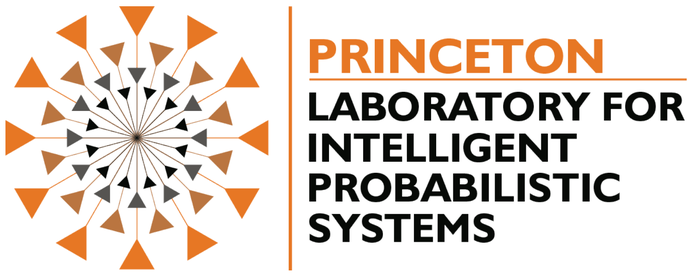
Paper: https://arxiv.org/abs/1907.08268
Code: https://github.com/PrincetonLIPS/reversible-inductive-construction

NeurIPS 2019

Discrete Object Generation with Reversible Inductive Construction

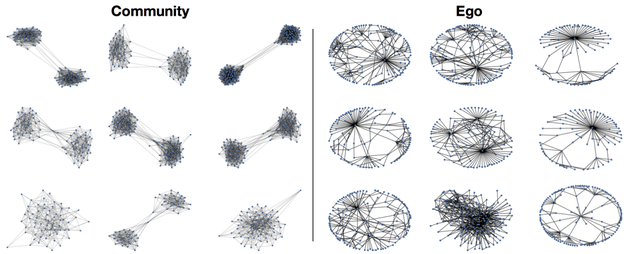
Ari Seff, Wenda Zhou, Farhan Damani, Abigail Doyle, Ryan P. Adams



PRINCETON
LABORATORY FOR
INTELLIGENT
PROBABILISTIC
SYSTEMS

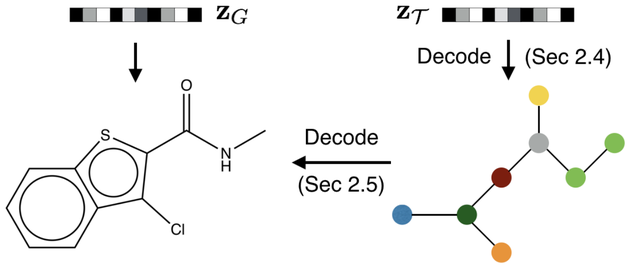
Modeling Discrete Domains

- Goal: Approximate (often high-dimensional) distributions of discrete data
- Surge of interest due to success on continuous domains



You et al., 2018

Networks/Graphs



Jin et al., 2018

Molecules

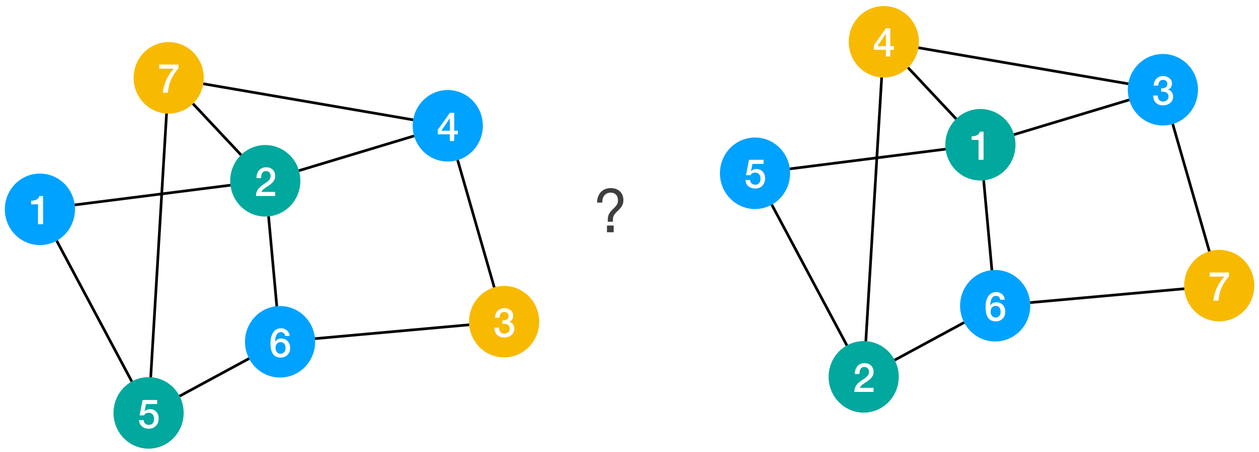
```
String s;
BufferedReader br;
FileReader fr;
try {
 fr = new FileReader($String);
 br = new BufferedReader(fr);
 while ((s = br.readLine()) != null) {}
 br.close();
} catch (FileNotFoundException _e) {
} catch (IOException _e) {
}
```

