Deep Generative Models for Graphs

VAEs, GANs, and reinforcement learning for *de novo* drug discovery

Nicola De Cao

with Thomas Kipf as supervisor

Informatics Institute
University of Amsterdam

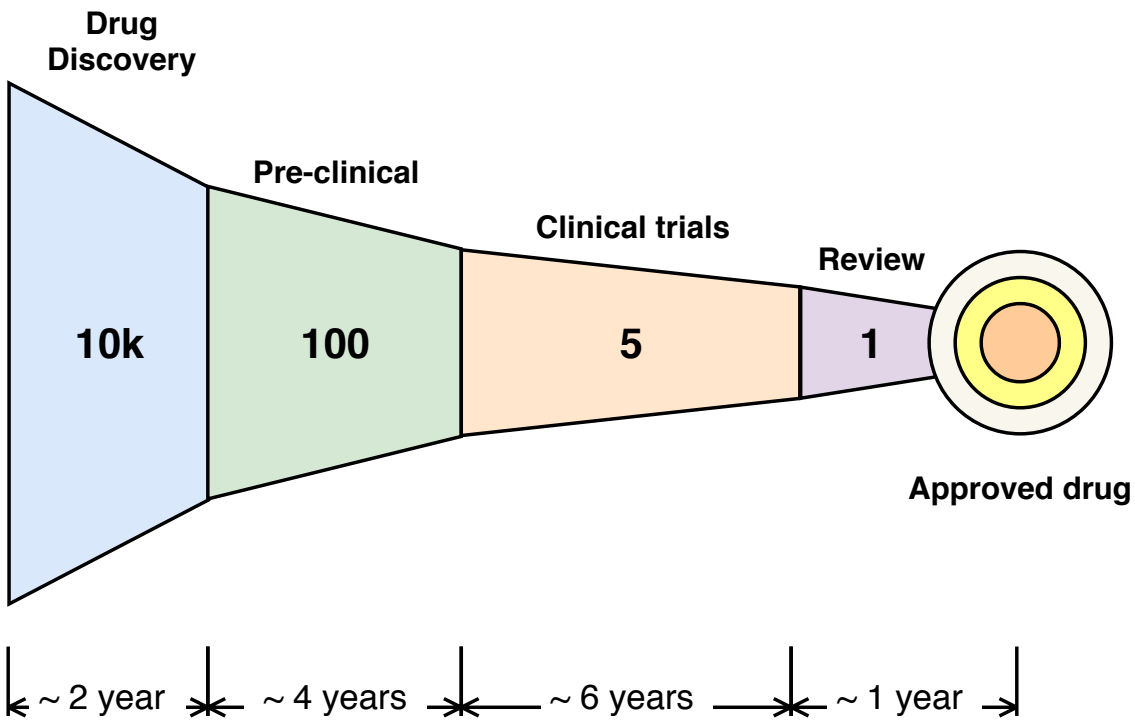
MSc Thesis presentation

Outline

The drug design processes

Which problem do we want to solve?

- **Drug design** is the process of finding new drugs



- The first step is **Drug Discovery**
  - screening large compound libraries
  - designing of **new unknown molecules** (*de novo*)

Our contributions

How others proposed to study the problem?

- Generating **SMILES** representations [Gómez-Bombarelli et al., 2016]
- Generating **labeled graphs** [Simonovsky and Komodakis, 2018]

