

## RESEARCH ARTICLE

## HAT: Hypergraph analysis toolbox

Joshua Pickard<sup>1,2</sup>, Can Chen<sup>3</sup>, Rahmy Salman<sup>4</sup>, Cooper Stansbury<sup>1</sup>, Sion Kim<sup>4</sup>, Amit Surana<sup>5</sup>, Anthony Bloch<sup>6</sup>, Indika Rajapakse<sup>1,2,6\*</sup>

**1** Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, United States of America, **2** iReprogram, Inc., Ann Arbor, Michigan, United States of America, **3** Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States of America, **4** Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan, United States of America, **5** Raytheon Technologies Research Center, East Hartford, Connecticut, United States of America, **6** Department of Mathematics, University of Michigan, Ann Arbor, Michigan, United States of America

\* [indikar@umich.edu](mailto:indikar@umich.edu)



## OPEN ACCESS

**Citation:** Pickard J, Chen C, Salman R, Stansbury C, Kim S, Surana A, et al. (2023) HAT: Hypergraph analysis toolbox. *PLoS Comput Biol* 19(6): e1011190. <https://doi.org/10.1371/journal.pcbi.1011190>

**Editor:** Mark Alber, University of California Riverside, UNITED STATES

**Received:** December 21, 2022

**Accepted:** May 16, 2023

**Published:** June 5, 2023

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1011190>

**Copyright:** © 2023 Pickard et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant code and documentation may be found at the software main page <https://hypergraph-analysis-toolbox.readthedocs.io/> Additional open source distributions of HAT are available on GitHub

## Abstract

Recent advances in biological technologies, such as multi-way chromosome conformation capture (3C), require development of methods for analysis of multi-way interactions. Hypergraphs are mathematically tractable objects that can be utilized to precisely represent and analyze multi-way interactions. Here we present the Hypergraph Analysis Toolbox (HAT), a software package for visualization and analysis of multi-way interactions in complex systems.

## Author summary

Classical networks typically focus on pairwise interactions and may overlook the intricate higher-order, multi-way interactions that occur among groups of nodes within a network. Our research has delved into the structural and dynamic characteristics of hypergraphs, which can effectively capture multi-way network interactions across various domains and data types. In this article, we introduce the Hypergraph Analysis Toolbox (HAT), a software package encompassing a range of techniques to identify, investigate, and visualize multi-way interactions in biological data.

This is a *PLOS Computational Biology* Software paper.

## Introduction

Network science is a powerful framework for studying complex systems. However, recent work highlights the limitations of classical methods in networks, which only consider pairwise interactions between nodes to describe group interactions. Use of hypergraphs, in which an edge can connect more than two nodes, has therefore emerged as a new frontier in network science [1–3].

(<https://github.com/Jpickard1/Hypergraph-Analysis-Toolbox>), PyPI (<https://pypi.org/project/HypergraphAnalysisToolbox/>), and MATLAB Central (<https://www.mathworks.com/matlabcentral/fileexchange/121013-hat-hypergraph-analysis-toolbox>), which are all linked on the main page. All issue and bug tracking for the software are handled through GitHub issues (<https://github.com/Jpickard1/Hypergraph-Analysis-Toolbox/issues>) which is also referenced from the main page. Tutorials on usage of the HAT are available through Google CoLab and MATLAB Online (<https://hypergraph-analysis-toolbox.readthedocs.io/en/latest/tutorials.html>).

**Funding:** This work is supported in part by the Air Force Office of Scientific Research (AFOSR) awards FA9550-18-1-0028 and FA9550-22-1-0215 (IR), NSF DMS2103026 (AB), and a MathWorks Fellowship to the Rajapakse Lab (IR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

Chromosome conformation capture (3C) methods identify physical interactions (“contacts”) between genomic loci [4, 5]. While classical 3C is pairwise, recent advancements capture multi-way chromatin interactions via proximity ligation (see Pore-C in [S1 File](#)) [6], Split-Pool Recognition of Interactions by Tag Extension (SPRITE) [7, 8], or multi-contact 3C (MC-3C) [9]. However, the investigation and biological interpretation of these multi-way contacts is hampered by scarcity of methods for multi-way data [6, 10]. Hypergraphs are a mathematically tractable extension of graph theory that precisely represent multi-way interactions (See Hypergraphs in [S1 File](#)) [2]. We introduce the Hypergraph Analysis Toolbox (HAT), a general purpose software for the analysis of multi-way interactions and higher-order structures. HAT contains both well-studied and novel mathematical methods for hypergraph analysis in both MATLAB and Python.

