Higher-Order Networks and Motif Analysis in Hypergraphs

A hitchhiker's guide to networks beyond pairwise interactions and their local structure

Quintino Francesco Lotito

University of Trento, Italy

quintino.lotito@unitn.it

@FraLotito

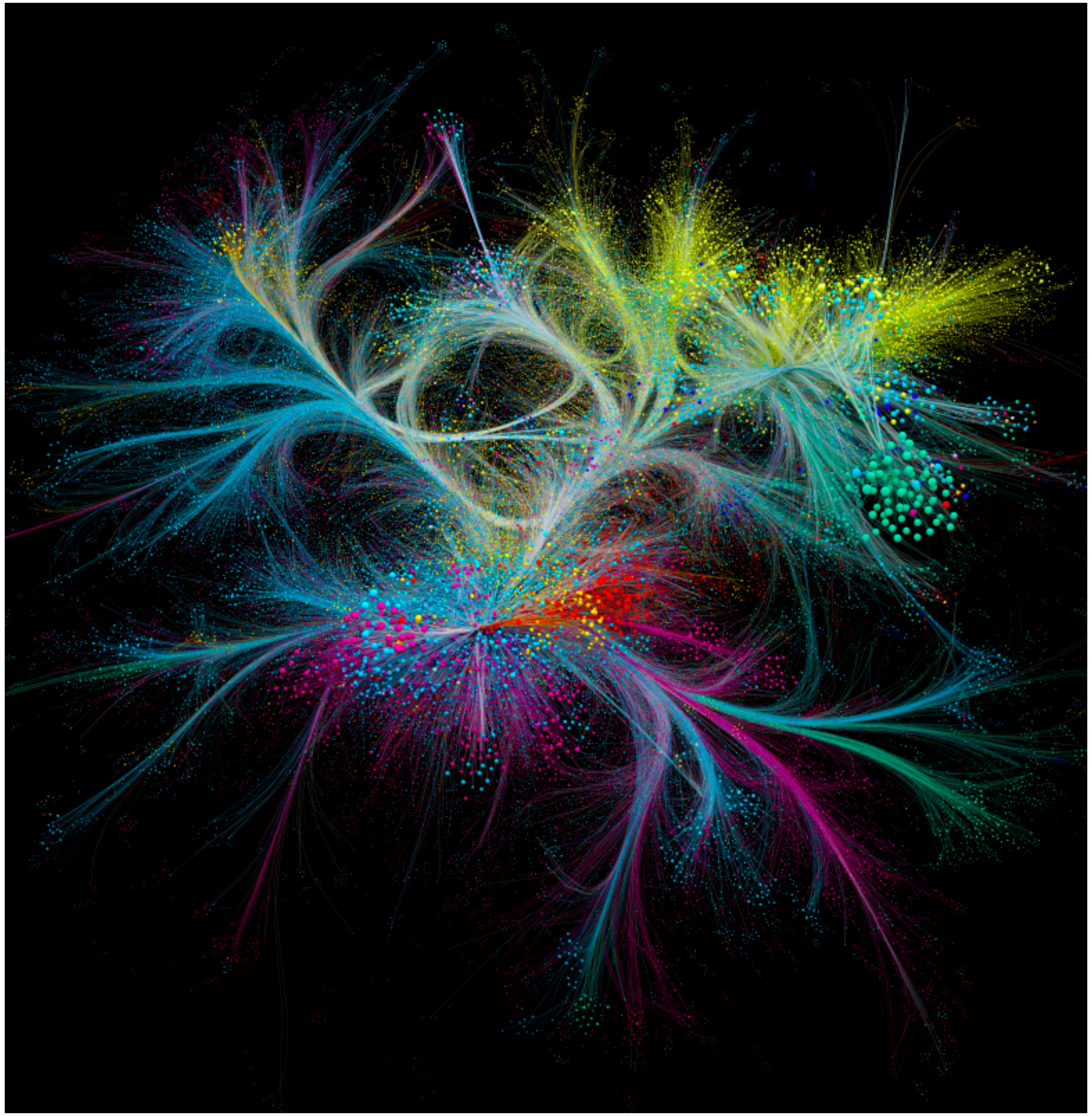
SML Journal Club

25/03/2022

Networks 101

Networks are all around us

- Networks model systems of *interacting* unities
- (Online) Social networks: whatever people do with other people (conversations, proximity, email exchange, being Facebook friends, retweets, …)
- Biological networks: protein-protein interactions, neural networks, …
- Information networks: WWW, co-authorship, co-citation, …
- **These systems are everywhere, and we want to understand them!**

