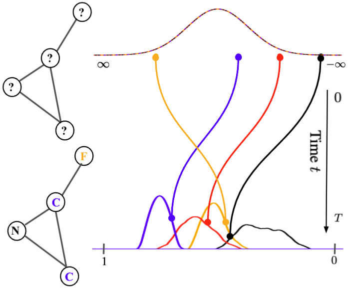


Modular Flows: Differential Molecular Generation

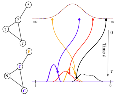
Yogesh Verma, Samuel Kaski, Markus Heinonen and Vikas Garg

Aalto University

36^{th} Conference on Neural Information Processing Systems (NeurIPS 2022)
New Orleans, USA

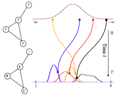
Challenge of Molecular Generation

- Molecular generation is fundamental for drug discovery, material synthesis, etc.
- Challenge: Generate valid molecules with various criteria



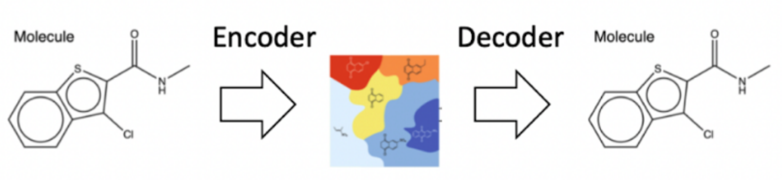
Challenge of Molecular Generation

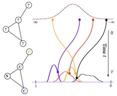
- Molecular generation is fundamental for drug discovery, material synthesis, etc.
- Challenge: Generate valid molecules with various criteria
- Can generative models can learn to achieve high generative validity intrinsically ??



Challenge of Molecular Generation

- Molecular generation is fundamental for drug discovery, material synthesis, etc.
- Challenge: Generate valid molecules with various criteria
- Can generative models can learn to achieve high generative validity intrinsically ??
- We resolve this dilemma, by proposing a method inspired by graph PDEs to reconcile local densities $\Rightarrow$ globally aligned densities





Representation

- Given a molecule, represent as a graph $G = (V, E)$, where vertex corresponds to atoms : $v \in \mathcal{A}$, while edges are bonds. We assume the following decomposition as

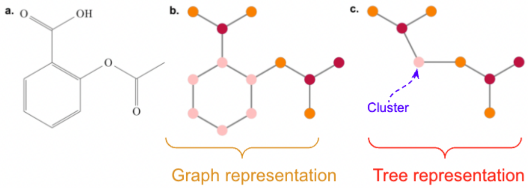
$$p(G) := p(V|E, \{z\}) = \prod_{i=1}^{M} \mathtt{Cat}(v_i|\sigma(\mathbf{z}_i)) \tag{1}$$

Representation

- Given a molecule, represent as a graph $G = (V, E)$, where vertex corresponds to atoms : $v \in \mathcal{A}$, while edges are bonds. We assume the following decomposition as

$$p(G) := p(V|E, \{z\}) = \prod_{i=1}^{M} \mathtt{Cat}(v_i|\sigma(\mathbf{z}_i)) \tag{1}$$

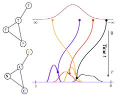
- **Tree representation** by decomposing a molecular graph into a tree, as in JT-VAE, but restricting these clusters to ring substructures.



Differential Modular Flows

- Model the node scores $\mathbf{z}_i$ as a CNF over time $t \in \mathrm{R}_+$, with $\mathbf{z}_i(0) \sim \mathcal{N}(0, I)$. The dynamics are parameterized by a coupled ODE, jointly represented as, where $\mathcal{N}_i$ is the set of neighbors, $\mathbf{x}$ is the spatial information.

$$\dot{\mathbf{z}}_i(t) = \begin{pmatrix} \dot{\mathbf{z}}_i(t)) \\ \vdots \\ \dot{\mathbf{z}}_M(t) \end{pmatrix} = \begin{pmatrix} f_\theta\big(t, \mathbf{z}_1(t), \mathbf{z}_{\mathcal{N}_1}(t), \mathbf{x}_i, \mathbf{x}_{\mathcal{N}_i}\big) \\ \vdots \\ f_\theta\big(t, \mathbf{z}_M(t), \mathbf{z}_{\mathcal{N}_M}(t), \mathbf{x}_i, \mathbf{x}_{\mathcal{N}_M}\big) \end{pmatrix} \tag{2}$$

