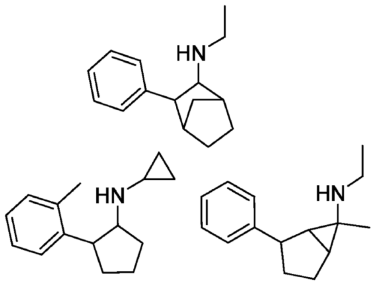
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

Prof. R. Wattenhofer

# Towards better graph generative models and globally accurate evaluation metrics




Graph generation is an interesting problem which is encountered when trying to generate novel molecules or proteins with desired properties or variations of social graphs that used to test hypotheses in social sciences. However, graph generation is a challenging problem due to the non-uniqueness of graphs and the complex non-local dependencies between their edges. Adding to this complexity, it is very challenging to compare different graph generation models. Typically, comparing graph distributions is performed by computing and comparing some statistics that are invariant to node permutations [10, 12, 14]. We believe that the existing statistics are insufficient because they are all based on local graph features and none of them are designed to capture non-local graph properties such as the diameter of a graph, its coarse structure or its general spectral properties.

In this project, we will investigate new measures that are effective for capturing global graph properties. To do so, we take inspiration from the unpublished work [11, Chapter 3.2] where general graph spectral features are defined. The student tasks will include 1. designing experiments and datasets emphasizing the shortcomings of existing metrics, 2. developing the theory of [11, Chapter 3.2] from the perspective of graph distribution comparison 3. developing better graph generative models and 4. an experimental comparison to existing models and metrics [3, 8, 15, 2, 16, 6, 13, 1, 7, 5, 9, 4].

To develop a better graph generation model, we will aim to create a denoising generative model, which can better capture those global properties of the graph. We hope to achieve this, by generating a graph in a latent space of a graph auto encoder, similarly to what is done for high-definition image generation. The latent space would likely be represented as a set of embeddings, those could potentially be made hierarchical. The developed models would be evaluated on larger molecule, protein and potentially other graph generation (e.g. object meshes) tasks, to confirm better performance on larger graphs compared to existing models.

**Requirements:** Strong motivation, knowledge in deep learning, or a solid background in machine learning. Experience with Python and PyTorch or TensorFlow is an advantage as well as knowledge in graph theory, generative models and graph neural networks.

**Interested? Please contact us for more details!**

## Contact

- Karolis Martinkus: martinkus@ethz.ch, ETZ G60.1

- Nathanaël Perraudin: nperraud@ethz.ch, ETZ G60.1

## References

[1] N. De Cao and T. Kipf. MolGAN: An implicit generative model for small molecular graphs. *ICML 2018 workshop on Theoretical Foundations and Applications of Deep Generative Models*, 2018.

[2] K. K. Haefeli, K. Martinkus, N. Perraudin, and R. Wattenhofer. Diffusion models for graphs benefit from discrete state spaces. In *NeurIPS 2022 Workshop: New Frontiers in Graph Learning*.

[3] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 2020.

[4] J. Jo, S. Lee, and S. J. Hwang. Score-based generative modeling of graphs via the system of stochastic differential equations. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2022.

[5] I. Krawczuk, P. Abranches, A. Loukas, and V. Cevher. Gg-gan: A geometric graph generative adversarial network. 2020.

[6] R. Liao, Y. Li, Y. Song, S. Wang, W. Hamilton, D. K. Duvenaud, R. Urtasun, and R. Zemel. Efficient graph generation with graph recurrent attention networks. In *Advances in Neural Information Processing Systems*, pages 4255–4265, 2019.

[7] J. Liu, A. Kumar, J. Ba, J. Kiros, and K. Swersky. Graph normalizing flows. *Advances in Neural Information Processing Systems*, 32:13578–13588, 2019.

[8] K. Martinkus, A. Loukas, N. Perraudin, and R. Wattenhofer. SPECTRE: Spectral conditioning helps to overcome the expressivity limits of one-shot graph generators. *Proceedings of the 39th International Conference on Machine Learning*, 2022.

[9] C. Niu, Y. Song, J. Song, S. Zhao, A. Grover, and S. Ermon. Permutation invariant graph generation via score-based generative modeling. In *International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 4474–4484, Online, 26–28 Aug 2020. PMLR.

[10] L. O'Bray, M. Horn, B. Rieck, and K. Borgwardt. Evaluation metrics for graph generative models: Problems, pitfalls, and practical solutions. In *International Conference on Learning Representations*, 2021.

[11] N. Perraudin. *Graph-based structures in data science: fundamental limits and applications to machine learning*. PhD thesis, Ecole Polytechnique Fédérale de Lausanne, 2017.

[12] H. Shirzad, K. Hassani, and D. J. Sutherland. Evaluating graph generative models with contrastively learned features. *arXiv preprint arXiv:2206.06234*, 2022.

[13] M. Simonovsky and N. Komodakis. Graphvae: Towards generation of small graphs using variational autoencoders. In *International conference on artificial neural networks*, pages 412–422. Springer, 2018.

[14] R. Thompson, B. Knyazev, E. Ghalebi, J. Kim, and G. W. Taylor. On evaluation metrics for graph generative models. In *International Conference on Learning Representations*, 2021.

[15] C. Vignac, I. Krawczuk, A. Siraudin, B. Wang, V. Cevher, and P. Frossard. Digress: Discrete denoising diffusion for graph generation. *arXiv preprint arXiv:2209.14734*, 2022.

[16] J. You, R. Ying, X. Ren, W. L. Hamilton, and J. Leskovec. Graphrnn: Generating realistic graphs with deep auto-regressive models. In *ICML*, 2018.