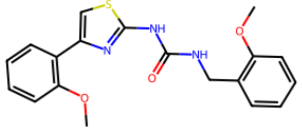
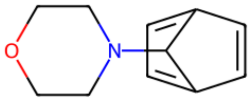
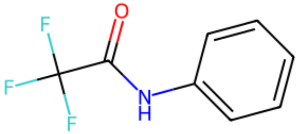
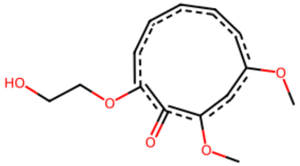
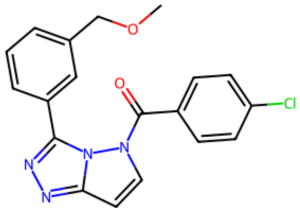






GUIDING DIFFUSION WITH LOGICAL CONSTRAINTS: MOLECULAR GRAPH GENERATION UNDER LIPINSKI'S RULES

Emma Meneghini
Supervised by Prof. Sergi Abadal Cavallé (UPC - BarcelonaTech)
and Prof. Nicolò Navarin (University of Padua)

Master's Degree in Data Science, UPC 2025

Part of this work was developed into a contribution presented at the
Special Session on Neural Networks for Graphs and Beyond at ICANN 2025, and at the
AI for Complex Systems Satellite at CCS 2025.

CONTEXT & MOTIVATION

De novo drug discovery

Generative diffusion models

Generating new drug-like molecules.

SLOW and COSTLY!

- Strong theoretical foundations,
- Efficiency,
- SOTA on molecular datasets,
- **Conditional variants.**

RESEARCH GAP

CURRENT GUIDANCE CAPABILITIES

Individual properties,

Conjunctions of properties.

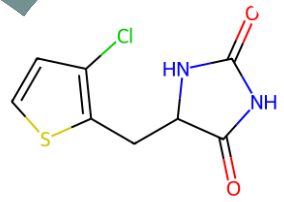
OPEN PROBLEMS

Complex constraints involving multiple properties are treated as black-boxes;

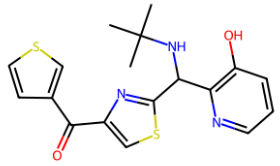
Guided diffusion models have not been systematically used to generate molecules at scale.

THESIS CONTRIBUTIONS

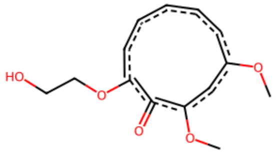
Framework for logical guidance in diffusion models



Preserve probabilistic and logical meaning of the rule.



Cardinality constraints in full DNF.



Per-property supervision.

THESIS METHODS

Diffusion Model

- DiGress [Vignac, 2023].

Dataset

- GuacaMol [Brown, 2019].

Constraint

- Lipinski's Rule of Five [Lipinski, 2000].

Evaluation Metrics

- Rule compliance,
- Validity,
- Distributional similarity to the training set,
- Diversity (within the generated set and relative to the training set).

Baselines

- Simple conjunction (all K properties),
- Black-box guidance.

DIGRESS: DIFFUSION FOR GRAPHS

Main idea of every diffusion model:

1. <u>Forward process</u>: gradually corrupt data with noise;
2. <u>Reverse process</u>: train a neural network (θ) to denoise step by step;
3. <u>Sampling</u>: generate a clean realistic sample from pure noise $p_\theta(G^{t-1}|G^t)$.