Motivated to investigate Pore-C data, HAT is designed as a versatile software for hypergraph analysis. While there are several robust libraries for graph analysis, most hypergraph software is not multi-faceted and targets specific problems, such as hypergraph partitioning or clustering ([Table 1](#)). As a general purpose tool, the algorithms implemented in HAT address hypergraph construction, visualization, and the analysis of structural and dynamic properties. HAT is the first software to utilize tensor algebra for hypergraph analysis [11–13], and it contains recently developed methods for hypergraph similarity measures [13]. HAT is open source, standardized across MATLAB (version 2021b onward) and Python (version 3.7 onward) implementations, and is documented at <https://hypergraph-analysis-toolbox.readthedocs.io>, where it will continue to be maintained and developed.

For ease of use, the MATLAB and Python implementations are functionally independent but syntactically similar. The software may be installed from the online documentation, GitHub, or via PIP and the MathWorks file exchange for the respective Python and MATLAB implementations.

## Materials and methods

HAT can visualize and analyze multi-way interactions. The incidence matrix is the primary representation of hypergraphs in HAT ([Fig 1b](#)) [10, 23]. HAT targets the following hypergraph features and problems: (1) construction from data [24–26], (2) expansion and numeric representation [27–29], (3) characteristic structural properties (such as entropy [11], centrality [30], distance [13], and clustering coefficients [11]), (4) controllability [12], and (5) similarity measures [13]. The workflow for using HAT is outlined in [Fig 1e](#).