Murali et al., 2018

Source Code

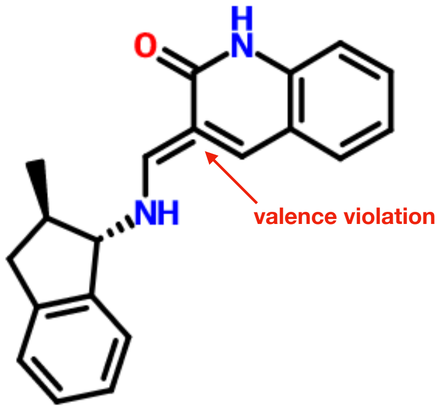
Challenges for Discrete Domains

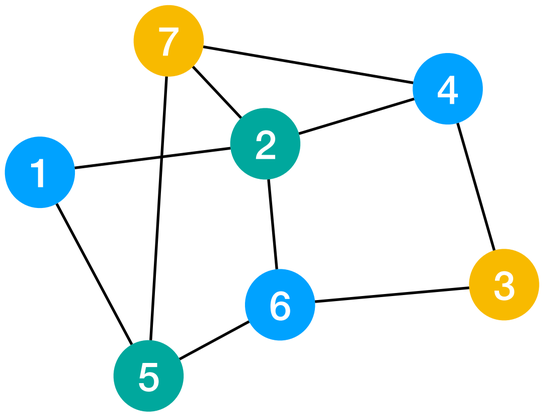
- Non-unique representations



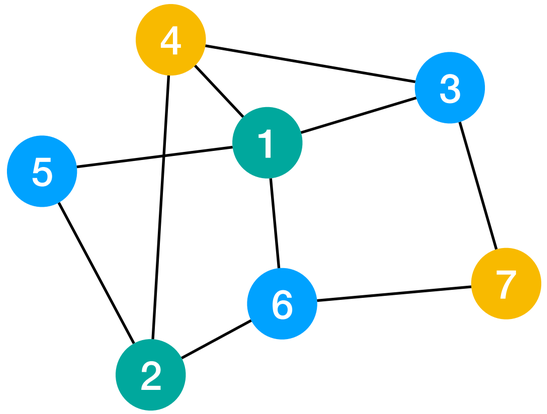
Challenges for Discrete Domains

- Non-unique representations
- Strict validity requirements





?



Reversible Inductive Construction

Proposed approach:

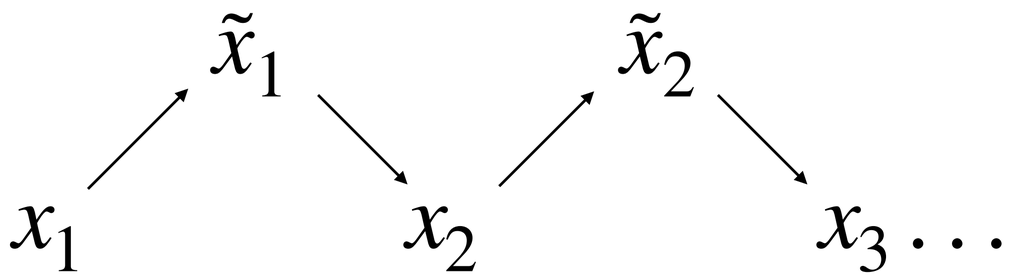
- Construct a Markov chain with equilibrium distribution approximating data distribution $p(x)$
- Restrict transitions to a set of *inductive* moves
  - Inductive moves: minimal insert and delete operations that maintain validity

$$x_1 \longrightarrow x_2 \longrightarrow x_3 \ldots$$

Reversible Inductive Construction

Proposed approach:

- Construct a Markov chain with equilibrium distribution approximating data distribution $p(x)$
- Restrict transitions to a set of *inductive* moves
  - Inductive moves: minimal insert and delete operations that maintain validity
- Transitions implemented via fixed corruption distribution $c(\tilde{x} \mid x)$ and learned reconstruction distribution $p_\theta(x \mid \tilde{x})$ (building off of Bengio et al., 2013)



Reversible Inductive Construction

Benefits:

- Constrains the generative model to only produce valid objects
- Requires the learner to only discover local modifications to the objects
- Avoids direct marginalization over an unknown and potentially large space of construction histories

Reversible Inductive Construction

Benefits:

- Constrains the generative model to only produce valid objects
- Requires the learner to only discover local modifications to the objects
- Avoids direct marginalization over an unknown and potentially large space of construction histories

Limitations:

- Expensive sampling, requiring Gibbs sampling at deployment time
- Inductive moves must be identified and specified for each new domain

Fixed Corrupter

1. Sample a number of moves *k* from a geometric distribution
2. For each move, sample a move type from {Insert, Delete}
3. Sample from among the legal operations for the given move type

$$\tilde{x}, s \sim c(\tilde{x}, s \mid x)$$

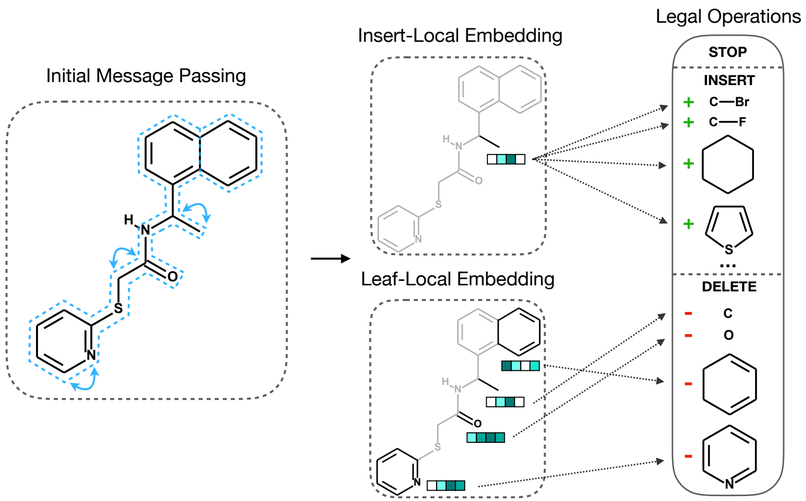
$$s = [s_1, s_2, ..., s_k]$$

Learned Reconstructor

- Target reconstruction distribution is Markov
- Factorize reconstruction distribution as product of memoryless transitions

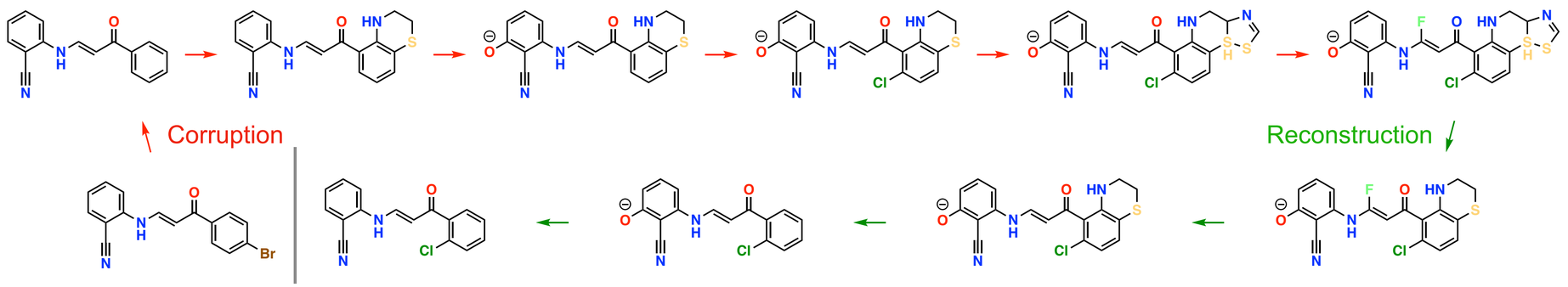
$$p_\theta(s_{\text{rev}} \mid \tilde{x}) = p_\theta(\text{stop} \mid x) p_\theta(x \mid \tilde{x}_1) \prod_{i=1}^{k-1} p_\theta(\tilde{x}_i \mid \tilde{x}_{i+1})$$