CLASSIFIER GUIDANCE WITH DIGRESS

Conditional denoising process:

$$G^{t-1}|G^t, y_G \sim \underbrace{p_\theta(G^{t-1}|G^t)}_{\text{unconditional denoising process}} \quad \underbrace{p_\eta(\hat{y}|G^{t-1})}_{\text{predicted probability of the final graph satisfying the property}}$$

Bayesian

equivalent perspective

KL-regularised variational problem:

- Maximise the similarity with the unconditional denoising distribution,
- Maximise the utility function i.e., linear approximation of the log-probability.

$$p_\eta(\hat{y}|G^{t-1}) \propto \exp\big(\underset{\text{guidance strength knob}}{\lambda}\underbrace{\langle \nabla_{G^t} \log \dot{q}_\eta(y_G|G^t), G^{t-1}\rangle}_{\text{utility}}\big)$$

LOGICAL GUIDANCE

Consider a logical formula defined on a set of Boolean variables $\mathcal{X} = \{X_1, \dots, X_K\}$.

Treat each X_i as a Bernoulli random variable.

The probability of satisfying the rule can then be modelled as a Bernoulli random variable, obtained by aggregating the Bernoulli's of the individual X_i's.

If:

- the satisfying assignments of the formula are known (i.e., its full DNF is known), and
- the relationship between the properties is known,

then the satisfaction probability can be computed.

for computational reasons

CARDINALITY CONSTRAINTS

Consider cardinality constraints of the form
$\sum_{i=1}^{K} X_i \geq r$.

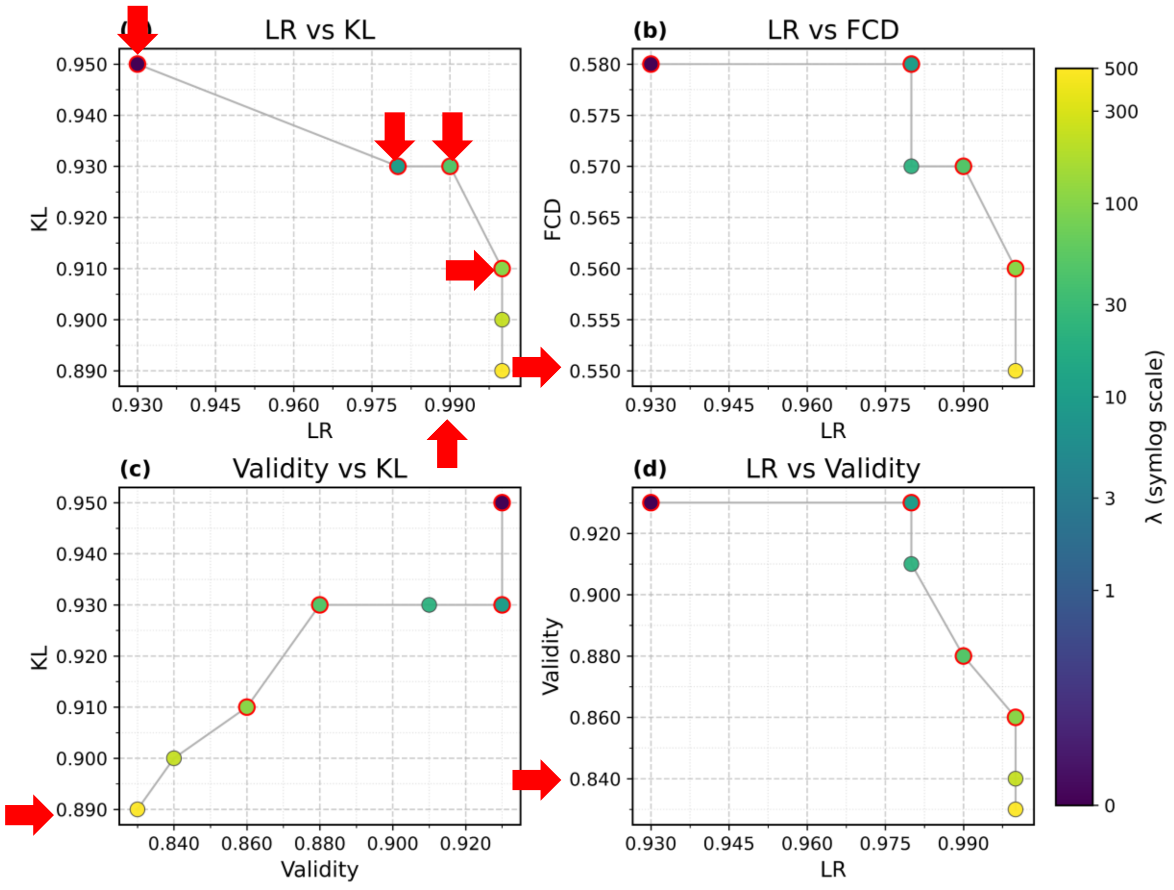
Assume that the variables (properties) are conditionally independent given G^t, and that their probabilities can differ.

