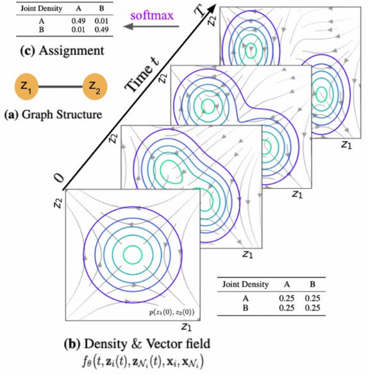


Joint Density	A	B
A	0.49	0.01
B	0.01	0.49

(c) Assignment

(a) Graph Structure

Joint Density	A	B
A	0.25	0.25
B	0.25	0.25

(b) Density & Vector field

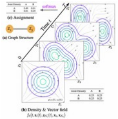
$f_\theta(t, \mathbf{z}_i(t), \mathbf{z}_{\mathcal{N}_i}(t), \mathbf{x}_i, \mathbf{x}_{\mathcal{N}_i})$

Modular Flows: Differential Molecular Generation

Yogesh Verma[*], Samuel Kaski[*,m], Markus Heinonen[*] and Vikas Garg[*,^]

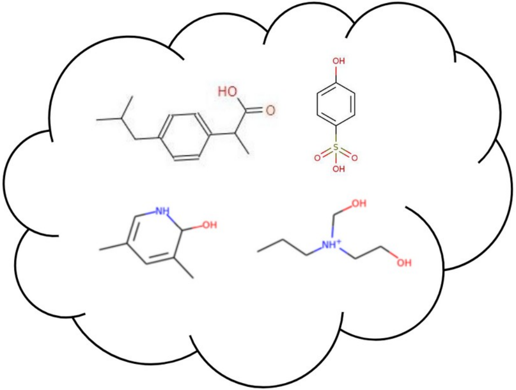
36th Conference on Neural Information Processing Systems (NeurIPS 2022)
New Orleans, USA

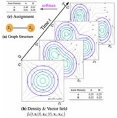
* Aalto University
^ YaiYai Ltd
m University of Manchester



Challenge of Molecular Generation

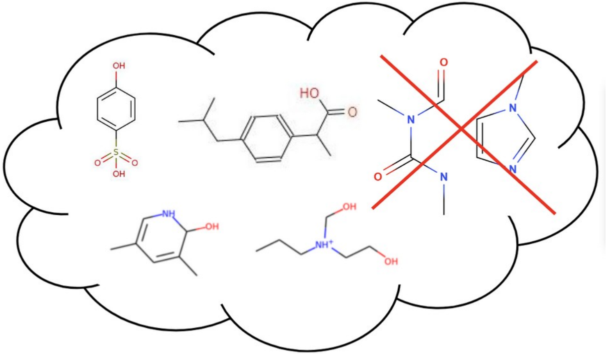
- Molecular generation is fundamental for drug discovery, material synthesis, etc.

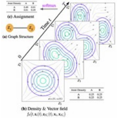




Challenge of Molecular Generation

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

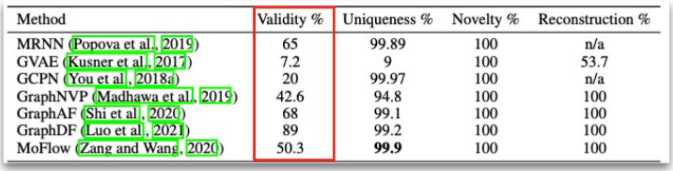




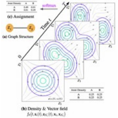
Challenge of Molecular Generation

- Molecular generation is fundamental for drug discovery, material synthesis, etc.
- Challenge: Generate valid molecules
- Can generative models achieve high intrinsic validity ?