- Message-passing neural network similar to Duvenaud et al., 2015 and Gilmer et al., 2017 is used to generate embeddings for each location and vocabulary element



Application: Molecules

- Validity constraints regarding valence and aromaticity
- Legal operations defined by an extracted vocabulary of valid substructures (bonds, rings and bridged compounds) as in Jin et al., 2018
- Train on ZINC dataset, 250K drug-like molecules



Application: Molecules

Distributional statistics over novel molecules:

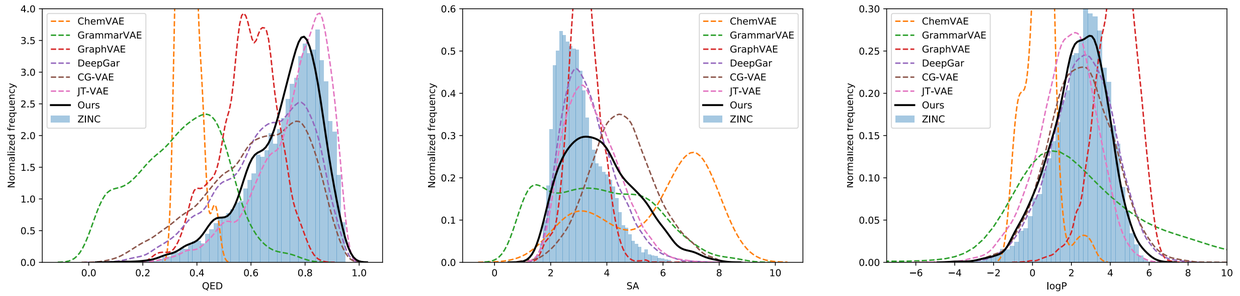
- Quantitative estimate of drug-likeness (QED) (Bickerton et al., 2012)
- Synthetic accessibility (SA) (Ertl and Schuffenhauer, 2009)
- Log octanol-water partition coefficient (logP) (Comer and Tam, 2007)

Application: Molecules

Distributional statistics over novel molecules:

- Quantitative estimate of drug-likeness (QED) (Bickerton et al., 2012)
- Synthetic accessibility (SA) (Ertl and Schuffenhauer, 2009)
- Log octanol-water partition coefficient (logP) (Comer and Tam, 2007)

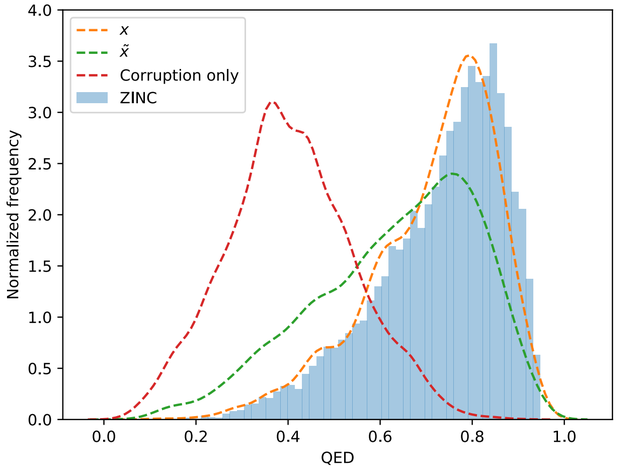
Source	QED KS	SA KS	logP KS	% valid
ChemVAE (Gómez-Bombarelli et al., 2018)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.7
GrammarVAE (Kusner et al., 2017)	0.94 (0.00)	0.95 (0.00)	0.95 (0.00)	7.2
GraphVAE (Simonovsky and Komodakis, 2018)	0.52 (0.00)	0.23 (0.00)	0.54 (0.00)	13.5
DeepGAR (Li et al., 2018)	0.20 (0.00)	0.15 (0.00)	0.062 (0.002)	89.2
JT-VAE (Jin et al., 2018)	0.090 (0.003)	0.21 (0.00)	0.20 (0.00)	100
CG-VAE (Liu et al., 2018)	0.27 (0.00)	0.56 (0.00)	0.064 (0.002)	100
GenRIC	0.045 (0.003)	0.28 (0.00)	0.057 (0.002)	100