$$q(y_G|G^t) = \text{Bernoulli}(\pi(G^t)_r)$$

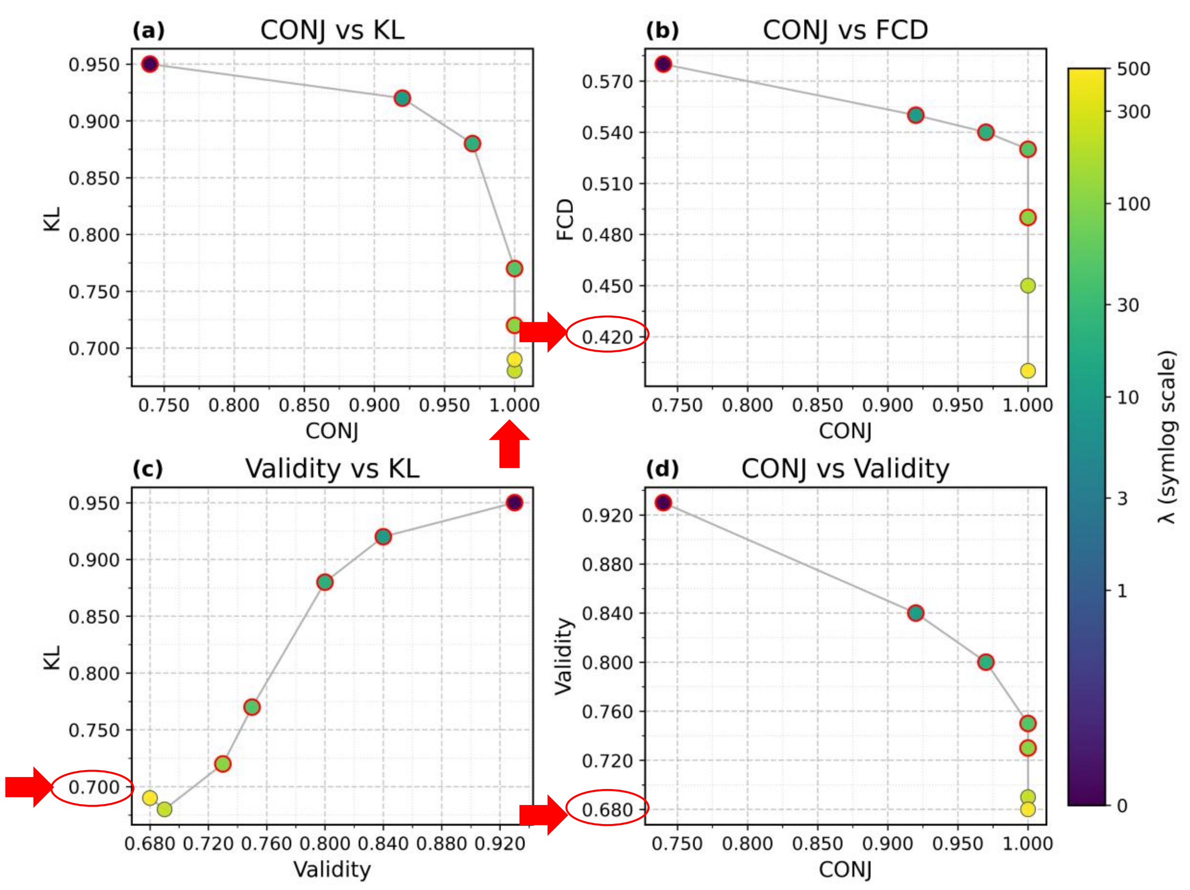
$$\pi(G^t)_r = \sum_{\substack{S \subseteq [K] \\ |S| \geq r}} \left(\prod_{i \in S} p_i \right) \left(\prod_{j \notin S} (1 - p_j) \right)$$

