploration % (y-axis) vs. random walk step for CORAML (left), CITESEER (middle), and POLBLOGS (right). For all the graphs, the larger SLOW model explores once the walk exceeds a certain threshold, whereas the lighter FAST model repeats the training data.

Graph	Max. degree	Assort-activity	Triangle Count	Power law exp.	Inter-comm. unity density	Intra-comm. unity density	Cluster-ing coeff.	Charac. path len.
CORA-ML	240	-0.075	2,814	1.860	4.3e-4	1.7e-3	2.73e-3	5.61
Netgan	<b>233</b>	-0.066	1,588	1.793	6.0e-4	<b>1.4e-3</b>	2.44e-3	5.20
FLOWGEN (ours)	224	<b>-0.080</b>	<b>2,123</b>	<b>1.857</b>	<b>5.4e-4</b>	1.3e-3	<b>2.50e-3</b>	<b>5.40</b>

Table 3. Comparison of FLOWGEN with Netgan for structural metrics for CORAML. The ground truth values are listed in the top-row, and the value closer to the ground truth is highlighted in bold. FLOWGEN closely matches the ground truth graph on a larger number of metrics.

Method	CORAML	CITESEER	POLBLOGS
Adamic/Adar	92.16	88.69	85.43
DC-SBM	96.03	94.77	95.46
node2vec	92.19	95.29	85.10
VGAE	95.79	95.11	93.73
NetGAN	95.19	96.30	<b>95.51</b>
FLOWGEN (ours)	<b>96.93</b>	<b>97.03</b>	95.00

Table 4. Area under the curve (AUC) for link prediction. FLOWGEN outperforms or matches strong baselines.

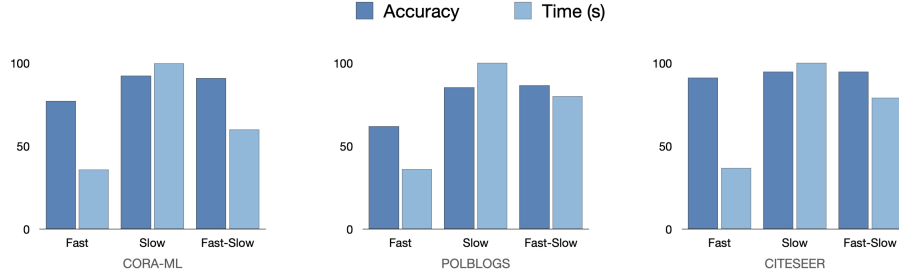


Figure 6. Average precision vs. time taken for the three graphs. The FAST and SLOW model speed-accuracy trade-off is apparent: FAST model is fast but less accurate (average precision  $\sim 75\%$ , compared to the SLOW model which is slower but has average precision of  $92\%$ . FLOW combines the strengths of the two modes: it achieves an accuracy of  $90\%$  while being  $\sim 50\%$  faster than the SLOW model. Note that the time is normalized relative to SLOW (SLOW takes  $100\%$  of the time).

Method	CORAML		CORA		CITSEER		DBLP		PUBMED		POLBLOGS	
	AUC	AP	AUC	AP	AUC	AP	AUC	AP	AUC	AP	AUC	AP
Adamic/Adar	92.16	85.43	93.00	86.18	88.69	77.82	91.13	82.48	84.98	70.14	85.43	92.16
DC-SBM	96.03	95.15	98.01	97.45	94.77	93.13	<b>97.05</b>	96.57	<b>96.76</b>	95.64	95.46	94.93
node2vec	92.19	91.76	<b>98.52</b>	<b>98.36</b>	95.29	94.58	96.41	96.36	96.49	95.97	85.10	83.54
VGAE	95.79	96.30	97.59	97.93	95.11	96.31	96.38	<b>96.93</b>	94.50	<b>96.00</b>	93.73	94.12
NetGAN (500K)	94.00	92.32	82.31	68.47	95.18	91.93	82.45	70.28	87.39	76.55	95.06	94.61
NetGAN (100M)	95.19	95.24	84.82	88.04	96.30	96.89	86.61	89.21	93.41	94.59	<b>95.51</b>	94.83
FLOWGEN (100M)	<b>96.93</b>	<b>97.22</b>	87.68	87.79	<b>97.03</b>	<b>97.45</b>	84.21	84.89	90.86	91.16	95.00	<b>95.05</b>

Table 5. Comparison of FLOWGEN with baselines on link prediction task for six different graphs.