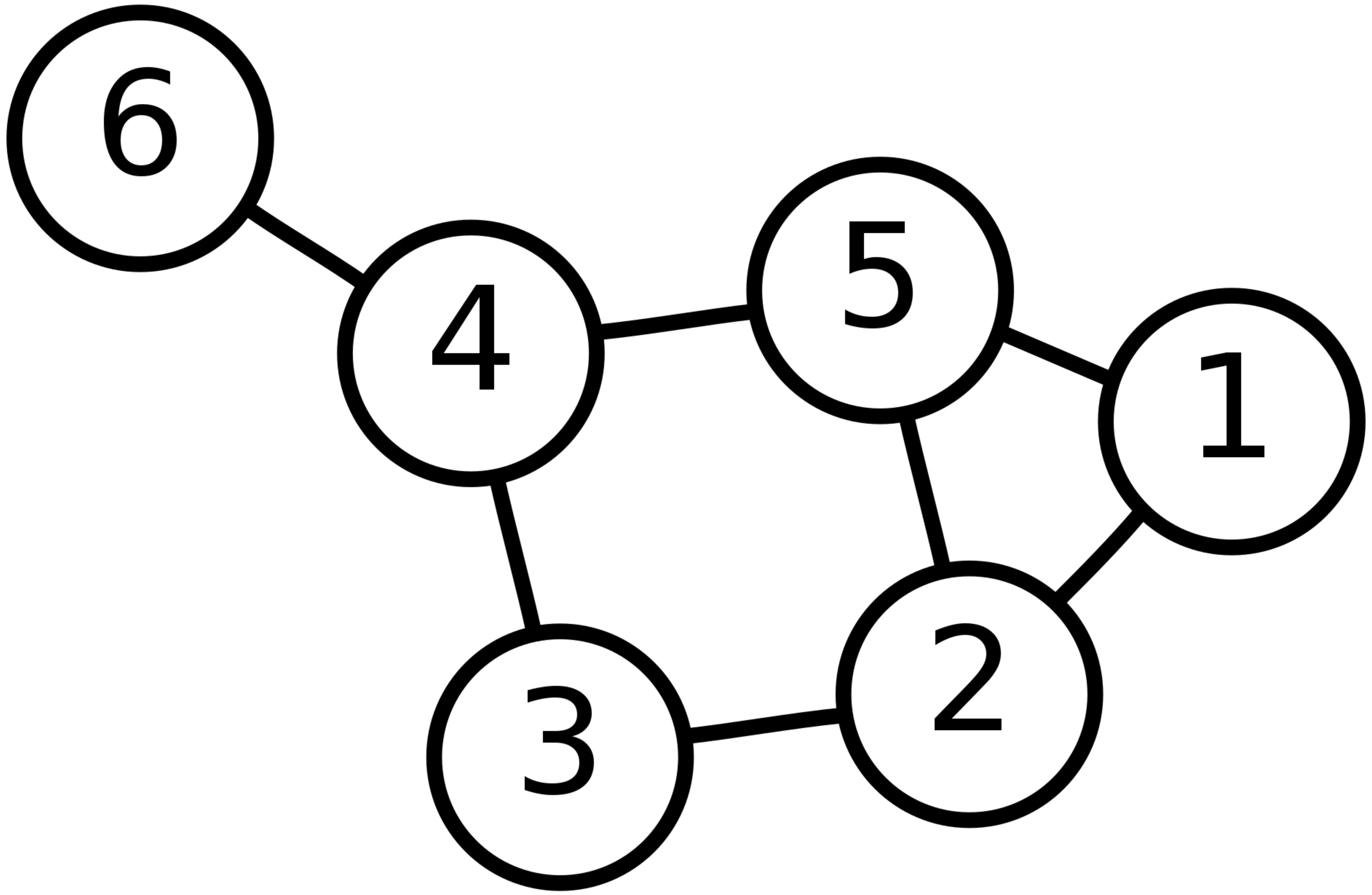


© BarabásiLab (Albert-László Barabási, Alexander Gates, Alice Grishchenko, Qing Ke, Mauro Martino, Onur Varol)

Networks 101

The mathematics of networks

- The mathematical backbone of networks are the graphs
- A graph is just a collection of dots (***the nodes***) connected by lines (***the edges***)
- So, the unities of a system are represented by the nodes of a graph, and the interactions are represented by the edges
- A lot of cool mathematics has been developed to handle graphs (Graph Theory), so we have lots of tools to study networks! 🤔