Lipinski's rule: at least 3 of logP ≤ 5, molWt ≤ 500 (Da), HBD ≤ 5, HBA ≤ 10.

LIPINSKI'S VS. CONJUNCTIVE RULES



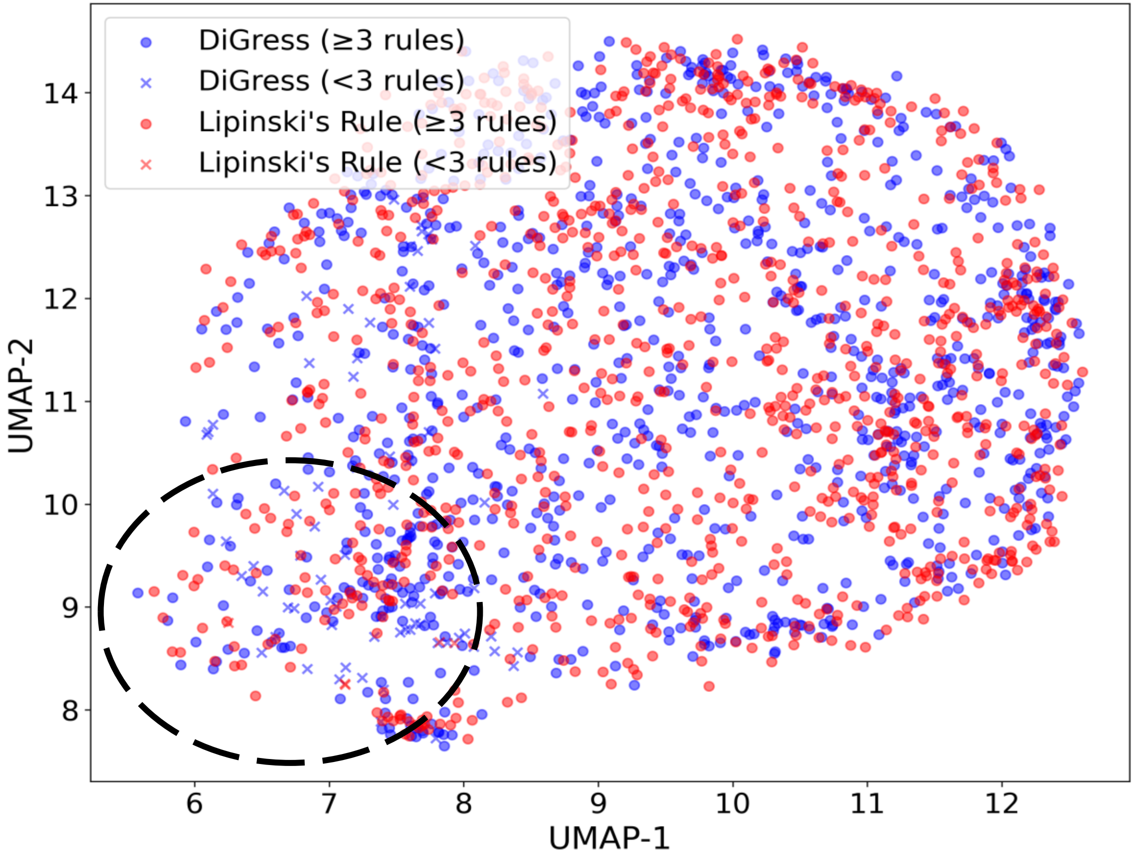
(a) Lipinski's rule



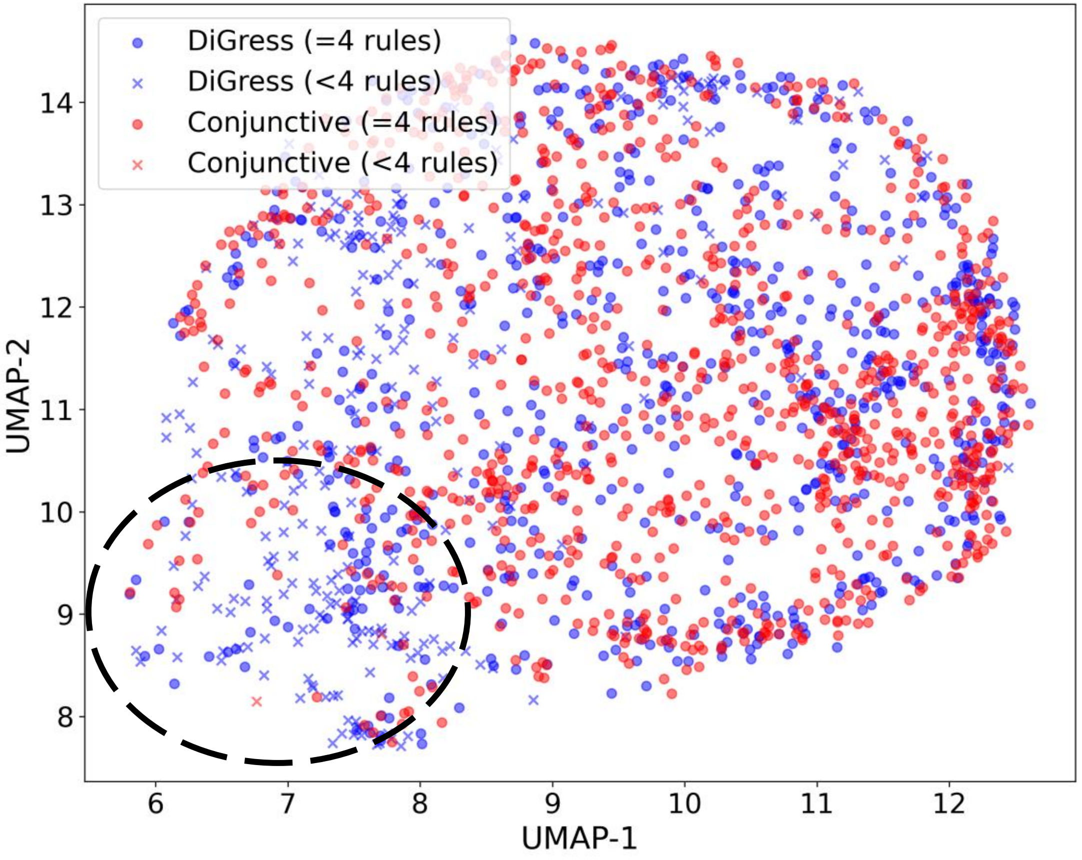
(b) Conjunctive rule

4 independent classifiers,
one for each property

LIPINSKI'S VS. CONJUNCTIVE RULES