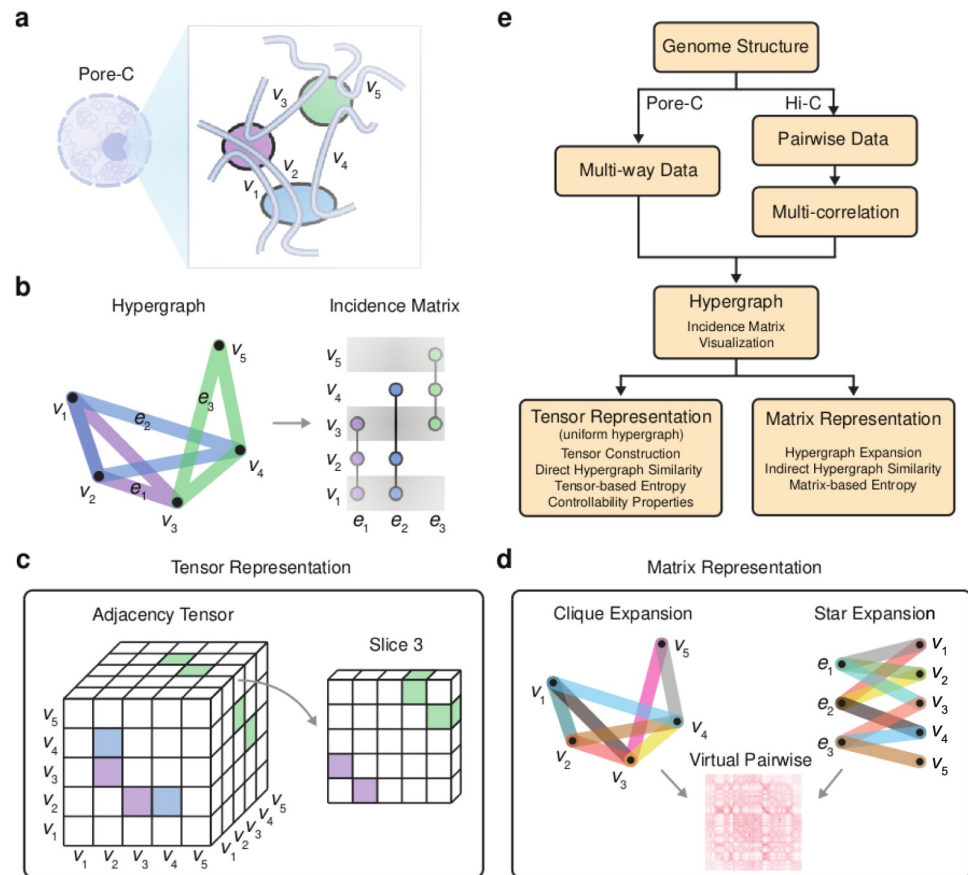
## Construction from data

There are two approaches for constructing a hypergraph from data (see Hypergraphs in [S1 File](#)). Data formats with explicit multi-way interactions, such as Pore-C are directly input to HAT for hypergraph construction. However, the vast majority of data are either pairwise

**Table 1. Comparison of HAT to well-documented hypergraph libraries.** There are several other notable hypergraph software not listed in the table [19–22].

Software	Language	Features
Hypergraph Analysis Toolbox	MATLAB and Python	construction, visualization, expansion, similarity measures, centrality, entropy, tensor- based analysis, controllability
HyperNetX [14]	Python	clustering, visualization, homology, clustering
HALP [15]	Python	directed hypergraphs, walks, partitioning
hMETIS [16]	C/C++	partitioning, implemented in parallel
Phoenix [17]	C/C++	clustering, implemented in parallel
HyperG [18]	R	hypergraph creation, clustering, graph based representations and calculations, visualization

<https://doi.org/10.1371/journal.pcbi.1011190.t001>



**Fig 1. HAT overview.** **a.** The Pore-C assay identifies multi-way chromatin strand colocalization within the nucleus [6]. **b.** Hypergraph representation of Pore-C is drawn where each chromatin strand is represented as a vertex and the multi-way contacts are hyperedges. This is depicted as both a hypergraph and an incidence matrix. **c.** For multi-way contacts of uniform size, hypergraphs are numerically represented as an adjacency tensor or multi-dimensional matrix. **d.** Multi-way structure are decomposed with clique and star expansions that generate virtual pairwise contacts [6]. **e.** The workflow of HAT to construct hypergraphs from data, visualize, represent numerically, and computations available for each representation are outlined as a flowchart.

<https://doi.org/10.1371/journal.pcbi.1011190.g001>

observations (e.g., Hi-C) or do not contain either pairwise or multi-way interactions (e.g., sequencing data), so we implemented three measures to infer multi-way relationships based on multi-correlation measures [24–26]. HAT constructs hyperedges by setting a minimum threshold for the multi-correlation to be considered a hyperedge.

### Expansion and numerical representation

For uniform hypergraphs, the adjacency, degree, and Laplacian tensors (Fig 1c) are provided and utilized in similarity, entropy, and controllability calculations (see Numeric Representations of Hypergraphs in S1 File). Such tensor based calculations are not currently supported for non-uniform hypergraphs and will be pursued in the future. However, both uniform and non-uniform hypergraphs expand to pairwise structures (Fig 1d, see Hypergraph Expansions in S1 File). HAT contains hypergraphs expansions to generate clique expansions, star expansions, and line graphs. These representations facilitate indirect hypergraph similarity and entropy measures for non-uniform hypergraphs. Each hypergraph expansion has unique adjacency, degree, Laplacian, and normalized Laplacian matrices [27–29].

## Characteristic structural properties

The following structural properties of hypergraphs are computed: average distance between nodes is computed based on [13] (see Hypergraph Structural Properties in [S1 File](#), Equation S1); the clustering coefficient is calculated based on [11] (Equation S2); hypergraph centrality is measured according to methods in [30, 31], which employ a variety of techniques to solve the nonlinear eigenvalue problem. For a uniform hypergraph, entropy is computed according to [11], which is defined based on the higher-order singular values of the Laplacian tensor (Equation S3). For non-uniform hypergraphs, standard graph entropy measures are applied to the aforementioned hypergraph expansions.

