Modeling and Mining High-order Interactions in Social Media Data

Alessia Antelmi, Daniele De Vinco, Andrea Failla,
Giulio Rossetti, and Carmine Spagnuolo

International AAAI Conference on Web and Social Media
June 23rd, 2025, Copenhagen, Denmark

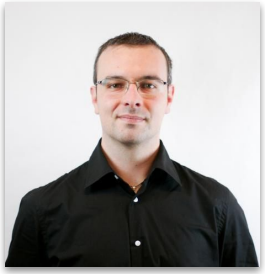
UNIVERSITÀ DI TORINO
UNIVERSITÀ DEGLI STUDI DI SALERNO
UNIVERSITÀ DI PISA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing
ISISLab
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" Consiglio Nazionale delle Ricerche

Tutorial Organizers

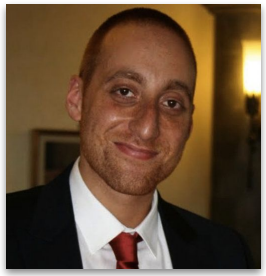
Alessia Antelmi
Assistant Professor
University of Turin

Daniele De Vinco
PhD Student
University of Salerno

Andrea Failla
PhD Student
University of Pisa



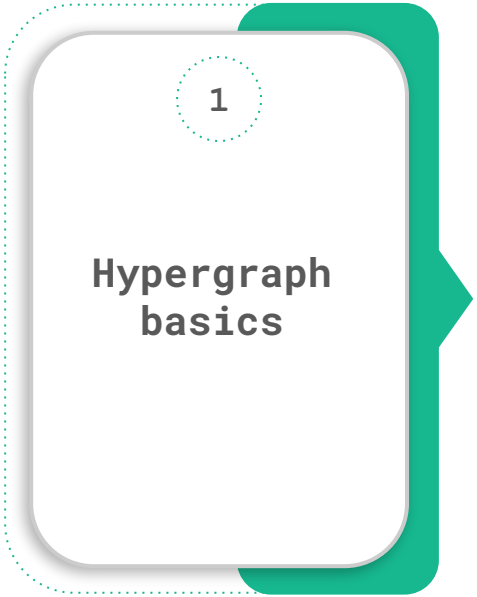
Giulio Rossetti
Senior Researcher
CNR-ISTI Pisa



Carmine Spagnuolo
Tenured Assistant Professor
University of Salerno

Tutorial Outline

Tutorial Outline - Part 1

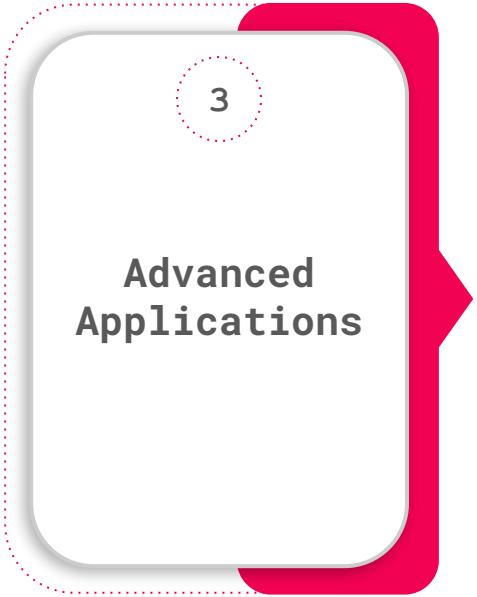


- Introduction to hypergraphs
- Using hypergraphs to model social media data
- Let's code: introduction to hypergraphx

Tutorial Outline - Part 2

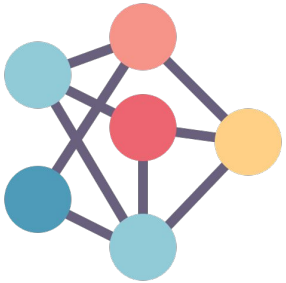
- Structural properties of hypergraphs
  - Degree distribution
  - s-paths
  - Centrality metrics
- Hypergraph communities

Tutorial Outline - Part 3



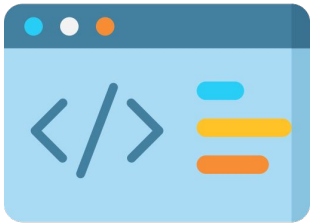
- Node attributed hypernetworks
  - Hyperedge purity
  - Star homogeneity
- High-order contagion dynamics

Prerequisites



- Familiarity with graph-related concepts (e.g., matrix representations, centrality metrics)

- Programming experience and basic knowledge of Python is beneficial, but not mandatory.



Hypergraphs

Let's start with an example

(Enrica, Amedeo, Martina)

Topic $_1$

(Amedeo, Martina)

Topic $_2$

(Mary, Simone, Michel, Martina)

Topic $_3$

(Mary)

Topic $_4$

(Alex)

-

Let's start with an example

(Enrica, Amedeo, Martina)

Topic 1

(Amedeo, Martina)

Topic 2

(Mary, Simone, Michel, Martina)

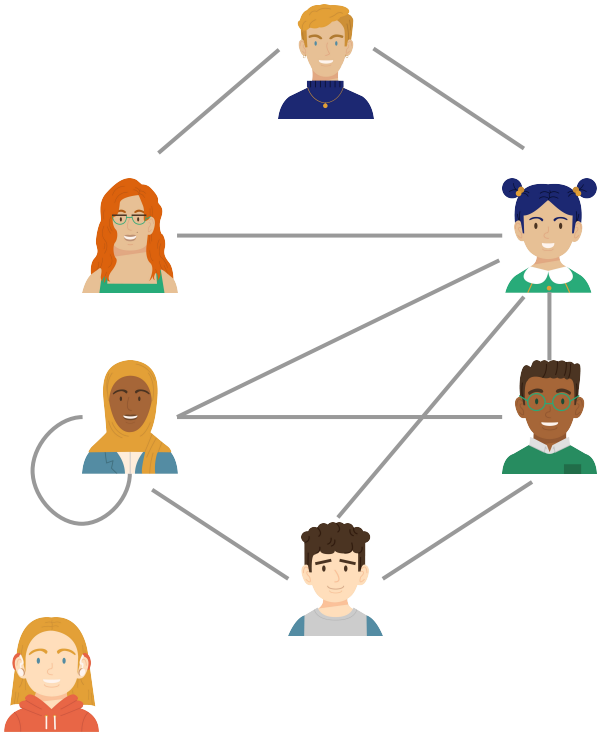
Topic 3

(Mary)