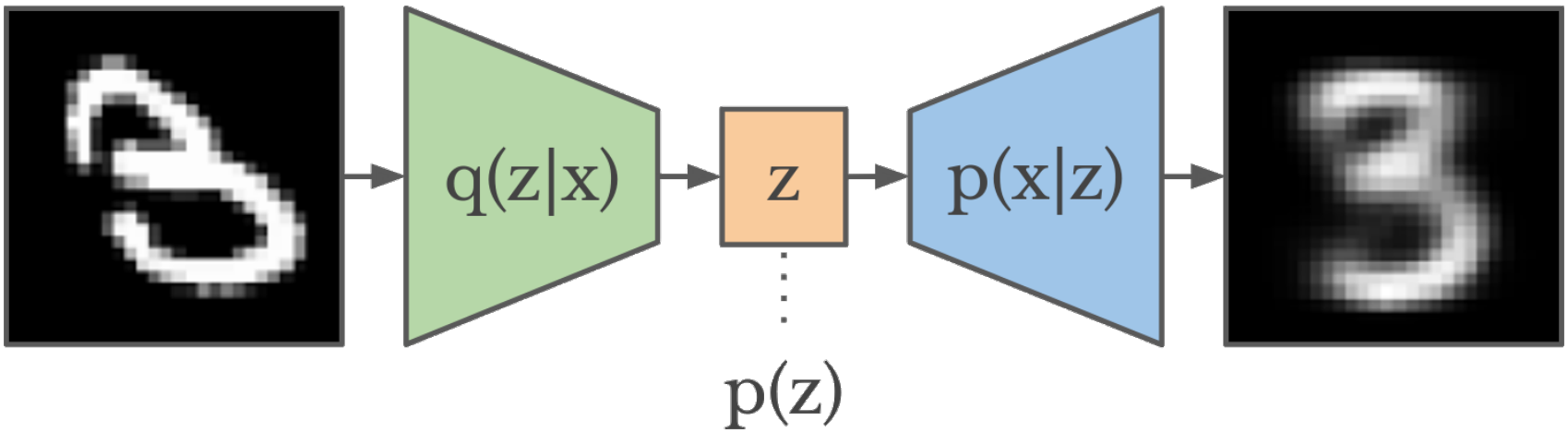
How do we study the problem?

- Using **labeled graphs**
- **Likelihood-based vs. likelihood-free** methods (VAE vs. GAN)
- Biasing the process using **reinforcement learning**

Background

Variational Auto-Encoders

Likelihood-based generative process [Kingma and Welling, 2013]



[Hafner, 2018]

Generative Adversarial Networks

Likelihood-free generative process [Goodfellow et al., 2014]

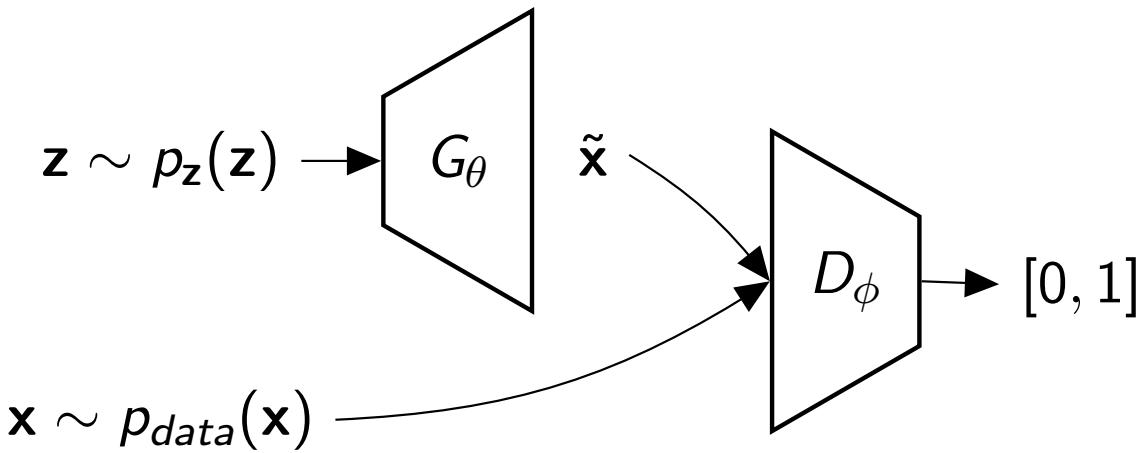


Figure: Schema of **GAN architecture**.