## Controllability

Hypergraph controllability refers to the ability to steer the underlying system of a hypergraph to a desired state by manipulating a subset of nodes (often referred to as driver nodes) [12]. For a uniform hypergraph, the minimum number of driver nodes required for controllability can be computed using the generalized Kalman's rank condition (see Hypergraph Controllability in [S1 File](#), Equation S7). HAT is the first software to analyze controllability properties of hypergraphs.

## Similarity measures

Hypergraph similarity is measured according to the recent work [13], which distinguishes direct and indirect hypergraph similarity measures. Direct measures utilize tensor representations of uniform hypergraphs; indirect measures utilize graph similarity measures applied to hypergraph expansions. A series of structural and feature-based hypergraph similarity measures, including the Hamming Distance, the Jaccard Index, spectral measures, and centrality measures are provided (see Hypergraph Similarity Measures in [S1 File](#), Equation S4 and S5). HAT is the first software to implement hypergraph similarity using a tensor representation based on the novel methods in [13].

## Results

Methods contained in HAT were utilized to examine Pore-C data ([Fig 1a](#), see Pore-C in [S1 File](#)) [10]. Hypergraphs were constructed from Pore-C data from multiple cell types. Hypergraph similarity measures were employed to compare the structural similarity between different regions of the genome and cell types. In terms of biological implications, hypergraph entropy of chromosome structure has identified bifurcation points that determined cell fate over the course of a cell reprogramming experiment, which remained unidentified with a graph-theoretic approach [11]. Hypergraph analysis was also integrated with other sequencing modalities to identify transcriptional clusters and elucidate the higher-order organization of the genome [10].

In addition to examining Pore-C data, HAT was also used to quantify the activity of hypothalamic neurons monitored during a mouse feeding, fasting, and re-feeding experiment [32]. When constructing graph and hypergraph representations of the neuronal network within the hypothalamus for each phase of the experiment, hypergraph entropy proved to be a better indicator of changes in neuronal activity compared to graph entropy [11]. A similar result was also observed from a controllability/observability perspective under the same setting [12, 33].

Other applications of HAT include detecting influential hubs in social networks [34, 35], gaining insights into the stability and robustness of biochemical reaction networks [36, 37], identifying keystone species in ecological networks [38], and pinpointing control targets in epidemiological networks [39].

## Discussion

Hypergraphs can represent multi-way relationships unambiguously. The computational methods provided in HAT include hypergraph controllability and similarity measures from a tensor-based perspective. Additionally, the inclusion of tensor-based hypergraph structural properties (i.e., entropy and centrality), the association of multi-correlations with hypergraphs, and the integration of previously implemented graph expansion and visualization techniques within one software is an advancement over previously disjoint implementations. Therefore, HAT can advance the study of multi-way interactions in the genome or other complex biological systems.

## Supporting information

**S1 File. Supplementary information for HAT.**  
(PDF)

## Acknowledgments

We would like to thank Dr. Frederick Leve at the Air Force Office of Scientific Research (AFOSR) for support and encouragement. We would also like to thank the two referees for their constructive comments, which led to a significant improvement of the article.

## Author Contributions

**Conceptualization:** Joshua Pickard, Indika Rajapakse.

**Formal analysis:** Joshua Pickard.

**Funding acquisition:** Indika Rajapakse.

**Investigation:** Can Chen, Indika Rajapakse.

**Methodology:** Joshua Pickard, Can Chen, Amit Surana, Anthony Bloch, Indika Rajapakse.

**Project administration:** Joshua Pickard, Indika Rajapakse.

**Resources:** Indika Rajapakse.

**Software:** Joshua Pickard, Can Chen, Rahmy Salman, Cooper Stansbury, Sion Kim, Amit Surana, Indika Rajapakse.

**Supervision:** Amit Surana, Anthony Bloch, Indika Rajapakse.

**Validation:** Joshua Pickard, Can Chen, Indika Rajapakse.

**Visualization:** Joshua Pickard, Can Chen, Rahmy Salman, Cooper Stansbury, Sion Kim, Indika Rajapakse.

**Writing – original draft:** Joshua Pickard, Can Chen, Cooper Stansbury, Sion Kim, Amit Surana, Anthony Bloch, Indika Rajapakse.