Topic 4

(Alex)

-



Let's start with an example

(Enrica, Amedeo, Martina)

Topic $_1$

(Amedeo, Martina)

Topic $_2$

(Mary, Simone, Michel, Martina)

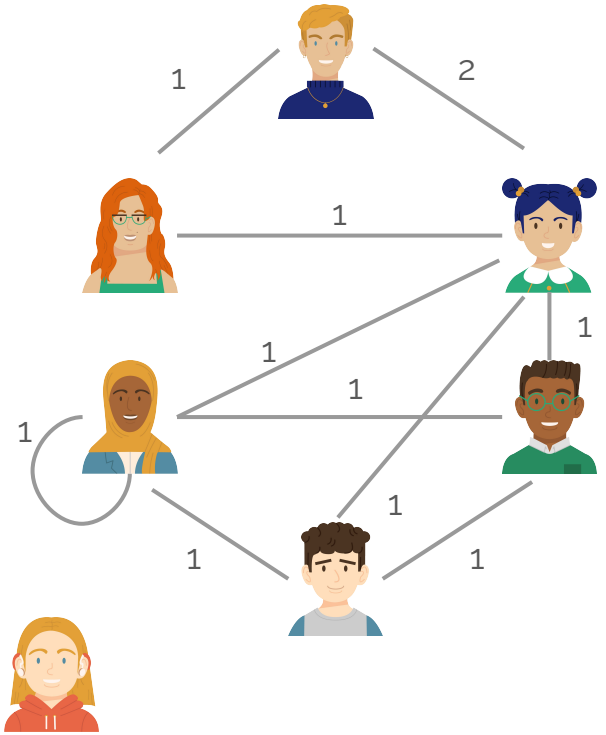
Topic $_3$

(Mary)

Topic $_4$

(Alex)

-



Let's start with an example

(Enrica, Amedeo, Martina)

Topic 1

(Amedeo, Martina)

Topic 2

(Mary, Simone, Michel, Martina)

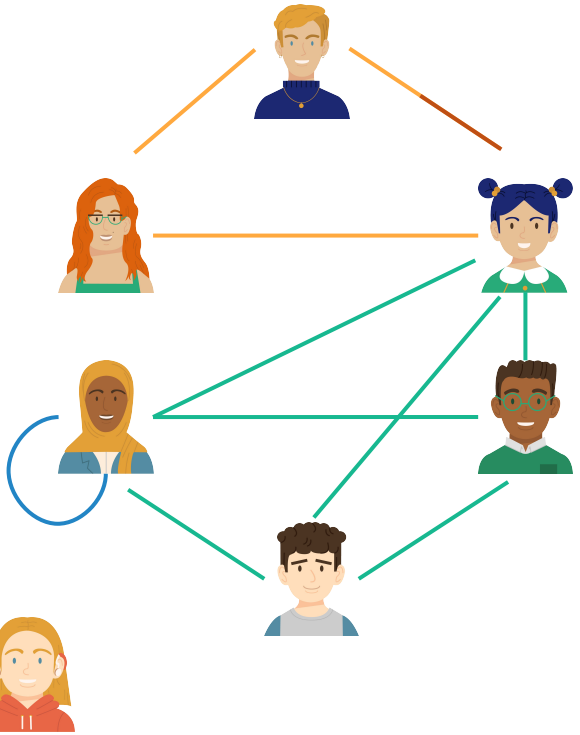
Topic 3

(Mary)

Topic 4

(Alex)

-



Hypergraphs

Generalization of graphs where a **hyperedge** can **connect more** than two vertices.

Hypergraphs: a formal definition

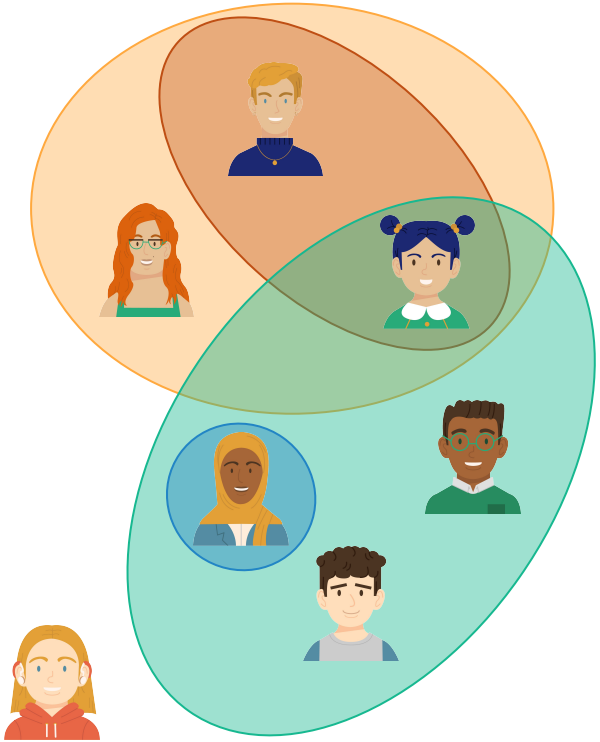
Definition 1.1 : Hypergraphs

A hypergraph H, denoted with $H = (\mathcal{V}, E = (e_i)_{i \in \mathcal{I}})$, on a finite set $\mathcal{V}$ and a finite set of indexes $\mathcal{I}$ is a family $(e_i)_{i \in \mathcal{I}}$ of subsets of $\mathcal{V}$ called hyperedges.

V = {Enri, Michel, Simo, Amedeo, Marti, Mary, Alex}

E = {$Topic_1$, $Topic_2$, $Topic_3$, $Topic_4$}, where

- $Topic_1$ = {Enrica, Amedeo, Martina}
- $Topic_2$ = {Amedeo, Martina}
- $Topic_3$ = {Martina, Michel, Simone, Mary}
- $Topic_4$ = {Mary}



A Bretto. 2013. *Hypergraph Theory: An Introduction.* Springer International Publishing.

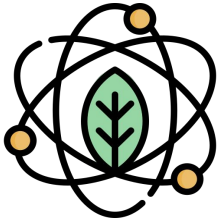
When can hypergraphs be useful??

When one should use hypergraphs

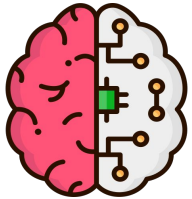
The system to examine exhibits
group/many-to-many/high-order
interactions.

