NetSciX2026: International School and Conference on Network Science
February 17^{th}-20^{th}, 2026 - Auckland, New Zealand

Multi-Order Hypergraph Stochastic Block Model

Keywords: hypergraphs, higher-order networks, stochastic block models, link prediction, network inference.

Extended Abstract

Complex systems often involve interactions among more than two entities, naturally represented as hypergraphs. Stochastic block models offer one approach to extracting community structures within complex systems, fitting generative models of hypergraphs that assume a latent community structure to real-world data [1, 2]. Existing models for hypergraphs (e.g., Refs. [3, 4]) assume a single, uniform interaction pattern (e.g., a global affinity matrix) regardless of hyperedge size. This simplification may obscure crucial qualitative differences in interaction dynamics across varying group sizes (e.g., a small group's dynamics versus a large group's), which can lead to suboptimal community detection and link prediction performance. Here, by accounting for the heterogeneity of community-specific higher-order interaction patterns across different hyperedge sizes, we propose the multi-order stochastic block model for hypergraphs, which we refer to as *HyperMOSBM*.

We model a hypergraph probabilistically, assuming an underlying K communities and soft community memberships for N nodes. The propensity of node v_i belonging to each community is specified by a K-dimensional vector $(u_{ik})_{1\leq k\leq K}$, where $u_{ik}\geq 0$ for all $i=1,\ldots,N$ and $k=1,\ldots,K$. We represent these membership vectors as an $N\times K$ matrix, denoted by $\boldsymbol{U}=(u_{ik})_{1\leq i\leq N,\ 1\leq k\leq K}$. We refer to the number of nodes contained in a hyperedge e as its size, and let $\mathscr{S}$ be the set of hyperedges' sizes observed in the hypergraph. We define a partition of $\mathscr{S}$ as $\mathscr{C}=\{C_1,C_2,\ldots,C_L\}$, where C_l are disjoint subsets of $\mathscr{S}$ such that $\bigcup_{l=1}^{L}C_l=\mathscr{S}$ and $L\geq 1$. Each subset $C_l\in\mathscr{C}$ corresponds to a set of hyperedges' sizes that are assumed to share a common interaction pattern. The strength of intra- and inter-community interactions among nodes is represented by a set of symmetric $K\times K$ affinity matrices, denoted by $\boldsymbol{W}^{(l)}=(w^{(l)}_{kq})_{1\leq k\leq K,\ 1\leq q\leq K}$, where $w^{(l)}_{kq}\geq 0$ for all $k=1,\ldots,K$, $q=1,\ldots,K$, and $l=1,\ldots,L$. Specifically, for any hyperedge e of size $s=|e|$, its interaction pattern is governed by the affinity matrix $\boldsymbol{W}^{(l)}$ where $s\in C_l$. HyperMOSBM is a generative model for a hypergraph, with the number of communities K, node membership matrix $\boldsymbol{U}$, and affinity matrices $\{\boldsymbol{W}^{(1)},\boldsymbol{W}^{(2)},\ldots,\boldsymbol{W}^{(L)}\}$. To investigate the community structure of a given hypergraph, we fit HyperMOSBM with a specified number of communities to that hypergraph, using a maximum likelihood approach to infer the model parameters $\boldsymbol{U}$ and $\{\boldsymbol{W}^{(1)},\boldsymbol{W}^{(2)},\ldots,\boldsymbol{W}^{(L)}\}$.

A central challenge for HyperMOSBM is determining the partition $\mathscr{C}$ of hyperedge sizes. We employ a data-driven strategy based on maximizing the area under the receiver operating characteristic curve (AUC) in link prediction. We begin with an initial clustering where all hyperedge sizes belong to a single cluster (i.e., $L=1$ and $C_1=\mathscr{S}$). In each subsequent step, we greedily explore splitting one existing hyperedge size cluster into two smaller clusters based on adjacent sizes. The split operation that yields the largest improvement in the AUC is chosen. This process continues until the AUC no longer improves by a predefined relative threshold (here we set 1%). For robust evaluation, we measure the AUC as the average over 100 independent training-test splits of the set of hyperedges.

