Physics-Constrained Predictive Molecular Latent Space Discovery with Graph Scattering VAE

Navid Shervani-Tabar, Nicholas Zabaras

Scientific Computing and Artificial Intelligence (SCAI) Laboratory
University of Notre Dame

UNIVERSITY OF NOTRE DAME | COLLEGE OF ENGINEERING

Problem Definition

Problem Definition

- The goal of a generative model is to learn a probability distribution from a set of training data $\mathscr{G} = \{\mathcal{G}^{(i)}\}_{i=1}^{K}$, with the help of latent variable $\boldsymbol{z}$.
- Graphs are mathematical objects, consisting of a set of vertices and a set of edges.
- Our focus is on training the generative model in small data regime.
- The objective is to reveal an underlying latent space Z that yields the observations.

$$p(\mathcal{G}) = \int p(\mathcal{G}, \boldsymbol{z}) d\boldsymbol{z} = \int p(\mathcal{G}|\boldsymbol{z}) p(\boldsymbol{z}) d\boldsymbol{z}$$

- We introduce a model $p(\mathcal{G}|\boldsymbol{\theta})$, parameterized by $\boldsymbol{\theta}$, to approximate $p_{target}(\mathcal{G})$.

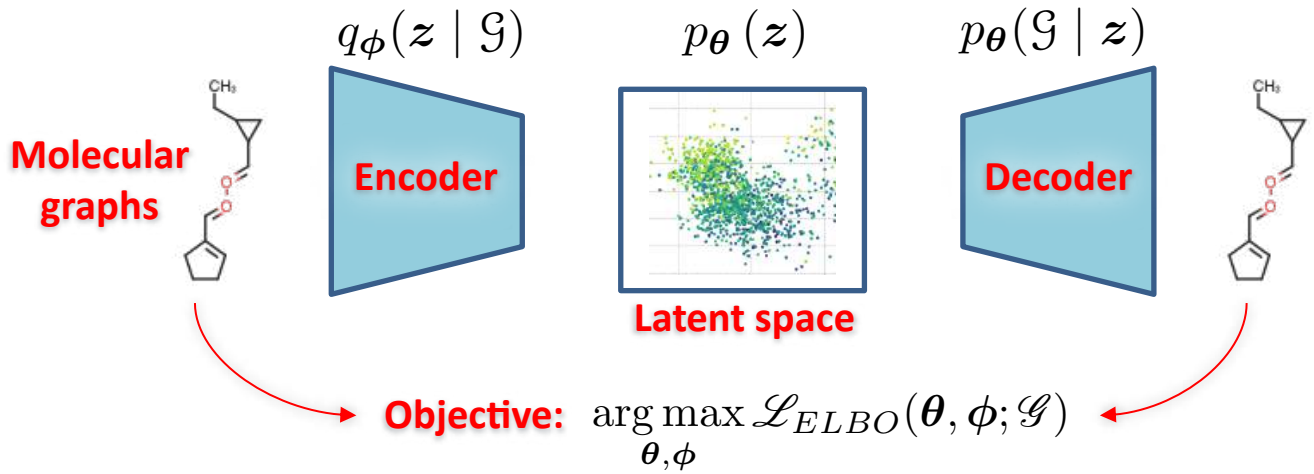
$$\boldsymbol{\theta}_{\mathrm{MLE}} = \arg\max_{\boldsymbol{\theta}} \log p(\mathscr{G} \mid \boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}} \sum_{k=1}^{K} \log p\left(\mathcal{G}^{(k)} \mid \boldsymbol{\theta}\right)$$

Problem Definition

- However, this involves an intractable integration.
- A common remedy is to use a variational posterior $q_{\boldsymbol{\phi}}(\boldsymbol{z} \mid \mathcal{G})$ and optimize the model by maximizing a lower-bound $\log p(\mathcal{G}|\boldsymbol{\theta})$.

$$\mathscr{L}_{ELBO}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathscr{G}) = \sum_{i=1}^{K} \mathscr{L}_{ELBO}\left(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathcal{G}^{(i)}\right)$$
$$= \sum_{i=1}^{K} \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z}^{(i)}|\mathcal{G}^{(i)})}\left[\log p_{\boldsymbol{\theta}}(\mathcal{G}^{(i)}|\boldsymbol{z}^{(i)})\right] - \sum_{i=1}^{K} \mathrm{D}_{\mathrm{KL}}\left[q_{\boldsymbol{\phi}}(\boldsymbol{z}^{(i)}|\mathcal{G}^{(i)}) \| p_{\boldsymbol{\theta}}(\boldsymbol{z}^{(i)})\right]$$

- Such a variational treatment lends itself to an autoencoder



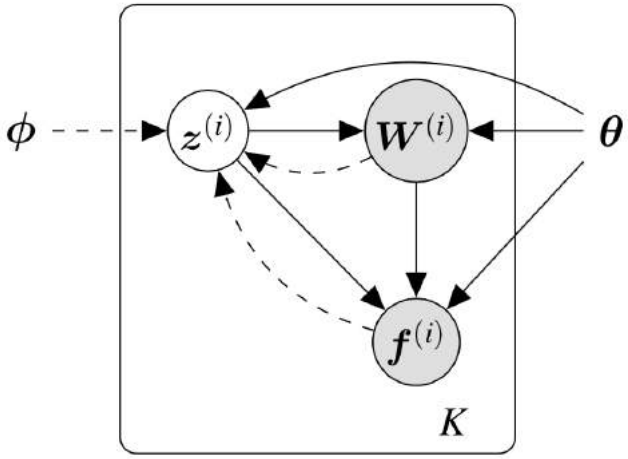
Graph

- A molecular graph represents atomic bonds with weighted edges and represents atom information as signals residing on graph vertices

$$\mathcal{G} = \{\boldsymbol{W}, \boldsymbol{f}\}$$

- We formulate the probabilistic mapping from the latent space to the molecular graph domain by conditioning atom types on molecular topology.

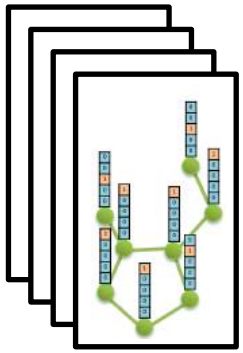
$$\begin{aligned} p_\theta(\mathcal{G}|\boldsymbol{z}) &= p_\theta(\boldsymbol{W}, \boldsymbol{f}|\boldsymbol{z}) \\ &= p_\theta(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{W}) p_\theta(\boldsymbol{W}|\boldsymbol{z}) \end{aligned}$$