Current SOTA



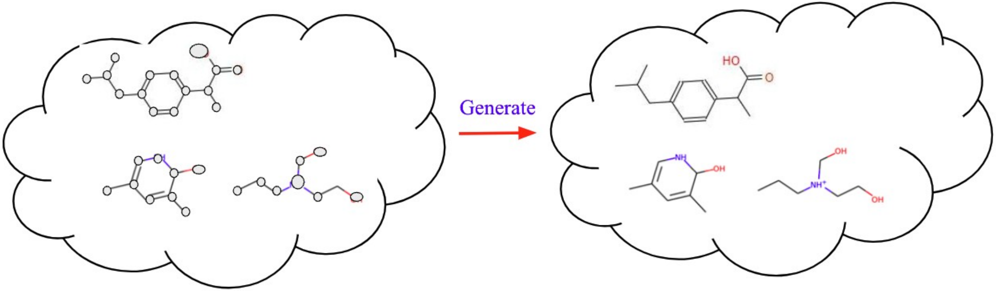
Method	Validity %	Uniqueness %	Novelty %	Reconstruction %
MRNN (Popova et al., 2019)	65	99.89	100	n/a
GVAE (Kusner et al., 2017)	7.2	9	100	53.7
GCPN (You et al., 2018a)	20	99.97	100	n/a
GraphNVP (Madhawa et al., 2019)	42.6	94.8	100	100
GraphAF (Shi et al., 2020)	68	99.1	100	100
GraphDF (Luo et al., 2021)	89	99.2	100	100
MoFlow (Zang and Wang, 2020)	50.3	**99.9**	100	100

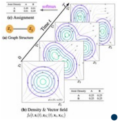


Problem Formulation

Molecular Generation:

- Given a molecular structure ⇒ assign atom labels

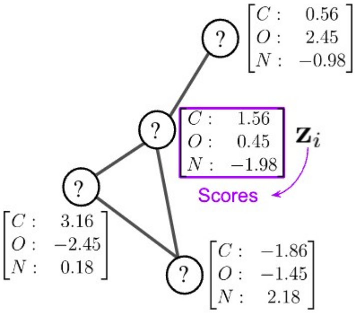


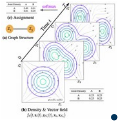


Representation

- Likelihood of atoms $\{v_i = \{ C,N,O,P....\}\}$ given edges *E* and atom scores $\{z_i\}$:

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

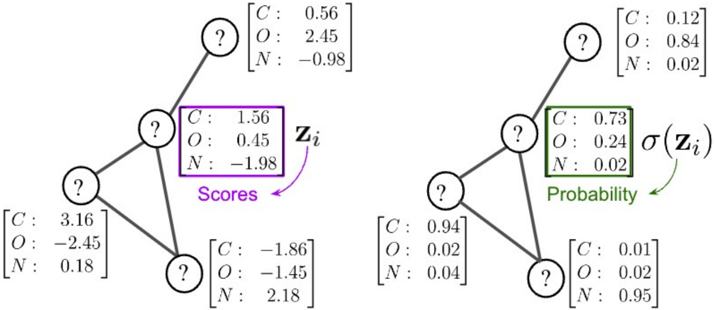


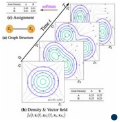


Representation

- Likelihood of atoms {v_i = { C,N,O,P….}} given edges *E* and atom scores {z_i}:

