# DiPol-GAN: Generating Molecular Graphs Adversarially with Relational Differentiable Pooling

**Pablo Rivas**
Department of Computer Science
Marist College
Poughkeepsie, NY 12601
Pablo.Rivas@Marist.edu

**Michael Guarino**
Department of Computer Science
Marist College
Poughkeepsie, NY 12601
michael.guarino1@marist.edu

**Alexander Shah**
Department of Computer Science
Marist College
Poughkeepsie, NY 12601
alexander.shah2@marist.edu

## 1 Problem Background and Methods

The extension of deep learning to graph data structures has become a popular research topic with compelling applications to the field of chemistry [1]. Of particular interest is the impact of deep generative models to *de novo* drug discovery. Drug discovery describes the process by which new candidate medications are found with the goal of identifying a new target molecule with specific desired properties. The key challenge presented is the vast size of the chemical space and the discrete nature of molecular structures [2]. The drug discovery process is expensive as even *de novo* approaches rely on some semblance of brute force. Improvements in methods to discover new drugs with desired properties would have great impact. Our work presents a method to generate previously undiscovered molecular graphs with specific desired properties.

Generative models have been making astounding progress in the field of computer vision and natural language processing showing significant improvements in the quality of generated samples. Prominent approaches to generative modeling are autoregressive models [3], variational autoencoders (VAEs) [4], and generative adversarial networks (GANS) [5], each excelling in solving very specific types of problems. For the task of image synthesis autoregressive models generate high quality images but can be slow to evaluate and do not have a latent representation. VAEs train via maximum likelihood estimation making them more numerically stable; however, they can produce lower quality results [6]. GANs implicitly learn model parameters without having to specify a likelihood [7] and are known to produce higher quality results but are susceptible to generating lower diversity samples (a.k.a. mode collapse), and are prone to numerical instability making them difficult to train even with recent advances in adversarial training theory [8, 9, 10, 11].

Generative models have started to make their way into the domain of graphs with works such as [7, 12, 13, 14, 15, 16]; however, learning to generate graphs presents problems for gradient based learning due to the discrete nature of graphs because of their arbitrary connectivity [16]. Likelihood-based models for graph generation are known to be stable but require expensive approximate graph matching procedures as seen in GraphVAE [16] or require some fixed ordered representations of the graph as seen in Junction Tree VAE (JT-VAE) [14], which often is not feasible. Autoregressive models suchs as presented in GraphRNN [15] offer an interesting approach building a graph by conditionally modeling sequences of nodes where the sequence generators are jointly aware of each other; however, they lack a latent representation that would allow us to optimize for a desired chemical property.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

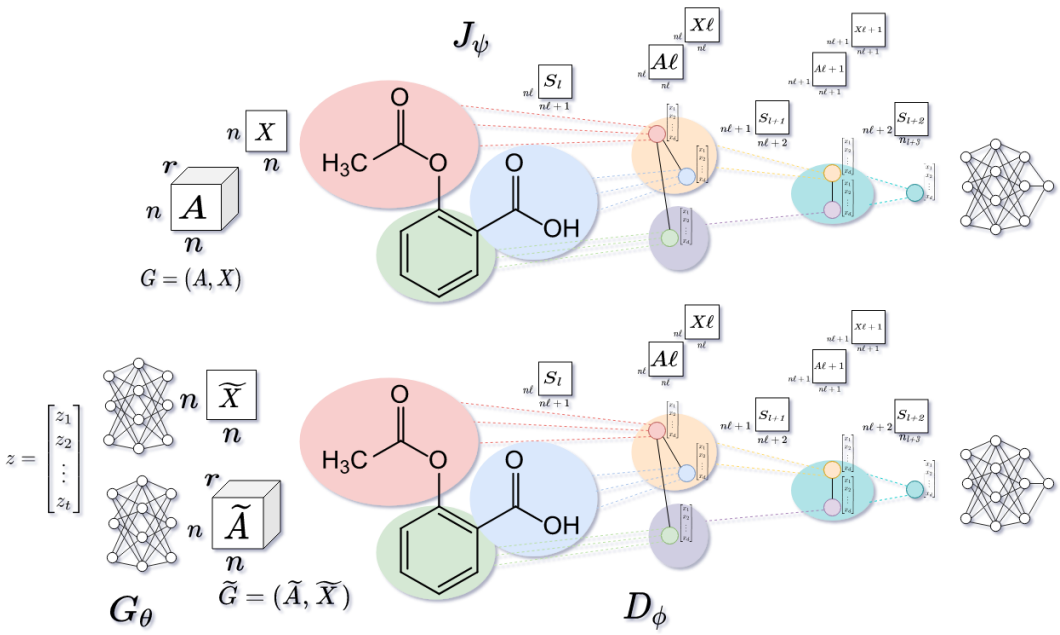


Figure 1: Proposed model architecture. $J_\psi$ is a policy network, $G_\theta$ is the generator, and $D_\phi$ is the discriminator in a graph-based generative adversarial network. The generator is a graph $G = (\mathbf{A}, \mathbf{X})$ where $\mathbf{A}$ is the adjacency matrix, and $\mathbf{X}$ is the features matrix. The model implements a reinforcement learning reward mechanism.

