Graph Neural Networks: Graph Generation

Renjie Liao

Shun Saito

Outline

- Introduction
- Classic Graph Generative Models
  - Erdős–Rényi Model
  - Stochastic Block Model
- Deep Graph Generative Models
  - Variational graph autoencoders (VAE)
    - Vanilla VAE
    - Graph VAE architecture
    - Graph VAE family

Introduction

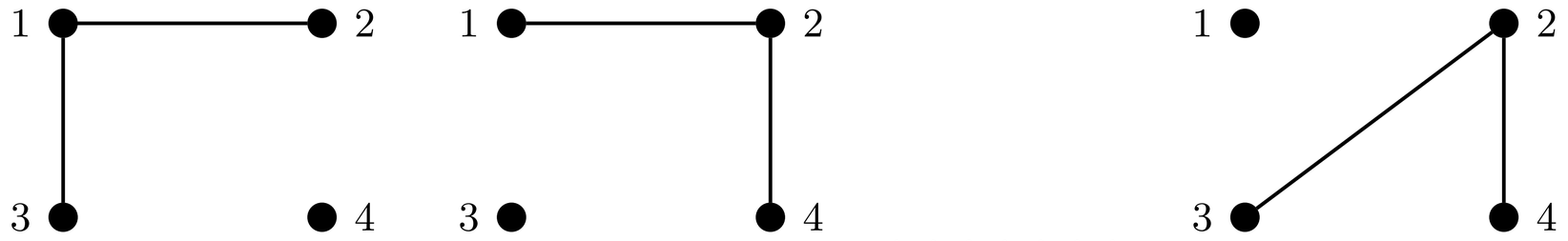
- Researchers are interested in building random graph models.
- Classic models are too simplistic for complex phenomenon in real-world graphs (e.g., highly-clustered, well-connected, scale-free).
- Deep graph generative models can build more expressive models.

Introduction

Random graph

Graph obtained by randomly sampling from a collection of graphs.

E.g., sampling uniformly from all graphs with 4 vertices and 2 edges:



(a) G_1 (b) G_2 (d) G_{15}

Geoffrey Fairchild and Jason Fries, Lecture Notes: Social Networks: Models, Algorithms, and Applications

Erdős–Rényi Model

G(n,p) Model

Randomly connecting an edge from one node to the other with probability p.

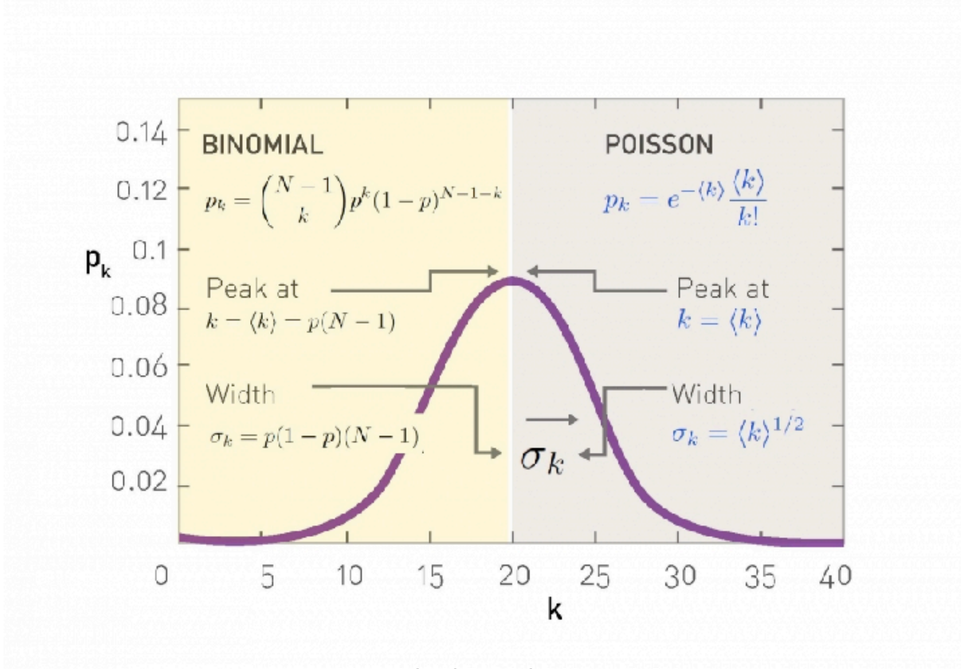
$$p(\text{a graph with } n \text{ nodes and } m \text{ edges}) = p^m(1-p)^{\binom{n}{2}-m}$$