Models

Vectorial representation of graphs

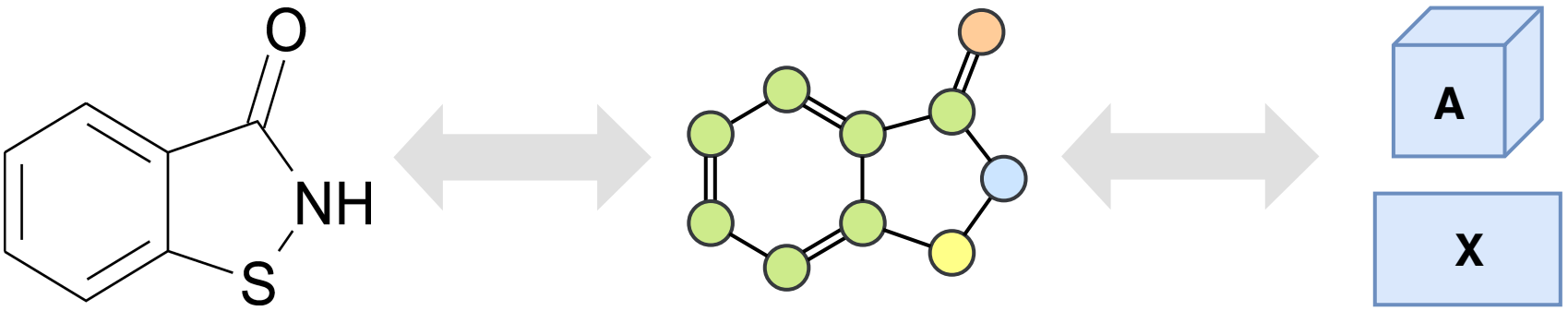
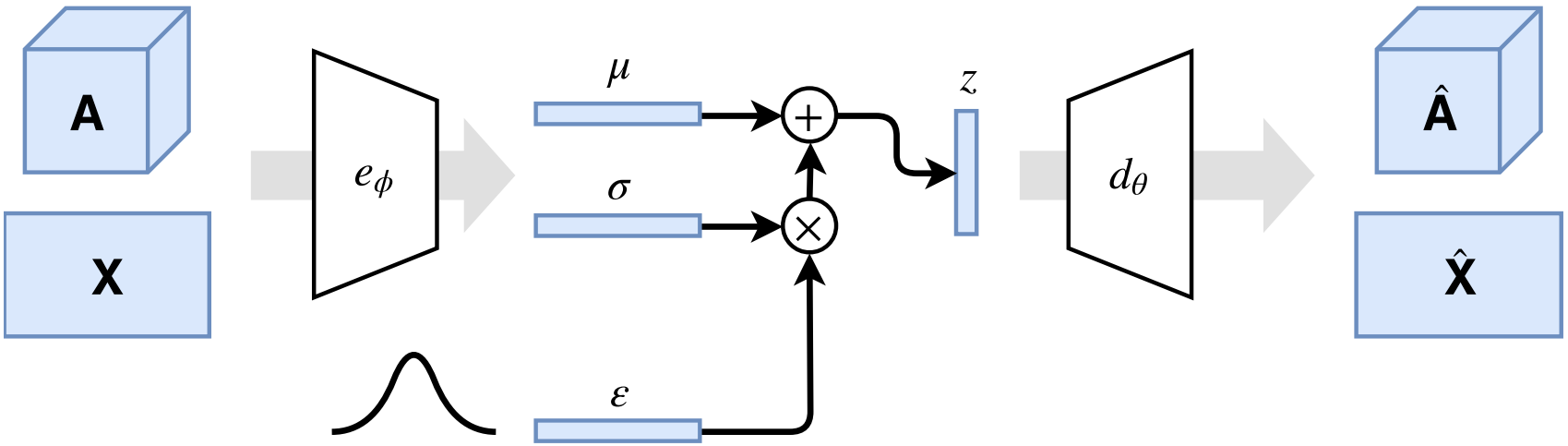


