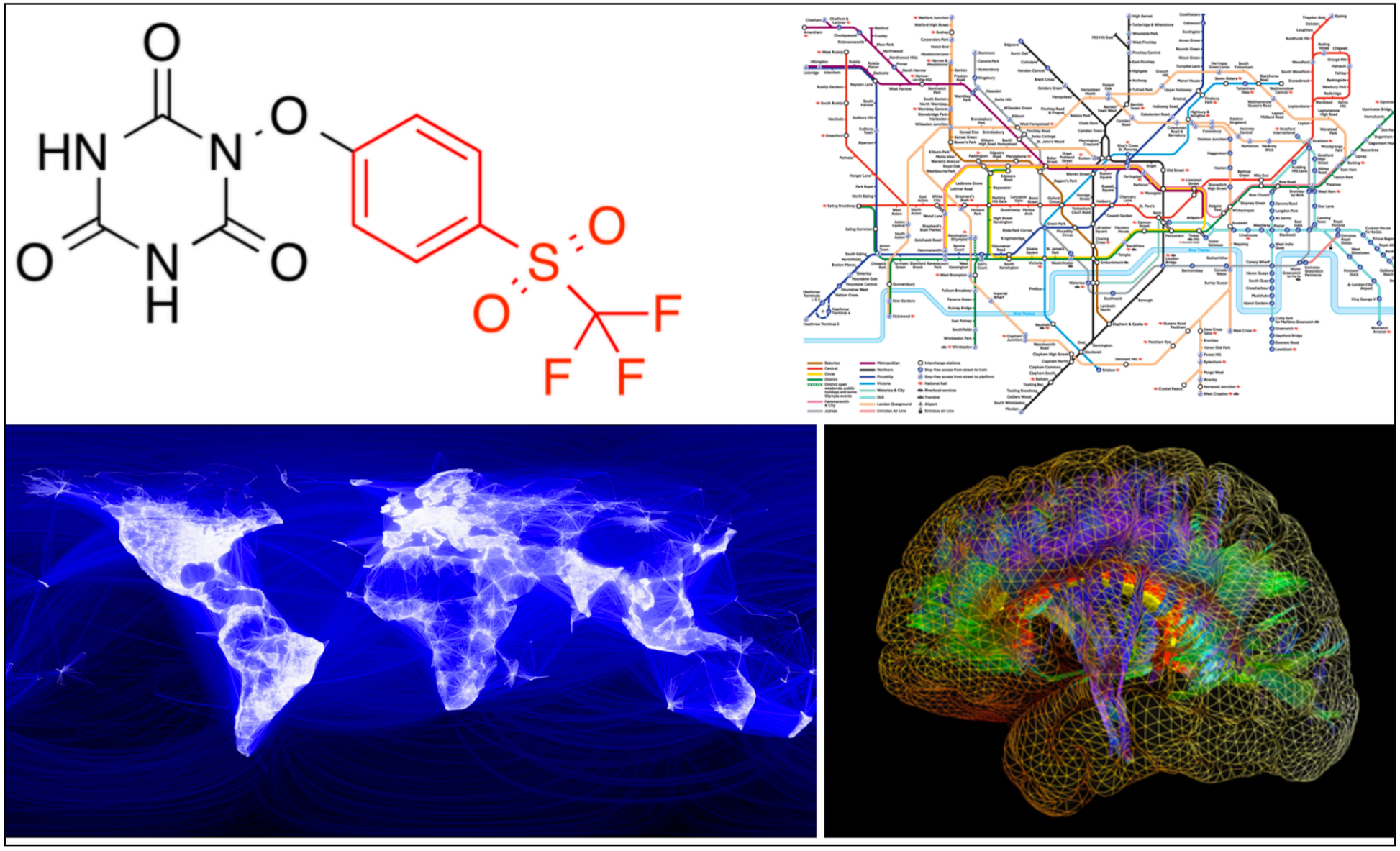
# Constrained Graph Variational Autoencoders for Molecule Design

Станислав Рыбин, 151

# Graph-structured inputs



Motivating examples of graph-structured inputs: molecular networks, transportation networks, social networks and brain connectome networks.

