# HYGENE: A Diffusion-based Hypergraph Generation Method

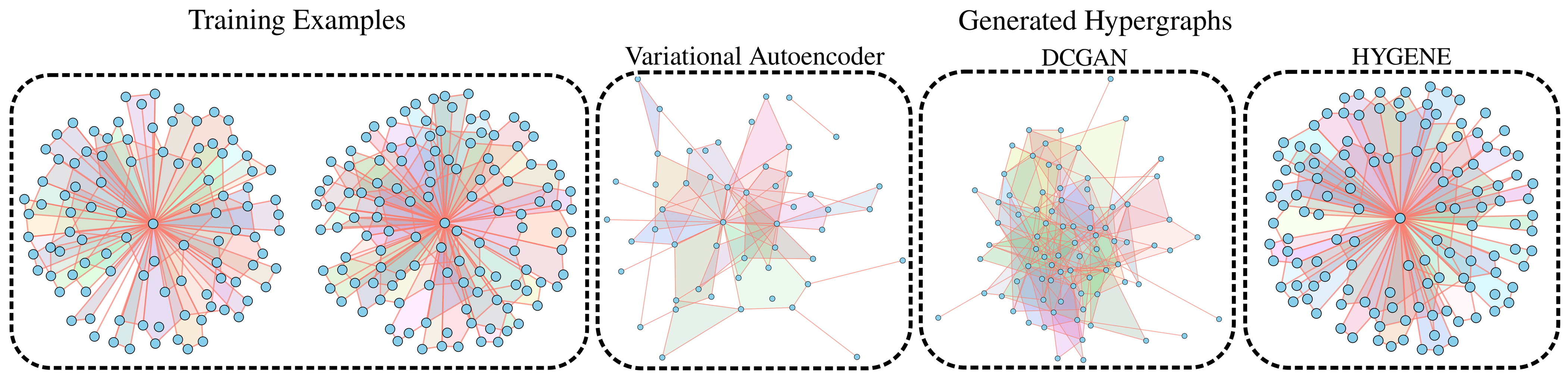
Dorian Gailhard   Enzo Tartaglione   Lirida Naviner de Barros   Jhony H. Giraldo