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

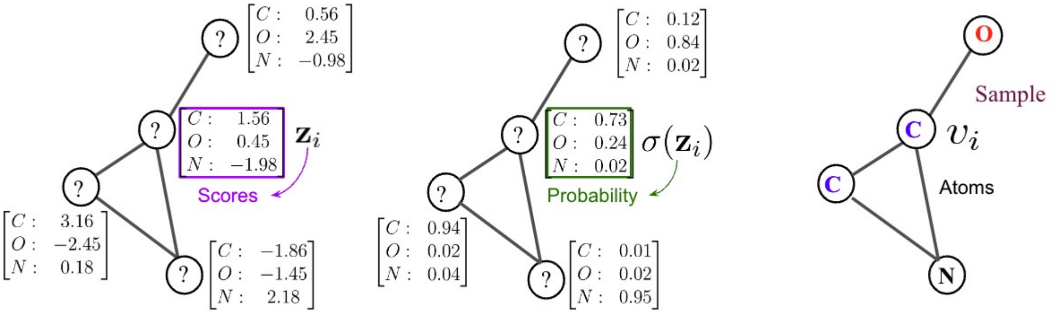


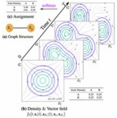


Representation

- Likelihood of atoms $\{v_i = \{ C,N,O,P....\}\}$ given edges *E* and atom scores $\{z_i\}$:

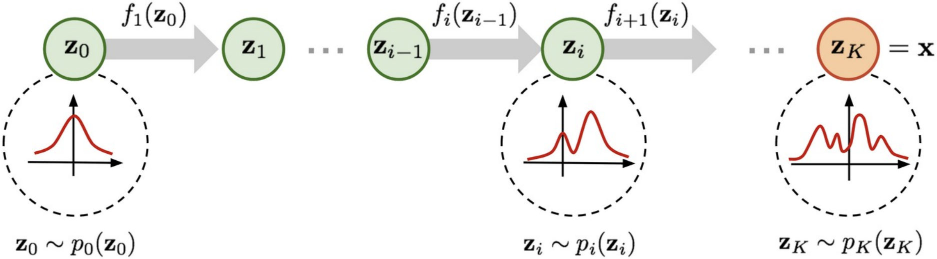
$$p(V|E,\{z\}) = \prod_{i=1}^{M} \texttt{Cat}(v_i|\sigma(\mathbf{z}_i)).$$





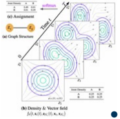
Differential Modular Flows

- Normalizing Flows*



$f_1(\mathbf{z}_0)$
$f_i(\mathbf{z}_{i-1})$
$f_{i+1}(\mathbf{z}_i)$
$\mathbf{z}_0$
$\mathbf{z}_1$
$\mathbf{z}_{i-1}$
$\mathbf{z}_i$
$\mathbf{z}_K = \mathbf{x}$
$\mathbf{z}_0 \sim p_0(\mathbf{z}_0)$
$\mathbf{z}_i \sim p_i(\mathbf{z}_i)$
$\mathbf{z}_K \sim p_K(\mathbf{z}_K)$

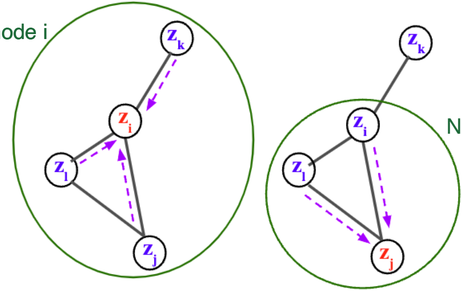
* FFJORD: Free-form Continuous Dynamics for Scalable Reversrible Generative Models



Differential Modular Flows

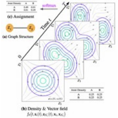
- Model the node scores z_i as a CNF over time $t \in \mathbb{R}_+$, with $z_i(0) \sim \mathcal{N}(0, I)$.
- The dynamics are parameterized by a coupled ODE over neighbors $\mathcal{N}_i$:

$$\dot{\mathbf{z}}_i(t) := \frac{\partial \mathbf{z}_i(t)}{\partial t} = f_\theta(t, \underbrace{\mathbf{z}_i(t), \mathbf{z}_{\mathcal{N}_i}(t)}_{\text{Scores}}, \underbrace{\mathbf{x}_i, \mathbf{x}_{\mathcal{N}_i}}_{\text{Spatial Information}}).$$