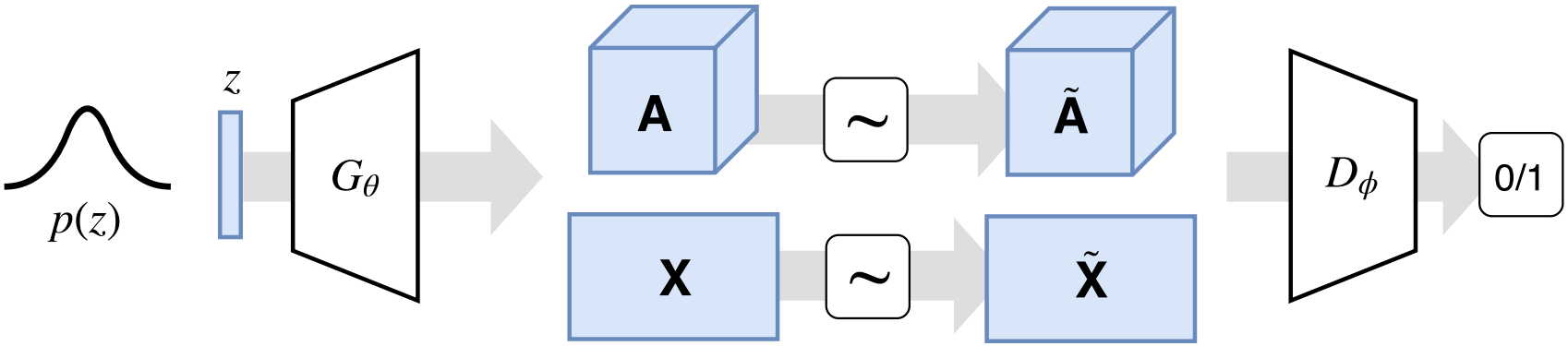
Figure: The **molecule** (a) is represented as an **labeled graph** (b) which can be encoded into an **adjacency tensor A** and an **annotation matrix X**.

Molecular graph VAE



The reconstruction loss is a sum of two **categorical cross entropy** losses.

Molecular graph GAN



From generator to discriminator with **differentiable sampling**.