Examples of application domains

Social Systems



Biology



Neuroscience

Ecology

[1] Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, Giovanni Petri, *Networks beyond pairwise interactions: Structure and dynamics*, Physics Reports, Volume 874, 2020.

[2] Alessia Antelmi, Gennaro Cordasco, Mirko Polato, Vittorio Scarano, Carmine Spagnuolo, and Dingqi Yang. *A Survey on Hypergraph Representation Learning*. ACM Comput. Surv. 56, 1, Article 24, 2023.

[3] Geon Lee, Fanchen Bu, Tina Eliassi-Rad, and Kijung Shin. *A Survey on Hypergraph Mining: Patterns, Tools, and Generators*. ACM Comput. Surv. 57, 8, Article 203, 2025.

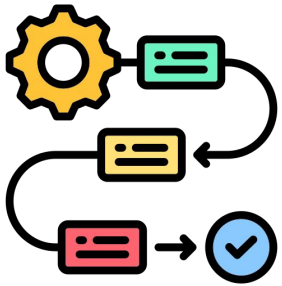
The expressiveness of hypergraphs - Example 1

How much **abstraction of group interactions** is sufficient in solving a hypergraph task?

Guideline on how to **trade off** between **complexity** and **accuracy** of solving a downstream task.

Se-eun Yoon, Hyungseok Song, Kijung Shin, and Yung Yi. 2020. **How Much and When Do We Need Higher-order Information in Hypergraphs? A Case Study on Hyperedge Prediction.** In Proceedings of The Web Conference 2020 (WWW '20). ACM., 2627–2633.

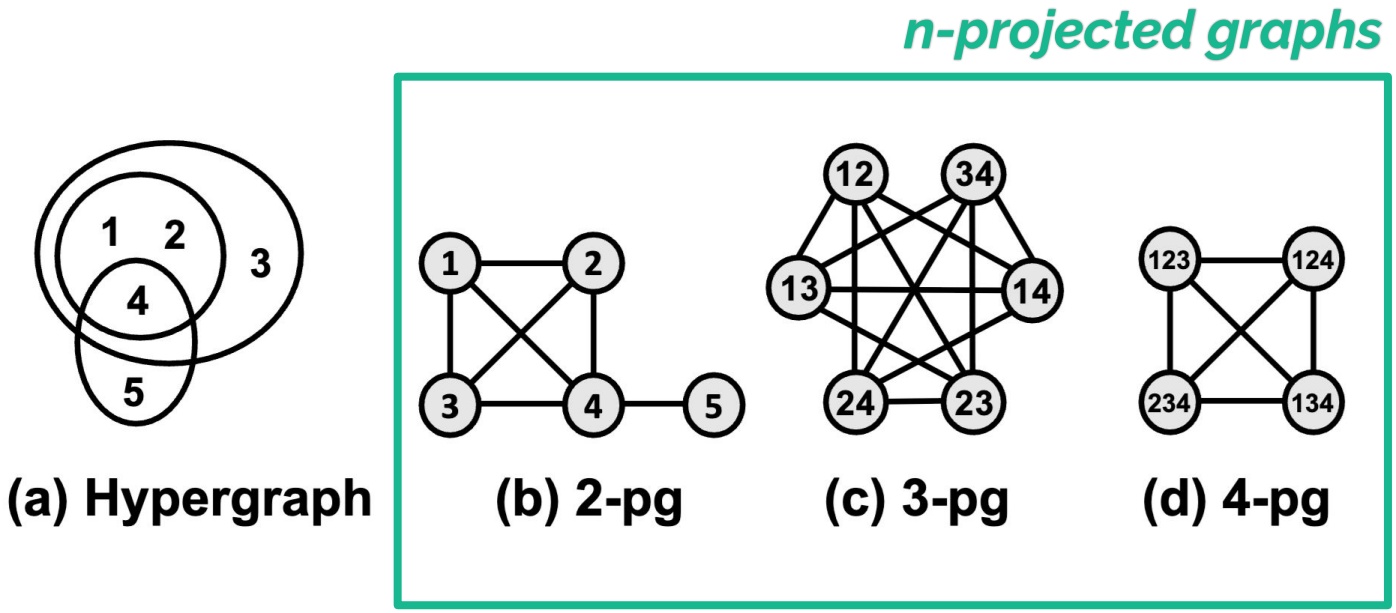
The expressiveness of hypergraphs - Example 1



1. Method for incrementally represent group interactions (*n-projected graphs*)
2. Quantify the accuracy of solving a task as n grows (*link prediction*)

Se-eun Yoon, Hyungseok Song, Kijung Shin, and Yung Yi. 2020. **How Much and When Do We Need Higher-order Information in Hypergraphs? A Case Study on Hyperedge Prediction.** In Proceedings of The Web Conference 2020 (WWW '20). ACM., 2627–2633.

The expressiveness of hypergraphs - Example 1

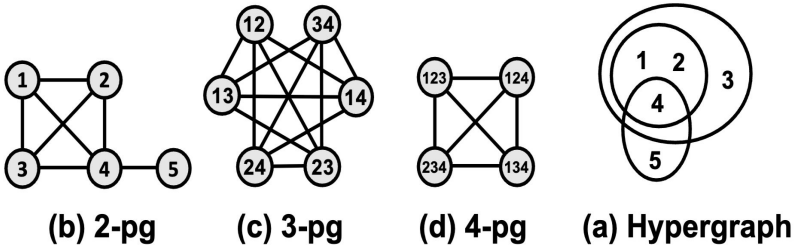


Se-eun Yoon, Hyungseok Song, Kijung Shin, and Yung Yi. 2020. **How Much and When Do We Need Higher-order Information in Hypergraphs? A Case Study on Hyperedge Prediction.** In Proceedings of The Web Conference 2020 (WWW '20). ACM., 2627–2633.

The expressiveness of hypergraphs - Example 1

High-order information are more useful when **group interactions**

- Are **more frequent** than pairwise relations, and
- **Share less information** with pairwise ones.



Better prediction quality (link prediction)

Se-eun Yoon, Hyungseok Song, Kijung Shin, and Yung Yi. 2020. **How Much and When Do We Need Higher-order Information in Hypergraphs? A Case Study on Hyperedge Prediction.** In Proceedings of The Web Conference 2020 (WWW '20). ACM., 2627–2633.

The expressiveness of hypergraphs - Example 2

How do **online interactions** within support communities **impact individuals'** psychological states?