**Writing – review & editing:** Joshua Pickard, Can Chen, Sion Kim, Indika Rajapakse.

## References

1. Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, et al. Networks beyond pairwise interactions: structure and dynamics. *Physics Reports*. 2020; 874:1–92. <https://doi.org/10.1016/j.physrep.2020.05.004>
2. Benson AR, Gleich DF, Higham DJ. Higher-order network analysis takes off, fueled by classical ideas and new data. *arXiv preprint arXiv:210305031*. 2021.



3. Chen C, Liu YY. A survey on hyperlink prediction. arXiv preprint arXiv:2207.02911. 2022.
4. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*. 2002; 295(5558):1306–11. <https://doi.org/10.1126/science.1067799> PMID: 11847345
5. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*. 2009 Oct; 326(5950):289–93. <https://doi.org/10.1126/science.1181369> PMID: 19815776
6. Deshpande AS, Ulahannan N, Pendleton M, Dai X, Ly L, Behr JM, et al. Identifying synergistic high-order 3D chromatin conformations from genome-scale nanopore concatemer sequencing. *Nature Biotechnology*. 2022:1–12. PMID: 35637420
7. Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, et al. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*. 2018 Jul; 174(3):744–57.e24. <https://doi.org/10.1016/j.cell.2018.05.024> PMID: 29887377
8. Quinodoz SA, Bhat P, Chovanec P, Jachowicz JW, Ollikainen N, Detmar E, et al. SPRITE: a genome-wide method for mapping higher-order 3D interactions in the nucleus using combinatorial split-and-pool barcoding. *Nature protocols*. 2022; 17(1):36–75. <https://doi.org/10.1038/s41596-021-00633-y> PMID: 35013617
9. Tavares-Cadete F, Norouzi D, Dekker B, Liu Y, Dekker J. Multi-contact 3C reveals that the human genome during interphase is largely not entangled. *Nature structural & molecular biology*. 2020; 27(12):1105–14. <https://doi.org/10.1038/s41594-020-0506-5>
10. Dotson GA, Chen C, Lindsly S, Cicalo A, Dilworth S, Ryan C, et al. Deciphering multi-way interactions in the human genome. *Nature Communications*. 2022 Sep; 13:5498. <https://doi.org/10.1038/s41467-022-32980-z> PMID: 36127324
11. Chen C, Rajapakse I. Tensor entropy for uniform hypergraphs. *IEEE Transactions on Network Science and Engineering*. 2020; 7(4):2889–900. <https://doi.org/10.1109/TNSE.2020.3002963>
12. Chen C, Surana A, Bloch AM, Rajapakse I. Controllability of hypergraphs. *IEEE Transactions on Network Science and Engineering*. 2021; 8(2):1646–57. <https://doi.org/10.1109/TNSE.2021.3068203>
13. Surana A, Chen C, Rajapakse I. Hypergraph Similarity Measures. *IEEE Transactions on Network Science and Engineering*. 2023; 10(2):658–74. <https://doi.org/10.1109/TNSE.2022.3217185>
14. Praggastis B, Arendt D, Yun JY, Liu T, Lumsdaine A, Joslyn C, et al. HyperNetX. Pacific Northwest National Laboratory. Available from: <https://github.com/pnnl/HyperNetX>.
15. Avent B, Ritz A, Murali TM, Cadena J, Keneshloo Y. Hypergraph Algorithms Package;. Available from: <https://murali-group.github.io/halp/>.
16. Karypis G. hMETIS 1.5: A hypergraph partitioning package. <http://www.cs.umn.edu/~metis>. 1998.
17. Kurte K, Imam N, Hasan S, Kannan R. Phoenix: A Scalable Streaming Hypergraph Analysis Framework. In: *Advances in Data Science and Information Engineering*. Springer; 2021. p. 3–25.
18. Marchette DJ. HyperG: Hypergraphs in R, Available from: <https://CRAN.R-project.org/package=HyperG>.
19. Aksoy S, Firoz J, Harun S, Jenkins L, Joslyn C, Lightsey C, et al. Chapel Hypergraph Library. Pacific Northwest National Laboratory. Available from: <https://pnnl.github.io/chgl/>.
20. Huang J, Zhang R, Yu JX. Scalable hypergraph learning and processing. In: *2015 IEEE International Conference on Data Mining*. IEEE; 2015. p. 775–80.
21. Lg A. HyperGraphLib. Available from: <https://alex-87.github.io/HyperGraphLib/>.
22. Karve V. Multihypergraph. Available from: <https://github.com/vaibhavkarve/multihypergraph>.
23. Valdivia P, Buono P, Plaisant C, Dufournaud N, Fekete JD. Analyzing Dynamic Hypergraphs with Parallel Aggregated Ordered Hypergraph Visualization. *IEEE Transactions on Visualization and Computer Graphics*. 2021; 27(1):1–13. <https://doi.org/10.1109/TVCG.2019.2933196> PMID: 31398121
24. Drezner Z. Multirelation—a correlation among more than two variables. *Computational Statistics & Data Analysis*. 1995; 19(3):283–92. [https://doi.org/10.1016/0167-9473\(93\)E0046-7](https://doi.org/10.1016/0167-9473(93)E0046-7)
25. Wang J, Zheng N. Measures of Correlation for Multiple Variables. arXiv preprint. 2014. Available from: <https://arxiv.org/abs/1401.4827>.
26. Taylor BM. A multi-way correlation coefficient. arXiv preprint arXiv:200302561. 2020.
27. Rodriguez JA. On the Laplacian spectrum and walk-regular hypergraphs. *Linear and Multilinear Algebra*. 2003; 51(3):285–97. <https://doi.org/10.1080/0308108031000084374>
28. Bolla M. Spectra, euclidean representations and clusterings of hypergraphs. *Discrete Mathematics*. 1993; 117(1-3):19–39. [https://doi.org/10.1016/0012-365X\(93\)90322-K](https://doi.org/10.1016/0012-365X(93)90322-K)
29. Zhou D, Huang J, Schölkopf B. Beyond Pairwise Classification and Clustering Using Hypergraphs. Max Planck Institute for Biological Cybernetics; 2005. 143.

