ETH zürich

SDSC

Prescient Design
A Genentech Accelerator

# Efficient and Scalable Graph Generation through Iterative Local Expansion

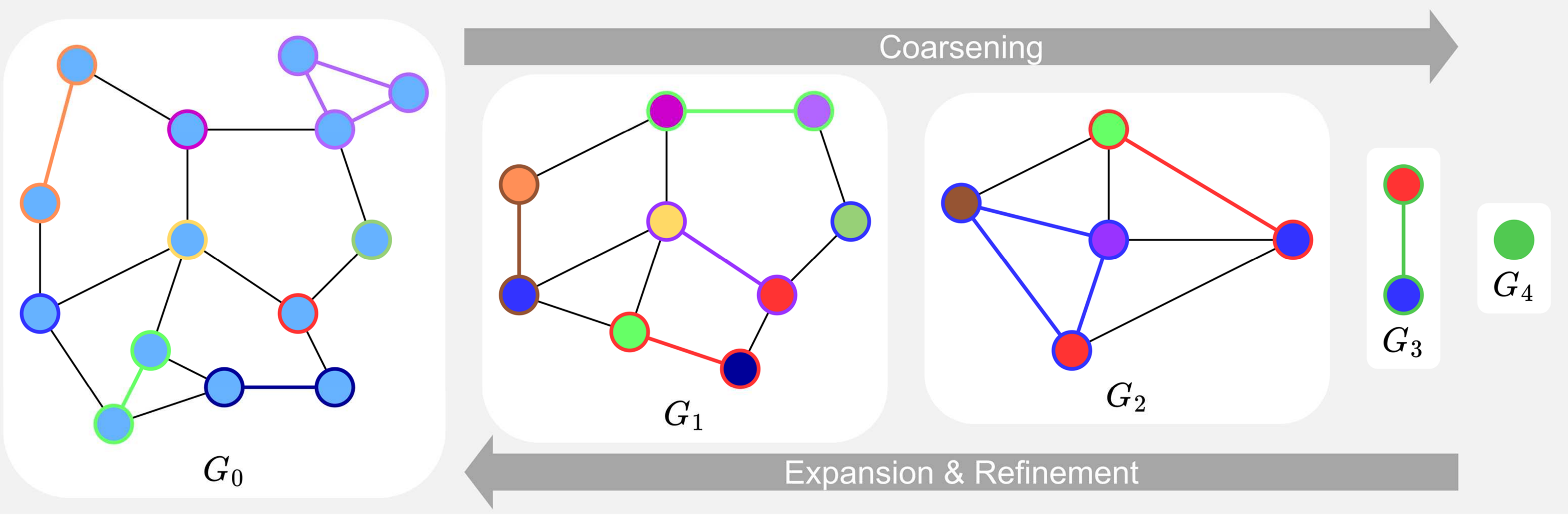
Andreas Bergmeister[1], Karolis Martinkus[2], Nathanaël Perraudin[1], Roger Wattenhofer[3]
[1]Swiss Data Science Center, ETH Zürich; [2]Prescient Design, Genetech; [3]DISCO, ETH Zürich

## 1 Excising Challenges

- **current graph generation methods**
  - model **joint distribution over all $\mathcal{O}(n^2)$ node pairs**
- **recent works to improve scalability**
  - make **simplifying assumptions** (community structure, edge independence, restricted bandwidth), sacrificing generality or sample fidelity

## 2 Our Idea

- model a **progressive expansion process** rather than the final graph directly
- equivalent to **inverting graph coarsening**
- achieve efficiency by only **locally modifying graphs**
- **no restrictive assumptions on the graph's structure**