Framework that combines **psycholinguistic** and **social network analysis** to investigate the **evolution of psychological states**

Virginia Morini, Salvatore Citraro, Elena Sajno, Maria Sansoni, Giuseppe Riva, Massimo Stella, Giulio Rossetti, **Online posting effects: Unveiling the non-linear journeys of users in depression communities on Reddit**, Computers in Human Behavior Reports, Volume 17, 2025.

The expressiveness of hypergraphs - Example 2

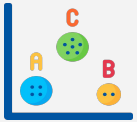
Data

150k active users
6 subreddits

Psycholinguistic dimensions

is represented by

Users' Psychological States



Four psychological clusters

Social Exposure

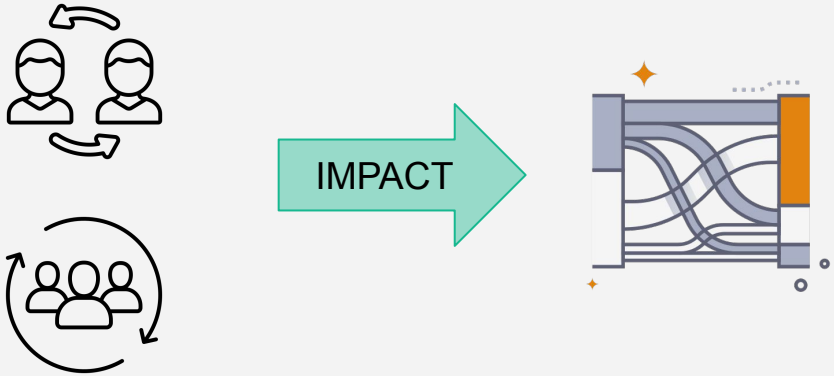
Direct

Indirect: same conversation

Virginia Morini, Salvatore Citraro, Elena Sajno, Maria Sansoni, Giuseppe Riva, Massimo Stella, Giulio Rossetti, **Online posting effects: Unveiling the non-linear journeys of users in depression communities on Reddit**, Computers in Human Behavior Reports, Volume 17, 2025.

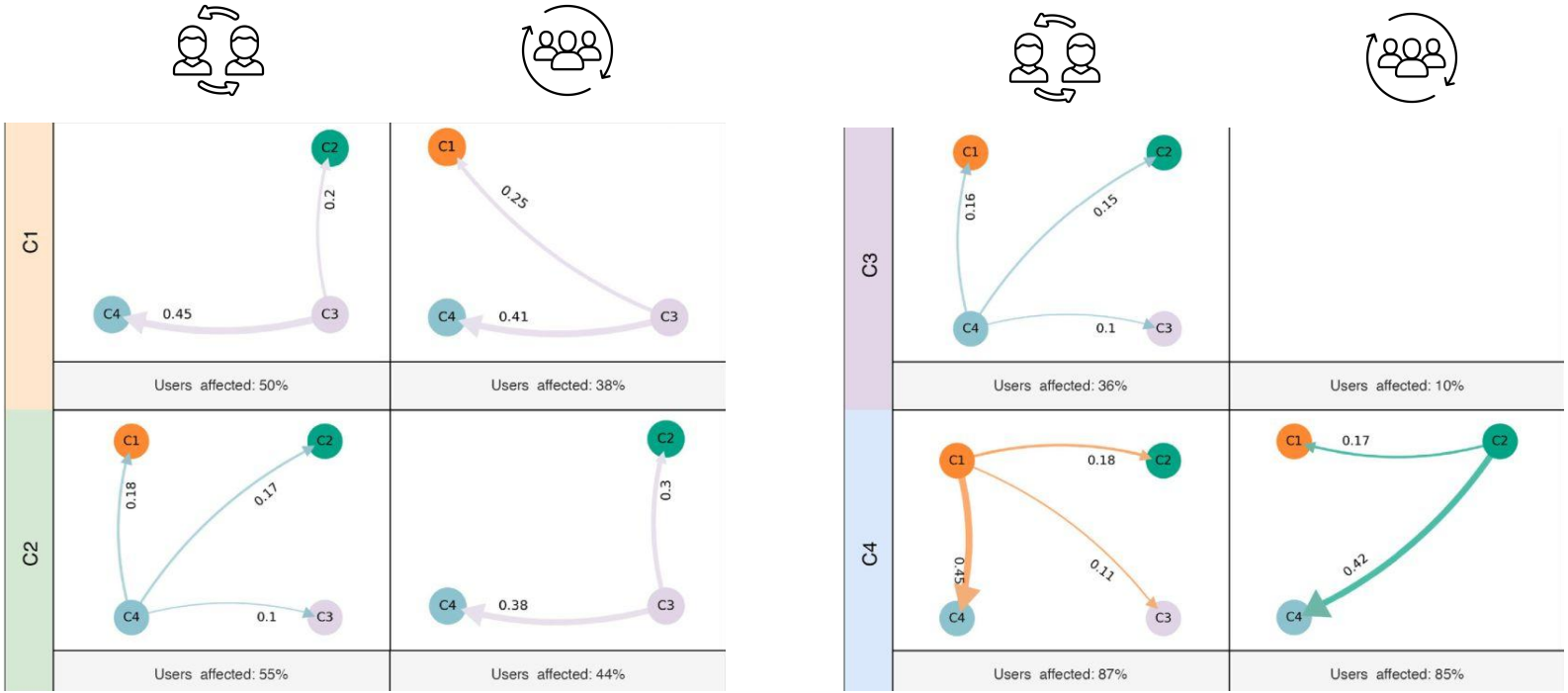
The expressiveness of hypergraphs - Example 2

How do **online interactions** within support communities **impact individuals'** psychological states?