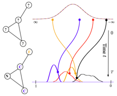


Differential Modular Flows

- Model the node scores $\mathbf{z}_i$ as a CNF over time $t \in \mathbb{R}_+$, with $\mathbf{z}_i(0) \sim \mathcal{N}(0, I)$. The dynamics are parameterized by a coupled ODE, jointly represented as, where $\mathcal{N}_i$ is the set of neighbors, $\mathbf{x}$ is the spatial information.

$$\dot{\mathbf{z}}_i(t) = \begin{pmatrix} \dot{\mathbf{z}}_i(t)) \\ \vdots \\ \dot{\mathbf{z}}_M(t) \end{pmatrix} = \begin{pmatrix} f_\theta\left(t, \mathbf{z}_1(t), \mathbf{z}_{\mathcal{N}_1}(t), \mathbf{x}_i, \mathbf{x}_{\mathcal{N}_i}\right) \\ \vdots \\ f_\theta\left(t, \mathbf{z}_M(t), \mathbf{z}_{\mathcal{N}_M}(t), \mathbf{x}_i, \mathbf{x}_{\mathcal{N}_M}\right) \end{pmatrix} \tag{2}$$

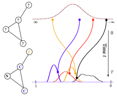
- **Equivariant Local differential**: To respect the natural equivariances of the molecule, we choose to use E(3)-Equivariant Graph Neural Networks as choice for f_θ



Training Objective

- We maximizing the score cross-entropy $\mathbb{E}_{\hat{p}_{\mathrm{data}}(\mathbf{z}(T))}[\log p_\theta(\mathbf{z}(T))]$, where we map the set of graphs $\{G_n\}$ into a set of scores $\{\mathbf{z}_n\}$ via:

$$\mathbf{z}_n(G_n; \epsilon) = (1-\epsilon)\,\mathrm{onehot}(G_n) \; + \; \frac{\epsilon}{|\mathcal{A}_f|}\mathbf{1}_{M(n)}\mathbf{1}_{|\mathcal{A}_f|}^\top \tag{3}$$

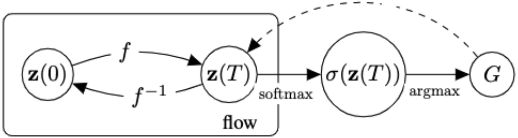


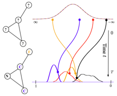
Training Objective

- We maximizing the score cross-entropy $\mathbb{E}_{\hat{p}_{\mathrm{data}}(\mathbf{z}(T))}[\log p_\theta(\mathbf{z}(T))]$, where we map the set of graphs $\{G_n\}$ into a set of scores $\{\mathbf{z}_n\}$ via:

$$\mathbf{z}_n(G_n;\epsilon) = (1-\epsilon)\,\mathrm{onehot}(G_n) \;+\; \frac{\epsilon}{|\mathcal{A}_f|}\mathbf{1}_{M(n)}\mathbf{1}_{|\mathcal{A}_f|}^\top \tag{3}$$

- We exploit the (non-reversible) composition of the `argmax` and `softmax` operations short-circuit in reverse direction as shown. This short-circuiting allows keep the forward and backward aligned.





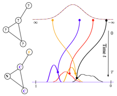
Training Objective

- We maximizing the score cross-entropy $\mathbb{E}_{\hat{p}_{\mathrm{data}}(\mathbf{z}(T))}[\log p_\theta(\mathbf{z}(T))]$, where we map the set of graphs $\{G_n\}$ into a set of scores $\{\mathbf{z}_n\}$ via:

$$\mathbf{z}_n(G_n; \epsilon) = (1 - \epsilon)\ \mathrm{onehot}(G_n)\ +\ \frac{\epsilon}{|\mathcal{A}_f|}\mathbf{1}_{M(n)}\mathbf{1}_{|\mathcal{A}_f|}^\top \tag{4}$$

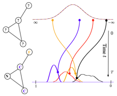
- We exploit the (non-reversible) composition of the `argmax` and `softmax` operations short-circuit in reverse direction as shown. This short-circuiting allows keep the forward and backward aligned.
- We thus maximize an objective over N training graphs,

$$\arg\max_\theta \qquad \mathcal{L} = \mathbb{E}_{\hat{p}_{\mathrm{data}}(\mathbf{z})} \log p_\theta(\mathbf{z}) \approx \frac{1}{N}\sum_{n=1}^{N} \log p_T\left(\mathbf{z}(T) = \mathbf{z}_n\right) \tag{5}$$