Molecular graph GAN with RL

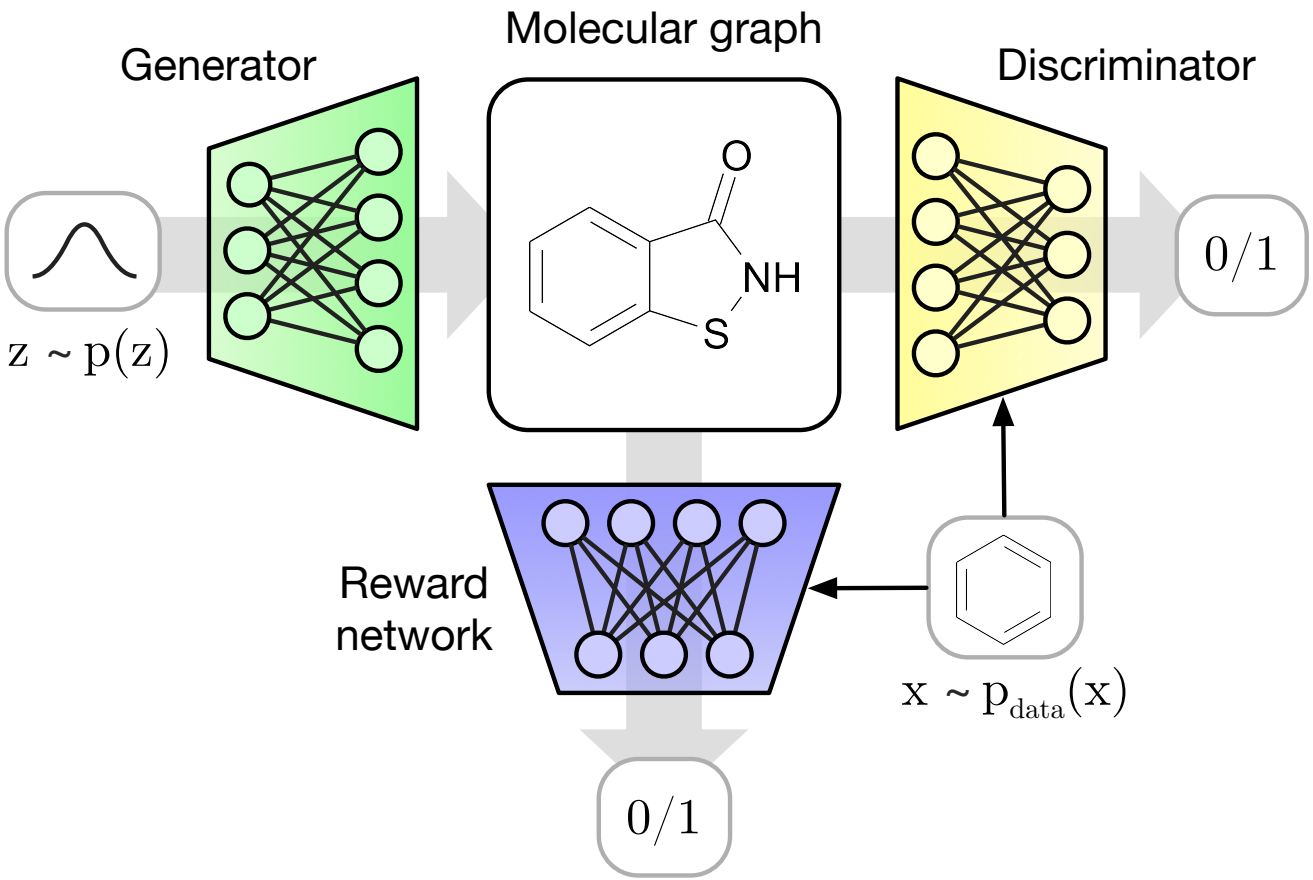


Figure: Schema of **MolGAN** from our previous work [De Cao and Kipf, 2018].

Experiments

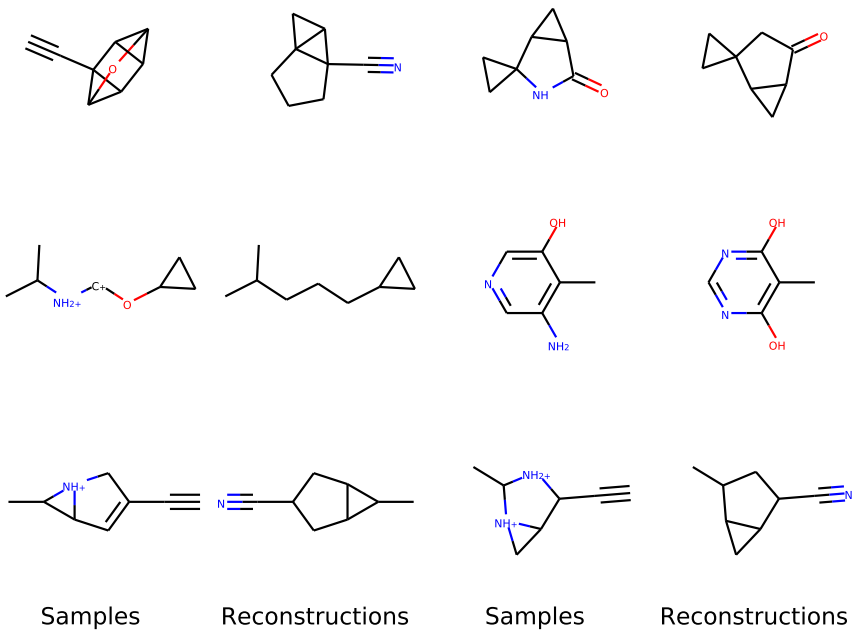
Experiments

Which questions we would like to answer?

- **likelihood-based vs. likelihood-free** (VAEs vs. GANs)
- the **effect of RL** towards chemical objectives
- Is generating a **graph better than a SMILES** representation?

VAE advantages and disadvantages

VAEs train an **encoder!**



- **VAE objective**: reconstruction loss and divergence
- **RL objective**: sampled molecules should maximize a score

There is a mismatch between these two!