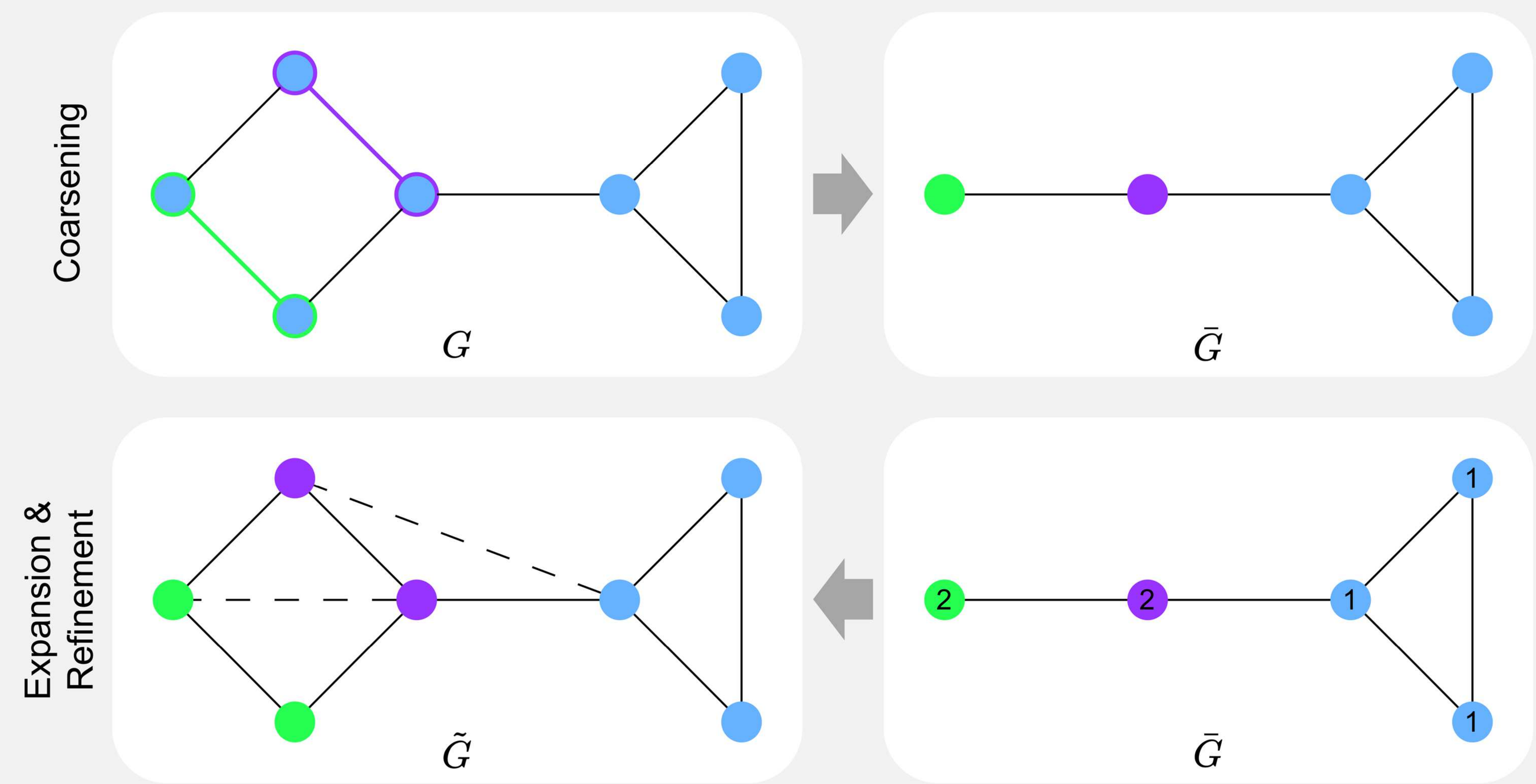



## 4 Local PPGN

- **novel** architecture to parameterize the diffusion model
- **efficient** on sparse graphs
- more **expressive** than message-passing GNNs
- $(\boldsymbol{h}')^{(i,j)} = \gamma\left(\boldsymbol{h}^{(i,j)}, \sum_{k \in N^-(i) \cap N^+(j)} \phi\left(\boldsymbol{h}^{(i,k)}, \boldsymbol{h}^{(k,j)}\right)\right)$