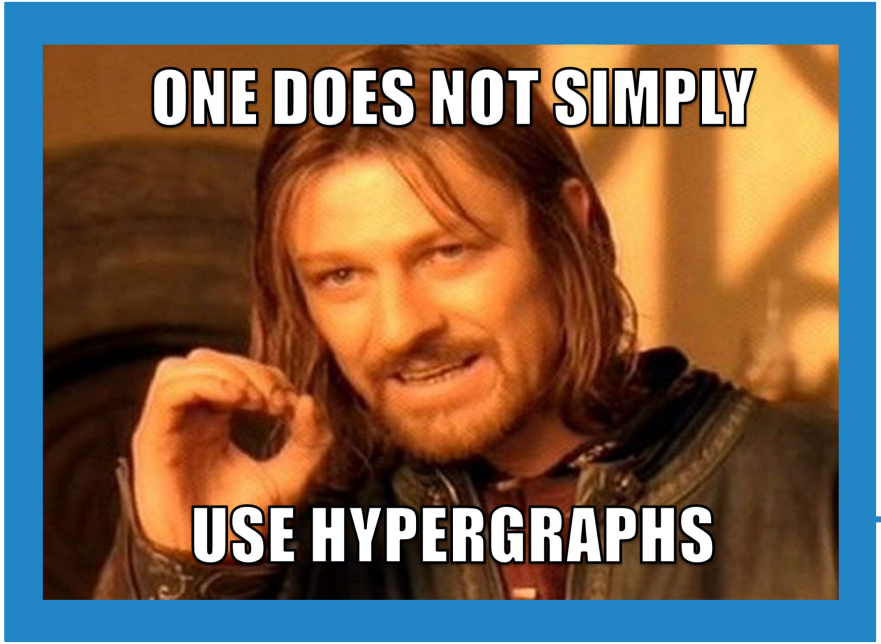
Virginia Morini, Salvatore Citraro, Elena Sajno, Maria Sansoni, Giuseppe Riva, Massimo Stella, Giulio Rossetti, **Online posting effects: Unveiling the non-linear journeys of users in depression communities on Reddit**, Computers in Human Behavior Reports, Volume 17, 2025.

The expressiveness of hypergraphs - Example 2



Conditioned transition matrices.

Virginia Morini, Salvatore Citraro, Elena Sajno, Maria Sansoni, Giuseppe Riva, Massimo Stella, Giulio Rossetti, **Online posting effects: Unveiling the non-linear journeys of users in depression communities on Reddit**, Computers in Human Behavior Reports, Volume 17, 2025.



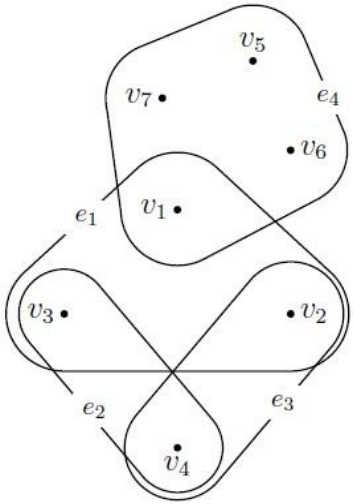
How can we leverage hypergraph representations?

Brace Yourself

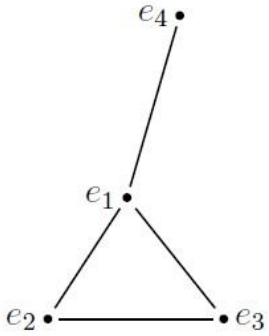
- Hypergraphs add **complexity** (e.g., exponential number of hyperedges);
- Need of **dedicate** algorithms and tools (e.g., hypergraph walks have length and width [1]).

[1] Aksoy, S.G., Joslyn, C., Ortiz Marrero, C. *et al.* **Hypernetwork science via high-order hypergraph walks**. *EPJ Data Sci.* 9, 16 (2020).

