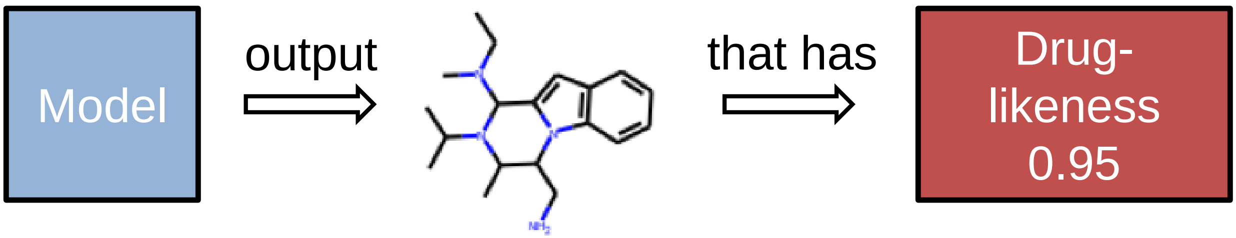
Graph Convolutional Policy Network for Goal-Directed Molecular Graph Generation

Jiaxuan You*, Bowen Liu*, Rex Ying, Vijay Pande, Jure Leskovec

Stanford University

Motivation

- Question:
  - Can we learn a model that can generate valid and realistic molecules with high value of a given chemical property?
  - Valid, Realistic, High scores



Goal-Directed Graph Generation

- Generating graphs that:
  - Optimize given objectives (High scores)
    - E.g., drug-likeness (black box)
  - Obey underlying rules (Valid)
    - E.g., chemical valency
  - Are learned from examples (Realistic)
    - E.g., Imitating a molecule graph dataset

Existing Approaches

- String representations + RL [Guimaraes et al, 2017]
  - “CCN(C)C1C2=CC3=C(C=CC=C3)N2C(CN)C”
  - Very likely to generate invalid strings

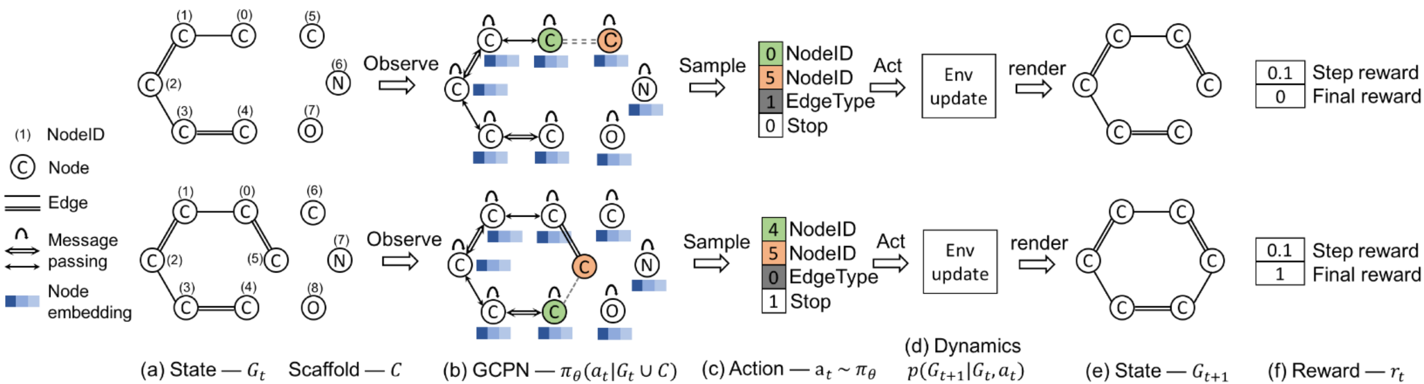
- Learned VAE-based vector representations + Bayesian optimization [Jin et al, 2018]
  - Depends on latent space, hand-coded decoder rules

GCPN

- Our Approach: Graph representation + RL
  - Graph representation enables validity check in each state transition (Valid)
  - Reinforcement learning optimizes intermediate and final rewards (High scores)
  - Adversarial training imitates examples in given datasets (Realistic)

GCPN

- Graph convolutional policy network (GCPN)



(1) Compute node embedding

$$H^{(l+1)} = \mathrm{AGG}(\mathrm{ReLU}(\{\tilde{D}_i^{-\frac{1}{2}} \tilde{E}_i \tilde{D}_i^{-\frac{1}{2}} H^{(l)} W_i^{(l)}\}, \forall i \in (1, \ldots, b)))$$

(2) Predict edge, edge type and stop token

(3) Optimize using PPO

Results

- Generating graphs from scratch:
  - Over 60% higher scores

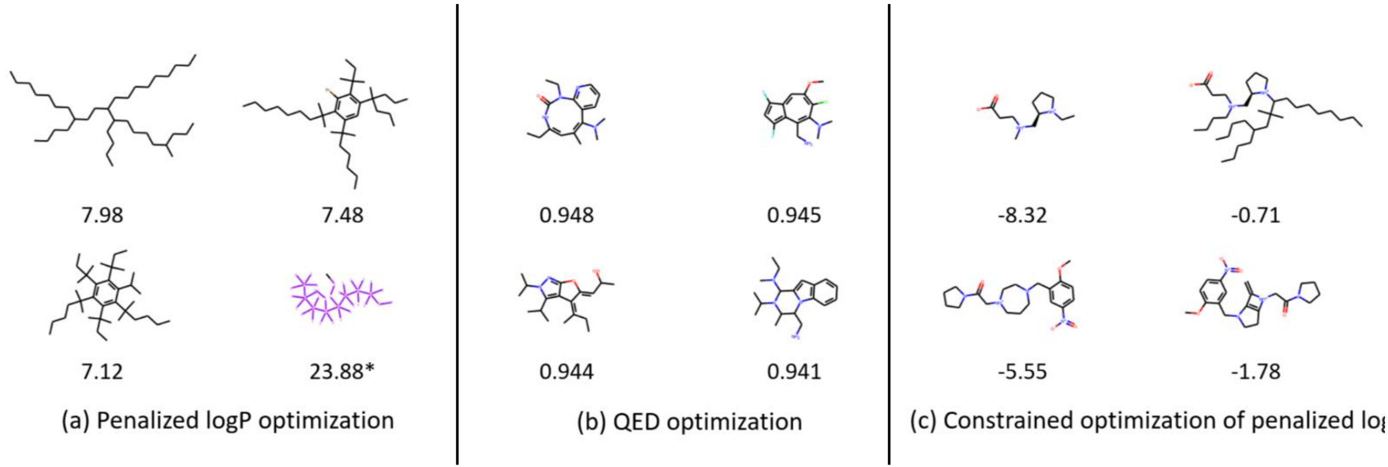
Table 1: Comparison of the top 3 property scores of generated molecules found by each model.

Method	Penalized logP				QED			
	1st	2nd	3rd	Validity	1st	2nd	3rd	Validity
ZINC	4.52	4.30	4.23	100.0%	0.948	0.948	0.948	100.0%
ORGAN	3.63	3.49	3.44	0.4%	0.896	0.824	0.820	2.2%
JT-VAE	5.30	4.93	4.49	100.0%	0.925	0.911	0.910	100.0%
GCPN	**7.98**	**7.85**	**7.80**	**100.0%**	**0.948**	**0.947**	**0.946**	**100.0%**

- Modifying existing graphs:
  - Over 180% higher scores improvement

Results

- Visualization



Results

- https://github.com/bowenliu16/rl_graph_generation

- Come to poster AB#140 for more results!