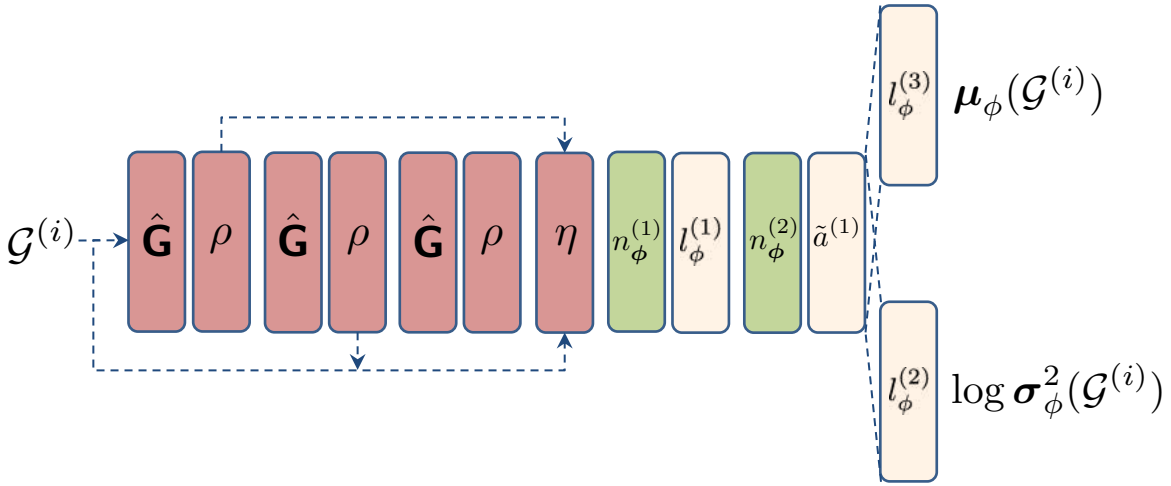


Encoding Network

Hybrid Encoder

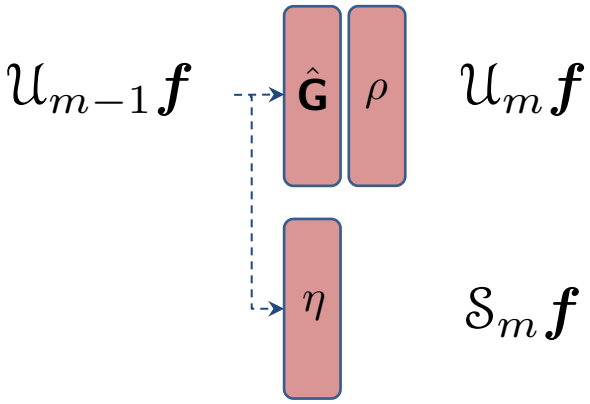
- We parameterize the variational approximate posterior with a Gaussian distribution, parameters of which are computed using an encoding network.
- In the present work, the encoding network has a hybrid architecture which extract features from the given input graphs using
  - Graph scattering network
  - Multilayer perceptron (MLP)





Scattering Network

- Graph scattering transform is a type of deep neural networks with predefined parameters to extract feature maps from input graphs.
- Scattering networks are constructed of a cascade of multiresolution filters, combined with modulus non-linearity to generate features.



Non-linearity

$$\mathcal{U}_m\boldsymbol{f} = \rho(\hat{\mathbf{G}}(\tilde{\mathcal{L}})\mathcal{U}_{m-1}\boldsymbol{f})$$

Convolution

$$\mathcal{S}_m\boldsymbol{f} = \eta\,(\mathcal{U}_{m-1}\boldsymbol{f})$$

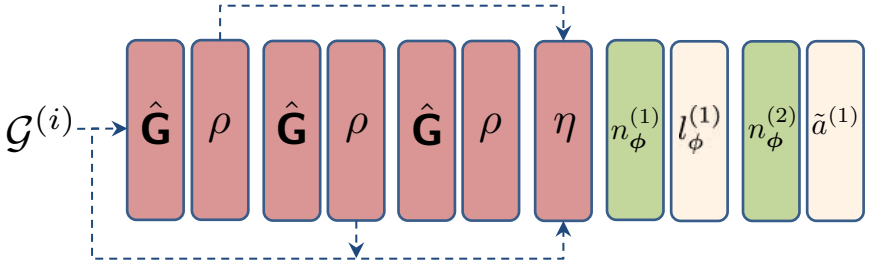
Average Pooling

$$\mathcal{A}(\mathcal{G}) = [\mathcal{S}_1\boldsymbol{f}, \ldots, \mathcal{S}_M\boldsymbol{f}]$$

- These features are invariant to permutation and stable with respect to graph and signal manipulation.

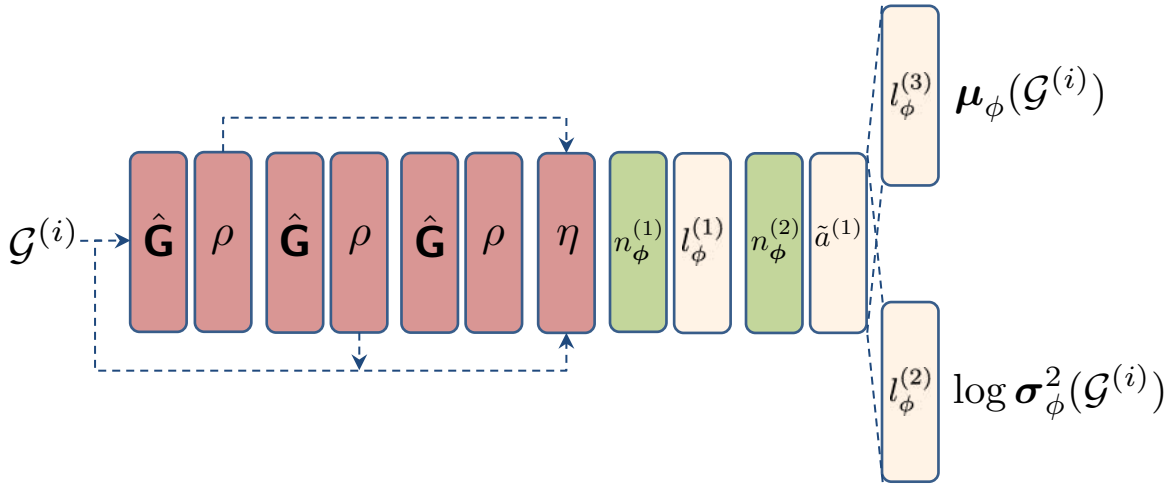
Linear Layers

- After extracting features in scattering layers, we use an additional linear layer followed by a non-linearity to learn features.



Linear Layers

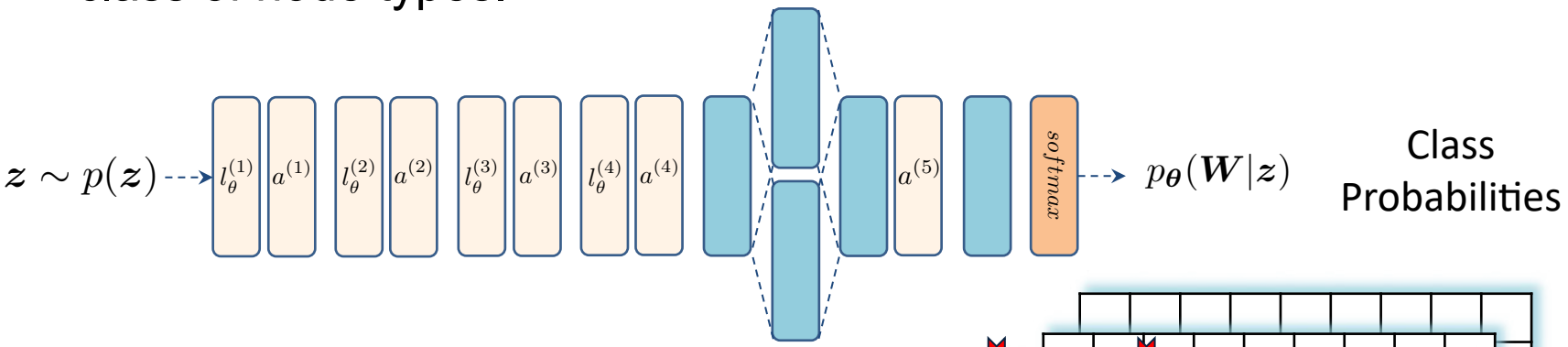
- After extracting features in scattering layers, we use an additional linear layer followed by a non-linearity to learn features.
- Furthermore, two linear layers are used to find the mean and variance of the variational approximate posterior q_φ of the the latent space variable z.