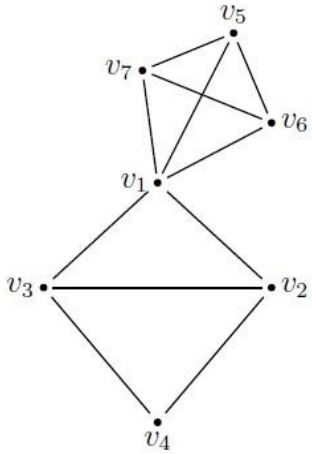
Hypergraph to graph transformations



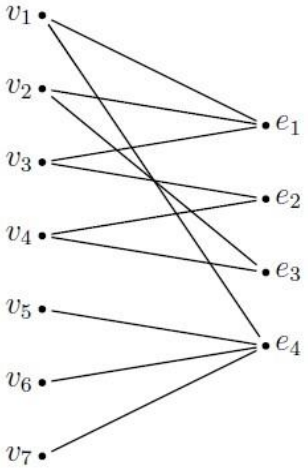
a) Hypergraph



b) Line graph

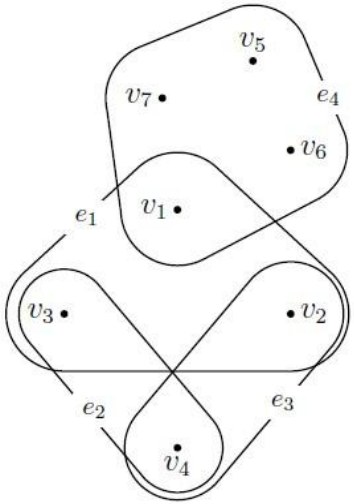


c) Clique graph

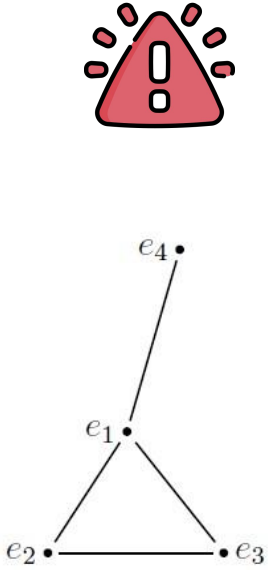


b) Bipartite graph

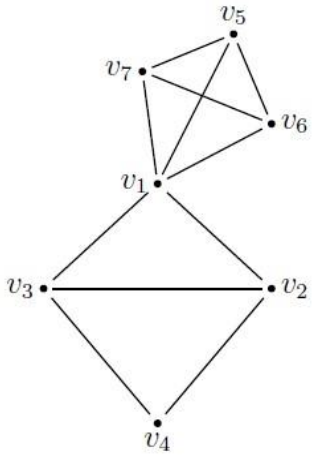
Hypergraph to graph transformations



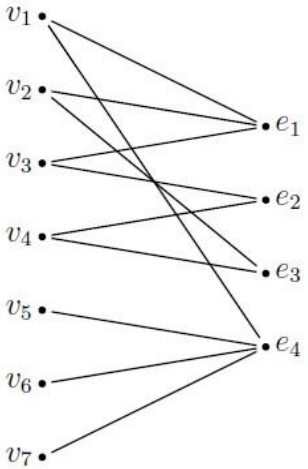
a) Hypergraph



b) Line graph

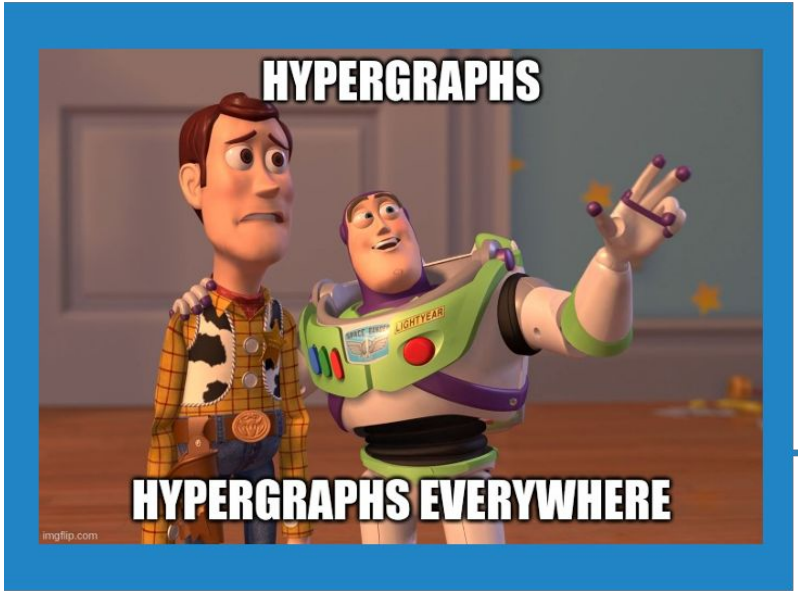


c) Clique graph



b) Bipartite graph

Hypergraphs & Social Media Data



High-order interactions in social media data

Examples of high-order interactions in social media data

Conversations

- Group conversations
- Comment threads

Content Sharing

- Hashtag sharing
- Content co-creation

User attributes

- Common interests
- Demographic characteristics

Multi-modal interactions

- Relations between users, content, and time

Our use case: Political debates on Reddit

- **Platform**: Reddit 
- **Data**:
  - Debate between Trump supporters and anti-Trump citizens
  - January 2017 - July 2019
  - Gun control, minorities discrimination, and political sphere

Dataset	# Subreddit	# Post	# User
GUN CONTROL	6	180,170	65,111
MINORITIES DISCRIMINATION	6	223,096	52,337
POLITICAL SPHERE	6	431,930	72,399

V. Morini, L. Pollacci, and G. Rossetti. **Toward a Standard Approach for Echo Chamber Detection: Reddit Case Study**. Applied Sciences, 11(12), 5390, 2021.

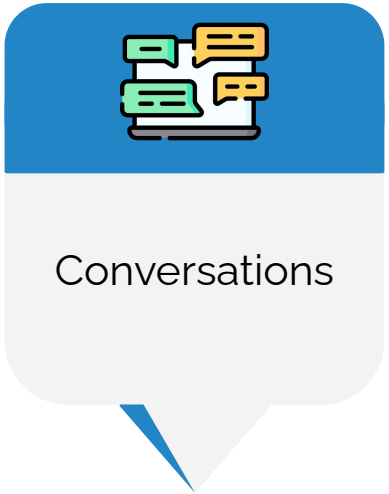
Our use case: Political debates on Reddit

- **Platform**: Reddit 
- **Data**:
  - Debate between Trump supporters and anti-Trump citizens
  - January 2017 - July 2019
  - Gun control, minorities discrimination, and political sphere

Dataset	# Subreddit	# Post	# User
GUN CONTROL	6	180,170	65,111
MINORITIES DISCRIMINATION	6	223,096	52,337
POLITICAL SPHERE	6	431,930	72,399

V. Morini, L. Pollacci, and G. Rossetti. **Toward a Standard Approach for Echo Chamber Detection: Reddit Case Study**. Applied Sciences, 11(12), 5390, 2021.

Our use case: Political debates on Reddit

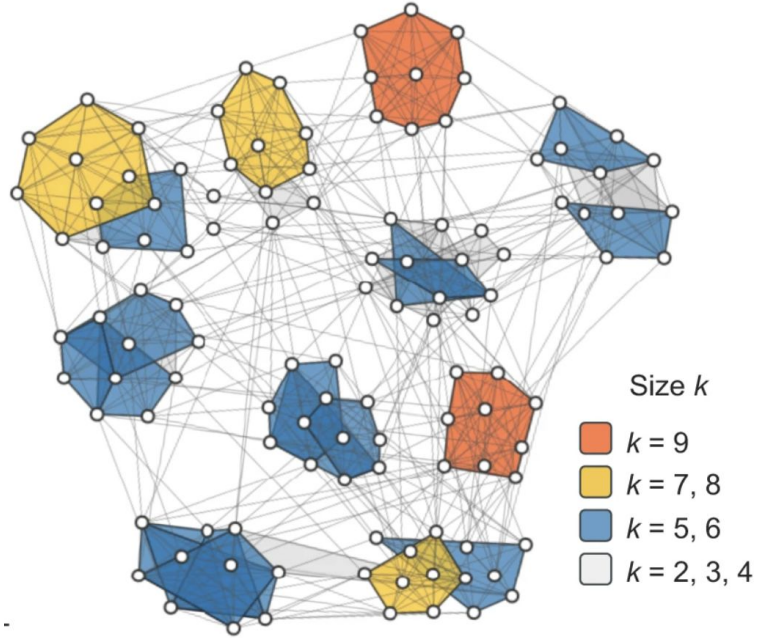


- Each hyperedge links all users who have directly interacted in a conversation (which does not have to be the same).
- The hypergraph structure is inferred by using the maximal cliques approach.

Failla, A., Citraro, S. & Rossetti, G. **Attributed Stream Hypergraphs: temporal modeling of node-attributed high-order interactions.** Applied Network Science 8, 31 (2023).

Our use case: Political debates on Reddit

- Let's suppose that…

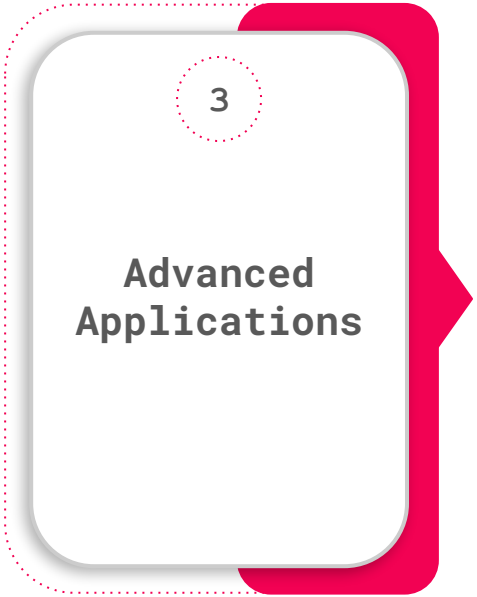


Young, JG., Petri, G. & Peixoto, T.P. **Hypergraph reconstruction from network data**. Commun Phys 4, 135 (2021).