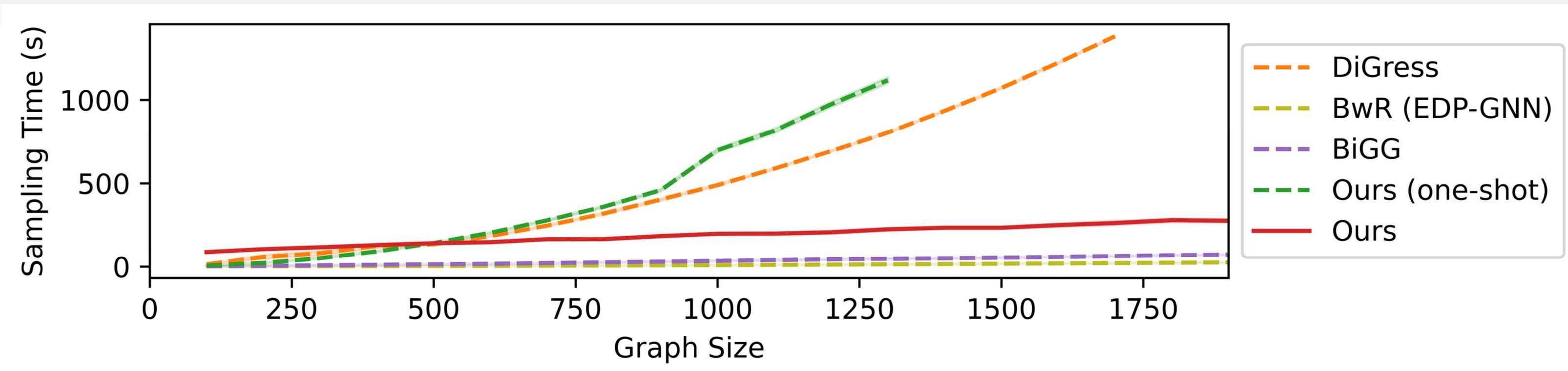
## 3 Details

- two-phase **inversion** of a single **graph coarsening** step
  1. **expansion** of the coarsened graph
  2. **refinement** of the resultant expanded graph
- **probabilistic inversion**
  - use **denoising diffusion** to model a distribution over
    - **node features**: size of expansion cluster
    - **edge features:** edge existence




## 5 Scalability

- theoretical **asymptotic complexity**: $\mathcal{O}(m \log n)$
- **scaling behaviour empirically validated on planar graphs**