Trade-off between WGAN and RL

Method	validity	uniqueness	QED
$\lambda = 0.0$ (full RL)	**100.00**	3.16	**0.61**
$\lambda = 0.125$	**100.00**	7.21	**0.61**
$\lambda = 0.25$	99.80	10.16	**0.61**
$\lambda = 0.375$	99.90	11.11	0.60
$\lambda = 0.5$	99.40	31.29	0.56
$\lambda = 0.625$	97.20	49.69	0.51
$\lambda = 0.75$	93.70	64.35	0.51
$\lambda = 0.875$	89.40	**69.69**	0.50
$\lambda = 1.0$ (no RL)	90.10	63.91	0.50

Table: WGAN and RL objectives trade-off.

Synthetic accessibility score (SAS) distributions I

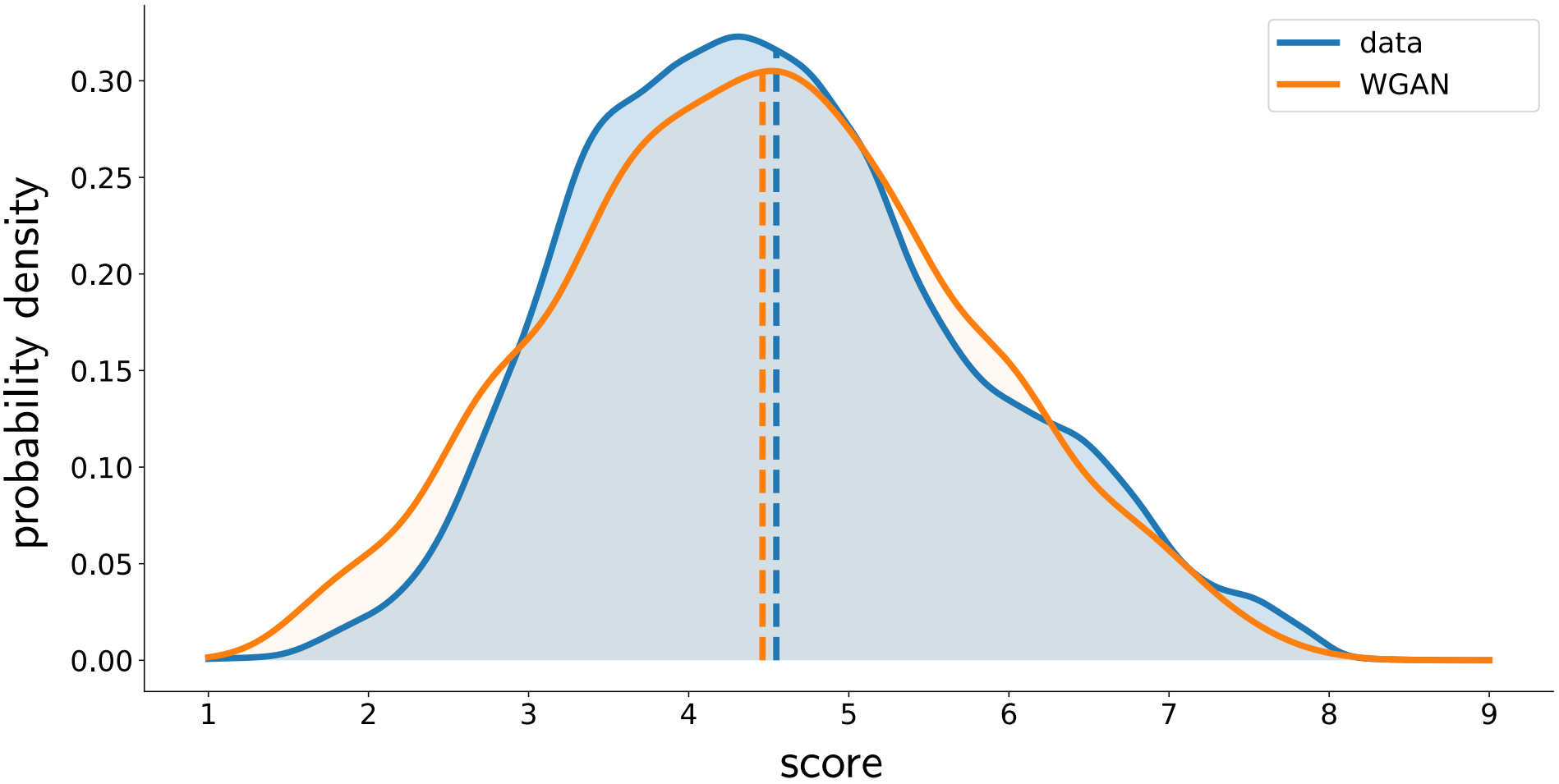


Figure: WGAN matches the data distribution of the synthetic accessibility score.