Our use case: Political debates on Reddit

- Identification of the most influential users and conversations
- Identification and analysis of user communities

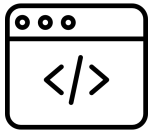
Our use case: Political debates on Reddit



- Homophilic behaviors in group political discussions on Reddit
- The impact of high-order interactions in spreading dynamics

Let’s code

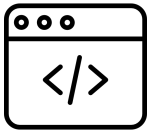
What do I need to run the code?



1. A browser installed on your device;
2. A Google account to run the Colab Notebooks we share.

No specific software licenses are required, and the setup should be almost immediate.

What do I need to run the code locally?



1. Python 3.11
2. hypergraphx v1.7.3

Plus

1. matplotlib
2. seaborn
3. networkx
4. numpy
5. ipython

Material

You can find all material at the following link:

https://dsh2025.github.io#material

Wrapping up

Take home message

Hypergraphs:

- Are effective **tools** to analyze and **mine** group interactions.
- Are complex objects and we need **efficient** and **easy-to-use** programming libraries to model and analyze them.

Take home message

Hypergraphs:

- Allow studying **group behavior** in complex systems with **different granularities**
  - *Microscale* → node level (e.g., s-centrality)
  - *Mesoscale* → hyperedge level (e.g., homophilic behaviors, communities)
  - *Macroscale* → entire structure (e.g., high-order spreading dynamics)

Take home message

Hypergraphs:

- Naturally model **group interactions** in complex systems and can be exploited to study **mixing behaviors** in phenomena like:
  - homophily
  - acrophily
  - polarization
  - information diffusion
  - …

Alessia Antelmi

alessia.antelmi@unito.it

@alessant.bsky.social

Daniele de Vinco

ddevinco@unisa.it

@ddevin96.bsky.social

Andrea Failla

andrea.failla@phd.unipi.it

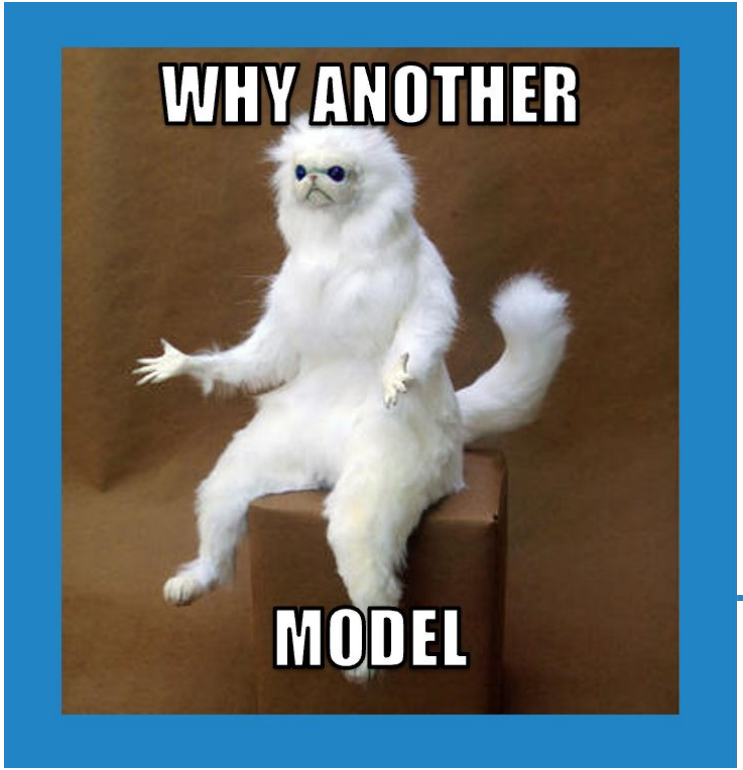
@andreajpg.bsky.social

Images from 'Flaticon.com'.

Takeaways

- Hypergraphs are effective **tools** to analyze and **mine** group interactions.
- Something about the libraries here → hg are complex objects and we need efficient tools to model and analyze them
- something about structural insights?
- Something from notebook 3

Additional slides



Why should we use hypergraphs?

Limitations of transforming hypergraphs to graphs

Line graphs and clique graphs

- We lose information about group interactions
  - In practice, we cannot go back to the original hypergraph once transformed into a graph…
  - …since different hypergraphs may have the same line/clique graph.
  - Further, we may materialize relations that do not exist.

Limitations of transforming hypergraphs to graphs

Line graphs and clique graphs

- We lose information about group interactions
  - In practice, we cannot go back to the original hypergraph once transformed into a graph…
  - …since different hypergraphs may have the same line/clique graph.
  - Further, we may materialize relations that do not exist.

- Need more space
  - Line graph: each vertex of size *d* yields to *d choose 2* edges;
  - Clique graph: each hyperedge of size k yields to $k(k-1)/2$.

Limitations of transforming hypergraphs to graphs

Line graphs and clique graphs

- We lose information about group interactions
  - In practice, we cannot go back to the original hypergraph once transformed into a graph…
  - …since different hypergraphs may have the same line/clique graph.
  - Further, we may materialize relations that do not exist.

- Need more space
  - Line graph: each vertex of size *d* yields to *d choose 2* edges;
  - Clique graph: each hyperedge of size k yields to $k(k-1)/2$.

