Challenger 2: DiGress: Discrete Denoising Diffusion for Graph Generation

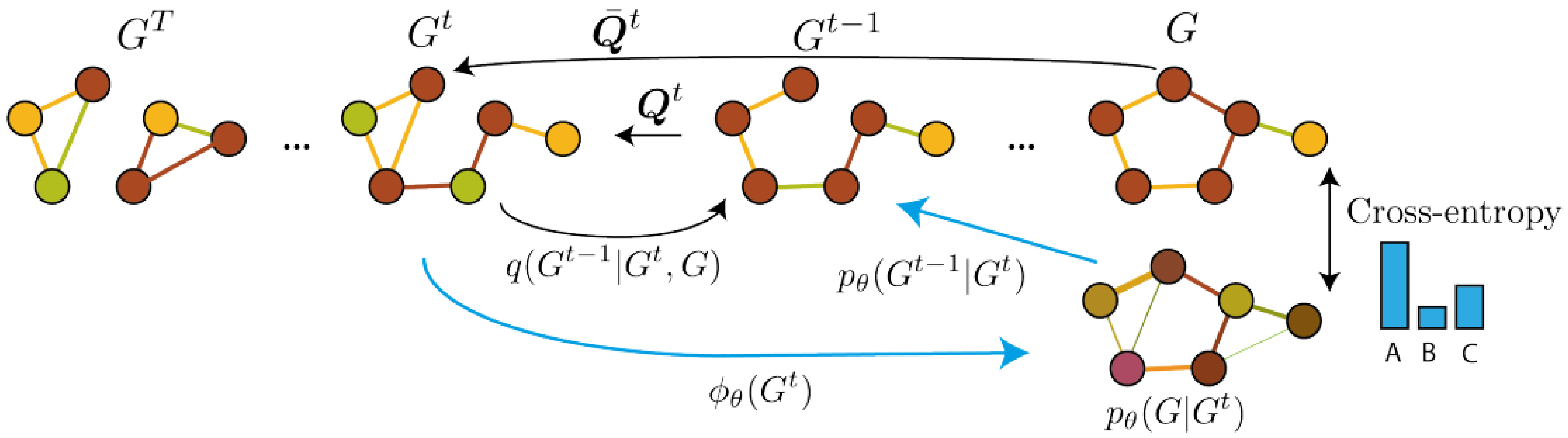
Jianpeng Chen

CS, Virginia Tech

Strengths

❑Impressive part of DiGress

1. Most novel part: Diffusion on discrete **node/edge types** rather than adj. or node embeddings.



Strengths

❑Impressive part of DiGress

1. Most novel part: Diffusion on discrete **node/edge types** rather than adj. or node embeddings.
2. **Theoretically solid**: DiGress theoretically achieved **exchangeability**, which can align with the natural property of graph learning: invariant to features or adj. permutations.

Lemma 3.3. *(Exchangeability)* *DiGress yields exchangeable distributions, i.e., it generates graphs with node features* $\boldsymbol{X}$ *and adjacency matrix* $\boldsymbol{A}$ *that satisfy* $\mathbb{P}(\boldsymbol{X}, \boldsymbol{A}) = P(\pi^T \boldsymbol{X}, \pi^T \boldsymbol{A} \pi)$ *for any permutation* π.

3. Comprehensive experiments: This paper experiments on various types of datasets, scaling from general graphs to large molecular datasets, as well experiments on conditioned generation, indicating that **DiGress can be scaled to different graph based fields**.
4. Easy learning task. Unlike other methods trying to learn the whole topology, the tasks of node and edge type prediction are relatively easy.

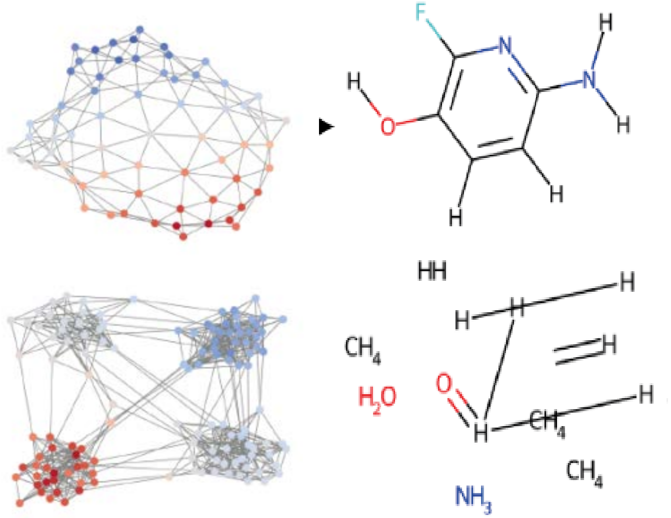
Strengths

❑Impressive part of DiGress

1. Most nov
   embeddi
2. Theoreti
   with the

Figure 4: Mean absolute error on conditional generation with discrete regression guidance on QM9.