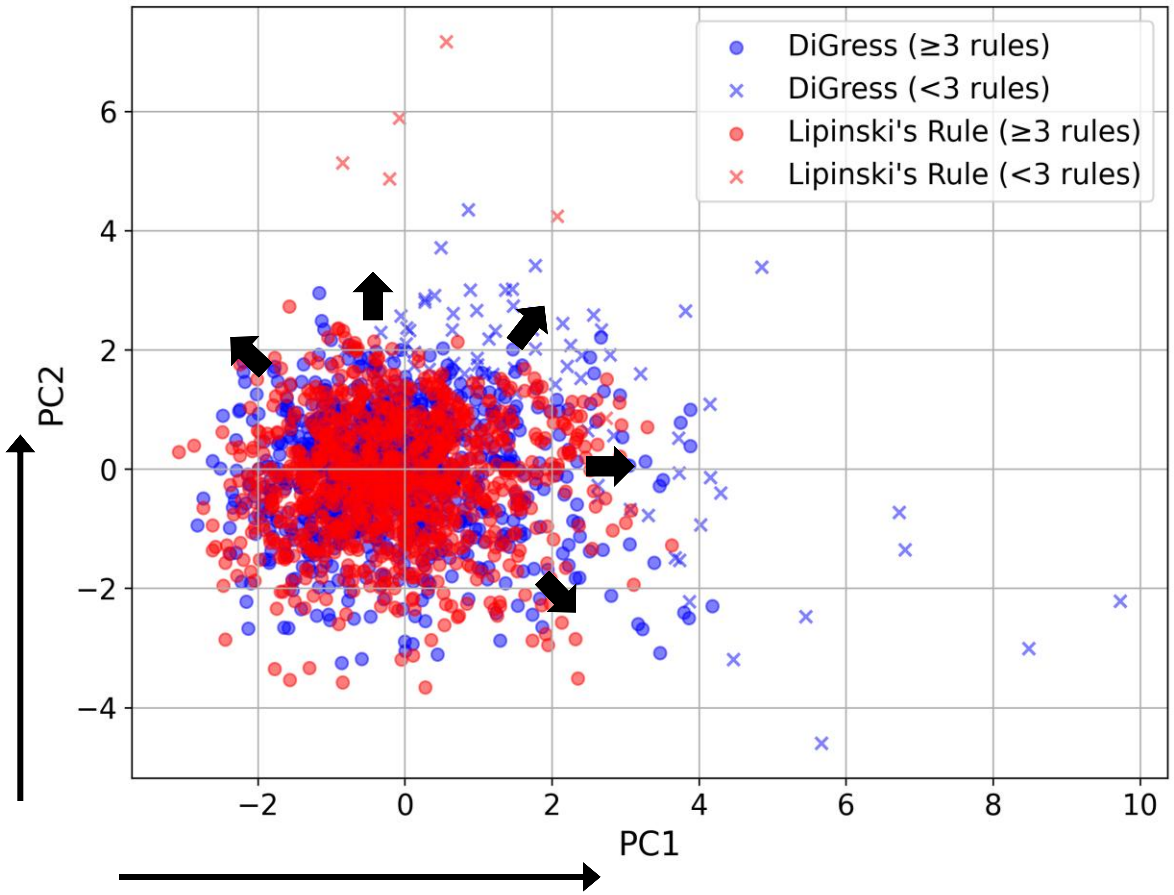
(a) Lipinski's rule



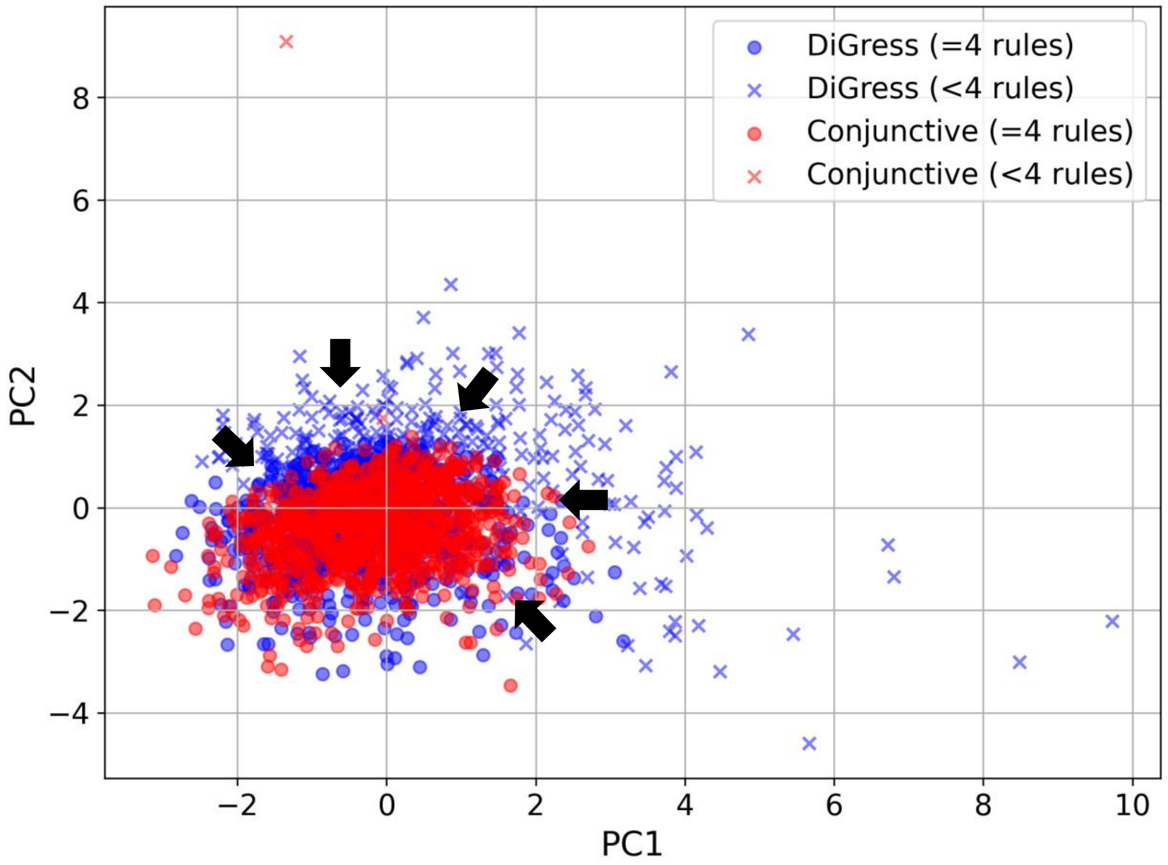
(b) Conjunctive rule

4 independent classifiers, one for each property; $\lambda = 47.82$

LIPINSKI'S VS. CONJUNCTIVE RULES



(a) Lipinski's rule



(b) Conjunctive rule

4 independent classifiers, one
for each property; $\lambda = 47.82$

THEORETICAL PROPERTIES

Assume that

i. the predictors are not negatively correlated under q_λ, i.e. $\mathrm{Cov}_{q_\lambda}(p_j, p_k) \geq 0$ for all j, k;

ii. the rule probability is non-decreasing in each p_j (true for cardinality rules).

1

Then, for any fixed reverse step t and any $\lambda \geq 0$ we have

$$\frac{d}{d\lambda}\mathbb{E}_{q_\lambda}[p_k] \geq 0.$$

Furthermore, let $S = \sum_{i=1}^{K} X_i$ with mean $\mu_S(p) = \sum_{i=1}^{K} p_i$. It follows that

$$\frac{d}{d\lambda}\mathbb{E}_{q_\lambda}[\mu_S] \geq 0.$$

THEORETICAL PROPERTIES

2

Define $p_r(p) \coloneqq \mathbb{P}(S = r | p)$ and assume that $\sigma^2_{\mu_S} = \mathbb{V}_{q_\lambda}[\mu_S] < +\infty$.

