HypeBoy: Generative Self-Supervised Representation Learning on Hypergraphs

**Summary**

- **TL;DR:** We study generative self-supervised representation learning (SSL) on hypergraphs.
- **Research question:** How can we train hypergraph neural networks, particularly when label supervision is insufficient?

**Introduction**

- **Hypergraph:** Higher-order interactions are prevalent; hypergraphs are widely used data structures for representing such interactions.
- **Hypergraph neural network (HNNs):** Neural networks for processing data in the form of hypergraphs.
- **Research question:** How can we train hypergraph neural networks, particularly when label-supervision is insufficient?

**Generative Task**

- **Task:** We formalize a hyperedge filling task.
- **Given:** A subset of a hyperedge where a single node is omitted.
- **Goal:** Correctly predict the missing node for the given subset.
- **Theory:** Can serve as an effective pre-training strategy.

**SSL Method**

- **SSL method:** We propose HypeBoy: hypergraph SSL method that performs hyperedge filling.
- **Step 1 [Augmentation]:** Augment the input hypergraph by feature- and hyperedge-augmentation.
- **Step 2 [Encoding]:** Obtain node and subset embeddings with an HNN and projection heads.
- **Step 3 [SSL Loss]:** Compute hyperedge filling loss and update parameters.

**Experiments**

- **Datasets:** We use 11 benchmark hypergraph datasets, expressing co-authorship, co-citation, movie-actor relationships, etc.
- **Baselines:** We use 16 baseline methods, including two SOTA hypergraph SSL methods (TriCL [3] and HyperGCL [4]).
- **Result 1:** HypeBoy is a better pre-training strategy than baselines (left table).
- **Result 2:** HypeBoy gives a better general-purpose embedding than baselines (right table).

**References**

- [2] He et al. (2022), Masked Autoencoders are Scalable Vision Learners, In CVPR.
- [3] Lee and Shin (2023), Tri-Directional Contrastive Learning on Hypergraphs, In AAAI.