Synthetic accessibility score (SAS) distributions II

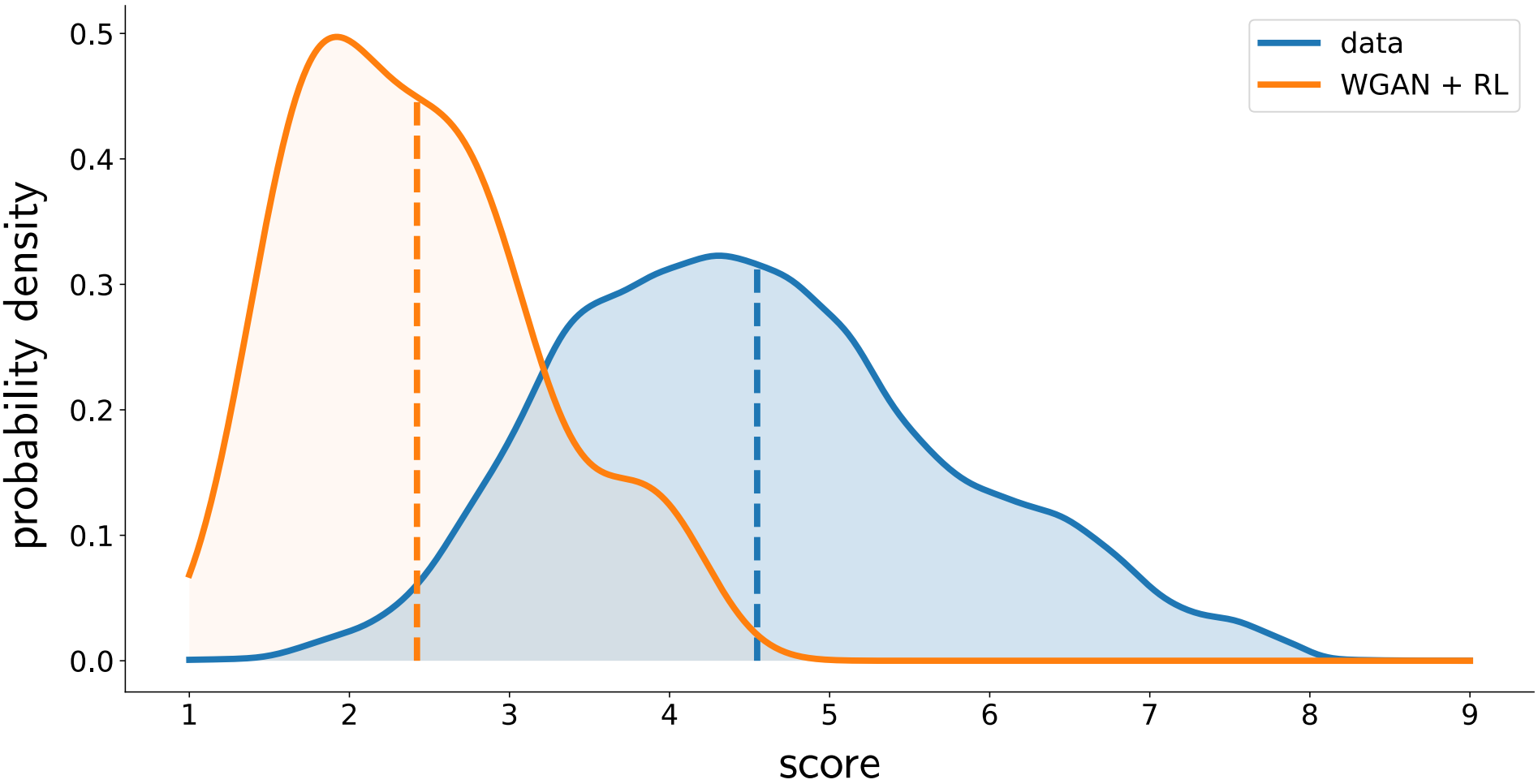


Figure: WGAN in combination with RL push the distribution of the synthetic accessibility score (SAS) to be as low as possible.