Degree distribution is binomial:

$$P\left(\deg(v) = k\right) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

If $n \to \infty$ and $np = \text{const.}$, degree distribution is Poisson:

$$P\left(\deg(v) = k\right) \to \frac{\left(np\right)^k e^{-np}}{k!}$$

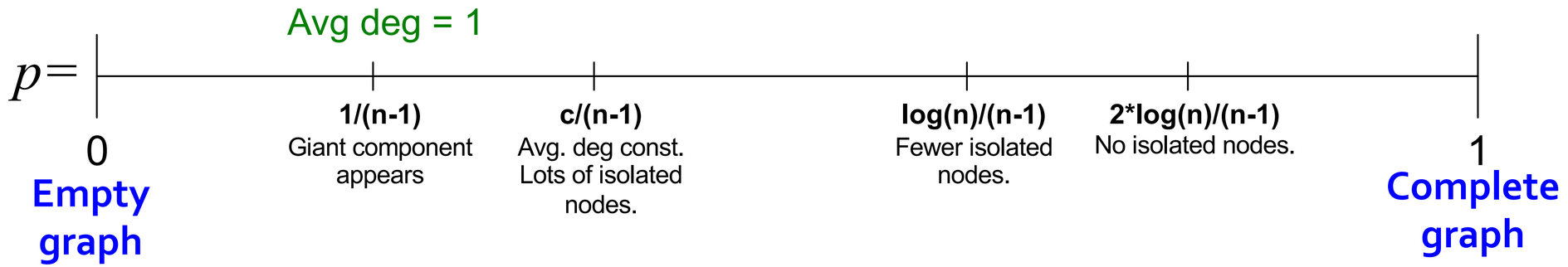


Albert-László Barabási, Network Science

Erdős–Rényi Model

G(n,p) Model

- Graph structure as p changes:



Jure Leskovec, Stanford CS224W: Machine Learning with Graphs

Erdős–Rényi Model

G(n,m) Model

Uniformly randomly choosing a graph from the set of all graphs with n nodes and m edges.