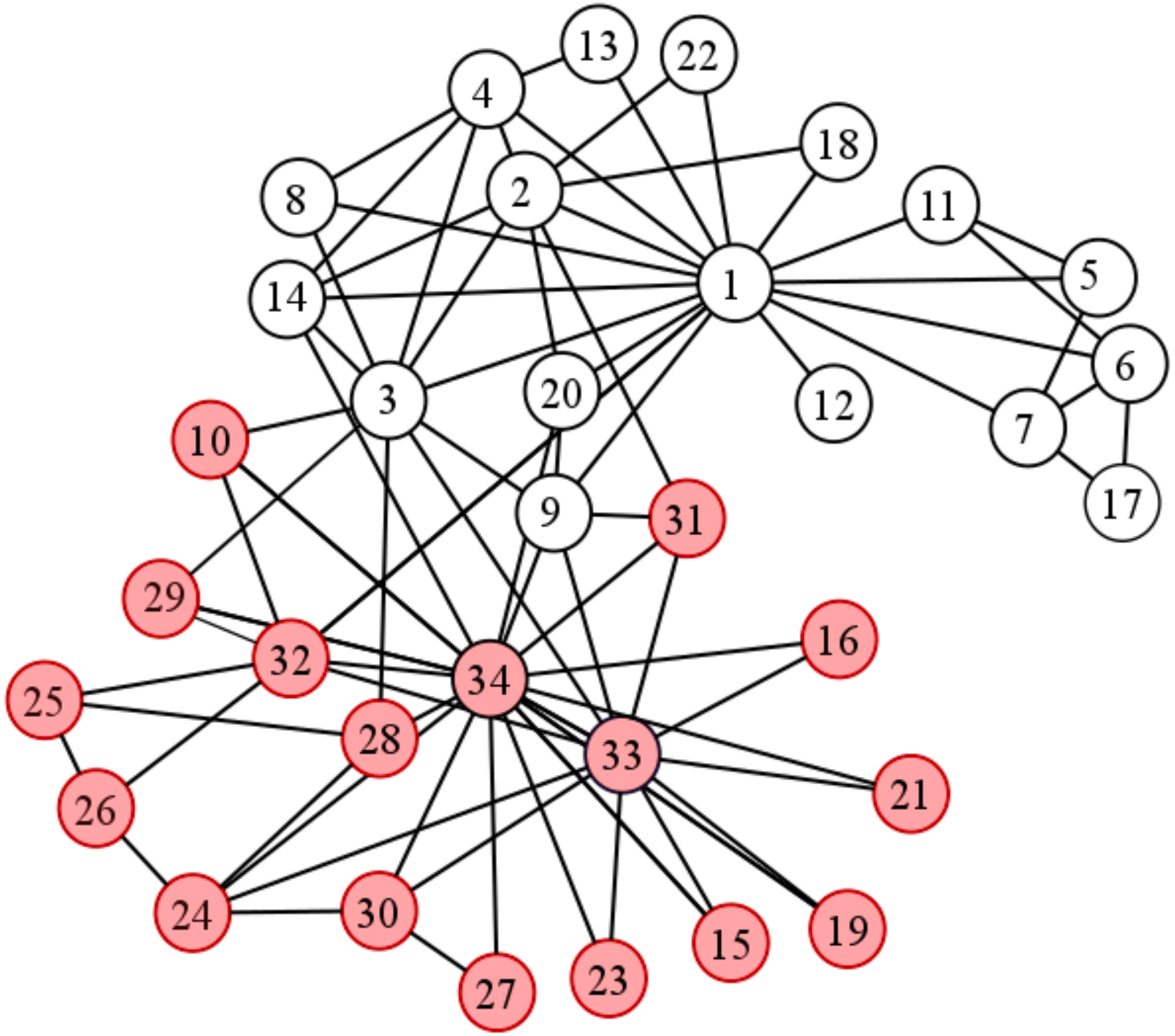


From Wikipedia "Graph Theory"

Networks 101

What can we do

- **Ranking:** what are the most important things in a network (The "important" nodes, frequent subgraphs)
- **Clustering:** what are the groups of related nodes (People with similar characteristics in a social network)
- **Dynamics:** how things "move" over a network (Spreading of diseases)



Social network of friendships between 34 members of a karate club in the 1970s
From Wikipedia