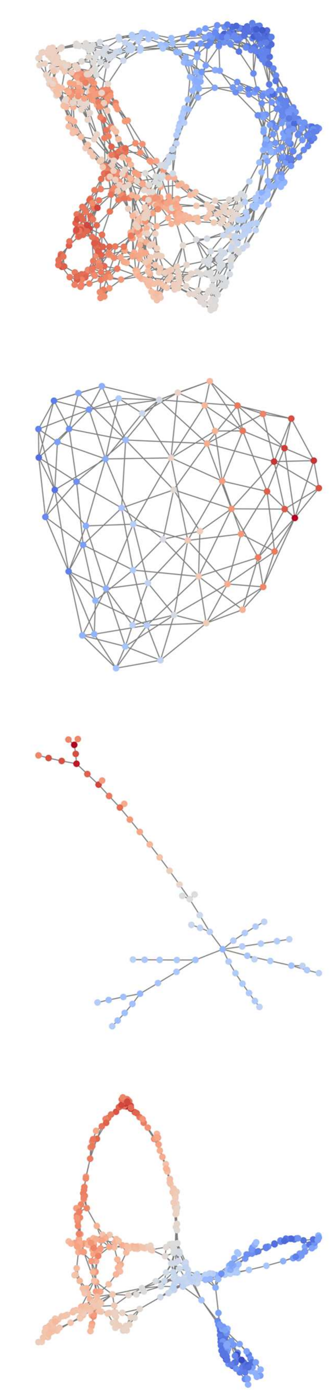



## 5 Generative Performance

| Planar Graphs ($n_{\max} = 64$, $n_{\text{avg}} = 64$) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Model | Deg. ↓ | Clus. ↓ | Orbit ↓ | Spec. ↓ | Wavelet ↓ | Ratio ↓ | Valid ↑ | Unique ↑ | Novel ↑ | V.U.N. ↑ |
| Training set | 0.0002 | 0.0310 | 0.0005 | 0.0038 | 0.0012 | 1.0 | 100 | 100 | — | — |
| GraphRNN | 0.0049 | 0.2779 | 1.2543 | 0.0459 | 0.1034 | 490.2 | 0.0 | 100 | 100 | 0.0 |
| GRAN | 0.0007 | 0.0426 | 0.0009 | 0.0075 | 0.0019 | 2.0 | 97.5 | 85.0 | 2.5 | 0.0 |
| SPECTRE | 0.0005 | 0.0785 | 0.0012 | 0.0112 | 0.0059 | 3.0 | 25.0 | 100 | 100 | 25.0 |
| DiGress | 0.0007 | 0.0780 | 0.0079 | 0.0098 | 0.0031 | 5.1 | 77.5 | 100 | 100 | 77.5 |
| EDGE | 0.0761 | 0.3229 | 0.7737 | 0.0957 | 0.3627 | 431.4 | 0.0 | 100 | 100 | 0.0 |
| BwR (EDP-GNN) | 0.0231 | 0.2596 | 0.5473 | 0.0444 | 0.1314 | 251.9 | 0.0 | 100 | 100 | 0.0 |
| BiGG | 0.0007 | 0.0570 | 0.0367 | 0.0105 | 0.0052 | 16.0 | 62.5 | 85.0 | 42.5 | 5.0 |
| GraphGen | 0.0328 | 0.2106 | 0.4236 | 0.0430 | 0.0989 | 210.3 | 7.5 | 100 | 100 | 100 |
| Ours (one-shot) | 0.0003 | 0.0245 | 0.0006 | 0.0104 | 0.0030 | **1.7** | 67.5 | 100 | 100 | 67.5 |
| Ours | 0.0005 | 0.0626 | 0.0017 | 0.0075 | 0.0013 | 2.1 | 95.0 | 100 | 100 | **95.0** |

| Stochastic Block Model ($n_{\max} = 187$, $n_{\text{avg}} = 104$) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Model | Deg. ↓ | Clus. ↓ | Orbit ↓ | Spec. ↓ | Wavelet ↓ | Ratio ↓ | Valid ↑ | Unique ↑ | Novel ↑ | V.U.N. ↑ |
| Training set | 0.0008 | 0.0332 | 0.0255 | 0.0027 | 0.0007 | 1.0 | 100 | 100 | — | — |
| GraphRNN | 0.0055 | 0.0584 | 0.0785 | 0.0065 | 0.0431 | 14.7 | 5.0 | 100 | 100 | 5.0 |
| GRAN | 0.0113 | 0.0553 | 0.0540 | 0.0054 | 0.0212 | 9.7 | 25.0 | 100 | 100 | 25.0 |
| SPECTRE | 0.0015 | 0.0521 | 0.0412 | 0.0056 | 0.0028 | 2.2 | 52.5 | 100 | 100 | 52.5 |
| DiGress | 0.0018 | 0.0485 | 0.0415 | 0.0045 | 0.0014 | **1.7** | 60.0 | 100 | 100 | 60.0 |
| EDGE | 0.0279 | 0.1113 | 0.0854 | 0.0251 | 0.1500 | 51.4 | 0.0 | 100 | 100 | 0.0 |
| BwR (EDP-GNN) | 0.0478 | 0.0638 | 0.1139 | 0.0169 | 0.0894 | 38.6 | 7.5 | 100 | 100 | 7.5 |
| BiGG | 0.0012 | 0.0604 | 0.0667 | 0.0059 | 0.0370 | 11.9 | 10.0 | 100 | 100 | 10.0 |
| GraphGen | 0.0550 | 0.0623 | 0.1189 | 0.0182 | 0.1193 | 48.8 | 5.0 | 100 | 100 | 5.0 |
| Ours (one-shot) | 0.0141 | 0.0528 | 0.0809 | 0.0071 | 0.0205 | 10.5 | 75.0 | 100 | 100 | **75.0** |
| Ours | 0.0119 | 0.0517 | 0.0669 | 0.0067 | 0.0219 | 10.2 | 45.0 | 100 | 100 | 45.0 |