Comparison with VAE based methods

Method	validity	uniqueness	novelty
CharacterVAE	10.3	67.5	90.0
GrammarVAE	60.2	9.3	80.9
GraphVAE	55.7	76.0	61.6
GraphVAE/imp	56.2	42.0	75.8
GraphVAE NoGM	81.0	24.1	61.0
Our VAE	61.5	**97.6**	69.1
Our VAE with RL	**89.1**	11.1	**92.3**
Our WGAN	**89.2**	26.5	55.7
Our WGAN with RL	**99.6**	14.5	**97.7**

Table: Baseline results from GraphVAE [Simonovsky and Komodakis, 2018].

Comparison with a GAN based method

Method	validity	SAS	time
ORGAN	96.5	0.83	8.7
OR(W)GAN	97.6	0.75	9.6
Naive RL	97.7	0.83	10.6
Our VAE with RL	89.6	0.71	**0.09**
Our VAE with RL (full QM9)	94.0	**0.86**	**2.2**
Our WGAN with RL	**100.0**	0.70	**0.15**
Our WGAN with RL (full QM9)	**99.8**	**0.92**	**3.3**

Table: Baseline results from ORGAN [Guimaraes et al., 2017].

Conclusion and future work

Conclusion

Considering experimental, we identify these further **contributions**:

- recurrent SMILES generation is more **computational expensive**
- likelihood-based models are **difficult to be optimized** with RL

... but keeping in mind and these **limitations**:

- we experimented using compounds of **at most 9 atoms**
- models are susceptible to **mode collapse**

Future work

We identify four principal directions for future work:

- address **mode collapse** [Srivastava et al., 2017]
- combine **variational** approaches with **adversarial** learning to benefit from both approaches [Mescheder et al., 2017, Rosca et al., 2017]
- train our models on **ChEMBL** [Gaulton et al., 2011]
- more **realistic reward** functions [Li et al., 2018]

References I

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

Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., et al. (2011).
ChEMBL: a large-scale bioactivity database for drug discovery.
Nucleic acids research, 40(D1):D1100–D1107.

References II

Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., and Aspuru-Guzik, A. (2016).
Automatic chemical design using a data-driven continuous representation of molecules.
ACS Central Science, 4(2):268–276.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014).
Generative adversarial nets.
In *Advances in neural information processing systems*, pages 2672–2680.

References III


Guimaraes, G. L., Sanchez-Lengeling, B., Farias, P. L. C., and Aspuru-Guzik, A. (2017).
Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models.
arXiv preprint arXiv:1705.10843.


Hafner, D. (2018).
Building variational auto-encoders in tensorflow.
Blog post.


Kingma, D. P. and Welling, M. (2013).
Auto-encoding variational bayes.
International Conference on Learning Representations.

References IV


Li, Y., Zhang, L., and Liu, Z. (2018).
Multi-objective de novo drug design with conditional graph generative model.
arXiv preprint arXiv:1801.07299.


Mescheder, L., Nowozin, S., and Geiger, A. (2017).
Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks.
International Conference on Machine Learning.


Rosca, M., Lakshminarayanan, B., Warde-Farley, D., and Mohamed, S. (2017).
Variational approaches for auto-encoding generative adversarial networks.
arXiv preprint arXiv:1706.04987.

References V


Simonovsky, M. and Komodakis, N. (2018).
GraphVAE: Towards generation of small graphs using variational autoencoders.
arXiv preprint arXiv:1802.03480.


Srivastava, A., Valkoz, L., Russell, C., Gutmann, M. U., and Sutton, C. (2017).
VEEGAN: Reducing mode collapse in GANs using implicit variational learning.
In *Advances in Neural Information Processing Systems*, pages 3308–3318.