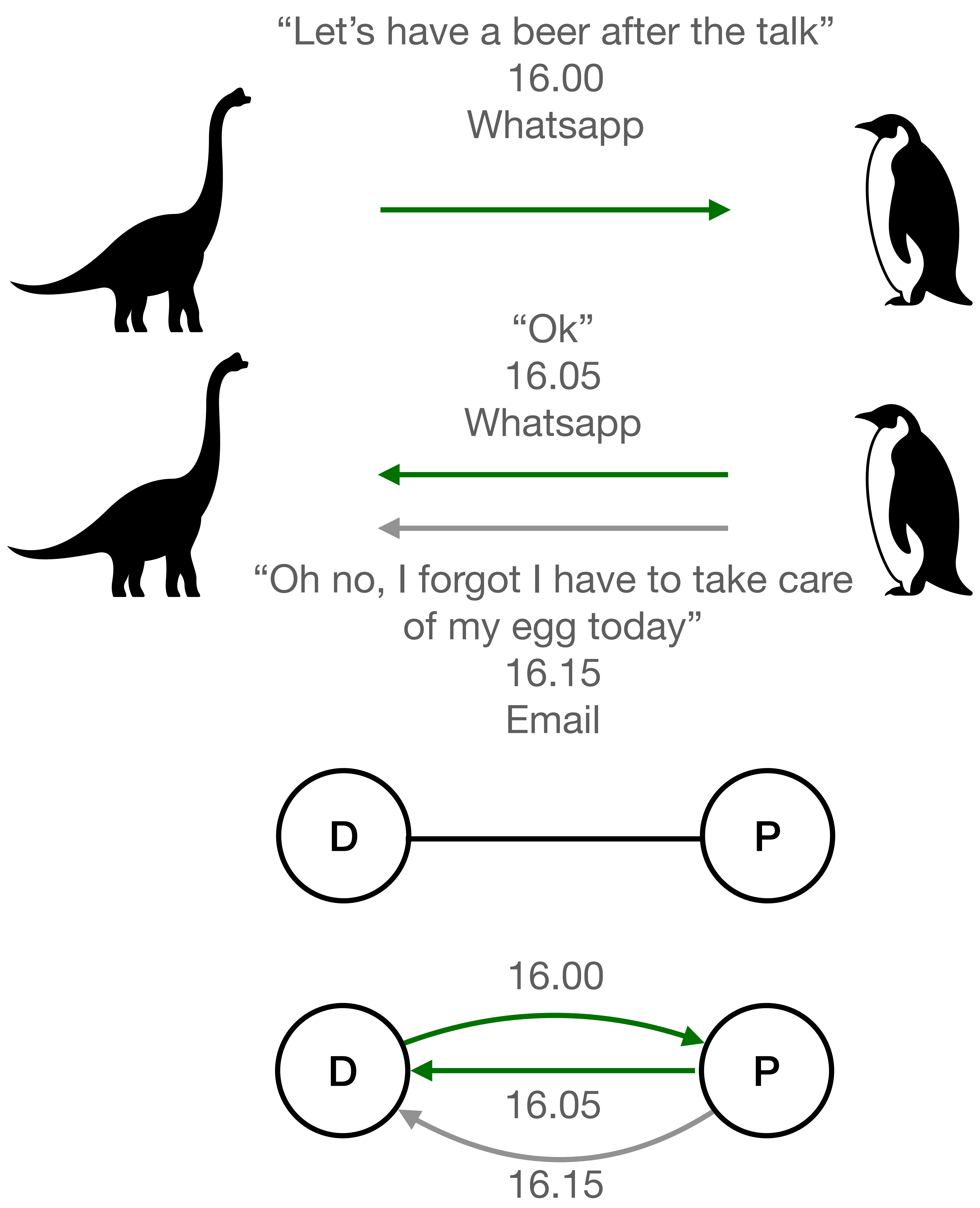

The power of the sun in the palm of my hand

But nature is not that simple

Nature is not that simple

Tuning the interactions 💪

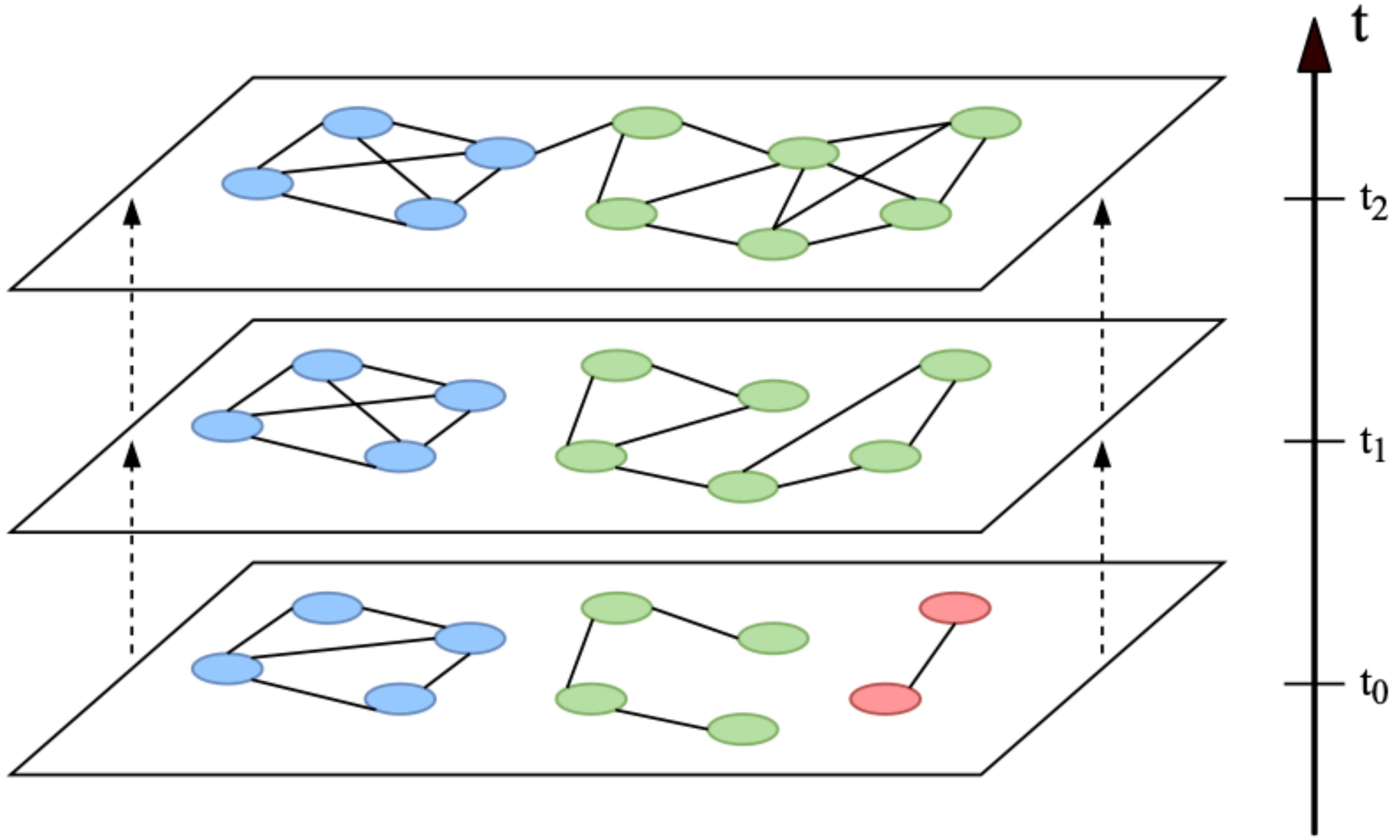
- There are situations in which “plain” graphs are a poor representation of a system
- Graphs model only simple interactions between two nodes (edges are just lines, right?)
- What if we care about the **direction** of an interaction? Or the **intensity**? Or we are interested in modeling **when** an interaction occurred? Or in which **modality**?