We applied HyperMOSBM to five contact hypergraphs. Each hypergraph consists of nodes representing individuals and hyperedges representing social contacts. HyperMOSBM achieves higher AUC values 30.8%, 26.7%, and 19.3% than Hy-MMSBM [4] in 'primary-school',

'high-school', and 'invs15' hypergraphs, respectively. This suggests that the interaction patterns within these contact hypergraphs exhibit specific heterogeneities across different group sizes. Indeed, in the 'high-school' dataset, HyperMOSBM largely recovered communities corresponding to distinct classrooms in a high school than Hy-MMSBM (see Fig. 1).

Taken together, HyperMOSBM explicitly models heterogeneous interaction patterns across varying group sizes. Future work includes exploring alternative strategies for efficiently determining the partition of the set of hyperedge sizes and applying it to a wider range of hypergraph datasets to further explore the nature of multi-order heterogeneity in real-world complex systems. This work contributes to the effective modeling, deeper understanding, and analysis of community and mesoscale structures in complex systems.

Ethical Considerations

Our study exclusively utilizes publicly available or anonymized datasets, ensuring that no personally identifiable information is processed. The insights gained are intended solely for academic and scientific understanding of complex systems, without direct applications that could pose immediate ethical concerns.

References

[1] Santo Fortunato. In: *Physics Reports* 486 (2010), pp. 75–174.

[2] Clement Lee and Darren J. Wilkinson. In: *Applied Network Science* 4 (2019), p. 122.

[3] Martina Contisciani, Federico Battiston, and Caterina De Bacco. In: *Nature Communications* 13.1 (2022), p. 7229.

[4] Nicolò Ruggeri et al. In: *Science Advances* 9 (2023), eadg9159.

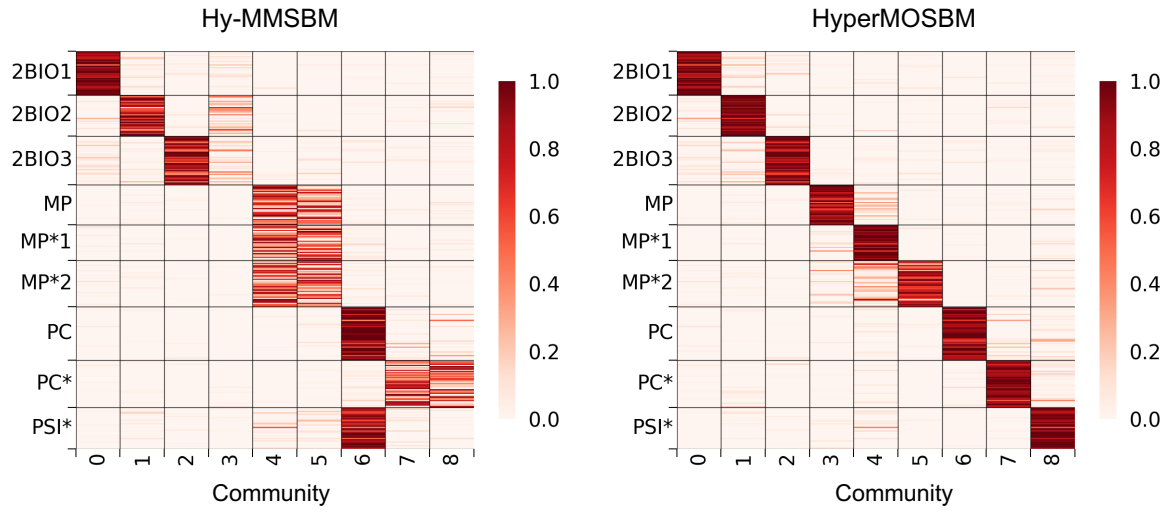

Figure 1: **Inferred community-membership matrix of size $N \times K$ in the high-school data.** The row indices are arranged according to the classroom. $N = 327$ and $K = 9$.