ISCHI**ARS'23**

Contribution ID : **36**                    Type : **Contribution in Organized Session**

# A Principled, Flexible and Efficient Framework for Hypergraph Benchmarking

*mercoledì 3 maggio 2023 09:30 (15)*

In recent years, hypergraphs have emerged as a powerful tool to study systems with multi-body interactions which cannot be trivially reduced to pairs. While highly structured benchmark models have proved fundamental for the standardized evaluation of algorithms and the statistical study of real-world networked data, these are scarcely available in the context of hypergraphs.

In this talk, we present a flexible and efficient framework for the generation of hypergraphs with many nodes and large hyperedges, which allows specifying general community structures and tune different local statistics. In particular, our model allows to sample synthetic hypergraphs with a variety of desired features, among which we list assortative or disassortative communities, and mixed or hard community assignments. Furthermore, we show how it is possible to condition on additional requirements, such as average degree, degree sequence, and hyperedge size sequence.

The availability of a highly structured benchmark model allows studies that were previously not feasible in the context of hypergraphs. Firstly, it allows comparing community detection algorithms against data generated with known ground truth. Secondly, by conditioning on various structural characteristics, it allows generating hypergraphs similar to real-world data. The samples produced via our benchmark can be then utilized for the replication, statistical study and validation of real data, as well as for statistical hypothesis testing.

To summarize, our work constitutes the first highly-structured, community-based benchmark model for the creation of synthetic higher-order data, and constitutes a substantial advancement in the statistical modeling of hypergraphs.

Presentation based on the following work:

A Principled, Flexible and Efficient Framework for Hypergraph Benchmarking Ruggeri N., Battiston F, De Bacco C.

Generalized Inference of Mesoscale Structures in Higher-order Networks Ruggeri N., Contisciani M., Battiston F, De Bacco C.

GitHub repository: github.com/nickruggeri/Hy-MMSBM

## Keywords

hypergraphs, benchmark models, synthetic data, probabilistic models

## Topics

• Statistical methods and models for network analysis

**Primary author(s) :**  RUGGERI, Nicolò (Max Planck Institute for Intelligent Systems / ETH);  Dr. DE BACCO, Caterina (Max Planck Institute for Intelligent Systems);  Ms. CONTISCIANI, Martina (Max Planck Institute for

Intelligent Systems);  Prof. BATTISTON, Federico (Central European University)

**Presenter(s) :**   RUGGERI, Nicolò (Max Planck Institute for Intelligent Systems / ETH)

**Session Classification :**  Statistical models for networks

**Track Classification :**  Statistical models for networks