| Tree Graphs ($n_{\max} = 64$, $n_{\text{avg}} = 64$) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Model | Deg. ↓ | Clus. ↓ | Orbit ↓ | Spec. ↓ | Wavelet ↓ | Ratio ↓ | Valid ↑ | Unique ↑ | Novel ↑ | V.U.N. ↑ |
| Training set | 0.0001 | 0.0000 | 0.0000 | 0.0075 | 0.0030 | 1.0 | 100 | 100 | — | — |
| GRAN | 0.1884 | 0.0080 | 0.0199 | 0.2751 | 0.3274 | 607.0 | 0.0 | 100 | 100 | 0.0 |
| DiGress | 0.0002 | 0.0000 | 0.0000 | 0.0113 | 0.0043 | **1.6** | 90.0 | 100 | 100 | 90.0 |
| EDGE | 0.2678 | 0.0000 | 0.7357 | 0.2247 | 0.4230 | 850.7 | 0.0 | 7.5 | 100 | 0.0 |
| BwR (EDP-GNN) | 0.0016 | 0.1239 | 0.0003 | 0.0480 | 0.0388 | 11.4 | 0.0 | 100 | 100 | 0.0 |
| BiGG | 0.0014 | 0.0000 | 0.0000 | 0.0119 | 0.0058 | 5.2 | 100 | 87.5 | 50.0 | 75.0 |
| GraphGen | 0.0105 | 0.0000 | 0.0000 | 0.0153 | 0.0122 | 33.2 | 95.0 | 100 | 100 | 95.0 |
| Ours (one-shot) | 0.0004 | 0.0000 | 0.0000 | 0.0080 | 0.0055 | 2.1 | 82.5 | 100 | 100 | 82.5 |
| Ours | 0.0001 | 0.0000 | 0.0000 | 0.0117 | 0.0047 | 4.0 | 100 | 100 | 100 | **100** |



| | Proteins ($n_{\max} = 500$, $n_{\text{avg}} = 258$) | | | | | | Point Clouds ($n_{\max} = 5037$, $n_{\text{avg}} = 1332$) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | Deg. ↓ | Clus. ↓ | Orbit ↓ | Spec. ↓ | Wavelet ↓ | Ratio ↓ | Deg. ↓ | Clus. ↓ | Orbit ↓ | Spec. ↓ | Wavelet ↓ | Ratio ↓ |
| Training set | 0.0003 | 0.0068 | 0.0032 | 0.0005 | 0.0003 | 1.0 | 0.0000 | 0.1768 | 0.0049 | 0.0043 | 0.0024 | 1.0 |
| GraphRNN | 0.004 | 0.1475 | 0.5851 | 0.0152 | 0.0530 | 91.3 | OOM | OOM | OOM | OOM | OOM | OOM |
| GRAN | 0.0479 | 0.1234 | 0.3458 | 0.0125 | 0.0341 | 87.5 | 0.0201 | 0.4330 | 0.2625 | 0.0051 | 0.0436 | 18.8 |
| SPECTRE | 0.0056 | 0.0843 | 0.0267 | 0.0052 | 0.0118 | 19.0 | OOM | OOM | OOM | OOM | OOM | OOM |
| DiGress | 0.0041 | 0.0489 | 0.1286 | 0.0018 | 0.0065 | 18.0 | OOM | OOM | OOM | OOM | OOM | OOM |
| EDGE | 0.1863 | 0.3406 | 0.6786 | 0.1075 | 0.2371 | 399.1 | 0.4441 | 0.3298 | 1.0730 | 0.4006 | 0.6310 | 143.4 |
| BwR (EDP-GNN) | 0.1262 | 0.4202 | 0.4939 | 0.0702 | 0.1199 | 245.4 | 0.4927 | 0.4690 | 1.0730 | 0.2912 | 0.5916 | 133.2 |
| BiGG | 0.0070 | 0.1150 | 0.4696 | 0.0067 | 0.0222 | 57.5 | 0.0994 | 0.6035 | 0.3633 | 0.1589 | 0.0994 | 38.8 |
| GraphGen | 0.0159 | 0.1677 | 0.3789 | 0.0181 | 0.0477 | 83.5 | OOT | OOT | OOT | OOT | OOT | OOT |
| Ours (one-shot) | 0.0015 | 0.0711 | 0.0396 | 0.0026 | 0.0086 | 13.3 | OOM | OOM | OOM | OOM | OOM | OOM |
| Ours | 0.0030 | 0.0309 | 0.0047 | 0.0013 | 0.0030 | **5.9** | 0.0139 | 0.5775 | 0.0780 | 0.0055 | 0.0186 | **7.0** |

## 7 Size extrapolation

- **generalizes to larger graphs** than seen during training