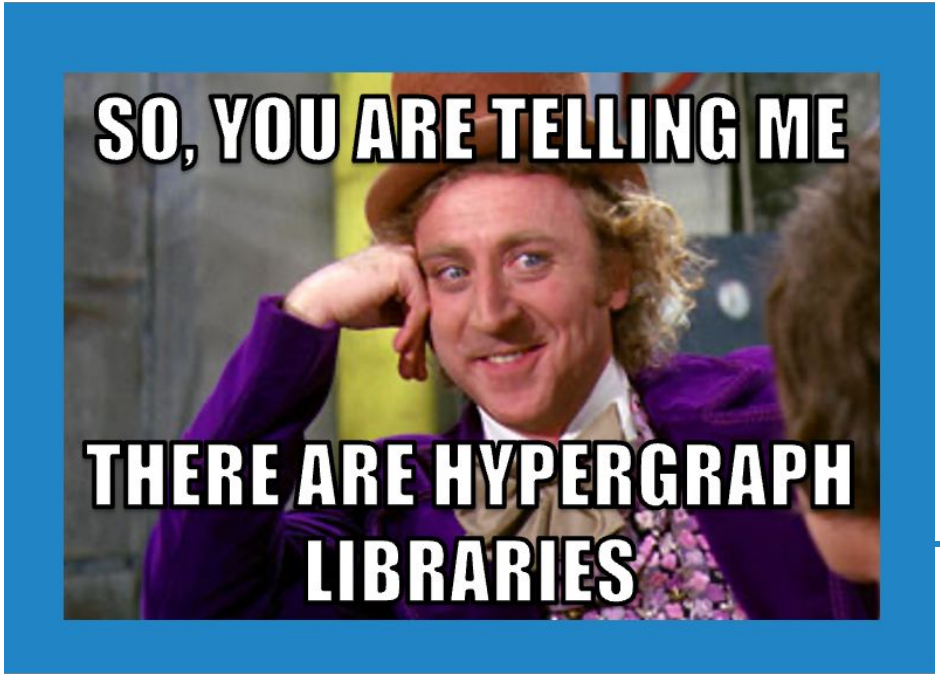
Bipartite graphs

- Vertices do not interact directly anymore.

The why of hypergraph-specific tools

- Hypergraph to graph transformations represent a **trade-off** between **computability** and **accuracy**
- An increasing number of systematic studies demonstrate why one should prefer hypergraphs over graphs
  - Clearly, **in presence of high-order relationships**!

Alessia Antelmi, Gennaro Cordasco, Mirko Polato, Vittorio Scarano, Carmine Spagnuolo, and Dingqi Yang. *A Survey on Hypergraph Representation Learning*. ACM Comput. Surv. 56, 1, Article 24, 2024,



Coding hypergraphs

Existing hypergraph software libraries

- Currently, we count **13 general-purpose** hypergraph software libraries
- Specifically designed to handle hypergraphs or expansion of existing graph libraries

Existing hypergraph software libraries

- Currently, we count **13 general-purpose** hypergraph software libraries
- Specifically designed to handle hypergraphs or expansion of existing graph libraries

- Programming language
  - Python
  - Julia
  - Chapel
  - Matlab
  - C/C++
  - Rust
  - R
  - JavaScript

Existing hypergraph software libraries

1 - Chapel HyperGraph Library

2 - Gspbox

3 - Halp

4 - Hygra

5 - Hypergraph

6 - HyperGraphLib

7 - hypergraphx

8 - HyperNetX

9 - HyperX

10 - Iper

11 - NetworkR

12 - Multihypergraph

13 - SimpleHypergraphs.jl

Existing hypergraph software libraries

1 - Chapel HyperGraph Library

2 - Gspbox

3 - Halp

4 - Hygra

5 - Hypergraph

6 - HyperGraphLib

7 - **hypergraphx**

8 - **HyperNetX**

9 - HyperX

10 - Iper

11 - NetworkR

12 - Multihypergraph

13 - **SimpleHypergraphs.jl**

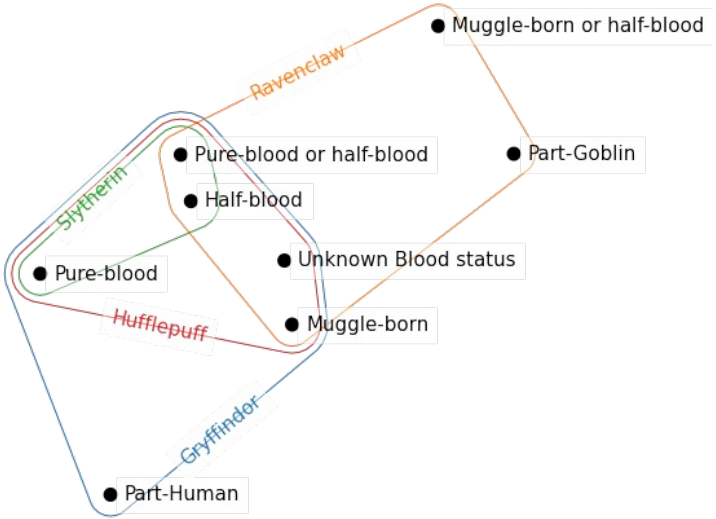
HyperNetX

- Python package to model, analyze, and visualize hypergraphs
- Developed by the Pacific Northwest National Laboratory since 2018
- Publicly available on a GitHub repository
  - https://github.com/pnnl/HyperNetX

C.A. Joslyn, S. Aksoy, D. Arendt, L. Jenkins, B. Praggastis, E. Purvine, and M. Zalewski. *Hypergraph analytics of domain name system relationships*. In Proceedings of Algorithms and Models for the Web Graph - 17th International Workshop (WAW'20), volume 12091 of Lecture Notes in Computer Science, pages 1–15. Springer, 2020.

HyperNetX

- Generalization of traditional graph metrics to hypergraphs
- Hypergraph-specific algorithms
- Visualization functionalities
- Add-on for providing optimized C++ implementations



S. G. Aksoy, C. Joslyn, C. Ortiz Marrero, B. Praggastis, and E. Purvine. *Hypernetwork science via high-order hypergraph walks.* EPJ Data Science, 9(1):16, 2020.

hypergraphx

- Python package to build, visualize, and analyze hypergraphs
- Joint project by University of Trento and Central European University
- Publicly available on a GitHub repository
  - https://github.com/HGX-Team/hypergraphx

Quintino Francesco Lotito, Martina Contisciani, Caterina De Bacco, Leonardo Di Gaetano, Luca Gallo, Alberto Montresor, Federico Musciotto, Nicolò Ruggeri, Federico Battiston, *Hypergraphx: a library for higher-order network analysis*, Journal of Complex Networks, Volume 11, Issue 3, June 2023, cnad019, https://doi.org/10.1093/comnet/cnad019

hypergraphx

- Different hypergraph representations
- Basic node and hyperedge statistics
- Centrality measures
- Motifs
- Mesoscale structures (e.g., communities)
- Filters
- Generative models
- Dynamical processes
- Visualization