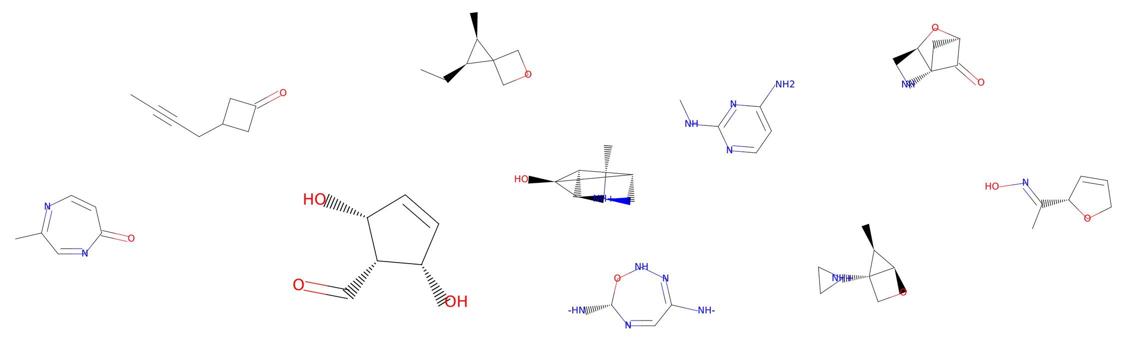
Discriminator-Driven Diffusion Mechanisms for Molecular Graph Generation