Neighborhood of node i

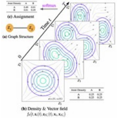
Neighborhood of node j



Differential Modular Flows

- Collecting all nodes, it can be represented as

$$\dot{\mathbf{z}}(t) = \underbrace{\begin{pmatrix} \dot{\mathbf{z}}_1(t) \\ \vdots \\ \dot{\mathbf{z}}_M(t) \end{pmatrix} = \begin{pmatrix} f_\theta\left(t, \mathbf{z}_1(t), \mathbf{z}_{\mathcal{N}_1}(t), \mathbf{x}_i, \mathbf{x}_{\mathcal{N}_i}\right) \\ \vdots \\ f_\theta\left(t, \mathbf{z}_M(t), \mathbf{z}_{\mathcal{N}_M}(t), \mathbf{x}_i, \mathbf{x}_{\mathcal{N}_i}\right) \end{pmatrix}}_{\text{Modular system of ODEs}}$$



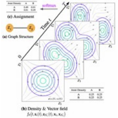
Differential Modular Flows

- Collecting all nodes, it can be represented as

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

$$\mathbf{z}(T) = \mathbf{z}(0) + \int_0^T \dot{\mathbf{z}}(t)dt$$

→ Solving the dynamics



Differential Modular Flows

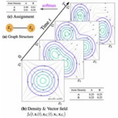
- Collecting all nodes, it can be represented as

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

$$\mathbf{z}(T) = \mathbf{z}(0) + \int_0^T \dot{\mathbf{z}}(t)dt$$

$$\frac{d \log p_t(\mathbf{z}_i(t))}{dt} = -\operatorname{tr}\left(\frac{\partial f_\theta\left(t, \mathbf{z}_i(t), \mathbf{z}_{\mathcal{N}_i}(t), \mathbf{x}_i, \mathbf{x}_{\mathcal{N}_i}\right)}{\partial \mathbf{z}_i}\right)$$

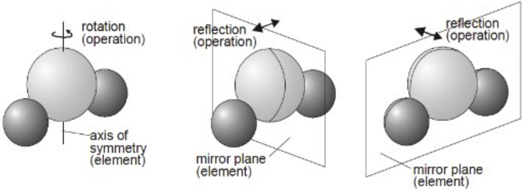
Change in density due to flow

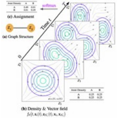


Equivariant local differential

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

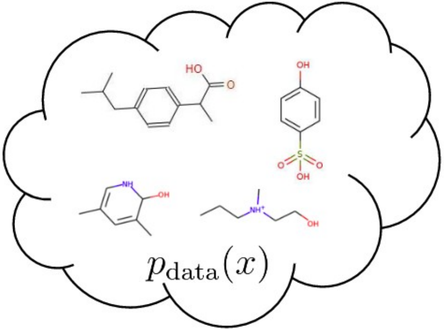
- Translation Equivariance
- Permutation Equivariance
- Rotation (and Reflection) Equivariance
- Size Invariance