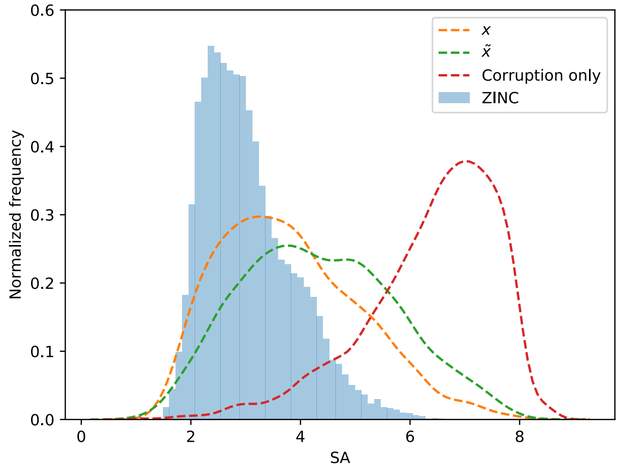


Application: Molecules

How do reconstructed and corrupted samples compare?

- Corrupted samples are less drug-like and less synthesizable than reconstruction counterparts (e.g., reconstructed molecule has 21% higher QED on average)
- Corruption only samples severely diverge from data distribution

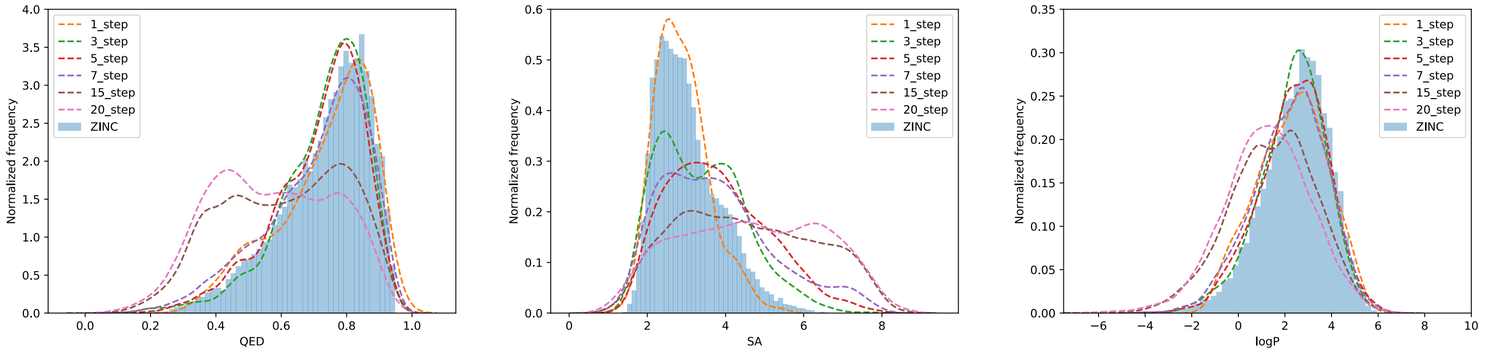




Application: Molecules

Varying the geometric distribution for corruption sequence length:

- In general, as sequence length increases (corruptions become less local) the model produces worse samples