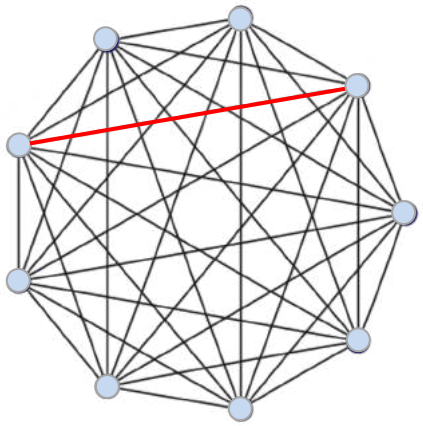


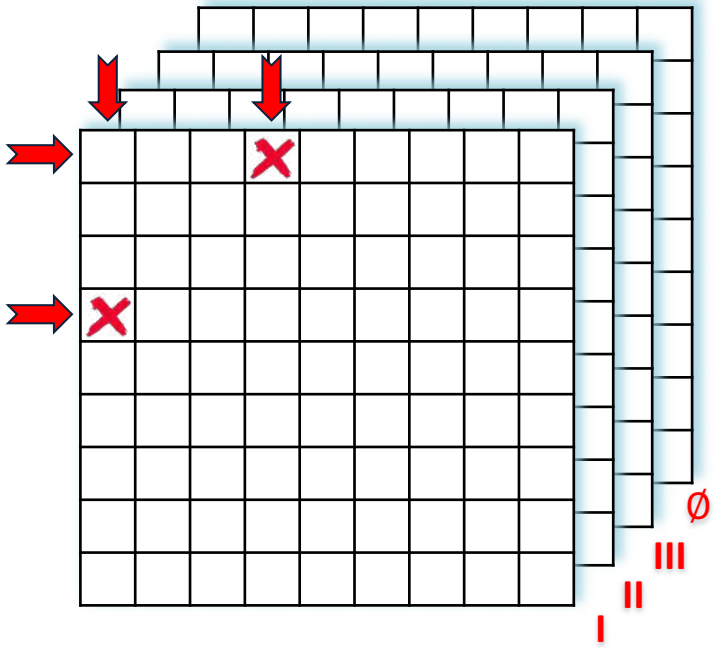
Decoding Network

Bond Generator

- Weighted adjacency matrix decoder is constructed of an MLP that projects samples from the latent space into score values for each class of node types.

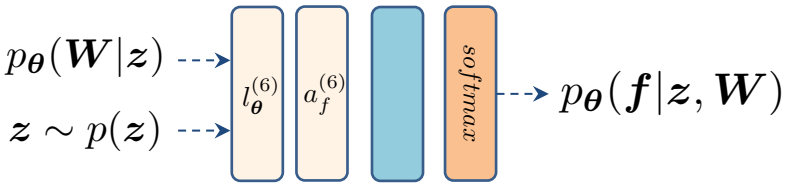




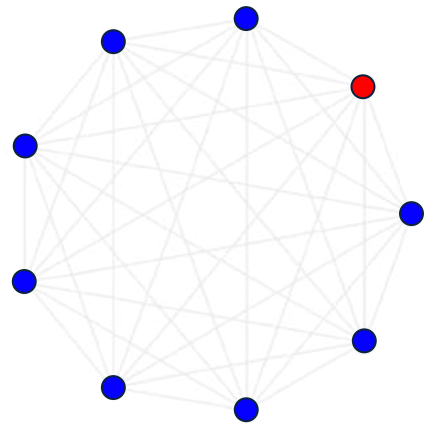


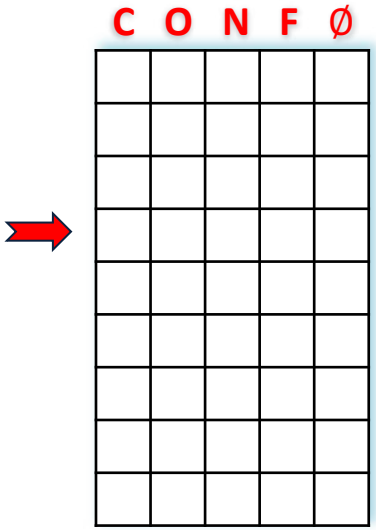
Atom Generator

- Atom type generator in the decoder is constructed of an MLP that projects samples from the latent space and molecular topology into discrete probabilities for classes of node types.



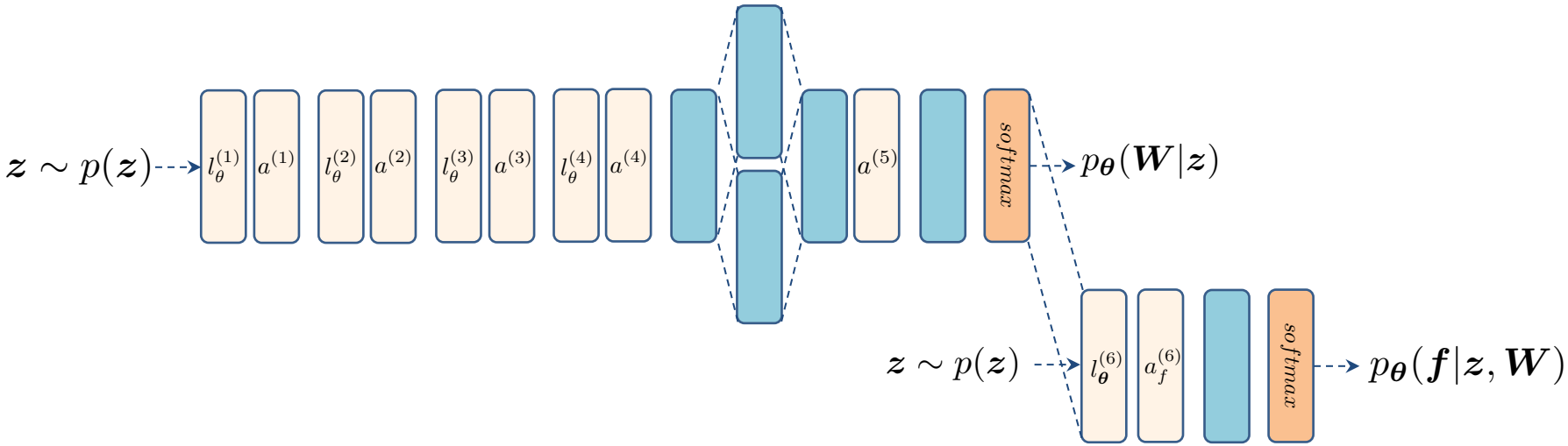
Class probabilities





Decoding Network

- As the two decoding networks generate correlated information on atoms and bonds of the molecular graph, we can have them share the input layers



- We train the model by minimizing the regularized expected negative log-likelihood for each node and each edge.

$$L^{(t)} = \sum_{i=1}^{N}\sum_{c=1}^{C_f} -t_c^{v_i}\log p_\theta\left(f_i = c|\boldsymbol{z},\boldsymbol{W}\right) + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j>i}}^{N}\sum_{c=1}^{C_W} -t_c^{e_{i,j}}\log p_{\boldsymbol{\theta}}\left(W_{i,j} = c|\boldsymbol{z}\right)$$

Physical Constraints

Regularized VAE

- In the regularization method, constraint terms are accompanied by regularization parameters that are tuned until we observe the desired output in terms of quality of the generated molecules.

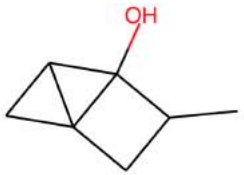
$$\mathscr{L}(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\mu}; \mathscr{G}) = -\sum_{i=1}^{K}\left(\frac{1}{L}\sum_{l=1}^{L}\log p_{\boldsymbol{\theta}}(\mathcal{G}^{(i)}|\boldsymbol{z}^{(i,l)})\right) + \sum_{i=1}^{K}\mathrm{D}_{\mathrm{KL}}\left[q_{\boldsymbol{\phi}}(\boldsymbol{z}^{(i)}|\mathcal{G}^{(i)})\|p(\boldsymbol{z}^{(i)})\right] + \sum_{i=1}^{K}\sum_{r=1}^{\mathcal{R}}\mu_r\left(\frac{1}{L}\sum_{l=1}^{L}\left(\mathfrak{C}_{+}^{(r)}(\boldsymbol{z}^{((i-1)L+l)}, \theta)\right)^2\right)^{\frac{1}{2}}.$$

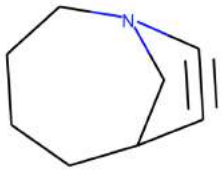
$$\boldsymbol{z}^{(i,l)} = \boldsymbol{\mu}_{\boldsymbol{\phi}}(\mathcal{G}^{(i)}) + \boldsymbol{S}_{\boldsymbol{\phi}}(\mathcal{G}^{(i)}) \odot \boldsymbol{\epsilon}^{(l)}$$

$$\boldsymbol{\epsilon}^{(l)} \sim \mathcal{N}(\boldsymbol{\epsilon}; \mathbf{0}, \boldsymbol{I})$$

$$\boldsymbol{z} \sim p(\boldsymbol{z})$$

- Connectivity
- Valency
- 3−member cycles
- Cycles with triple bonds





Model Uncertainty

Model Uncertainty

- In this work, we focus on obtaining a predictive model accounting for uncertainty.
- Given a set of K i.i.d. observations $\mathscr{G} = \{\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(K)}\}$ sampled from $p_{target}(\mathcal{G})$, we are interested in finding a model $p_{\boldsymbol{\theta}}(\mathcal{G})$ that closely mimics the target distribution

$$\boldsymbol{\theta}_{\text{MLE}}, \boldsymbol{\phi}_{\text{MLE}} = \arg\max_{\boldsymbol{\theta},\boldsymbol{\phi}} \mathscr{L}_{ELBO}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathscr{G}) = \arg\max_{\boldsymbol{\theta},\boldsymbol{\phi}} \sum_{k=1}^{K} \mathscr{L}_{ELBO}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathcal{G}^{(k)})$$

- Quantifying uncertainties in model parameters $\boldsymbol{\theta}$ enables us to capture the epistemic uncertainty induced by the limited training data.

$$p(\bar{\mathcal{G}} \mid \mathscr{G}) = \int p(\bar{\mathcal{G}} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathscr{G}) d\boldsymbol{\theta}$$

- The full posterior over model parameters is often approximated using a Gaussian distribution. However, this results in highly invalid molecular graphs.