Dr. Gerrit Großmann
gerritgrossmann.de

Problem Setting

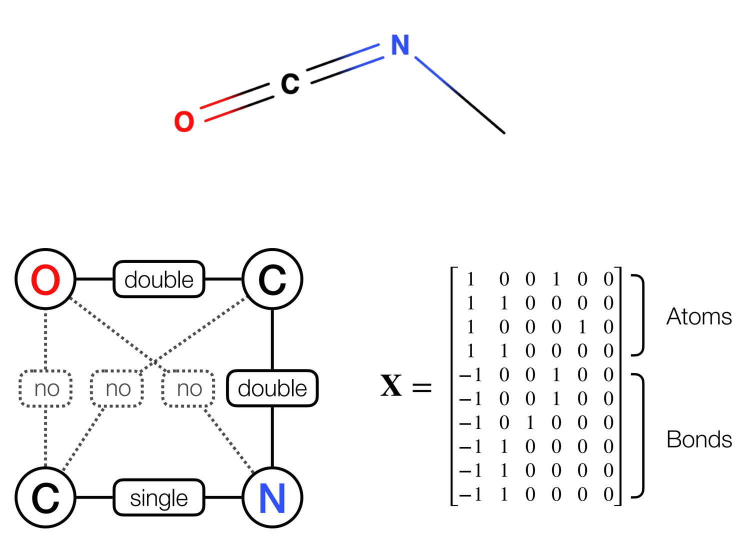


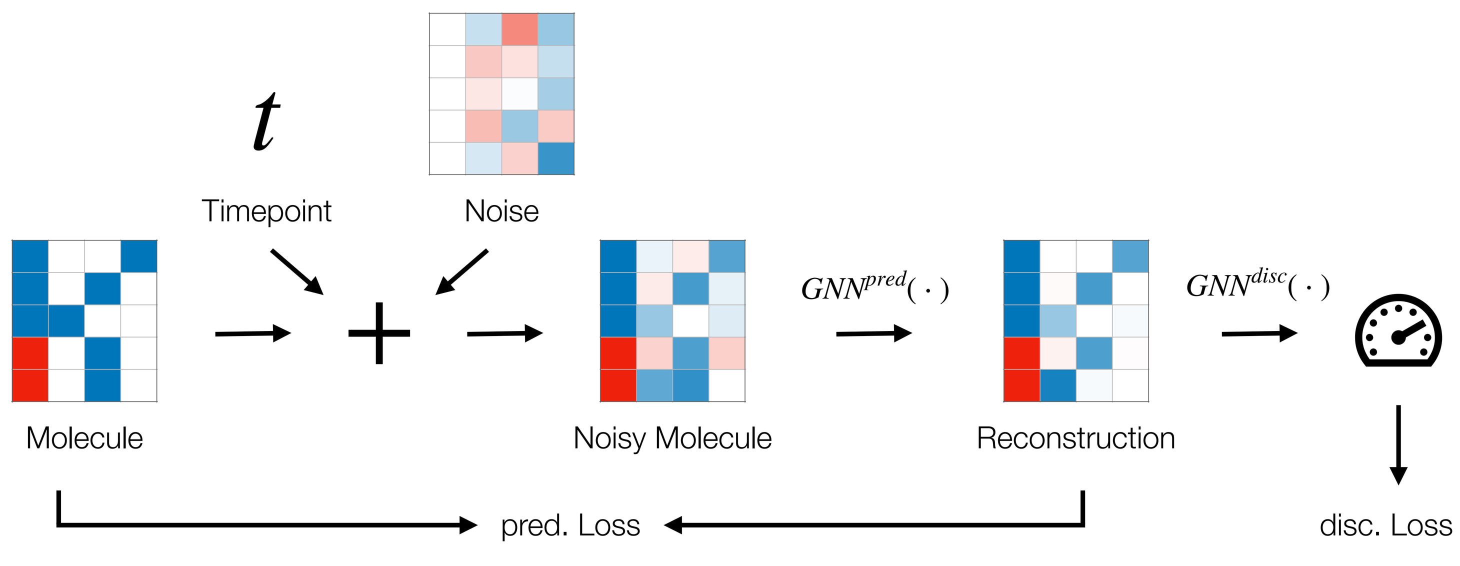
- Given: Set of trainings data (molecules) without knowledge of the underlying data distribution
- Wanted: Method to generate new molecules that appear to follow the data distribution.

Motivation

- Generating graphs and molecules is significant in numerous scientific fields.
- Diffusion models are state-of-the-art (SOTA) for various generative tasks.
- They struggle with complex constraints and the convex loss function is not well-suited for binary data.
- We propose that incorporating an adverbial loss could enhance the quality of generation.

Method





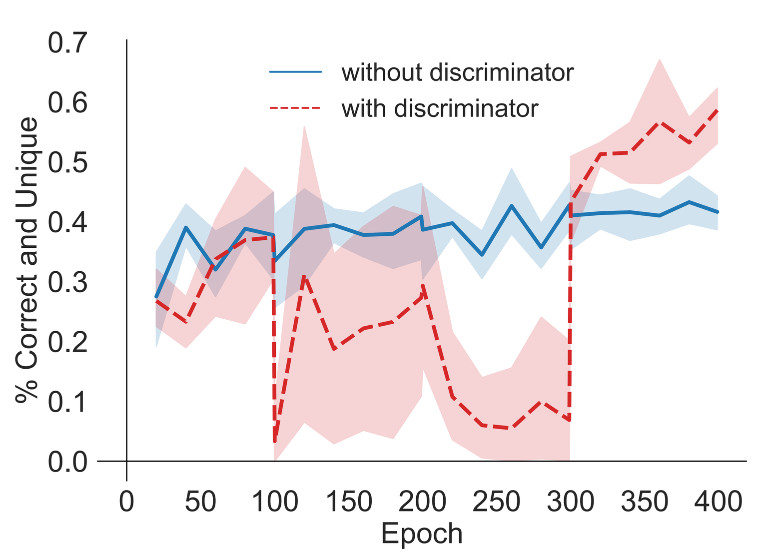
Representation:

- Molecules are represented as fully connected molecular graphs.
- Special type-indicator nodes on each edge indicate the absence, presence, or type of a covalent bond.
- Node features encode bond-type or the element.

Schematic Overview:

- Training: During training, molecules from the training data are sampled, noise is added, and a graph neural network (GNN) is trained to denoise these samples.
- Our method: In addition to measuring how well the loss is removed, we also measure how well the denoised molecule aligns with the data distribution, using a secondary GNN.
- Inference: To generate new molecules, the process begins with pure noise, which is incrementally removed step-by-step using the trained GNN.

Results



Conclusion: The integration of a discriminator loss results in a noticeable improvement in sample quality, albeit not to a large extent.

Future Work

- Adapt the method to other domains characterized by complex structural constraints.
- Explore diffusion models trained with a purely adversarial loss for the reverse process.
- Implement neuro-symbolic formal constraints as an alternative to a neural network-based discriminator.
- Utilize the discriminator loss for guidance during inference, in addition to its use during training.

Saarland Informatics Campus
German Research Center for Artificial Intelligence (DFKI)