LTCI, Télécom Paris, Institut Polytechnique de Paris, France

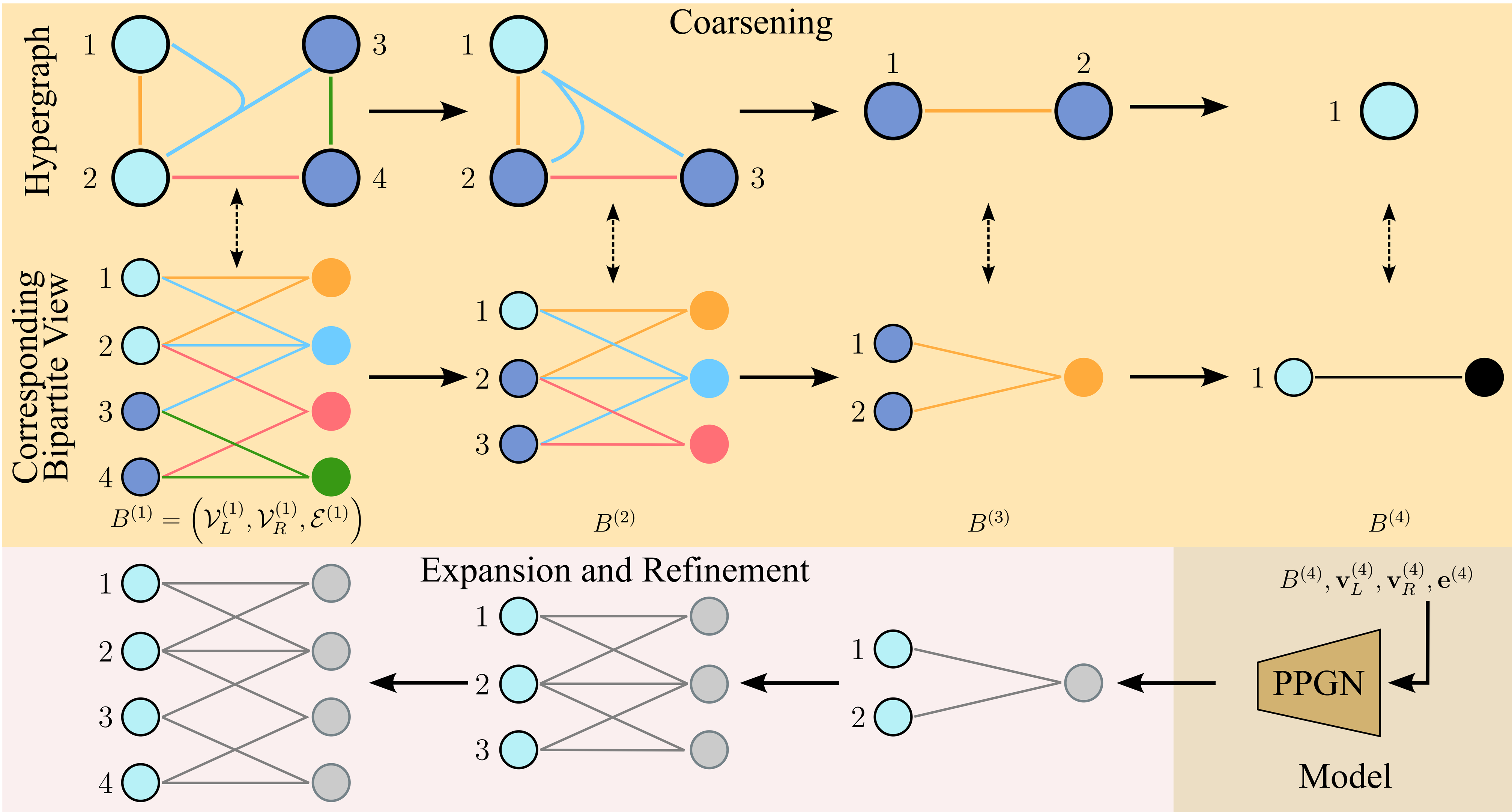
{name.surname}@telecom-paris.fr

## Motivation

- Hypergraphs (higher-order extension of graphs where *hyper*edges can contain more than two nodes) are known to capture subtler relationships compared to graphs.
- Broad field of application: 3D graphics (meshes), pharmaceutical research (molecules), electronics (chip design).
- No deep learning-based hypergraph generator exists and classical ML methods (VAE, GAN and regular diffusion models) fail.
- We aim to lay the groundwork for future research. Here we are interested in *unfeatured* hypergraphs.




## Pipeline




1. Compute increasingly coarse views of a sample hypergraph thanks to spectrum-preserving graph coarsening applied to the clique expansion (dark blue nodes are merged in the figure).

2. Maintain an equivalent bipartite representation at each coarsening step by merging corresponding nodes. Right side nodes representing the same hyperedge are merged.

3. Train two ML models to recover the bipartite representation of a coarsening level from its subsequent level: (a) predict cluster sizes (merged node counts) ; (b) identify edges for removal. In our implementation, we used a denoising diffusion model framework and employ a PPGN as the denoiser architecture.

## Ideas

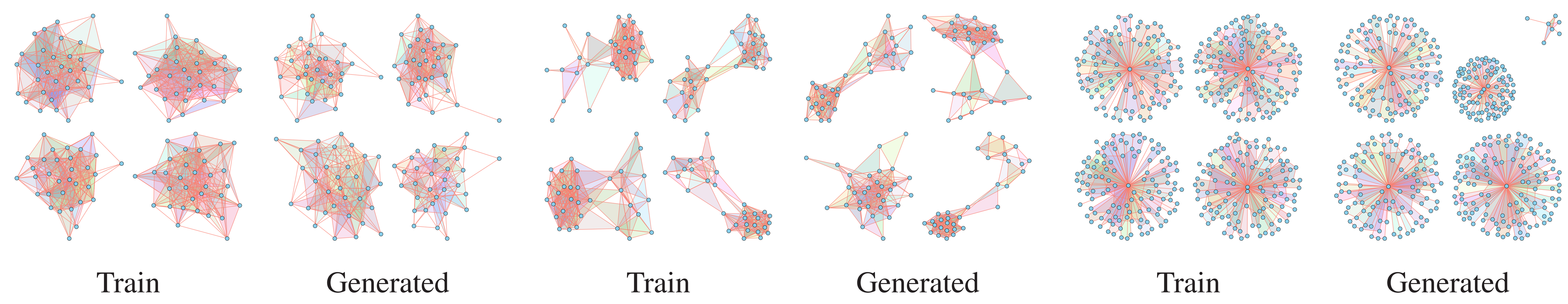
Similarly to images, graphs can be generated by gradually adding details: starting from a "blurry" version (from afar, every graph is a single node), gradually "zooming" and adding details.[1]

We generalize this process to hypergraphs thanks to two projections, depending on which properties we want to preserve:

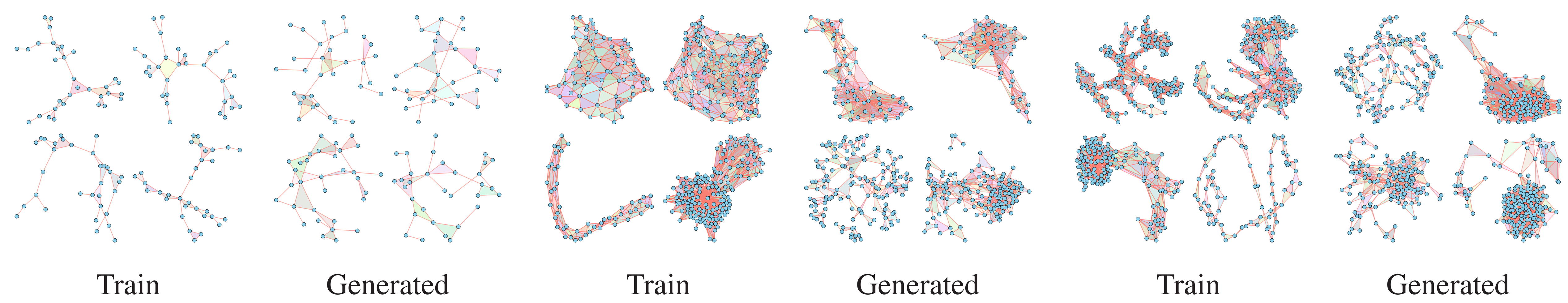
1. **Clique expansion:** hyperedges $e$ are collapsed into a weighted clique (all of the nodes of the hyperedge are connected by regular edges) with weight $1/|e|$. This preserves spectral properties, but recovering the hypergraph is *NP-hard*.
2. **Star expansion:** the hypergraph is converted to a bipartite graph with a partition for the nodes and the other for the hyperedges, each node being connected to the hyperedges containing it. This allows to easily recover the original hypergraph.

1. Efficient and Scalable Graph Generation through Iterative Local Expansion by Bergmeister et al., 2024

## Generated Examples




(a) Erdos-Renyi   (b) Stochastic Block Model   (c) Ego




(d) Tree   (e) ModelNet40 Plant meshes   (f) ModelNet40 Piano meshes

## Numerical Results

| Model | SBM Hypergraphs ($n_{avg} = 31.73$, $std = 0.55$) | | | | Ego Hypergraphs ($n_{avg} = 109.71$, $std = 10.23$) | | | | Tree Hypergraphs ($n_{avg} = 32$, $std = 0$) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Valid SBM ↑ | Node Deg ↓ | Edge Size ↓ | Spectral ↓ | Valid Ego ↑ | Node Deg ↓ | Edge Size ↓ | Spectral ↓ | Valid Tree ↑ | Node Deg ↓ | Edge Size ↓ | Spectral ↓ |
| HyperPA | 2.5% | 4.062 | 0.407 | 0.273 | 0% | 2.590 | 0.423 | 0.237 | 0% | 0.315 | 0.284 | 0.159 |
| VAE | 0% | 1.280 | 1.059 | 0.024 | 0% | 0.803 | 1.458 | 0.133 | 0% | 0.072 | 0.480 | 0.124 |
| GAN | 0% | 2.106 | 1.203 | 0.059 | 0% | 0.917 | 1.665 | 0.230 | 0% | 0.151 | 0.469 | 0.089 |
| Diffusion | 0% | 1.717 | 1.390 | 0.031 | 0% | 3.984 | 2.985 | 0.190 | 0% | 1.718 | 1.922 | 0.127 |
| HYGENE | **65%** | **0.321** | **0.002** | **0.010** | **90%** | **0.063** | **0.220** | **0.004** | **77.5%** | **0.059** | **0.108** | **0.012** |

| Model | Erdos-Renyi Hypergraphs ($n_{avg} = 32$, $std = 0.07$) | | | ModelNet40 Piano ($n_{avg} = 177.29$, $std = 57.11$) | | | ModelNet40 Plant ($n_{avg} = 124.86$, $std = 87.88$) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Node Deg ↓ | Edge Size ↓ | Spectral ↓ | Node Deg ↓ | Edge Size ↓ | Spectral ↓ | Node Deg ↓ | Edge Size ↓ | Spectral ↓ |
| HyperPA | 5.530 | 0.183 | 0.177 | 9.254 | **0.023** | **0.067** | 6.566 | 0.046 | 0.061 |
| VAE | 2.140 | 0.540 | 0.035 | 8.060 | 1.686 | 0.396 | 3.895 | 1.573 | 0.205 |
| GAN | 2.560 | 0.657 | 0.048 | 409.0 | 86.38 | 0.697 | 378.1 | 56.35 | 0.364 |
| Diffusion | 2.225 | 0.781 | 0.014 | 20.90 | 4.192 | 0.113 | 21.03 | 3.439 | 0.069 |
| HYGENE | **0.445** | **0.012** | **0.006** | **6.290** | 0.027 | 0.117 | **2.428** | **0.027** | **0.034** |

**Acknowledgments:** The authors acknowledge the ANR – FRANCE (French National Research Agency) for its financial support of the System On Chip Design leveraging Artificial Intelligence (SODA) project under grant ANR-23-IAS3-0004. This project has also been partially funded by the Hi!PARIS Center on Data Analytics and Artificial Intelligence.