30. Tudisco F, Higham DJ. Node and Edge Eigenvector Centrality for Hypergraphs. arXiv preprint arXiv:210106215. 2021.
31. Benson AR. Three hypergraph eigenvector centralities. *SIAM Journal on Mathematics of Data Science*. 2019; 1(2):293–312. <https://doi.org/10.1137/18M1203031>
32. Sweeney P, Chen C, Rajapakse I, Cone RD. Network dynamics of hypothalamic feeding neurons. *Proceedings of the National Academy of Sciences*. 2021; 118(14). <https://doi.org/10.1073/pnas.2011140118> PMID: 33795520
33. Pickard J, Surana A, Bloch A, Rajapakse I. Observability of Hypergraphs. arXiv preprint arXiv:230404883. 2022.
34. Luqman A, Akram M, Smarandache F. Complex neutrosophic hypergraphs: new social network models. *Algorithms*. 2019; 12(11):234. <https://doi.org/10.3390/a12110234>
35. Arya D, Worring M. Exploiting relational information in social networks using geometric deep learning on hypergraphs. In: *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*; 2018. p. 117–25.
36. Jost J, Mulas R. Hypergraph Laplace operators for chemical reaction networks. *Advances in mathematics*. 2019; 351:870–96. <https://doi.org/10.1016/j.aim.2019.05.025>
37. Chen C, Liao C, Liu YY. Teasing out missing reactions in genome-scale metabolic networks through hypergraph learning. *Nature Communications*. 2023; 14:2375. <https://doi.org/10.1038/s41467-023-38110-7> PMID: 37185345
38. Golubski AJ, Westlund EE, Vandermeer J, Pascual M. Ecological networks over the edge: hypergraph trait-mediated indirect interaction (TMII) structure. *Trends in ecology & evolution*. 2016; 31(5):344–54. <https://doi.org/10.1016/j.tree.2016.02.006> PMID: 26924738
39. Bodó Á, Katona GY, Simon PL. SIS epidemic propagation on hypergraphs. *Bulletin of mathematical biology*. 2016; 78(4):713–35. <https://doi.org/10.1007/s11538-016-0158-0> PMID: 27033348