Model Uncertainty

- We use Bayesian bootstrap method to approximate the Bayesian predictive distribution.
- Bootstrapping mimics sampling data from the population by random sampling, with replacement, from sampled data.
- Given a dataset $\mathscr{G} = \{\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(K)}\}$, we generate B samples of size K by resampling from the original dataset.
- A resampled dataset can be denoted by the associated resampling weights $\tilde{\mathscr{G}} = (\mathscr{G}, \boldsymbol{\pi})$.

$$p(\boldsymbol{\pi}|\mathscr{G}) = \mathcal{D}ir(\boldsymbol{\pi}; \boldsymbol{\alpha}'), \quad \text{with } \boldsymbol{\alpha}' = [1, \ldots, 1] \in \mathbb{R}^K$$

- The MLE objective can be reformulated as maximizing the weighted likelihood estimate

$$\boldsymbol{\theta}_{\text{MWLE}}(\boldsymbol{\pi}) = \arg\max_{\boldsymbol{\theta}} \sum_{k=1}^{K} \pi_k \log p(\mathcal{G}^{(k)}|\boldsymbol{\theta})$$

Model Uncertainty

- Since resampling weights $\boldsymbol{\pi} \sim p(\boldsymbol{\pi}|\mathcal{G})$ are positive and since

$$\log p(\mathcal{G}^{(k)}|\boldsymbol{\theta}) \geq \mathscr{L}_{ELBO}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathcal{G}^{(k)})$$

  we can define a lower-bound on the weighted marginal log-likelihood

$$\boldsymbol{\theta}_{\text{MWLE}}, \boldsymbol{\phi}_{\text{MWLE}} = \arg\max_{\boldsymbol{\theta}, \boldsymbol{\phi}} \sum_{k=1}^{K} \pi_k \mathscr{L}_{ELBO}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathcal{G}^{(k)})$$

- We can simulate approximately from the posterior distribution over $\boldsymbol{\theta}$ by repeated sampling from the posterior distribution $p(\boldsymbol{\pi}|\mathcal{G})$ and maximizing a weighted likelihood to calculate $\boldsymbol{\theta}_{\text{MWLE}}$.

$$p_{BB}(\bar{\mathcal{G}} \mid \mathscr{G}) = \int p\left(\bar{\mathcal{G}} \mid \boldsymbol{\theta}_{\text{MWLE}}(\boldsymbol{\pi})\right) p(\boldsymbol{\pi} \mid \mathscr{G}) d\boldsymbol{\pi}$$

- Using Monte Carlo estimation, we approximation this predictive distribution

$$p(\bar{\mathcal{G}}|\mathscr{G}) = \frac{1}{B} \sum_{b=1}^{B} p(\bar{\mathcal{G}}|\boldsymbol{\theta}_{\text{MWLE}}(\boldsymbol{\pi}^b))$$

Results

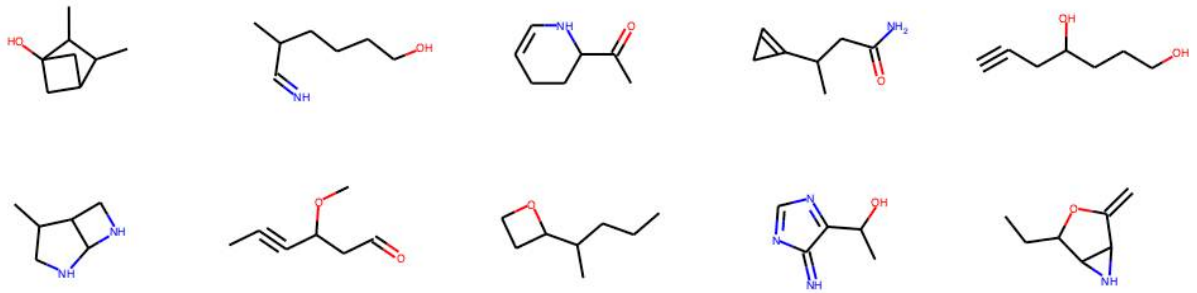
Molecule Sampling

- We train the model on QM9 dataset ($K = 600$), containing small organic molecules with maximum 9 heavy atoms.
- Atoms include carbon, nitrogen, oxygen, and fluorine.

Algorithm 1 Predicting molecules using ancestral sampling.

Input Trained model $p_{\boldsymbol{\theta}}(\boldsymbol{z})p_{\boldsymbol{\theta}}(\mathcal{G}|\boldsymbol{z})$ and L the number of samples to be drawn.
for $l = 1, \ldots, L$ **do**
  Generate a sample $\boldsymbol{z}^{(l)}$ from $p_{\boldsymbol{\theta}}(\boldsymbol{z})$.
  Draw sample $\mathcal{G}^{(l)}$ from distribution $p_{\boldsymbol{\theta}}(\mathcal{G}|\boldsymbol{z})$.
end for
Return $\boldsymbol{\mathcal{G}}^{1:L}$.

- Example of molecules sampled using the generative model.



Quality Metrics

- The validity, uniqueness, and novelty of the generated molecules are basic quality metrics* of molecular generative models:

$$\text{Validity} = \frac{|valid(S)|}{|S|}$$

$$\text{Uniqueness} = \frac{|S^*|}{|S|}$$

$$\text{Novelty} = \frac{|S| - |T \cap S|}{|S|}$$

T = Set of training molecules
S = Set of sampled molecules (with repeat)
S^* = Set of sampled molecules

Quality metrics for graph scattering VAE

Type	H_{val}			3−member cycle	Triple bond cycle	H_{unq}	H_{nvl}
	Total	Valency	Connectivity				
Base	77.8%	15.9%	7.2%	3910	113	77.7%	86.3%
Constrained	89.9%	5.6%	4.6%	1745	35	71.1%	95.8%

* Samanta, Bidisha, Abir De, Niloy Ganguly, and Manuel Gomez-Rodriguez. "Designing random graph models using variational autoencoders with applications to chemical design." *arXiv preprint arXiv:1802.05283* (2018).

Quality Metrics

- We have compared performance of our model with baseline methods.

Method	H_{val}	H_{unq}	H_{nvl}
CVAE	10.3%	67.5%	90.0%
GVAE	60.2%	9.3%	80.9%
GraphVAE	55.7%	76.0%	61.6%
MolGAN	98.1%	10.4%	94.2%
$K = 600$ ➡ GSVAE	77.8%	77.7%	86.3%
$K = 10000$ ➡ GSVAE	87.1%	82.6%	92.7%

✦ Rafael Go´mez-Bombarelli, Jennifer N. Wei, David Duvenaud, Jose´ Miguel Herna´ndez-Lobato, Benjam´ın Sa´nchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Ala´n Aspuru-Guzik. Automatic chemical design using a data-driven contin- uous representation of molecules. ACS central science, 4(2):268–276, 2018.