Generalized networks

New mathematical structures

- **Directed graphs:** edges are allowed to have a direction
- **Weighted graphs:** edges are allowed to have a length or a weight
- **Temporal graphs:** edges are allowed to have temporal labels
- **Multigraphs:** edges are allowed to have the same end nodes



A temporal graph with time evolving communities

Quintino F. Lotito and Alberto Montresor. 2020. Efficient Algorithms to Mine Maximal Span-Trusses From Temporal Graphs. In MLG '20: 16th Inter- national Workshop on Mining and Learning With Graphs, 2020

Generalized networks

What can we do

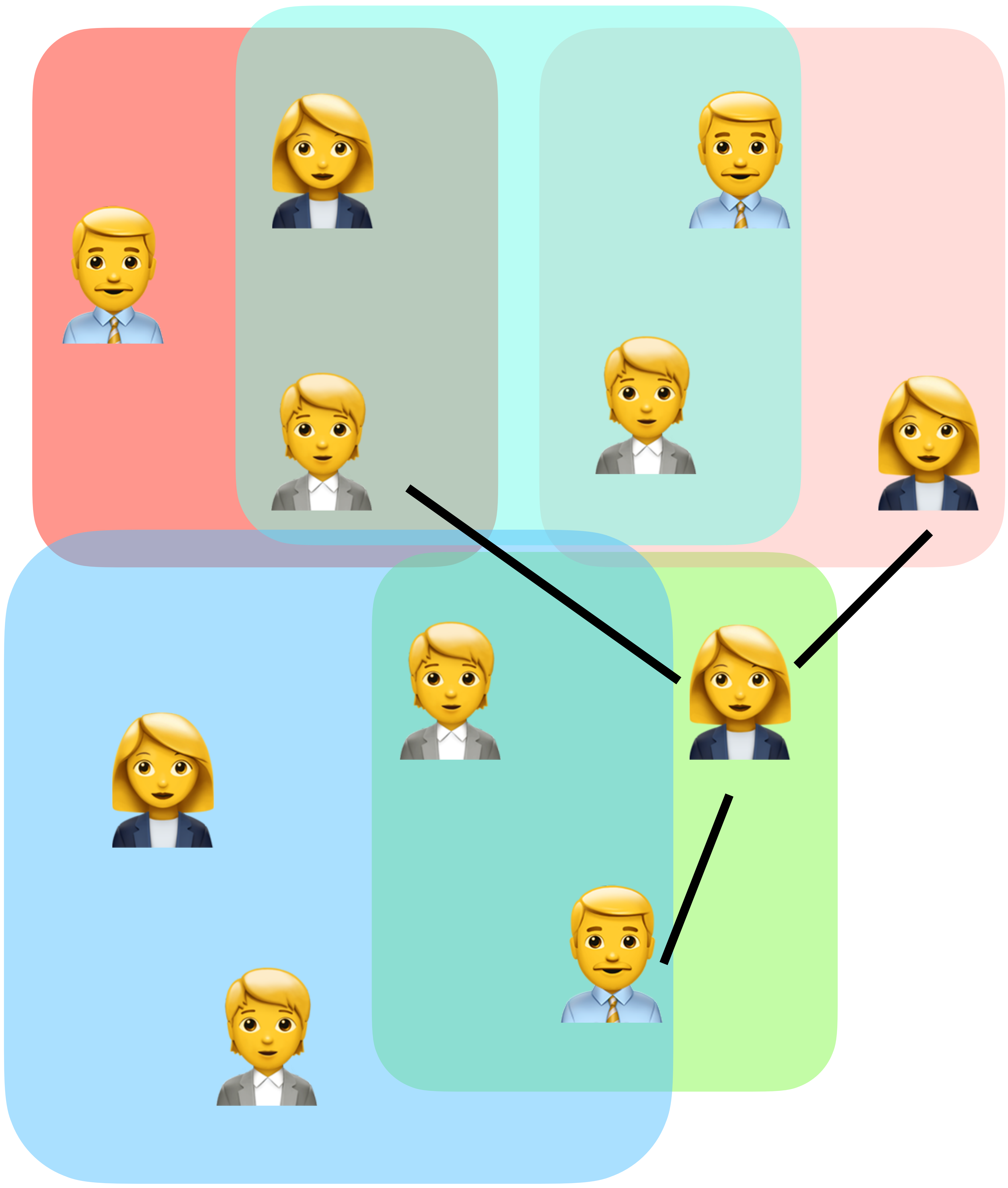
- Generalization of traditional network notions (centralities, communities, …)
- New notions (temporal patterns in the evolution of a network, edge overlap in multilayer networks, …)
- We gain expressive power 👍
- We add complexity to the system… and our algorithms / frameworks should cope with this (can be hard) 👎

What about group interactions?