The probability of choosing each graph : $\binom{\binom{n}{2}}{m}^{-1}$

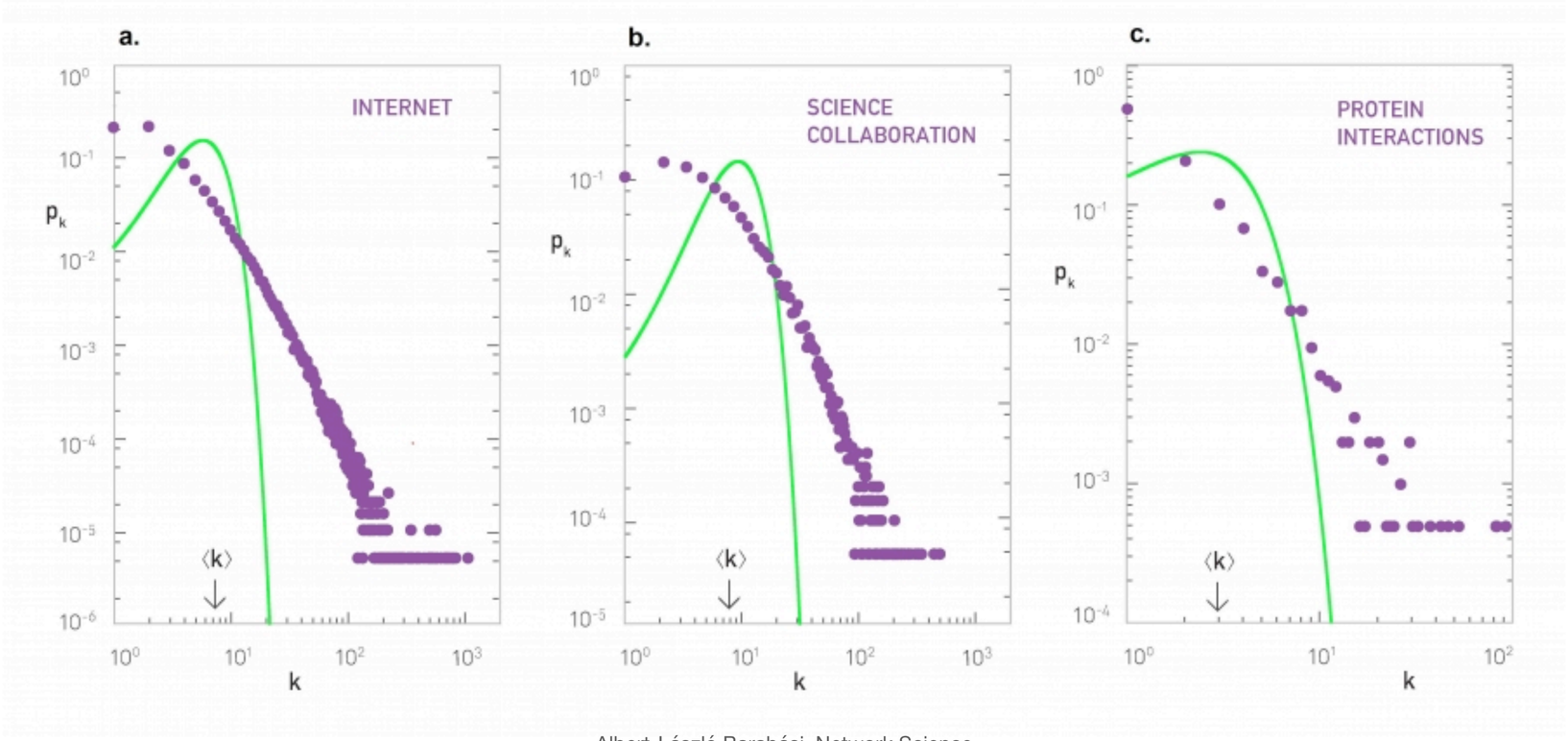
E.g., for $G(4,2)$, each of the possible graphs are included with probability $\frac{1}{15}$.

$G(n,p)$ model is easier to analyze because of independence of edges.

Erdős–Rényi Model

Discussion

The assumptions are too strong for real-world networks.



Albert-László Barabási, Network Science

Stochastic Block Model (SBM)

Simple random graph model with clusters of nodes for tasks like community detection.

Provide ground-truth cluster memberships.

n: number of nodes

k: number of communities/clusters

p: prior probability over the k clusters

W: $k \times k$ matrix with entries $[0,1]$

Stochastic Block Model (SBM)

Procedures

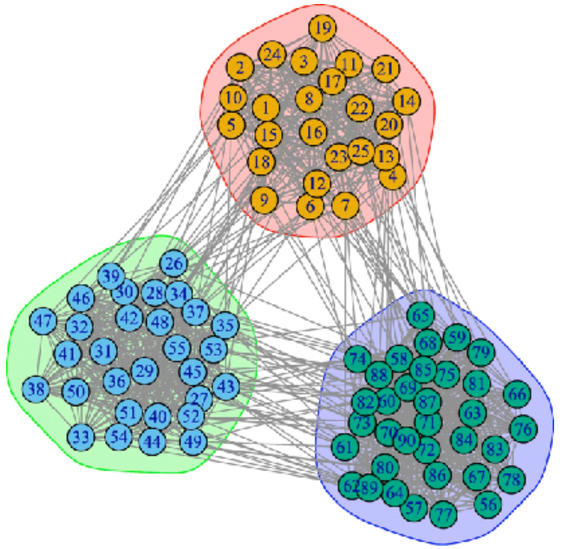
1. For each node, generate its community label ($\{1,\ldots,k\}$) by independently sampling from p.

2. For each pair of nodes, generate an edge by independently sampling with probability $W_{i,j}$, where i and j denote community labels.