GANs seem to present a clear advantage for the task of graph generation because they allow us to implicitly learn the data generation distribution without having to formalize a likelihood that necessitates a graph matching strategy as in the likelihood-based models (*e.g.* VAEs) while also allowing for the optimization of the learned distribution towards a desired chemical property.

In this work we present DiPol-GAN, summarized in Figure 1, which is a GAN-based approach to generating complete graph structures that are resistant to graph isomorphisms by learning to predict discrete connections. Through the use of a reinforcement learning objective DiPol-GAN's learned generation distribution is encouraged towards a desired chemical property. DiPol-GAN implements Differentiable Pooling (DIFFPOOL) [17] in its discriminator, which learns to aggregate nodes on the molecular graph in a hierarchical way improving classification accuracy and the quality of learned graph embeddings. We believe that, through improving the discriminator's architecture, it will encourage the generator to learn higher quality graph representations.

## 2 Results and Conclusions

Table 1 shows the results obtained with the proposed model on the QM9 dataset [18]. Reportedly, the proposed solutions are highly unique and diverse, which means that the molecules produced are not simply regurgitations of the existing (and known to be) valid molecules. However, in spite of the great results obtained, there are a few performance remarks that need to be noted. Using the QM9 dataset for the task of goal directed molecular graph generation can be problematic in that there are 134k molecular graph structures lacking software to calculate chemical properties on a GPU in a timely manner. However, nearly all related work reports lengthy training times with GraphVAE.

Table 1: Comparisons with related molecular graph generation work, presented are test results.

| Algorithm | Valid (%) | Uniqueness | Diversity | Druglikeliness | Time (s) |
|---|---|---|---|---|---|
| MolGAN | $77.5 \pm 42.0$ | $0.098 \pm 0.16$ | $0.877 \pm 0.21$ | $0.513 \pm 0.22$ | $16473 \pm 651$ |
| DiPo (WGAN) | $100 \pm 0.0$ | $0.883 \pm 0.01$ | $0.999 \pm 0.45$ | $0.648 \pm 0.21$ | $19484 \pm 134$ |

In summary, DiPOLGAN presents a method for generating molecular graphs with specific chemical properties via hierarchical differentiable pooling to improve discriminator performance while also optimizing towards a a reinforcement learning objective via the policy network. Through learning hierarchical graph representations it is our intuition from representation learning that our discriminator and policy networks are learning to hierarchically filter learned representations toward information present to improve each specific network's objective and therefore implicitly encourage the generation of realistic molecular graphs.

## References


[1] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1263–1272. JMLR. org, 2017.

[2] Pavel G Polishchuk, Timur I Madzhidov, and Alexandre Varnek. Estimation of the size of drug-like chemical space based on gdb-17 data. *Journal of computer-aided molecular design*, 27(8):675–679, 2013.

[3] Aaron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. *arXiv preprint arXiv:1601.06759*, 2016.

[4] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[6] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

[7] Nicola De Cao and Thomas Kipf. Molgan: An implicit generative model for small molecular graphs. *arXiv preprint arXiv:1805.11973*, 2018.

[8] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In *Advances in neural information processing systems*, pages 2234–2242, 2016.

[9] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems*, pages 5767–5777, 2017.

[10] David Berthelot, Thomas Schumm, and Luke Metz. Began: Boundary equilibrium generative adversarial networks. *arXiv preprint arXiv:1703.10717*, 2017.

[11] Naveen Kodali, Jacob Abernethy, James Hays, and Zsolt Kira. On convergence and stability of gans. *arXiv preprint arXiv:1705.07215*, 2017.

[12] Bidisha Samanta, Abir De, Gourhari Jana, Pratim Kumar Chattaraj, Niloy Ganguly, and Manuel Gomez-Rodriguez. Nevae: A deep generative model for molecular graphs. *arXiv preprint arXiv:1802.05283*, 2018.

[13] Jiaxuan You, Bowen Liu, Zhitao Ying, Vijay Pande, and Jure Leskovec. Graph convolutional policy network for goal-directed molecular graph generation. In *Advances in Neural Information Processing Systems*, pages 6410–6421, 2018.

[14] Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction tree variational autoencoder for molecular graph generation. *arXiv preprint arXiv:1802.04364*, 2018.

[15] Jiaxuan You, Rex Ying, Xiang Ren, William L Hamilton, and Jure Leskovec. Graphrnn: Generating realistic graphs with deep auto-regressive models. *arXiv preprint arXiv:1802.08773*, 2018.

[16] Martin Simonovsky and Nikos Komodakis. Graphvae: Towards generation of small graphs using variational autoencoders. In *International Conference on Artificial Neural Networks*, pages 412–422. Springer, 2018.

[17] Zhitao Ying, Jiaxuan You, Christopher Morris, Xiang Ren, Will Hamilton, and Jure Leskovec. Hierarchical graph representation learning with differentiable pooling. In *Advances in Neural Information Processing Systems*, pages 4800–4810, 2018.

[18] Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1, 2014.