Molecular Experiments

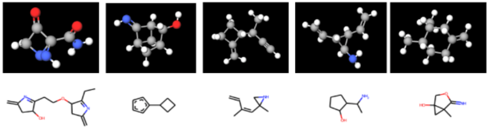
- We trained the model on QM9 and ZINC250K dataset, and evaluated in terms of:
    - **Validity**: Fraction of molecules that satisfy chemical valency rule
    - **Uniqueness**: Fraction of non-duplicate generations
    - **Novelty**: Fraction of molecules not present in training data
    - **Reconstruction**: Fraction of molecules that can be reconstructed from their encoding



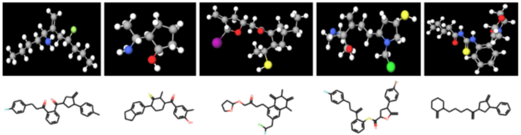
Molecular Experiments

Method	Validity %	Uniqueness %	Novelty %	Reconstruction %
GVAE	60.2	9.3	80.9	96.0
GraphNVP*	83.1	99.2	58.2	100
GRF*	84.5	66	58.6	100
GraphAF*	67	94.2	88.8	100
GraphDF*	82.7	97.6	98.1	100
MoFlow*	89.0	98.5	96.4	100
`ModFlow` (2D-EGNN)	**96.2** ± 1.7	**99.5**	**100**	100
`ModFlow` (3D-EGNN)	**98.3** ± 0.7	99.1	**100**	100
`ModFlow` (JT-2D-EGNN)	**97.9** ± 1.2	99.2	**100**	100
`ModFlow` (JT-3D-EGNN)	**99.1** ± 0.8	99.3	**100**	100

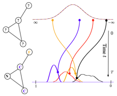
Method	Validity %	Uniqueness %	Novelty %	Reconstruction %
MRNN	65	99.89	100	n/a
GVAE	7.2	9	100	53.7
GCPN	20	99.97	100	n/a
GraphNVP*	42.6	94.8	100	100
GRF*	73.4	53.7	100	100
GraphAF*	68	99.1	100	100
GraphDF*	89	99.2	100	100
MoFlow*	50.3	**99.9**	100	100
`ModFlow` (2D-EGNN)	**94.8** ± 1.0	99.4	100	100
`ModFlow` (3D-EGNN)	**95.4** ± 1.2	99.7	100	100
`ModFlow` (JT-2D-EGNN)	**97.4** ± 1.4	99.1	100	100
`ModFlow` (JT-3D-EGNN)	**98.1** ± 0.9	99.3	100	100



QM9



ZINC250K



Property-targeted Molecular Optimization

- We used a linear regression model g_σ to regress the latent embeddings computed via pre-trained `ModFlow` model, to property scores (y) , and interpolate via,

Property-targeted Molecular Optimization

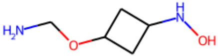
- We used a linear regression model g_σ to regress the latent embeddings computed via pre-trained `ModFlow` model, to property scores (y) , and interpolate via,

$$Z = f_\theta(\mathcal{M}), \; y = g_\sigma(Z) \tag{6}$$

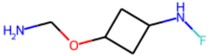
$$Z' = Z + \lambda * \frac{dy}{dZ} \tag{7}$$

- Method is conducted for K steps, where λ is the search step and Z' is decoded back into molecule via $\mathcal{M}' = f^{-1}(\mathcal{Z}')$

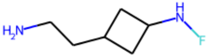
Property-targeted Molecular Optimization



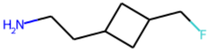
QED: 0.355



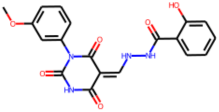
QED: 0.419



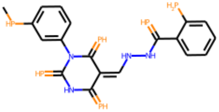
QED: 0.552



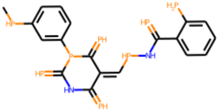
QED: 0.614



QED: 0.332



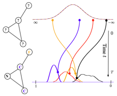
QED: 0.428



QED: 0.557

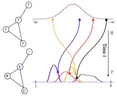
Conclusion

- We propose Physics-inspired co-evolving continuous-time flows, inspired by graph PDEs $\Rightarrow$ modular coupled ODE system



Conclusion

- We propose Physics-inspired co-evolving continuous-time flows, inspired by graph PDEs $\Rightarrow$ modular coupled ODE system
- Resulting in accurate modeling of graph densities and high-quality molecular generation without any validity checks or corrections.



Conclusion

- We propose Physics-inspired co-evolving continuous-time flows, inspired by graph PDEs $\Rightarrow$ modular coupled ODE system
- Resulting in accurate modeling of graph densities and high-quality molecular generation without any validity checks or corrections.

More information on the website and visit the **poster** for in-person discussion!

Thank you for listening!