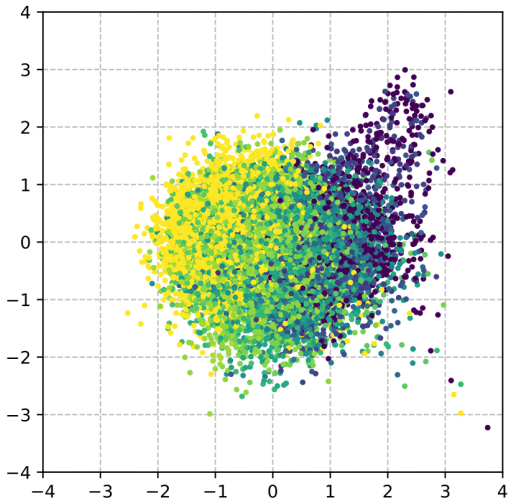
✶ Matt J. Kusner, Brooks Paige, and Jose´ Miguel Herna´ndez-Lobato. Grammar variational autoen- coder. In Doina Precup and Yee Whye Teh (eds.), Proceedings of the 34th International Con- ference on Machine Learning, volume 70 of Proceedings of Machine Learning Research, pp. 1945–1954, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR. URL http://proceedings.mlr.press/v70/kusner17a.html.

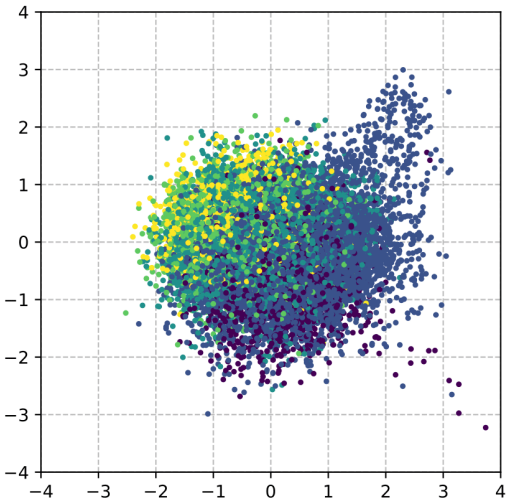
★ Martin Simonovsky and Nikos Komodakis. GraphVAE: Towards generation of small graphs using variational autoencoders. arXiv preprint arXiv:1802.03480, 2018.

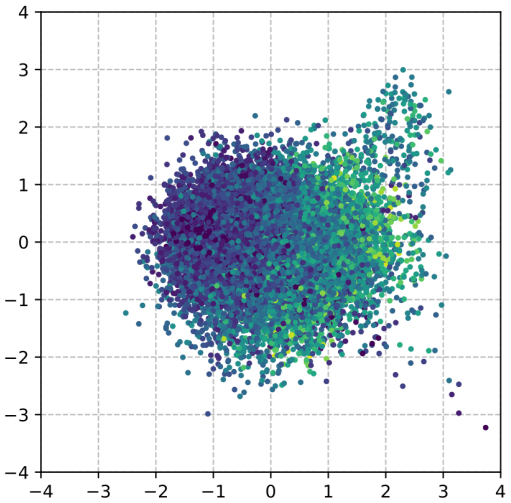
✣ Nicola De Cao and Thomas Kipf. MolGAN: An implicit generative model for small molecular graphs. arXiv preprint arXiv:1805.11973, 2018.

VAE Latent Space

- Properties of of interest include
  - Fraction of carbons with SP3 hybridization
  - Number of rings
  - Total Polar surface area

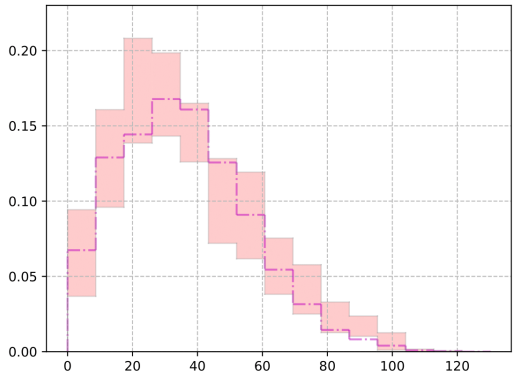




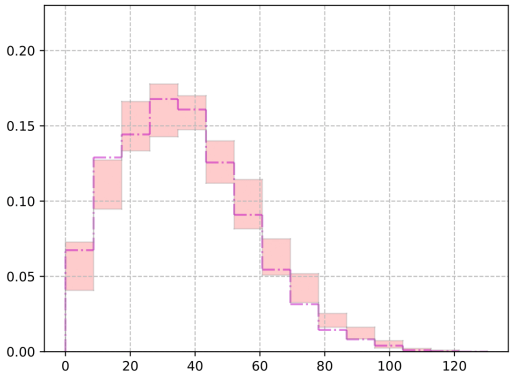


Predictive Results

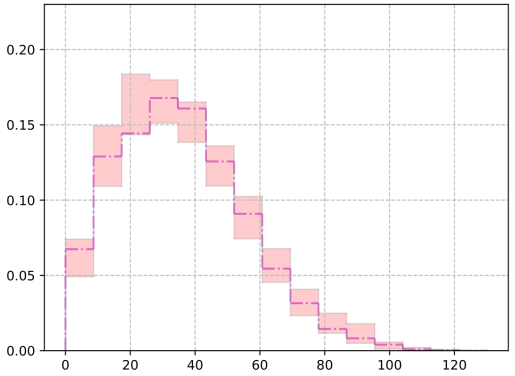
- The credible interval envelopes the reference solution ($K = 10\text{k}$).
- Polar Surface area