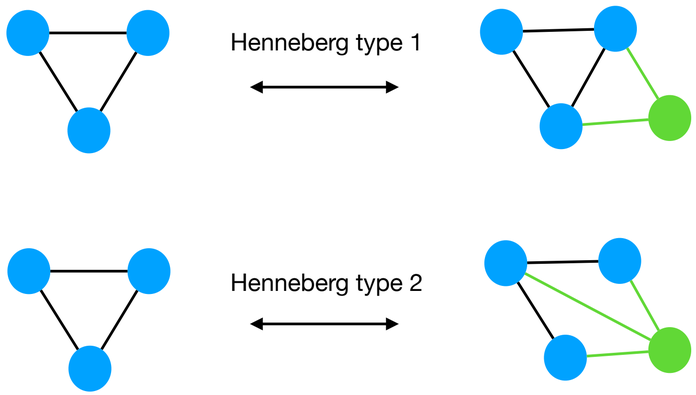


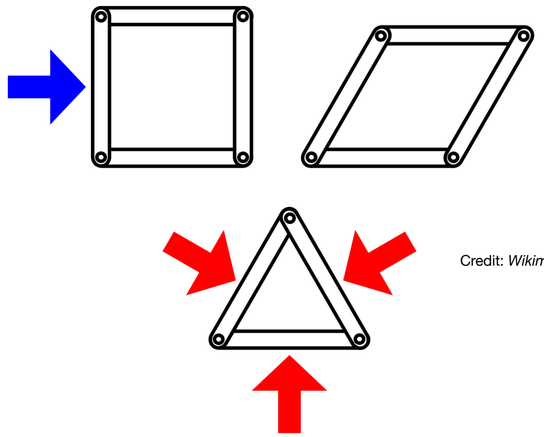
Application: Laman Graphs

- Geometric constraint graphs are widely employed in computer-aided design, molecular modeling, and robotics
- Nodes represent geometric primitives (e.g., lines, circles) and edges represent constraints (e.g., perpendicularity)

Application: Laman Graphs

- Laman graphs describe minimally rigid 2D geometry where primitives have 2 DOF and edges restrict 1 DOF (system of rods and joints)
- Minimally rigid if the Laman conditions are met
  - Must have exactly $2n$ - 3 edges
  - Each subgraph of k noes can have no more than $2k$ - 3 edges
- Inductive moves based off of Henneberg construction





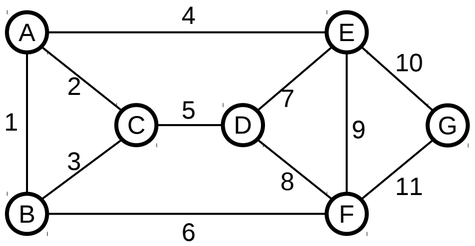
Credit: *Wikimedia Commons*

Structural rigidity

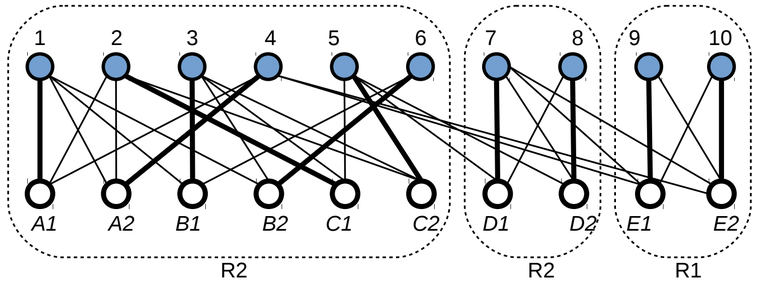
Application: Laman Graphs

Distributional statistics:

- Degree of decomposability (DoD) (Moussaoui, 2016)
  - Indicates extent to which a Laman graph is composed of well-constrained subgraphs



(a) A Geometric constraint graph



(b) The corresponding bipartite graph of fig. 4.8a after pinning the edge (F, G), its perfect matching and the set of irreducible subgraphs

Moussaoui, 2016

Application: Laman Graphs

Distributional statistics:

- Degree of decomposability (DoD) (Moussaoui, 2016)
  - Indicates extent to which a Laman graph is composed of well-constrained subgraphs
- Difficult for baseline models to learn strict topological constraints

Source	DoD KS	% valid
E-R (Erdös and Rényi, 1959)	0.95 (0.03)	0.08 (0.02)
GraphRNN (You et al., 2018)	0.96 (0.00)	0.15 (0.03)
GenRIC	0.33 (0.01)	100 (0.00)

Conclusion and Future Work

Conclusion:

- Introduced a new method for modeling distributions of discrete objects
- The model is only trained to undo a series of local corrupting operations
  - Both the corrupting and reconstructing operations preserve possibly-complicated validity constraints
- This simple approach can capture relevant distributional statistics over complex and highly structured domains while always producing value structures

Future work:

- Learn inductive moves from data
  - For example, restrict Markov chain's transitions to learned chemical reactions when generating molecules
- Enforce additional hard constraints such as the presence of a desired core structure
  - For example, generation conditioned on a particular drug scaffold