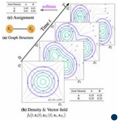




Training Objective

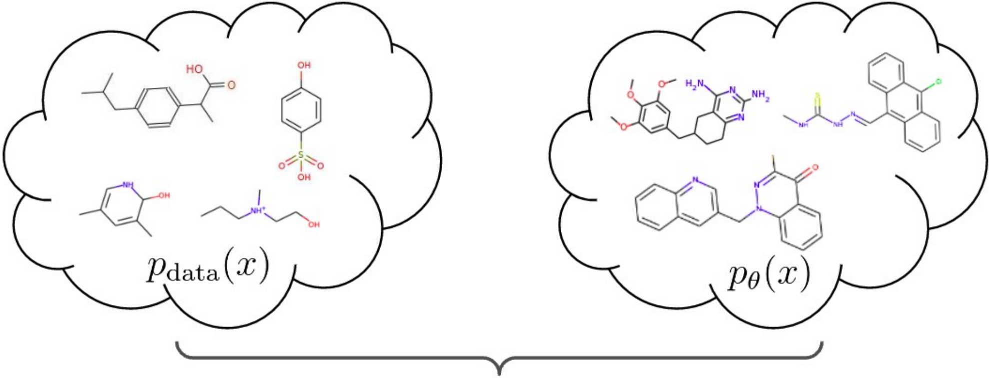
- Data consists of molecular graphs



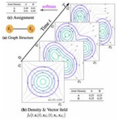


Training Objective

- The minimization of $\mathrm{KL}[p_{\text{data}}||p_\theta]$ is equivalent maximizing the cross entropy $\mathrm{E}_{p_{\text{data}}}[\log p_\theta]$



$$\min \mathrm{KL}[p_{\text{data}}||p_\theta] \propto \max \mathrm{E}_{p_{\text{data}}}[\log p_\theta]$$

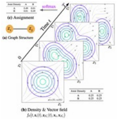


Training Objective

- We map the set of graphs {Gₙ} into a set of scores {zₙ} via a noisy one-hot encoding and thus maximize an objective over *N* training graphs,

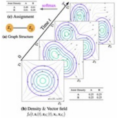
$$\arg\max_{\theta} \quad \frac{1}{N}\sum_{n=1}^{N} \log p_T\left(\mathbf{z}(T) = \mathbf{z}_n^{obs}\right)$$

$$\underbrace{\log p(\mathbf{z}(0))}_{\text{Starting from normal distribution}} - \underbrace{\log p(\mathbf{z}(T))}_{\text{Final target distribution}} = \underbrace{\int_0^T \mathrm{tr}\left(\frac{\partial f}{\partial z(t)}\right) dt}_{\text{Divergence term by flow}}$$

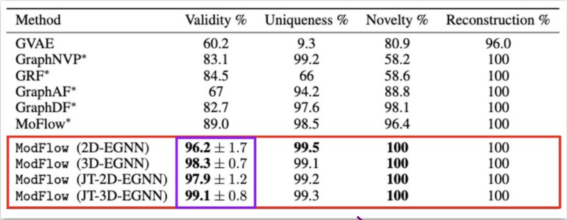


Molecular Experiments

- Data: (i) QM9: 134k small organic molecules, (ii) ZINC250K: 250k drug-like molecules.
- Metrics:
  - **Validity:** Fraction of molecules that satisfy chemical valency rule
  - **Uniqueness:** Fraction of non-duplicate generations
  - **Novelty:** Fraction of novel molecules
  - **Reconstruction:** Fraction of molecules that can be reconstructed from their encoding
  - **FCD:** measures diversity and chemical and biological property alignment
  - **SNN:** quantifies closeness of generated molecules to true molecule manifold
  - **Frag:** measures distance between the fragment frequencies generated and reference
  - **IntDiv:** diversity by computing pairwise similarity of the generated molecules



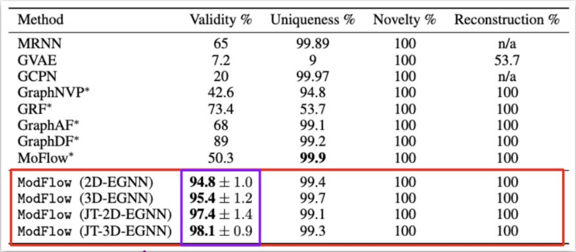
Molecular Experiments



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

High Validity

QM9

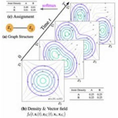


Method	Validity %	Uniqueness %	Novelty %	Reconstruction %
MRNN	65	99.89	100	n/a
GVAE	7.2	9	100	53.7
GCPN	20	99.97	100	n/a
GraphNVP*	42.6	94.8	100	100
GRF*	73.4	53.7	100	100
GraphAF*	68	99.1	100	100
GraphDF*	89	99.2	100	100
MoFlow*	50.3	**99.9**	100	100
ModFlow (2D-EGNN)	**94.8** ± 1.0	99.4	100	100
ModFlow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

