Research Projects on Hypergraphs

In recent years, hypergraphs have emerged as a generalization of graph-type data [Battiston et al., 2020]. Their popularity has been growing thanks to their ability to represent higher-order interactions, which are often observed in a variety of contexts. Some examples are individuals interacting in groups, co-voting patterns and scientific collaborations, as well as specific applications, such as protein interactions. As a consequence, recent works expand several classical concepts from the graph literature to hypergraphs, e.g. percolation, syncronization or contagion dynamics. The following proposals deal with two under-explored areas in the hypergraph literature.

Skills required: linear algebra, statistics and probability, basic coding (Python main libraries).
Supervisors: Caterina De Bacco (website), Nicoló Ruggeri (website), Martina Contisciani (website).

Community Detection on Hypergraphs: Incorporating Covariate Information

Community detection, the problem of finding hidden clusters of similar nodes, has been only recently investigated for the case of hypergraphs [Contisciani et al., 2022, Chodrow et al., 2021, Chodrow et al., 2022], including ongoing work performed in our research group [Ruggeri et al., 2022a, Ruggeri et al., 2022b, Ruggeri et al., 2023, *in preparation*]. These works suggest that it is possible to build efficient and effective inference models for community detection based only on the structural properties of hypergraphs. However, usually hypergraph data come with extra information, as node attributes, which is not accounted in all these recent methods. Similarly to previous approaches on graph-type data [Contisciani et al., 2020], it is believed that incorporating covariate information could potentially boost the quality of inference. Hence, the need for approaches that are capable of incorporating this extra information.

In this project, we aim at combining the intuitions from previous works [Contisciani et al., 2022, Contisciani et al., 2020] to perform community detection on hypergraphs incorporating additional covariate information.
The project will require developing the mathematical model, deriving a procedure for performing parameters' inference, coding the algorithmic implementation and anything related to data analysis of both synthetic and real data.

Node and Hypergraph Embedding

A standard way to perform downstream tasks on graphs is to perform node or graph embedding [Hamilton, 2020]. These two tasks consist of extracting a numerical vector representation (*embedding*) for every node in the graph, or a single representation for the whole graph, respectively. Famous examples algorithms that perform these tasks are Node2Vec [Grover and Leskovec, 2016] and Graph2Vec [Narayanan et al., 2017].

In this project, we aim at performing the first studies on the feasibility, utility, and effectivity of node embedding on hypergraphs, and/or of hypergraph embedding. The starting point for developing our method will be current work on graph embedding [Hamilton, 2020], random walks and inference on hypergraphs [Battiston et al., 2020, Contisciani et al., 2022], as well as the existing literature on hypergraph embedding [Zhou et al., 2006, Sun et al., 2021, Maleki et al., 2022].
The project will require exploring the current literature on the topic, with the goal of developing the first mathematical model, as well as the training/inference procedure, coding the algorithmic implementation and anything related to data analysis of both synthetic and real data.

References

[Battiston et al., 2020] Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.-G., and Petri, G. (2020). Networks beyond pairwise interactions: structure and dynamics. *Physics Reports*, 874:1–92.

[Chodrow et al., 2022] Chodrow, P., Eikmeier, N., and Haddock, J. (2022). Nonbacktracking spectral clustering of nonuniform hypergraphs. *arXiv preprint arXiv:2204.13586*.

[Chodrow et al., 2021] Chodrow, P. S., Veldt, N., and Benson, A. R. (2021). Generative hypergraph clustering: From blockmodels to modularity. *Science Advances*, 7(28):eabh1303.

[Contisciani et al., 2022] Contisciani, M., Battiston, F., and De Bacco, C. (2022). Principled inference of hyperedges and overlapping communities in hypergraphs. *arXiv preprint arXiv:2204.05646*.

[Contisciani et al., 2020] Contisciani, M., Power, E. A., and De Bacco, C. (2020). Community detection with node attributes in multilayer networks. *Scientific reports*, 10(1):1–16.

[Grover and Leskovec, 2016] Grover, A. and Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864.

[Hamilton, 2020] Hamilton, W. L. (2020). Graph representation learning. *Synthesis Lectures on Artifical Intelligence and Machine Learning*, 14(3):1–159.

[Maleki et al., 2022] Maleki, S., Saless, D., Wall, D. P., and Pingali, K. (2022). Hypernetvec: Fast and scalable hierarchical embedding for hypergraphs. In *International Conference on Network Science*, pages 169–183. Springer.

[Narayanan et al., 2017] Narayanan, A., Chandramohan, M., Venkatesan, R., Chen, L., Liu, Y., and Jaiswal, S. (2017). graph2vec: Learning distributed representations of graphs. *arXiv preprint arXiv:1707.05005*.

[Sun et al., 2021] Sun, X., Yin, H., Liu, B., Chen, H., Cao, J., Shao, Y., and Viet Hung, N. Q. (2021). Heterogeneous hypergraph embedding for graph classification. In *Proceedings of the 14th acm international conference on web search and data mining*, pages 725–733.

[Zhou et al., 2006] Zhou, D., Huang, J., and Schölkopf, B. (2006). Learning with hypergraphs: Clustering, classification, and embedding. *Advances in neural information processing systems*, 19.