    E.g., 90 nodes, 1192 edges

        groups 1, 2 and 3 containing 25, 30 and 35 nodes

$$W_{i,j} = \begin{pmatrix} 0.8 & 0.05 & 0.05 \\ 0.05 & 0.8 & 0.05 \\ 0.05 & 0.05 & 0.8 \end{pmatrix}$$



Lee *et al*. A review of stochastic block models and extensions for graph clustering. *Appl Netw Sci* **4**, 1–50 (2019).

Stochastic Block Model (SBM)

Recover the community label

For recovered label $X \in \mathbb{R}^{n\times 1}$ and ground-truth label $Y \in \mathbb{R}^{n\times 1}$,

$$R(X,Y) = \max_{P\in\Pi} \quad \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left[X_i = (PY)_i\right]$$

Permutation matrix

i-th element of *X*

i-th element of *PY*

Agreement considers the best possible reshuffle between two sequences of labels.

Stochastic Block Model (SBM)

Discussion

Pros:

- More realistic for community structures than Erdős–Rényi.

Cons:

- Number of communities is unknown in advance.
- Graphs may not have clear community structures.

Limitations of traditional models

- Rely on a fixed, handcrafted generation process.
- Lack the ability to learn a generative model from data.

Deep Graph Generative Models

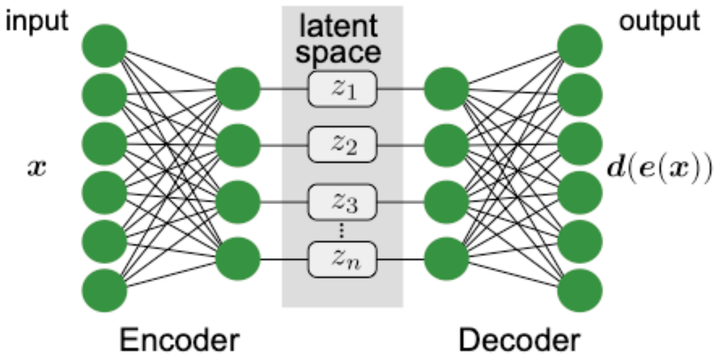
- Variational autoencoders (VAEs)
- Deep autoregressive models
- Generative adversarial networks (GANs)

Graph representation

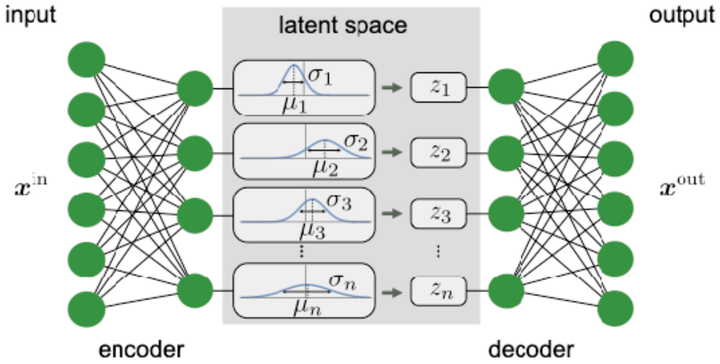
- Graph: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- Adjacency matrix: A_π (with node ordering π)
- If π is changed to π', adjacency matrix is permuted: $A_{\pi'} = PA_\pi P^\top$
- For simplicity, $\mathcal{G} \equiv A = \{PAP^\top \,|\, P \in \Pi_\mathcal{G}\}$.
- With node features X, $\mathcal{G} \equiv \{(PAP^\top, PX \,|\, P \in \Pi_\mathcal{G}\}$.