$K = 50$

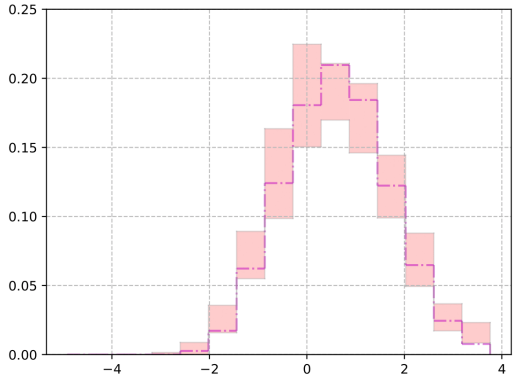


$K = 200$

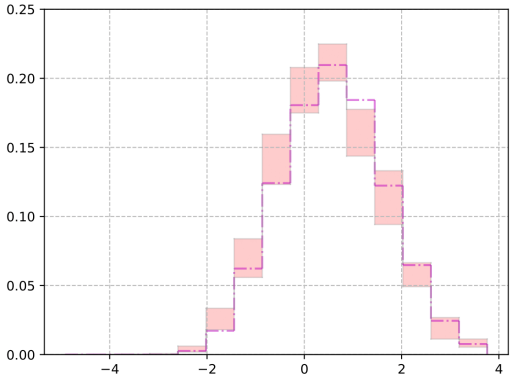


$K = 500$

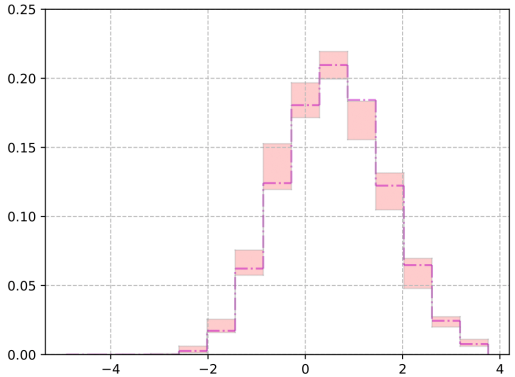
- LogP



$K = 50$



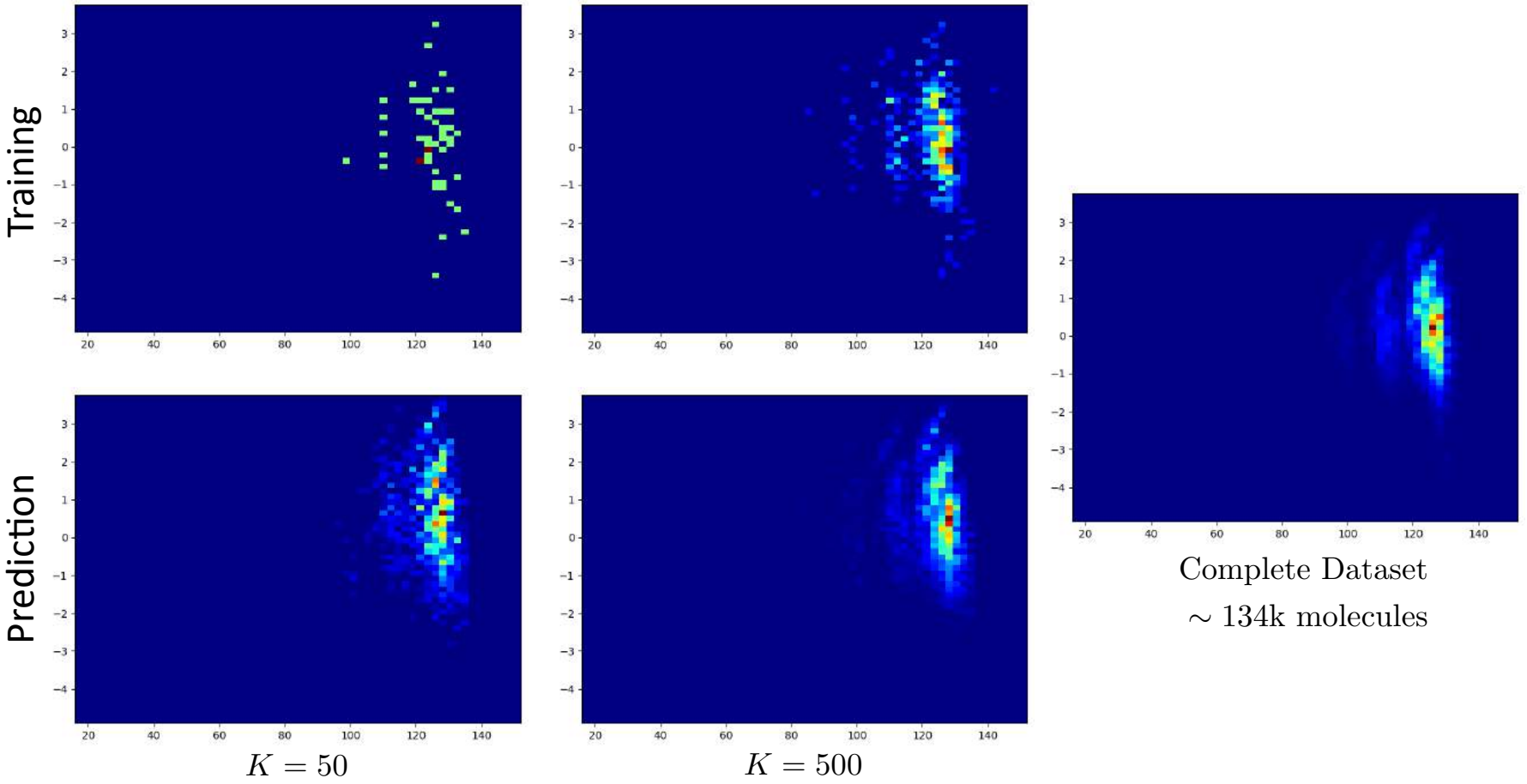
$K = 200$



$K = 500$

Predictive Results

- LogP vs. MolWt chemical space



Conditional Design

- One of the popular applications of generative models, is to design molecules with desired properties.
- In a conditional generative task, we simultaneously train the model on the molecules and their corresponding properties.

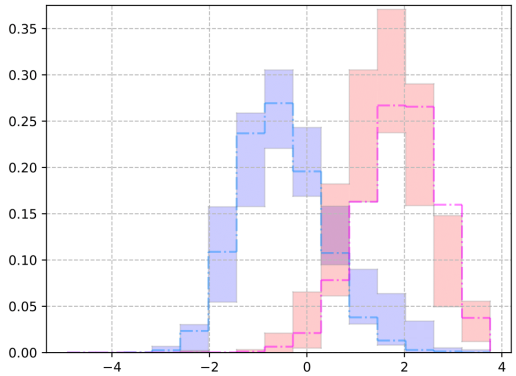
$$q_{\boldsymbol{\phi}}(\boldsymbol{z}|\mathcal{G},\boldsymbol{y}) = \mathcal{N}(\boldsymbol{z};\boldsymbol{\mu}_{\boldsymbol{\phi}}(\mathcal{G},\boldsymbol{y}),\boldsymbol{S}_{\boldsymbol{\phi}}(\mathcal{G},\boldsymbol{y}))$$

$$p_\theta(\mathcal{G}|\boldsymbol{z},\boldsymbol{y}) = p_\theta(\boldsymbol{W},\boldsymbol{f}|\boldsymbol{z},\boldsymbol{y}) = p_\theta(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{W},\boldsymbol{y})p_\theta(\boldsymbol{W}|\boldsymbol{z},\boldsymbol{y})$$

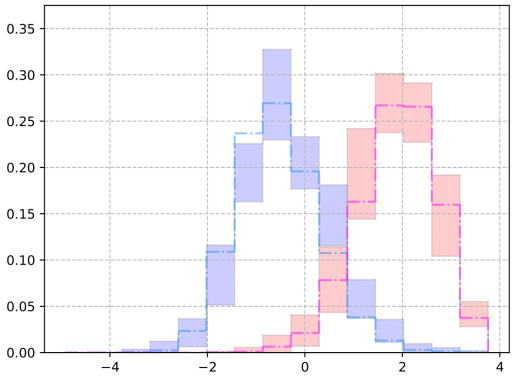
- New molecules are generated by sampling from $p(\boldsymbol{y}) = \mathcal{N}(\boldsymbol{y};\boldsymbol{\mu}_{\boldsymbol{y}},\boldsymbol{S}_{\boldsymbol{y}})$ and $p(\boldsymbol{z})$.
- Given a target value y_i, we sample the rest of the property values from $p(y_{l,l\neq i}|y_i)$.

Conditional Design

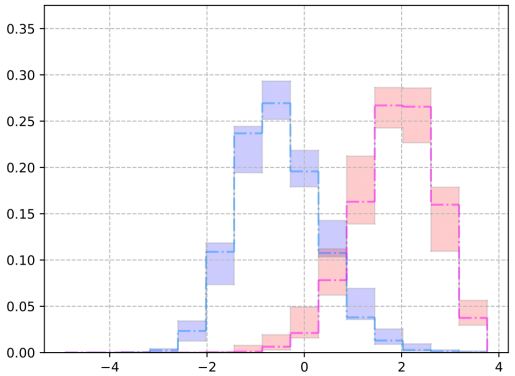
- Target 1: LogP, $y_1 = -1$ (blue), $y_1 = 2$ (red).



$K = 50$

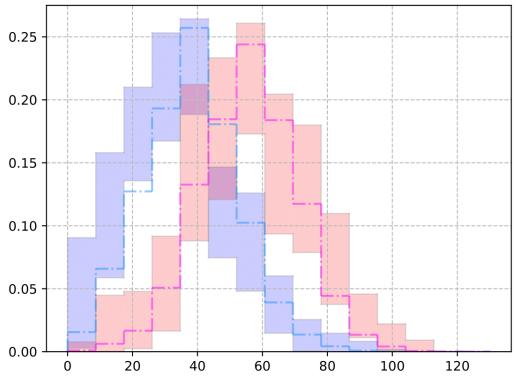


$K = 200$

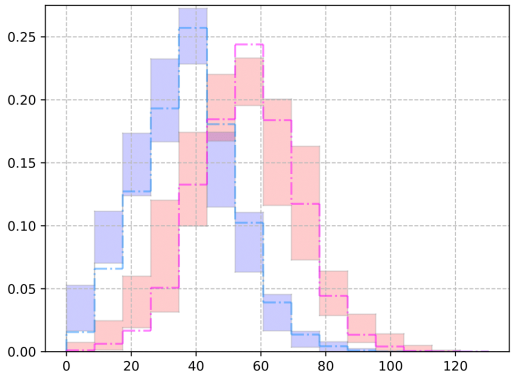


$K = 500$

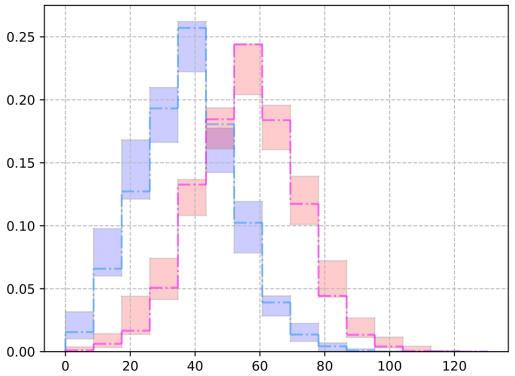
- Target 2: PSA, $y_2 = 40$ (blue), $y_2 = 60$ (red).