Beyond pairwise interactions

Group interactions are everywhere too

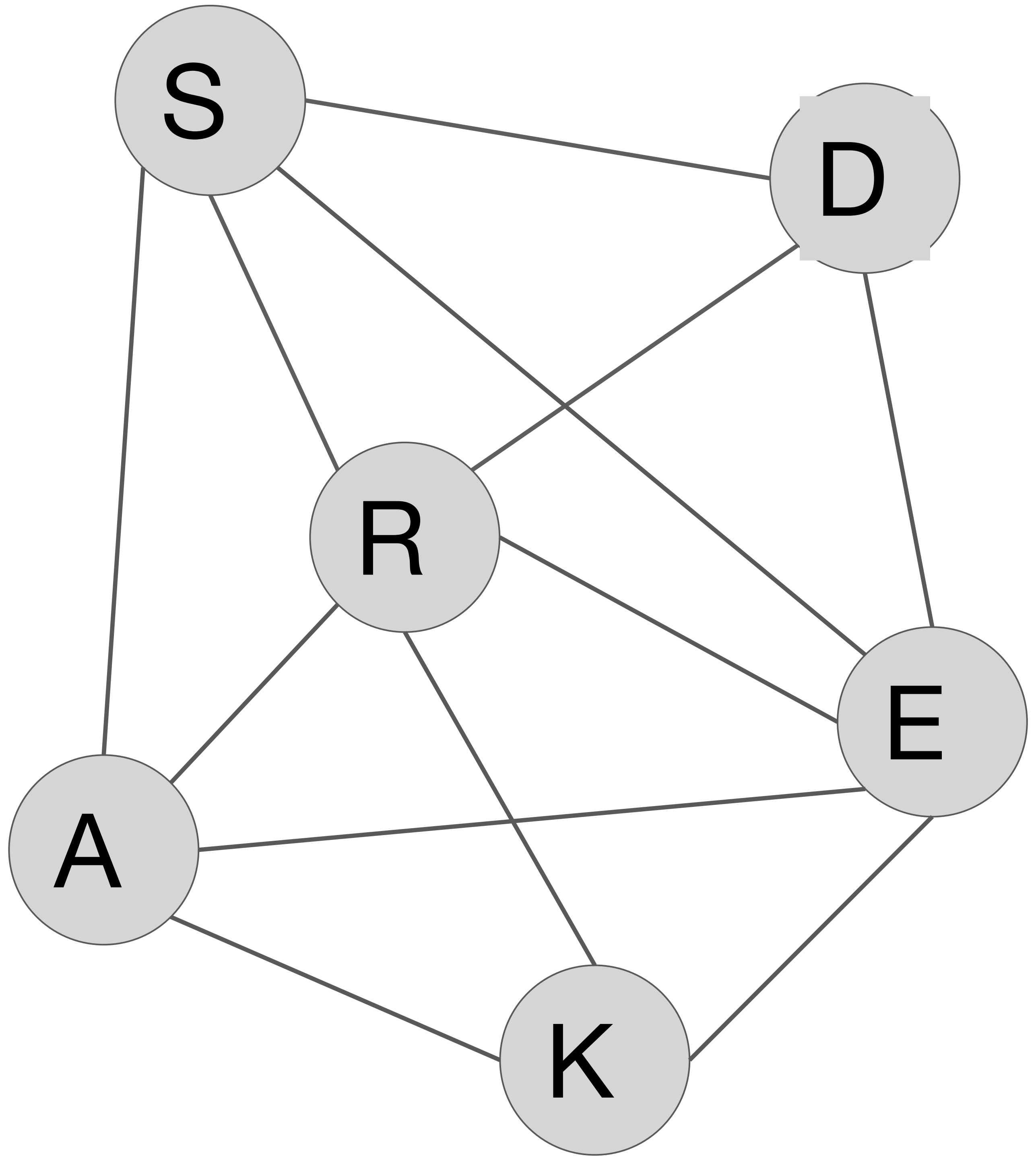
- Co-authorship networks
- Affiliation networks
- Prey / predator networks
- Complexes of proteins