Vanilla VAE

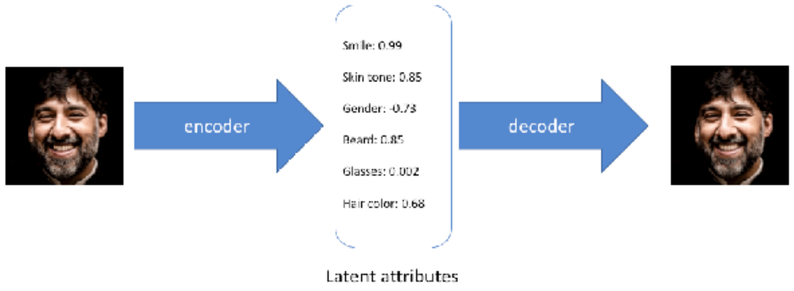
General autoencoder

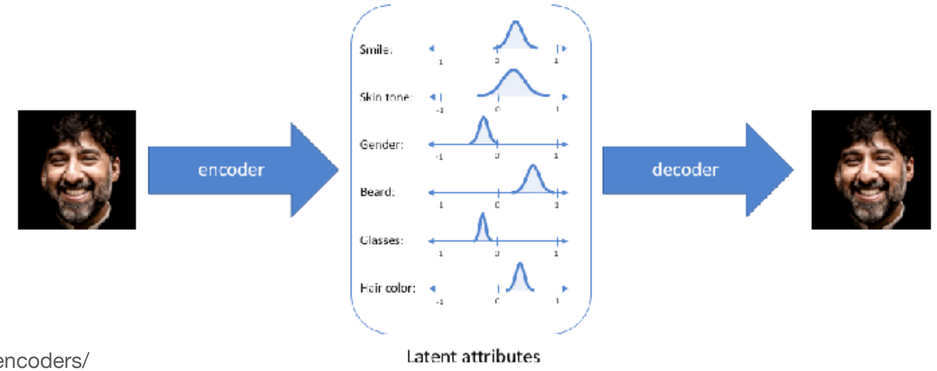


Variational autoencoder



https://ml-lectures.org/docs/unsupervised_learning/ml_unsupervised-3.html





https://www.jeremyjordan.me/variational-autoencoders/

Variational Auto-Encoder Methods

Encoder

Using GCN, $H = \tilde{A}\sigma(\tilde{A}XW_1)W_2$, where $\tilde{A} = D^{-\frac{1}{2}}(A + I)D^{-\frac{1}{2}}$

Variational distribution: $q_\phi(Z|A,X) = \prod_{i=1}^{n} q\left(\mathbf{z}_i|A,X\right)$

$$q(\mathbf{z}_i|A,X) = \mathcal{N}(\mu_i, \sigma_i \mathrm{I})$$

$$\mu = \mathrm{MLP}_\mu(H)$$

$$\log\sigma = \mathrm{MLP}_\sigma(H)$$

Variational Auto-Encoder Methods

Encoder

- Probability distribution is invariant w.r.t. the permutation of nodes.

$$q(P_1 Z \mid P_1 A P_1^\top, P_1 X) = q(P_2 Z \mid P_2 A P_2^\top, P_2 X), \text{ where } P_1, P_2 \in \Pi$$

- Gaussian are so powerful that the conditional distributions are expressive in capturing the latent variables.
- Less computationally expensive.

Variational Auto-Encoder Methods

Prior

Fixed prior distribution with node-independent Gaussian:

$$p(Z) = \prod_{i=1}^{n} p\left(\mathbf{z}_i\right)$$

$$p(\mathbf{z}_i) = \mathcal{N}(0, \mathrm{I})$$

Variational Auto-Encoder Methods

Decoder

Need to consider all possible node orderings:

$$p(\mathcal{G}\,|\,Z) = \sum_{p \in \Pi_{\mathcal{G}}} p(PAP^{\top}, PX\,|\,Z)$$

Construction:

$$p(A, X\,|\,Z) = \prod_{i,j} p(A_{ij}\,|\,Z) \prod_{i=1}^{n} p\left(\mathbf{x}_i\,|\,Z\right)$$

$$p(A_{ij}\,|\,Z) = \text{Bernoulli}(\Theta_{ij}),\, p(\mathbf{x}_i\,|\,Z) = \mathcal{N}(\tilde{\mu}_i, \tilde{\sigma}_i)$$

$$\Theta_{ij} = \text{MLP}_{\Theta}([\mathbf{z}_i \| \mathbf{z}_j])\ ,\ \tilde{\mu}_i = \text{MLP}_{\tilde{\mu}}(\mathbf{z}_i),\, \tilde{\sigma}_i = \text{MLP}_{\tilde{\sigma}}(\mathbf{z}_i)$$