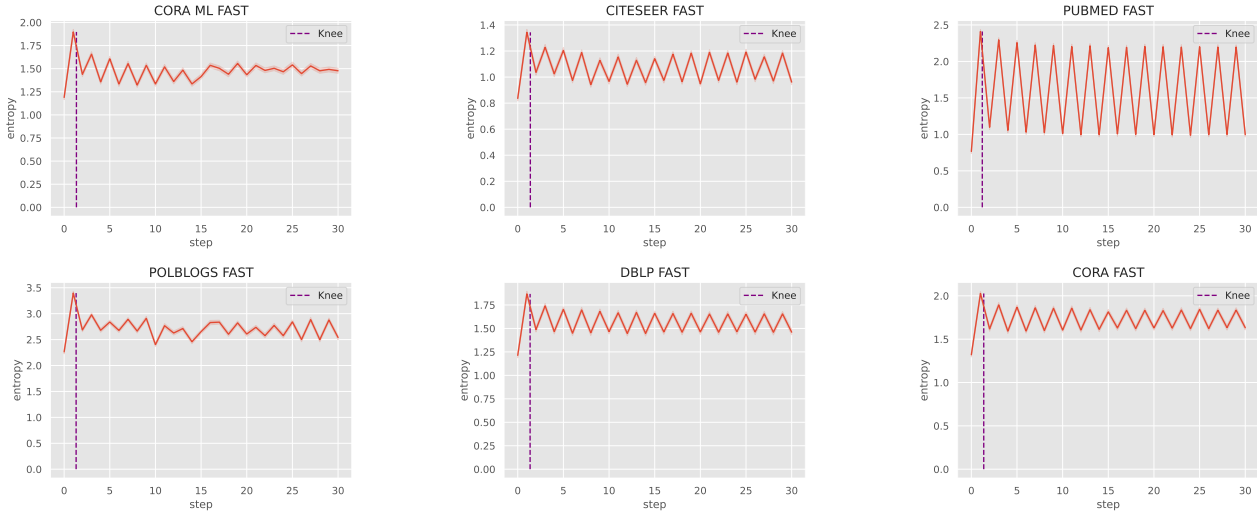


Figure 7. Entropy analysis for the FAST models

Graph	Max. degree	Assortativity	Triangle Count	Power law exp.	Inter-comm. unity density	Intra-comm. unity density	Clustering coeff.	Charac. path len.	Average rank
CORA-ML	240	-0.075	2,814	1.860	4.3e-4	1.7e-3	2.73e-3	5.61	
Netgan VAL	199	-0.060	1,410	1.773	6.5e-4	1.3e-3	2.33e-3	5.17	3.00
Netgan EO	233	-0.066	1,588	1.793	6.0e-4	1.4e-3	2.44e-3	5.20	1.75
FLOWGEN	224	-0.080	2,123	1.857	5.4e-4	1.3e-3	2.50e-3	5.40	

Table 6. Comparison of FLOWGEN with baselines for CORAML on structural metrics.

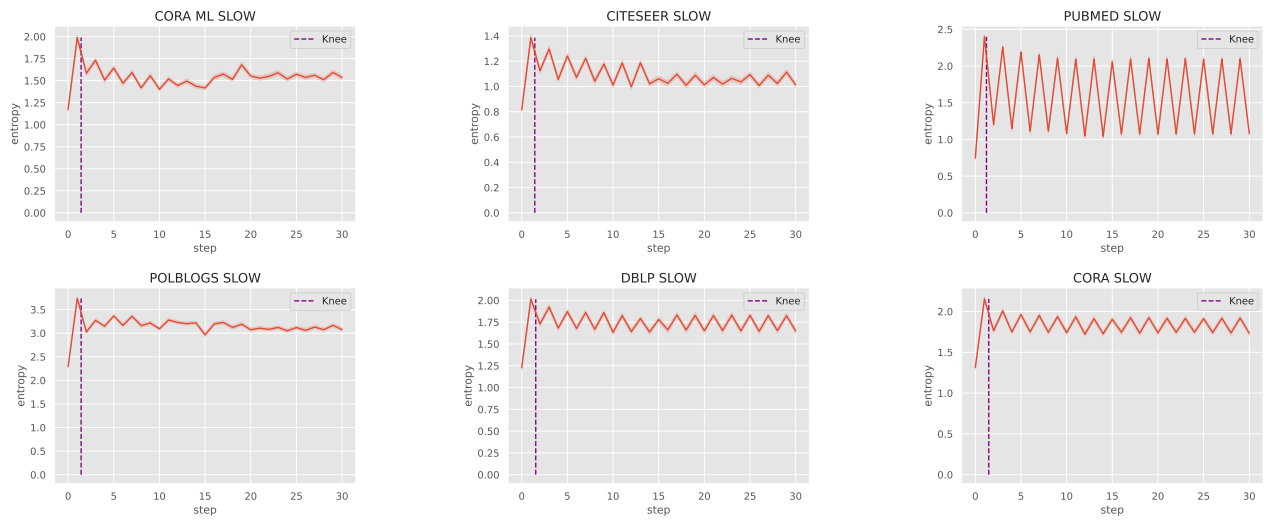


Figure 8. Entropy analysis for the SLOW models

	FAST	SLOW	FLOWGEN
CORAML (500k)	76.8 (288)	<b>92.2</b> (806)	90.8 (484)
CORAML (100M)	91.5 (50k)	96.7 (180k)	<b>96.9</b> (110k)
CITeseer (500k)	90.9 (313)	<b>94.6</b> (862)	93.3 (687)
CITeseer (100M)	96.1 (62k)	<b>96.8</b> (172k)	96.5 (137k)
POLBLOGS (500k)	61.9 (309)	85.4 (854)	<b>86.5</b> (686)
POLBLOGS (100M)	66.2 (48k)	<b>93.8</b> (156k)	<b>93.8</b> (108k)

*Table 7. Main results:* AUC for FAST, SLOW, and FLOWGEN, a combination of FAST-SLOW models. The total time (seconds) taken by each setup is in parentheses. FLOWGEN closely matches or outperforms the larger model SLOW while taking a fraction of time.