Generative Diffusion Models on Graphs: Methods and Applications

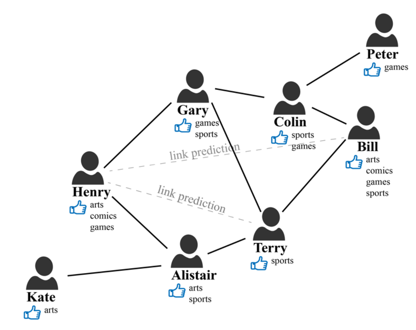
Chengyi LIU[1], Wenqi Fan[1], Yunqing Liu[1], Jiatong Li[1], Hang Li[2], Hui Liu[2], Jiliang Tang[2], Qing Li[1]

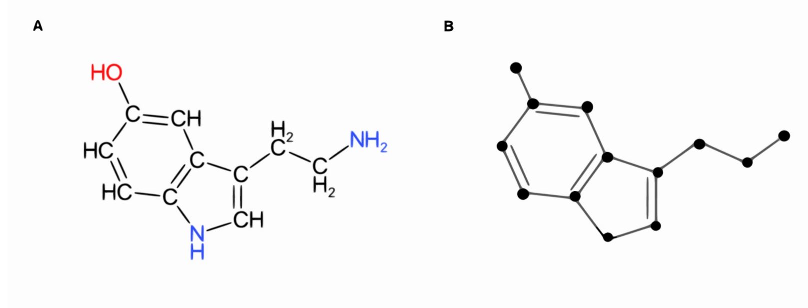
[1]The Hong Kong Polytechnic University, [2]Michigan State University

Introduction

Graphs

Graphs can represent the rich variety of relationships between real-world entities. They have been widely used in a diversity of domains, aiming to model association information and structural patterns among various real-world objects.





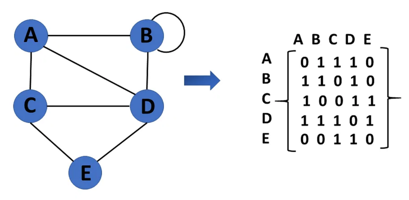
Graph Generation

Graph generative models, with the goal of learning the given graph distributions and generating novel graphs, can be categorized into two generation patterns:

- autoregressive generation
- one-shot generation

In general, graph generation faces three fundamental challenges:

- ☐ *Discreteness*
- ☐ *Complex Intrinsic Dependencies*
- ☐ *Permutation Invariant*



Diffusion Models

Three representative diffusion frameworks:

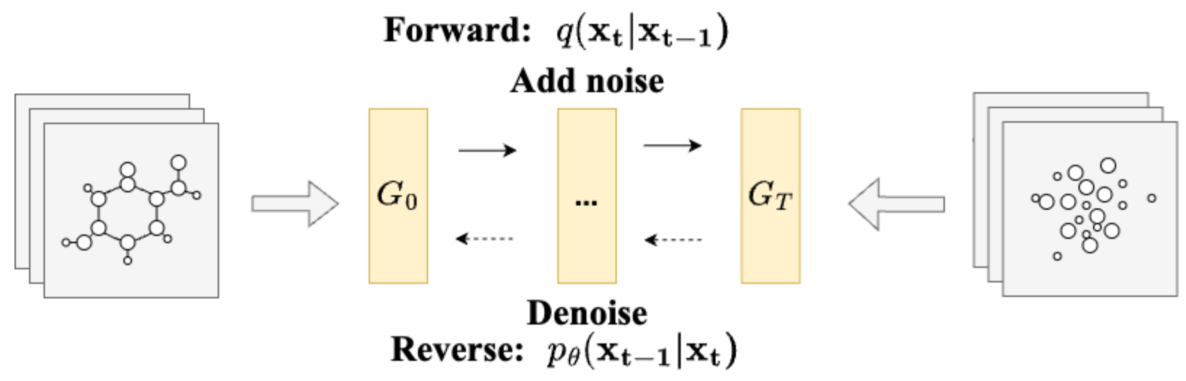
- ➢ Score Matching with Langevin Dynamics (SMLD)
- ➢ Denoising Diffusion Probabilistic Model (DDPM) on Graphs
- ➢ Diffusion Models

Two main stages:

- ◆ Forward diffusion: perturb the original data by adding random noise (generally Gaussian noise)
- ◆ Reverse diffusion: recover the original input data from the random noise.