Target	μ	HOMO	μ & HOMO
Uncondit.	1.71±.04	0.93±.01	1.34±.01
Guidance	0.81±.04	0.56±.01	0.87±.03



3. Comprehensive experiments: This paper experiments on various types of datasets, scaling from general graphs to large molecular datasets, as well experiments on conditioned generation, indicating that **DiGress can be scaled to different graph based tasks**.
4. Easy learning target: Unlike other methods trying to learn the whole topology, the tasks of node and edge type prediction are relatively easy.

Strengths

❑Impressive part of DiGress

1. Most novel part: Diffusion on discrete **node/edge types** rather than adj. or node embeddings.
2. Theoretically solid: DiGress theoretically achieved **exchangeability**, which can align with the natural property of graph learning: invariant to features or adj. permutations.
3. Comprehensive experiments: This paper experiments on various types of datasets,

$$l(\hat{p}^G, G) = \sum_{1\le i\le n} \text{cross-entropy}(x_i, \hat{p}_i^X) + \lambda \sum_{1\le i,j\le n} \text{cross-entropy}(e_{ij}, \hat{p}_{ij}^E)$$

4. Easy learning target: Unlike other methods trying to learn the whole topology, the tasks of node and edge type prediction are relatively easier.

Strengths

❑Impressive part of DiGress

1. Most novel part: Diffusion on discrete **node/edge types** rather than adj. or node embeddings.
2. Theoretically solid: DiGress theoretically achieved **exchangeability**, which can align with the natural property of graph learning: invariant to features or adj. permutations.
3. Comprehensive experiments: This paper experiments on various types of datasets, scaling from general graphs to large molecular datasets, as well experiments on conditioned generation, indicating that **DiGress can be scaled to different graph based tasks**.
4. Easy learning target: Unlike other methods trying to learn the whole topology, the tasks of node and edge type prediction are relatively easier.

Weaknesses

❑ Weaknesses worth to be further explored

1. **Scalability to large graphs**: The diffusion on edge types in DiGress requires the representation of all edges, including non-existent ones, making it difficult to scale to large graphs.
2. Discrete is not always effective. The diffusion process may result in the loss of information if we roughly categorize naturally continuous node or edge properties into another category. For example, when cutting or connecting edges, there is a probability that two nodes should be connected, but discrete transitions ignore these continuous changes.
3. The uniform samples used to add noise may violate expert knowledge laws, such as deleting edges of a molecule ring, resulting in the loss of knowledge-related information like aroma.

Weaknesses

❑Weaknesses worth to be further explored

1. Scalability to large graphs: The diffusion on edge types in DiGress requires the representation of all edges, including non-existent ones, making it difficult to scale to large graphs.
2. **Discrete is not always effective**. The diffusion process may result in the loss of information if we roughly categorize naturally continuous node or edge properties into another category. For example, when cutting or connecting edges, there is a probability that two nodes should be connected, but discrete transitions ignore these continuous changes.
3. The uniform samples used to add noise may violate expert knowledge laws, such as deleting edges of a molecule ring, resulting in the loss of knowledge-related information like aroma.

Weaknesses

❑Weaknesses worth to be further explored

1. Scalability to large graphs: The diffusion on edge types in DiGress requires the representation of all edges, including non-existent ones, making it difficult to scale to large graphs.
2. Discrete is not always effective. The diffusion process may result in the loss of information if we roughly categorize naturally continuous node or edge properties into another category. For example, when cutting or connecting edges, there is a probability that two nodes should be connected, but discrete transitions ignore these continuous changes.
3. The **uniform samples** used to add noise may violate expert knowledge laws, such as deleting edges of a molecule ring, resulting in the loss of knowledge-related information like aroma.

Weaknesses

❑ Weaknesses worth to be further explored

1. **Scalability to large graphs**: The diffusion on edge types in DiGress requires the representation of all edges, including non-existent ones, making it difficult to scale to large graphs.
2. **Discrete is not always effective**. The diffusion process may result in the loss of information if we roughly categorize naturally continuous node or edge properties into another category. For example, when cutting or connecting edges, there is a probability that two nodes should be connected, but discrete transitions ignore these continuous changes.
3. The **uniform samples** used to add noise may violate expert knowledge laws, such as deleting edges of a molecule ring, resulting in the loss of knowledge-related information like aroma.

Thanks