# Three classes of graph generation models

- **Uncorrelated generation** - ignores the correlations & models the existence and label of each edge with *independent* random variables

- **Sequential generation** - factor the distribution into a *sequence* of discrete decisions in a graph construction trace

- **New class (our model)** - the learned component is conditioned only on the current state of generation and not on the arbitrarily chosen path to that state

# Generative model

- Stages:

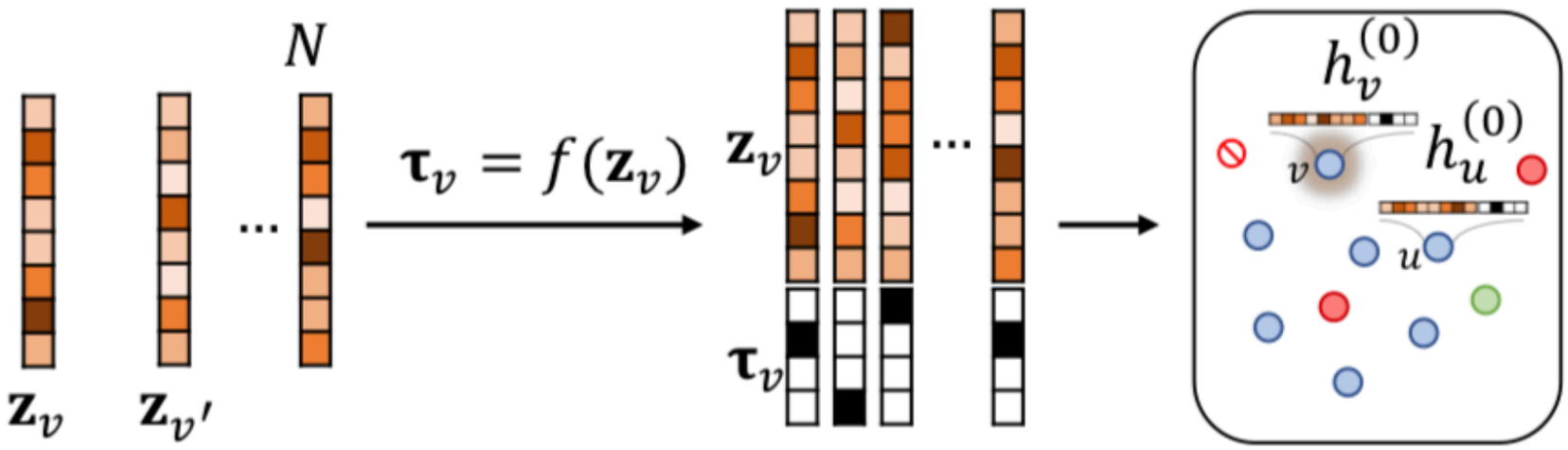
    1. Node Initialization

    2. Node Update

    3. Edge selection & labelling

    4. Termination

# Node Initialization



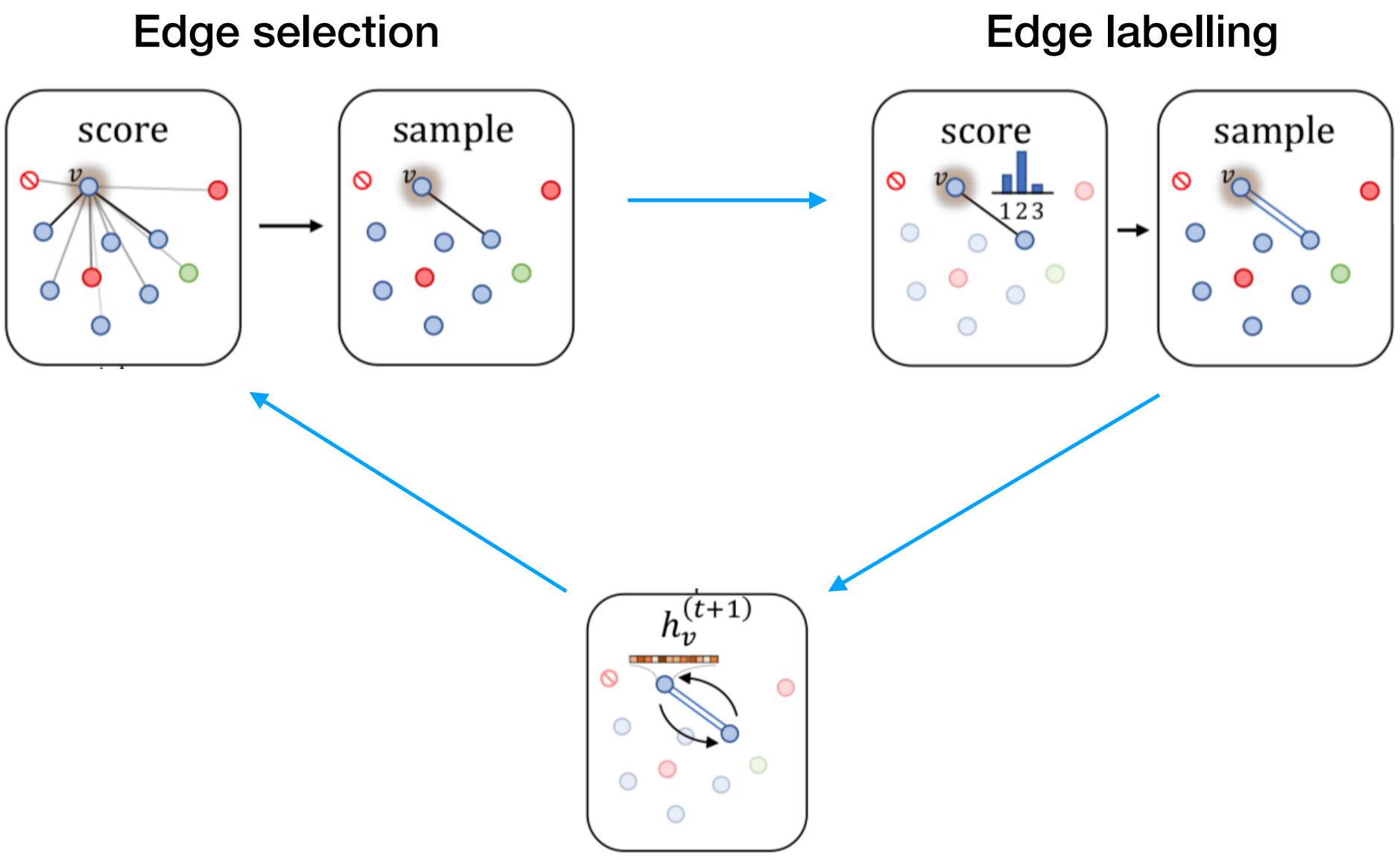
# Generative model

- Generation of edges between nodes proceeds using two decision functions: **focus** and **expand**

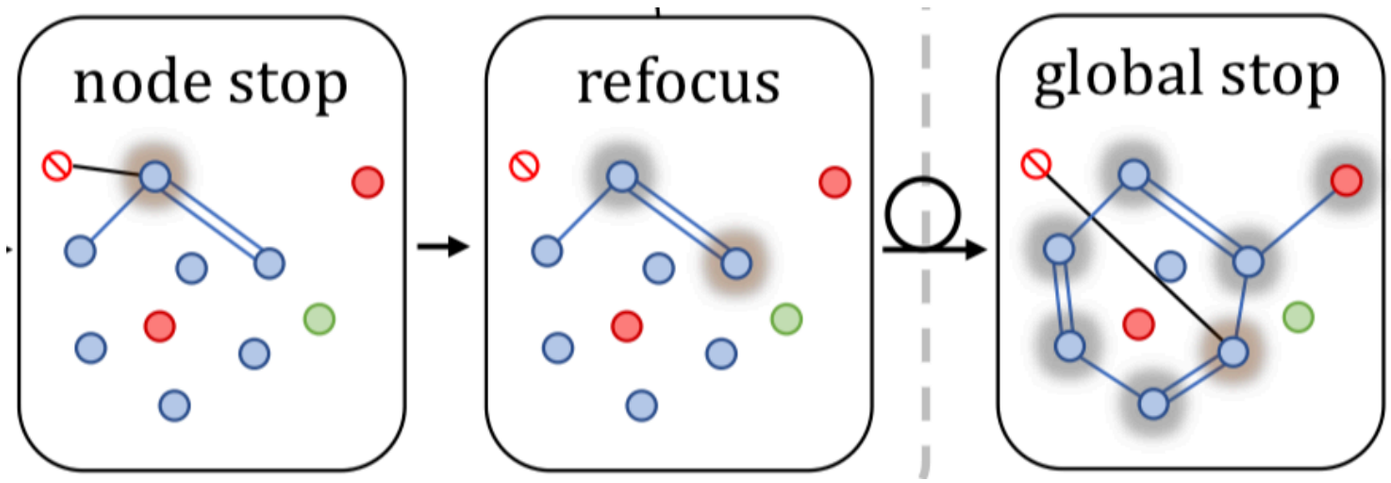
1. **Focus** function chooses a focus node to visit