$K = 50$



$K = 200$

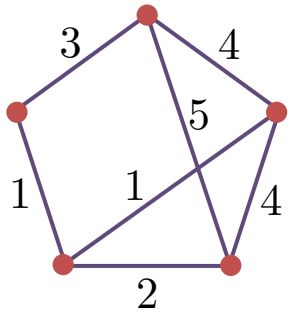


$K = 500$

Thanks!

Spectral Graph Filters

- In a weighted graph, in addition, a number (weight) is assigned to each edge, which shows the relation of the nodes depending on the given problem.



$$G = (V, E, W)$$

$$W = \begin{bmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{bmatrix}$$

$$W_{i,j} \in \mathbb{R}^+$$

$$d(i) = \sum_{j \in \mathcal{N}(i,1)} W_{i,j}$$

$$L = D - W$$

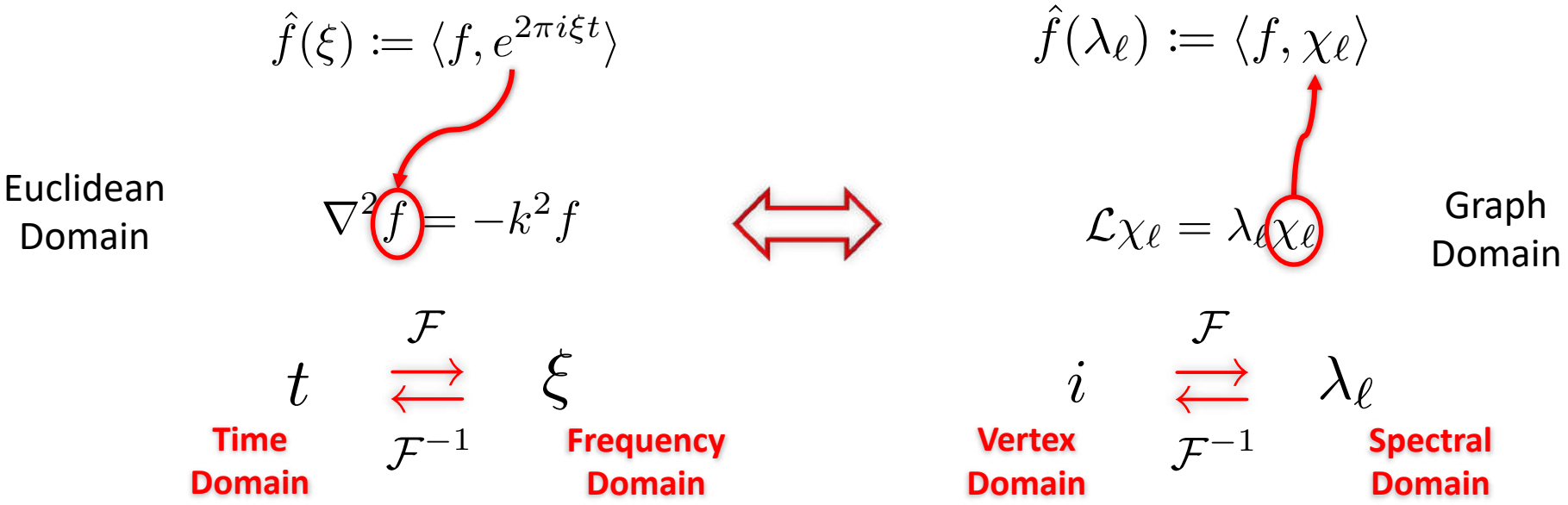
Laplacian

Diagonal degree matrix

Weight matrix

Spectral Graph Filters

- Fourier transform on graphs is defined as as an expansion in the Laplacian eigenvector basis



✣ Hammond, David K., Pierre Vandergheynst, and Rémi Gribonval. "Wavelets on graphs via spectral graph theory." *Applied and Computational Harmonic Analysis* 30, no. 2 (2011): 129-150.

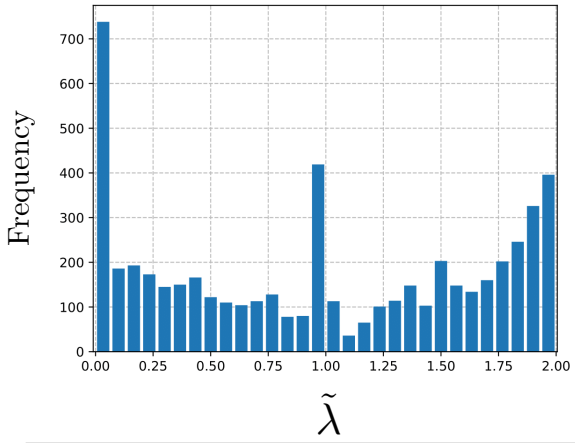
Spectral Graph Filters

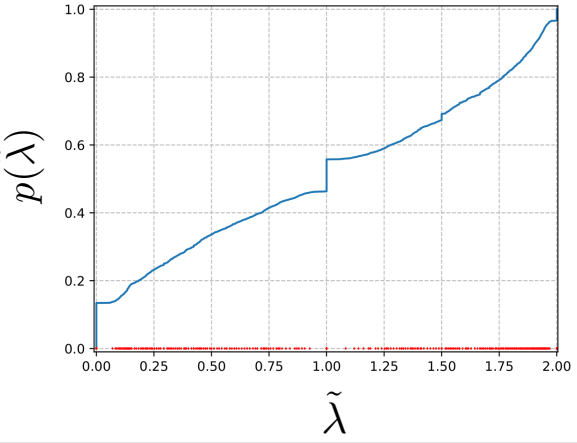
- Spectrum of each graph consists of eigenvalues of its Laplacian.

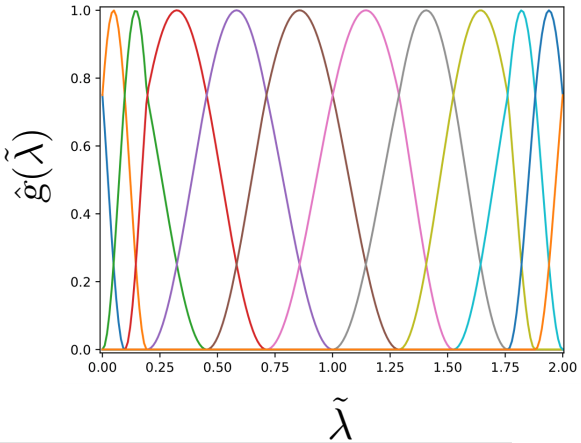
$$\mathcal{L} = \chi \Lambda \chi^* \qquad \tilde{\mathcal{L}} = \Delta^{-1/2} \mathcal{L} \Delta^{-1/2}$$

- We form spectrum adapted kernels✤ to increases the discriminatory power of the encoding network
- Filters are adapted based on cumulative spectral density function

$$P_{\tilde{\lambda}}(s) := \frac{1}{NK} \sum_{i=1}^{K} \sum_{\ell=0}^{N-1} \mathbb{1}\{\tilde{\lambda}_\ell^{(i)} \leq s\}$$







✤ Shuman, David I., Christoph Wiesmeyr, Nicki Holighaus, and Pierre Vandergheynst. "Spectrum-adapted tight graph wavelet and vertex-frequency frames." *IEEE Transactions on Signal Processing* 63, no. 16 (2015): 4223-4235.