Variational Auto-Encoder Methods

Decoder

Decoder is not permutation invariant.

$$p(P_1 A P_1^\top, P_1 X \,|\, P_1 Z) \neq p(P_2 A P_2^\top, P_2 X \,|\, P_2 Z)$$

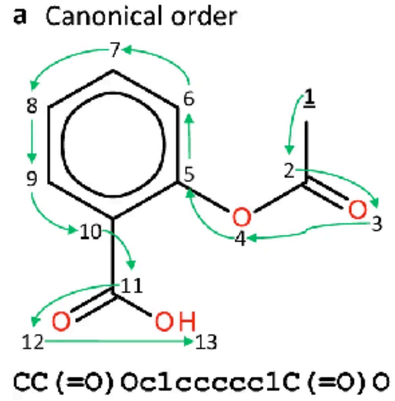
Number of permutation matrices is very large.

Possible approaches

$$p(\mathcal{G} \,|\, Z) = \sum_{P \in \Pi_{\mathcal{G}}} p(PAP^\top, PX \,|\, Z) \approx \max_{P \in \Pi_{\mathcal{G}}} p(PAP^\top, PX \,|\, Z)$$

For canonical node ordering (e.g., simplified molecular-input line-entry system (SMILES)

$$p(\mathcal{G} \,|\, Z) = \sum_{P \in \Pi_{\mathcal{G}}} p(PAP^\top, PX \,|\, Z) \approx p(\tilde{P} A \tilde{P}^\top, \tilde{P} X \,|\, Z)$$



Arús-Pous *et al*. Randomized SMILES strings improve the quality of molecular generative models. *Journal of Cheminformatics* **11**, 71 (2019).

Variational Auto-Encoder Methods

Training objective

$$\mathscr{L} = \mathbb{E}_{q_\phi(Z|A,X)}[\log p_\theta(\mathscr{G} \mid Z)] - KL(q_\phi(Z \mid A, X) \| p(Z))$$

Reconstruction Loss

How well the reconstructed graph matches the original graph

Kullback-Leibler (KL) Divergence

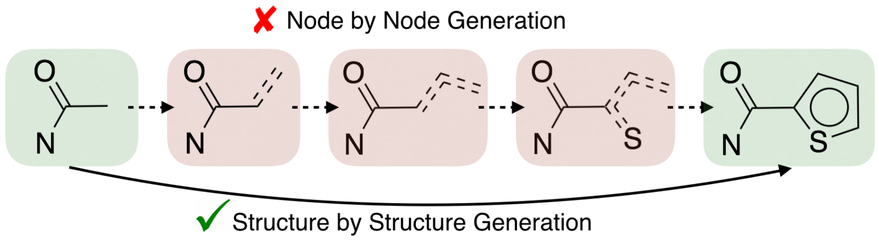
how similar the distribution of the latent variable and the target distribution.