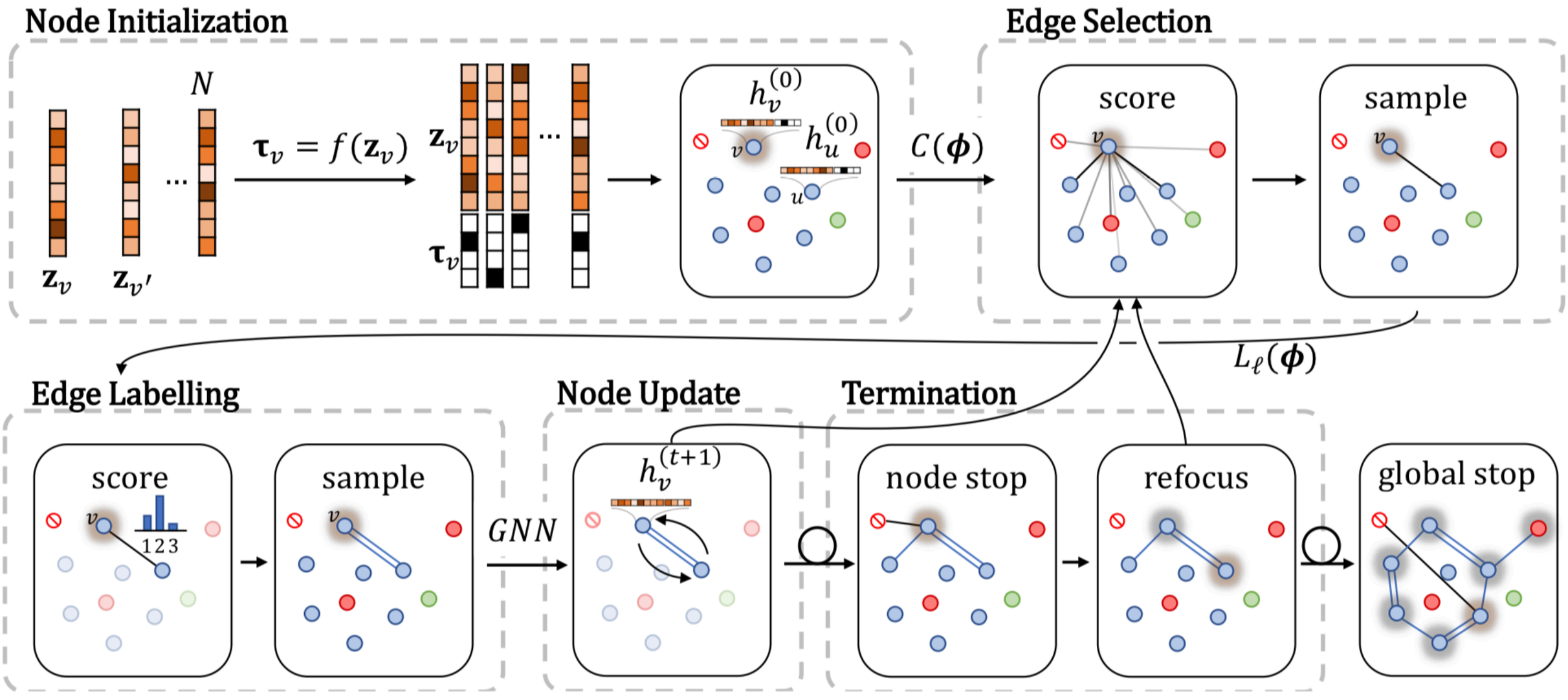
2. **Expand** function chooses edges to add from the focus node.

# Node Update & Edge selection & labelling



# Termination




Node Initialization
N
$\mathbf{z}_v$
$\mathbf{z}_{v'}$
$\boldsymbol{\tau}_v = f(\mathbf{z}_v)$
$\mathbf{z}_v$
$\boldsymbol{\tau}_v$
$h_v^{(0)}$
$h_u^{(0)}$
$v$
$u$
$C(\boldsymbol{\phi})$
Edge Selection
score
sample
$L_\ell(\boldsymbol{\phi})$
Edge Labelling
score
1 2 3
sample
GNN
Node Update
$h_v^{(t+1)}$
Termination
node stop
refocus
global stop

# Training the Generative model

- Encoder
- Decoder
- Node Initialization

$$p(\mathcal{G}^{(0)} \mid \mathbf{z}) = \sum_{\mathcal{P}} p(\boldsymbol{\tau} = \mathcal{P}(\boldsymbol{\tau}^*) \mid \mathbf{z}) > \prod_{v} p(\boldsymbol{\tau}_v = \boldsymbol{\tau}_v^* \mid \mathbf{z}_v).$$

- Edge Selection and Labelling
-

$$\log p(\pi \mid \mathcal{G}^{(0)}) = \sum_{(t,v,\epsilon)\in\pi} \Big\{ \log p(v \mid \pi, t) + \log p(\epsilon \mid \mathcal{G}^{(t-1)}, v) \Big\}$$

# Application: Molecule Generation

- Datasets: QM9, ZINC, CEPDB