Using the assumptions in the previous slide it follows that, for every r and $c > 0$,

- $\frac{d}{d\lambda}\mathbb{E}_{q_\lambda}[p_r] \geq 0$ whenever $\mathbb{E}_{q_\lambda}[\mu_S] \leq r - 1 - c\sigma_{\mu_S}$, and
- $\frac{d}{d\lambda}\mathbb{E}_{q_\lambda}[p_r] \leq 0$ whenever $\mathbb{E}_{q_\lambda}[\mu_S] \geq r + 1 + c\sigma_{\mu_S}$

up to a tail probability of order ${}^1/_{c^2}$ by Chebyshev's inequality.

3

Let $p_K(p) = \prod_{j=1}^K p_j$. Using the assumptions in the previous slide it follows that

$$\frac{d}{d\lambda}\mathbb{E}_{q_\lambda}[p_K] \geq 0.$$

COMPOSITION OF THE PROPERTY DISTRIBUTIONS

- Increase in 4-of-4 proportion.

- 3-of-4 proportion not guaranteed to vanish;
- 3-of-4 proportion decreases as $\mu_S \geq 3$.

Experimental results:

- Independent-2L: $[77.9, 82.9]$
- Shared-1L: $[79.2, 90.0]$
- Shared-2L: $[77.7, 88.0]$

- Independent-2L $[17.1\%, 19.9\%]$
- Shared-1L: $[9.9\%, 20\%]$
- Shared-2L: $[11.9\%, 21\%]$

Better preserved when $\lambda \in [10, 104]$.

LOGICAL GUIDANCE VS. BLACK-BOX GUIDANCE

- For **Lipinski's rule**, it achieves the same compliance as our logical method.
- The proportion of molecules that satisfies exactly 3 properties increases with λ: $[19.4\%, 26.1\%]$.
- The proportion of molecules that satisfies exactly 4 properties decreases with λ: $[73.9\%, 79.6\%]$.

Shortcut learning!

Learns to satisfy molWt, HBD and HBA, which are the most frequent in the training samples that satisfy the rule.

This neglects part of the satisfying set!

CONCLUSIONS & FUTURE WORK

Key Takeaways

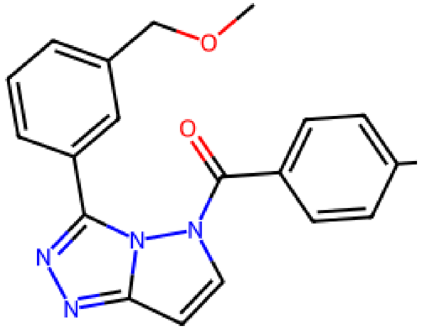
- First framework for embedding logical rules (full DNF cardinality constraints) into graph diffusion models for conditional generation at scale.

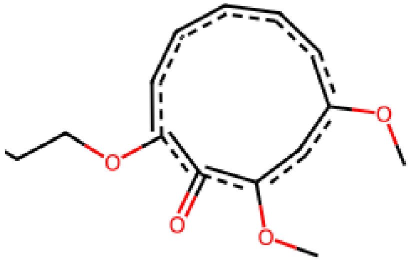
Future Directions

- Test and build scalability for larger rules;
- Investigate and mitigate constraint overfitting;
- Relax the conditional independence assumption to reduce bias;
- Extend beyond DiGress (e.g. score-based SDEs).
- Experimental:
  - latent variable control of k-of-K satisfaction patterns.

Vision

- Probabilistic-symbolic integration in graph generative modelling beyond molecular design.





THANK YOU!

QUESTIONS?

Clement Vignac et al. "DiGress: Discrete Denoising Diffusion for Graph Generation". In: *The 11th International Conference on Learning Representations*. 2023.

Nathan Brown et al. "GuacaMol: Benchmarking Models for de novo Molecular Design". In: *Journal of Chemical Information and Modelling* 59.3 (2019). DOI: 10.1021/acs.jcim.8b00839.

Christopher A. Lipinski. "Drug-like Properties and the Causes of Poor Solubility and Poor Permeability". In: *Journal of Pharmacological and Toxicological Methods* 44.1 (2000). DOI: 10.1016/S1056-8719(00)00107-6.