Let’s talk about parties

Graph modeling

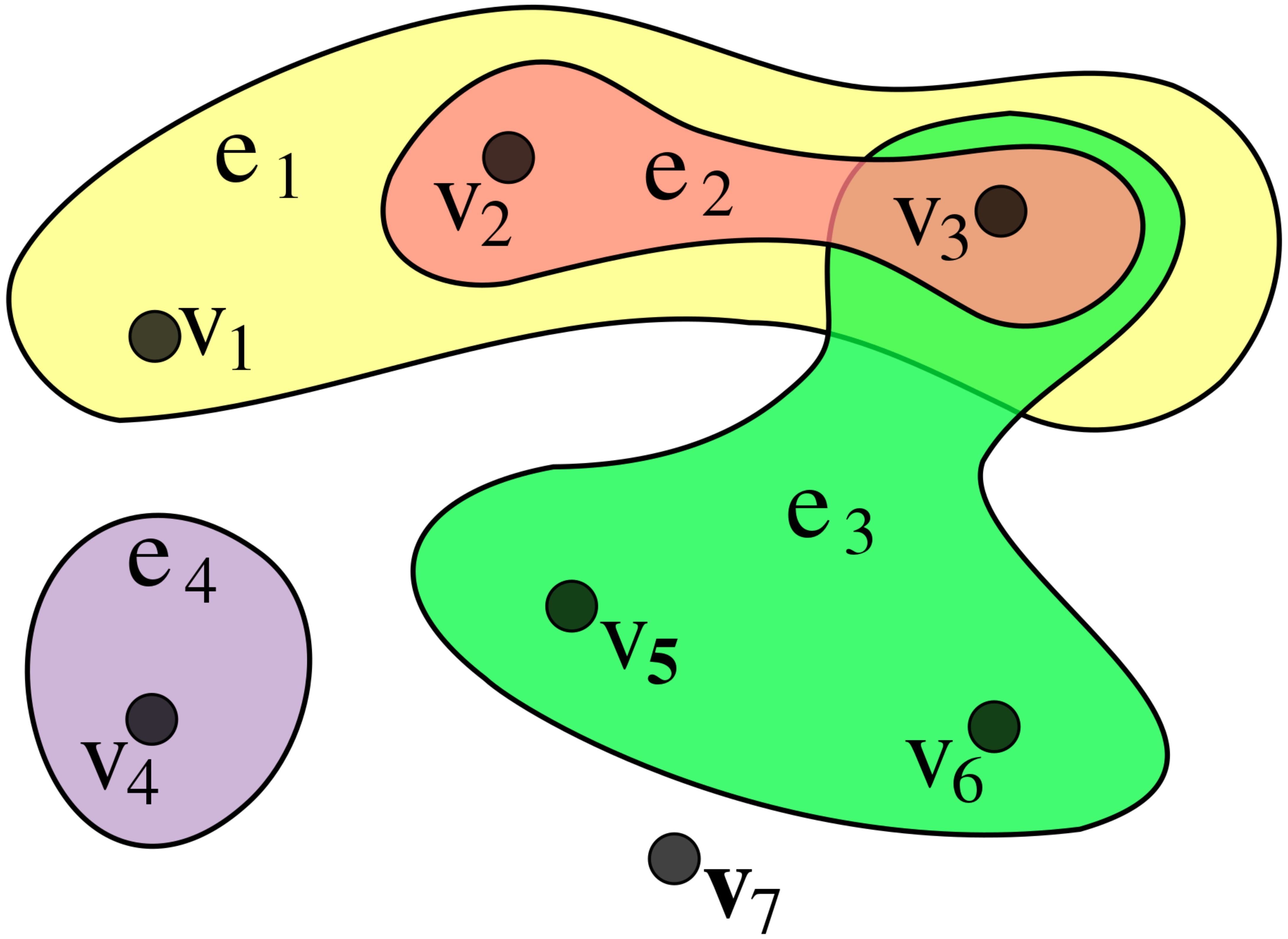
- Suppose there have been three parties this week
    - Sarah, Ross and Allison went to Party 1
    - Sarah, Ross, Drew and Eliot went to Party 2
    - Ross, Eliot, Keith and Allison went to Party 3
- Let’s model this with a graph
    - We can use “Went to party together” as a relation



Battiston et al., Networks beyond pairwise interactions: Structure and dynamics, Physics Reports, Volume 874, 2020.

The mathematics of group interactions
Hypergraphs

- Hypergraphs are generalizations of graphs
- In an hypergraph, any hyperedge can link any number of nodes
- Basically, each hyperedge is a non-empty set of nodes
- Graphs are just hypergraphs in which every hyperedge has cardinality equal to 2



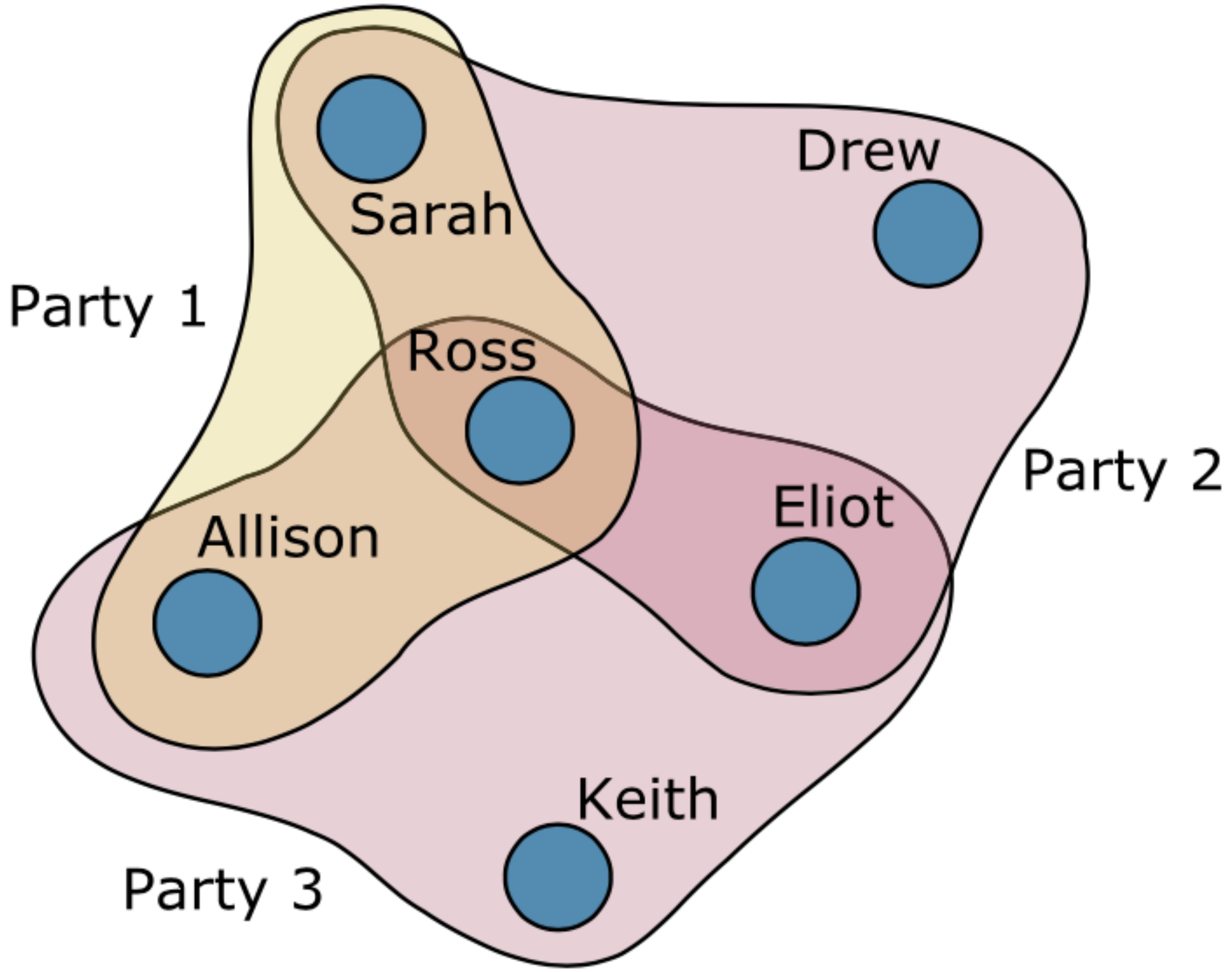
From Wikipedia “Hypergraph”