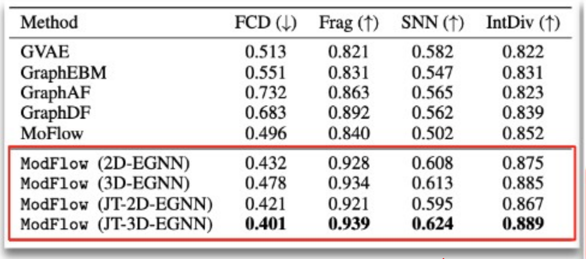
High Validity

ZINC250K

Near to 100% Validity

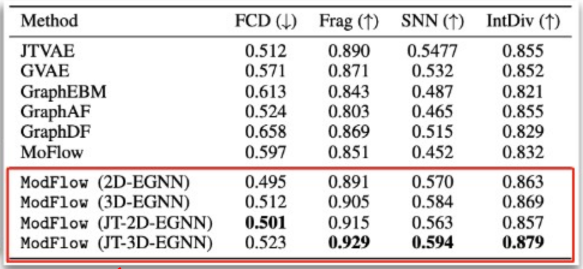


Molecular Experiments



Method	FCD (↓)	Frag (↑)	SNN (↑)	IntDiv (↑)
GVAE	0.513	0.821	0.582	0.822
GraphEBM	0.551	0.831	0.547	0.831
GraphAF	0.732	0.863	0.565	0.823
GraphDF	0.683	0.892	0.562	0.839
MoFlow	0.496	0.840	0.502	0.852
ModFlow (2D-EGNN)	0.432	0.928	0.608	0.875
ModFlow (3D-EGNN)	0.478	0.934	0.613	0.885
ModFlow (JT-2D-EGNN)	0.421	0.921	0.595	0.867
ModFlow (JT-3D-EGNN)	**0.401**	**0.939**	**0.624**	**0.889**

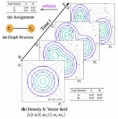
QM9



Method	FCD (↓)	Frag (↑)	SNN (↑)	IntDiv (↑)
JTVAE	0.512	0.890	0.5477	0.855
GVAE	0.571	0.871	0.532	0.852
GraphEBM	0.613	0.843	0.487	0.821
GraphAF	0.524	0.803	0.465	0.855
GraphDF	0.658	0.869	0.515	0.829
MoFlow	0.597	0.851	0.452	0.832
ModFlow (2D-EGNN)	0.495	0.891	0.570	0.863
ModFlow (3D-EGNN)	0.512	0.905	0.584	0.869
ModFlow (JT-2D-EGNN)	**0.501**	0.915	0.563	0.857
ModFlow (JT-3D-EGNN)	0.523	**0.929**	**0.594**	**0.879**

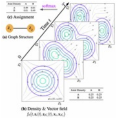
ZINC250K

Improvement across all metrics



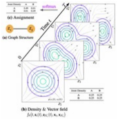
Conclusion

- We propose physics-inspired co-evolving continuous-time flows, inspired by graph PDEs. We also extend this framework to work with Junction Trees.



Conclusion

- We propose physics-inspired co-evolving continuous-time flows, inspired by graph PDEs. We also extend this framework to work with Junction Trees.
- Accurate modeling of graph densities and high-quality molecular generation with improvement across all metrics.



Conclusion

- We propose physics-inspired co-evolving continuous-time flows, inspired by graph PDEs. We also extend this framework to work with Junction Trees.
- Accurate modeling of graph densities and high-quality molecular generation with improvement across all metrics.
- Website: `https://yogeshverma1998.github.io/ModFlow/`
- Visit the **poster** for in-person discussion!

Thank you for listening!