Physical Constraint

- The constrained optimization problem is

$$\begin{aligned} \min_{\boldsymbol{\theta},\boldsymbol{\phi}} \ & -\mathscr{L}_{ELBO}(\boldsymbol{\theta},\boldsymbol{\phi};\mathscr{G}) \\ \text{subject to} \ & \text{for almost all } \boldsymbol{z} \sim p(\boldsymbol{z}) \\ & \mathcal{C}_1(\boldsymbol{\theta},\boldsymbol{z}) \leq 0, \ldots, \mathcal{C}_R(\boldsymbol{\theta},\boldsymbol{z}) \leq 0 \end{aligned}$$

VAE objective

Constraints

- This can be transformed to an unconstrained optimization problem using Lagrange multipliers methods

$$\min_{\boldsymbol{\theta},\boldsymbol{\phi}} \ -\mathscr{L}_{ELBO}(\boldsymbol{\theta},\boldsymbol{\phi};\mathscr{G}) + \sum_{r=1}^{\mathcal{R}} \mu_r \left(\int \mathcal{C}_+^{(r)}(\boldsymbol{z},\boldsymbol{\theta})^2 p(\boldsymbol{z}) d\boldsymbol{z} \right)^{\frac{1}{2}}$$

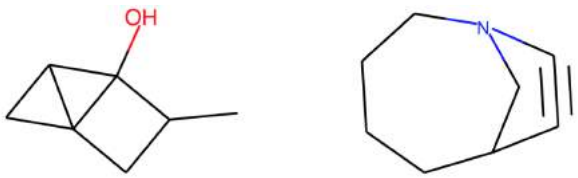
- In the constraint term, marginalization is performed to generalize the Lagrangian function for an infinite constraint set.

$$\int \mathcal{C}_+(\boldsymbol{z},\theta)^2 p(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \approx \frac{1}{L} \sum_{l=1}^{L} \mathcal{C}_+(\boldsymbol{z}^{(l)},\theta)^2, \quad \text{with } \boldsymbol{z}^{(l)} \sim p(\boldsymbol{z})$$

- The regularization method differs in that rather than solving for multipliers, regularization coefficients are changed to get desired result in terms of validity of the molecular graphs.

Physical Constraint

- Imposed constraints include:
  - **Connectivity**
  - **Valency**
  - While the above constraints result in valid Lewis structures, not all such molecules are feasible. One aspect that is overlooked in the literature is that many of these valid combinations would be energetically unstable.



  - **3−member cycles**
  - **Cycles with triple bonds**

Physical Constraint

- **Valence**
- This regularization term penalizes the network when the sum of the orders of the bonds connected to an atom exceeds the valency of that atom.

Capacity of bond type k

Probability of edge ε_{ij} having type k

$$Y_{m,n} = \sum_k y_k \tilde{W}_{m,n,k}$$

Expected capacity of edge ε_{ij}

Valence of atom type k

Probability of node v_i having type k

$$U_m = \sum_k u_k \tilde{f}_{m,k}$$

Expected valency of node v_i

- Since only the positive values of the term $\sum_n Y_{m,n} - U_m$ are not desirable, we define the regularization term as

$$\mathscr{G}_m^{(1)}(\boldsymbol{\theta}, \boldsymbol{z})_+ = \max\left(\sum_{n=1, n\neq m}^{N} Y_{m,n} - U_m, 0 \right)$$

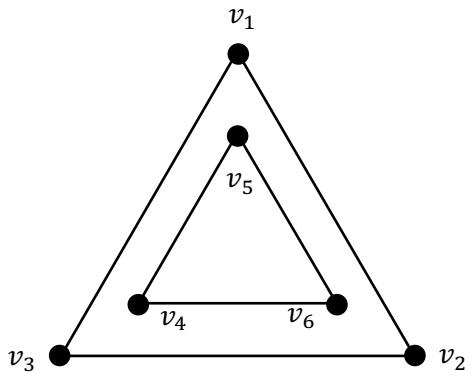
Physical Constraint

- **Connectivity**
- This regularization term penalizes the network if the output of the network is not a single connected graph.
- The n-th power of an adjacency matrix shows if there is a path of length n between nodes v_i and v_j.
- For a connected graph, the sum of all powers of matrix up to $N-1$ should only have non-zero elements.

$$A_{i,j} = 1 - \tilde{W}_{i,j,0}$$

$$A^0 = I,\ A^1 = A,\ A^{n+1} = \sigma(A^n A),\ n = 1, \ldots, N-2$$

$$C = \sigma\left(\sum_{n=0}^{N-1} A^n\right)$$



- This constraint encourages that the output is a single connected graph

$$\mathcal{G}^{(2)}_{m,n}(\boldsymbol{\theta}, \boldsymbol{z}) = q_m q_n \cdot [1 - 2C_{m,n}] + C_{m,n}\ , \quad \text{where} \quad q_i = 1 - \tilde{f}_{i,0}$$

Physical Constraint

- Let $\boldsymbol{A}$ represent the adjacency matrix of graph $\mathcal{G}$.
- We can find the number of the 3−member cycles of graph by summing the diagonal elements

$$\frac{1}{6}\mathrm{tr}(\boldsymbol{A}^3) = \frac{1}{6}\sum_m (\boldsymbol{A}^3)_{m,m}$$

- Using the generative model, instead of an adjacency matrix, we have a probabilistic representation of the weight matrix $p_{\boldsymbol{\theta}}(\boldsymbol{W}^{(l)}|\boldsymbol{z}^{(l)})$, where $\boldsymbol{z}^{(l)} \sim p(\boldsymbol{z})$

$$\bar{A}_{m,n} = 1 - p_{\boldsymbol{\theta}}(W_{m,n} = \varnothing|\boldsymbol{z})$$

- Using the relation for the number of the 3−member cycles, we can formulate the constraint term

$$\mathcal{C}^{(1)}(\boldsymbol{\theta}, \boldsymbol{z}) = \frac{1}{6}\mathrm{tr}(\bar{\boldsymbol{A}}^3)$$

Physical Constraint

- This constraint is specified by two terms:
  - A function to inspect membership in a cycle $\boldsymbol{B}_{\varepsilon_{m,n}} := (\boldsymbol{I} - \frac{1}{N}\boldsymbol{A}_{\varepsilon_{m,n}})^{-1}$.
  - An indicator function for triple bonds $\mathfrak{I}_{III}(\boldsymbol{W})$.
- Using these, we can formulate the constraint as

$$\mathfrak{I}_{III}(W_{m,n})(B_{\varepsilon_{m,n}})_{m,n} \leq 0, \quad \forall m \neq n$$

- We can define a probabilistic version of $\mathfrak{I}_{III}(\boldsymbol{W})$ as

$$\bar{D}_{m,n} := p_{\boldsymbol{\theta}}(W_{m,n} = III|\boldsymbol{z})$$

- Using these matrices, we can reformulate and obtain the constraint

$$\mathfrak{C}^{(2)}(\boldsymbol{\theta}, \boldsymbol{z}) = \sum_{m=1}^{N} \sum_{\substack{n=1 \\ n>m}}^{N} \bar{D}_{m,n}(\bar{B}_{\varepsilon_{m,n}})_{m,n}$$

Model Uncertainty

- $\boldsymbol{\pi}$ has prior distribution

$$p(\boldsymbol{\pi}) = \mathcal{D}ir(\boldsymbol{\pi}; \boldsymbol{\alpha}), \quad \text{with } \boldsymbol{\alpha} = [0, \ldots, 0] \in \mathbb{R}^K$$

- Bootstrap methods work under the assumption that all distinct values in $\mathcal{G}$ have been observed.

$$\begin{aligned} p(\boldsymbol{\pi}|\mathcal{G}) &\propto p(\mathcal{G}|\boldsymbol{\pi})p(\boldsymbol{\pi}) \\ &\propto \prod_k \pi_k^{n_k} \prod_k \pi_k^{\alpha_k - 1} \\ &\propto \prod_k \pi_k^{n_k + \alpha_k - 1}. \end{aligned}$$

- This leads to the posterior

$$p(\boldsymbol{\pi}|\mathcal{G}) = \mathcal{D}ir(\boldsymbol{\pi}; \boldsymbol{\alpha}'), \quad \text{with } \boldsymbol{\alpha}' = [1, \ldots, 1] \in \mathbb{R}^K$$