Advancement:

- ✓ Solid theoretical foundation
- ✓ Easy-to-tractable probabilistic parameters



Diffusion Model

Generative Diffusion Models on Graphs

Score Matching with Langevin Dynamics (SMLD) on Graphs

- ◆ Forward: a sequence of incremental noise: $q_\sigma(\tilde{x}|x) \coloneqq \mathcal{N}(\tilde{x}|x,\sigma^2 I)$
- ◆ Reverse: learning the gradient of the data distribution $\nabla_x \log p(x)$

EDP-GNN:

- The very first score-matching diffusion method, which is for undirected graph.

Denoising Diffusion Probabilistic Model (DDPM) on Graphs

Constructs two parameterized Markov chains:

- ◆ Forward: $q(x_t|x_{t-1}) = \mathcal{N}(x_t;\ \sqrt{1-\beta_t}x_{t-1},\ \beta_t I)$
- ◆ Reverse: $p_\theta(x_t|x_{t-1}) = \mathcal{N}(x_{t-1};\ \mu_\theta(x_t,t), \Sigma_\theta(x_t,t))$

DiGress

- Simplify the task to a sequence of classification by incorporating the cross entropy.

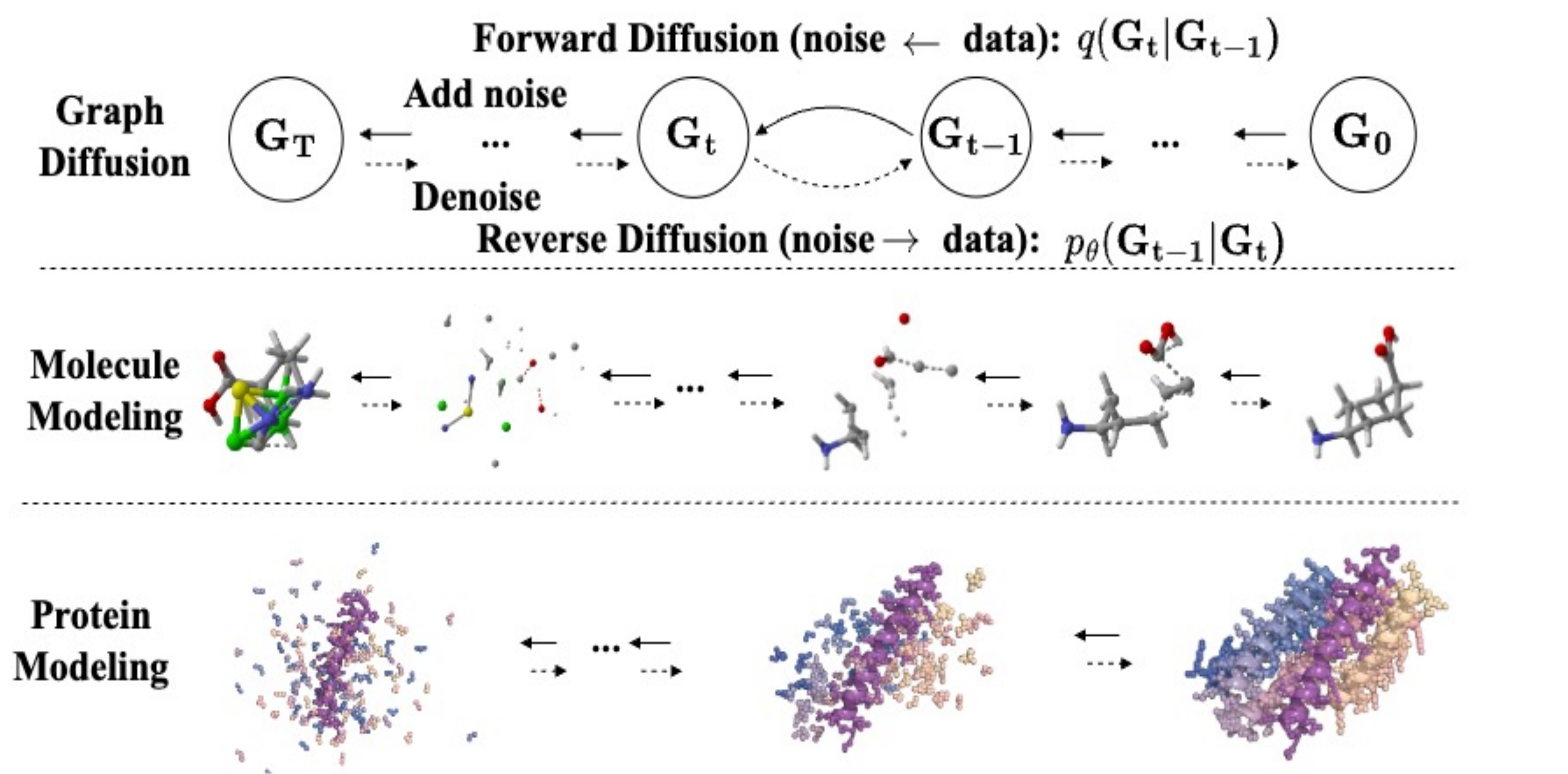
Score-based Generative Models (SGM) on Graphs

The score SDE formula describes the diffusion process in continuous time steps.