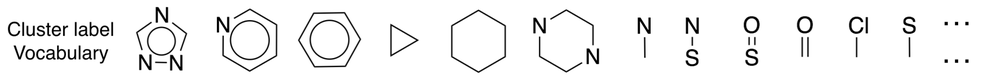
GraphVAE family

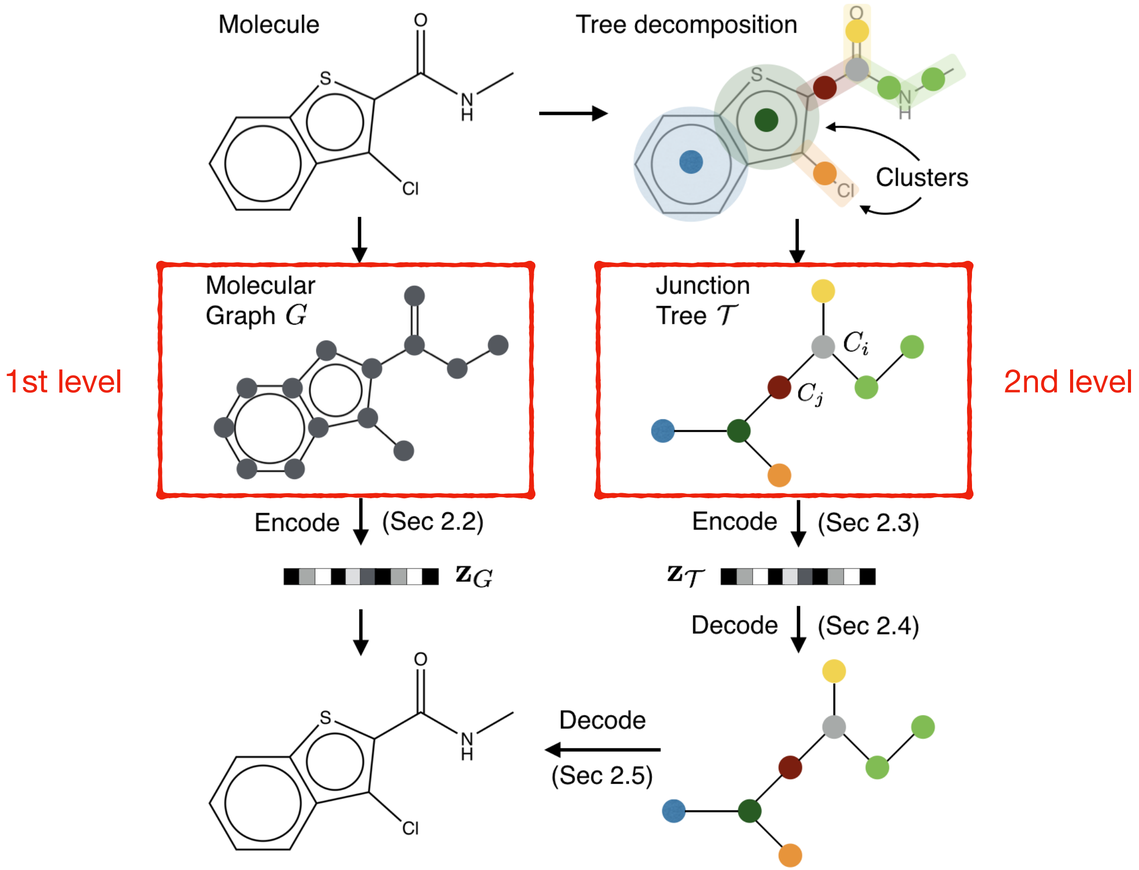
- Hierarchical GraphVAEs
- Constrained GraphVAEs

Hierarchical GraphVAEs

Junction Tree VAEs







Jin *et al*. Junction Tree Variational Autoencoder for Molecular Graph Generation. Preprint at https://doi.org/10.48550/arXiv.1802.04364 (2019).

Constrained GraphVAEs

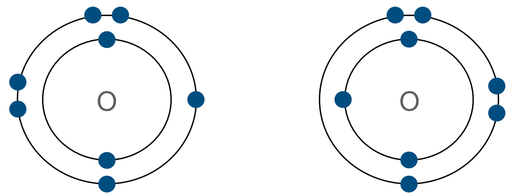
Constraints for molecules

Constraints on the generated graphs to ensure semantic validity.
E.g., configuration of chemical bonds (edges) must meet the valency rules of atoms (nodes).

Decoder generates graphs satisfying the constraints by construction.

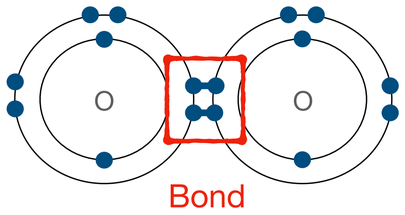
E.g., Valency masking [1]:

Generation process will be terminated when the number of bonds exceeds the valency of the node (invalid graph).



Soft constraints by Lagrangian-based regularizers [2]:

Transforming constrained optimization problem to regularized unconstrained one.



[1] Liu *et al*. Constrained Graph Variational Autoencoders for Molecule Design. Preprint at https://doi.org/10.48550/arXiv.1805.09076 (2019).
[2].Ma *et al*. Constrained Generation of Semantically Valid Graphs via Regularizing Variational Autoencoders. Preprint at https://doi.org/10.48550/arXiv.1809.02630 (2018).

Thank you!

Any questions?