- ◆ Forward: $dx = f(x,t)dt + g(t)$
- ◆ Reverse: $dx = [f(x,t) - g(t)^2 \nabla_x \log p_t(x)]\, dt + g(t)d\overline{w}$

GDSS:

- Model the nodes and edges simultaneously
- The very first diffusion framework that enables the generation of a whole graph



Tasks	Applications	Frame	Representative Methods
Molecule Modeling	Molecule Conformation Generation	SMLD	MDM [Huang *et al.*, 2022b]
		DDPM	GeoDiff [Xu *et al.*, 2022], EDMs [Hoogeboom *et al.*, 2022], EEGSDE [Bao *et al.*, 2023], DiGress [Vignac *et al.*, 2023]
		SGM	Torsional Diffusion [Jing *et al.*, 2022], MOOD [Lee *et al.*, 2022], GDSS [Jo *et al.*, 2022], DGSM [Luo *et al.*, 2021a], DiffBridges [Wu *et al.*, 2022b]
	Molecular Docking	DDPM	FragDiff [Peng *et al.*, 2023], DiffLink [Igashov *et al.*, 2022], TargetDiff [Guan *et al.*, 2023], DiffBP [Lin *et al.*, 2022]
		SGM	DiffDock [Corso *et al.*, 2022]
Protein Modeling	Protein Generation	DDPM	SMCDiff [Trippe *et al.*, 2023], SiamDiff [Zhang *et al.*, 2022], DiffFold [Wu *et al.*, 2022a], ProSSDG [Anand and Achim, 2022], DiffAntigen [Luo *et al.*, 2022a], RFdiffusion [Watson *et al.*, 2022]
		SGM	ProteinSGM [Luo *et al.*, 2022a]
	Protein-ligand Complex Structure Prediction	DDPM	DiffEE [Nakata *et al.*, 2022]
		SGM	NeuralPLexer [Qiao *et al.*, 2022]

Application

Molecule Modelling

To employ graph learning techniques for the purpose of representing to better perform downstream tasks.

Molecule conformation generation

- ➢ The biological and physical characteristics of the molecule are significantly influenced by its 3-D structure.

Molecule docking

- ➢ A computational method for predicting the preferred orientation of one molecule to a second molecule (typically a protein)
- ➢ Drug discovery

Protein Modeling

- ➢ Generate and predict the structure of proteins with specific structural and functional properties.
- ➢ Predict the protein-ligand complex structure.

4 Future Challenges and Opportunities

Conditional Generation for Graph Diffusion Models.

- Incorporating conditions into generative models

Trustworthiness for Graph Diffusion Models.

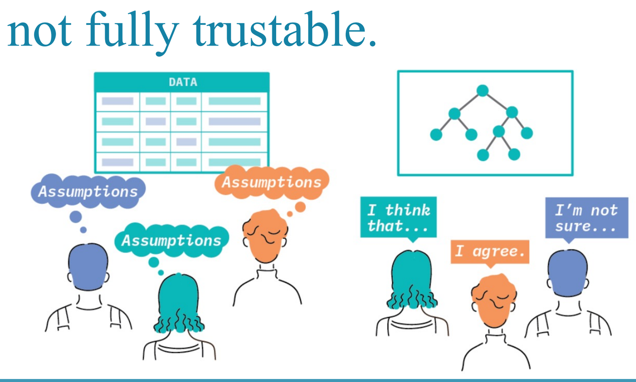
- Unintentional harm to users and society in various real-world tasks
- Safety-critical fields such as drug discovery.

Evaluation Metrics.

- Metrics based on graph statistics and properties are not fully trustable.
- Validity and diversity for graph generation

Graph Diffusion Applications.

- *Recommender Systems*
- *Graph Anomaly Detection*
- *Causal Graph Generation*



For arXiv version and reference

The Hong Kong Polytechnic University 香港理工大學
Department of Computing 電子計算學系
MICHIGAN STATE UNIVERSITY
IJCAI/2023 MACAO