THAT'S SUSPICIOUS...

Let’s talk about parties… again

Hypergraph modeling

- Ok maybe we can model the party thing with hypergraphs, right?
- We have more expressive power
- We do not lose information

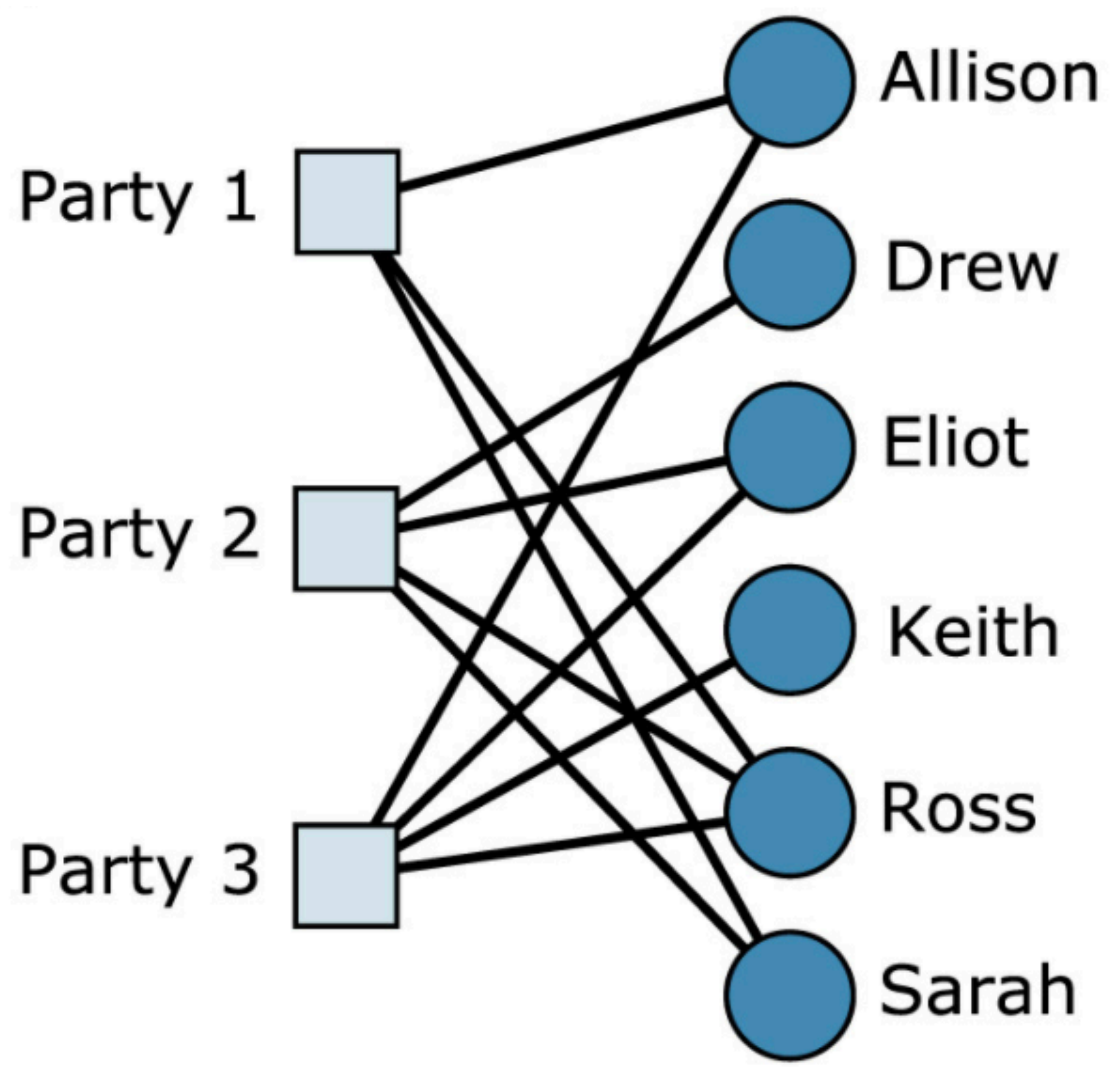


Battiston et al., Networks beyond pairwise interactions: Structure and dynamics, Physics Reports, Volume 874, 2020.

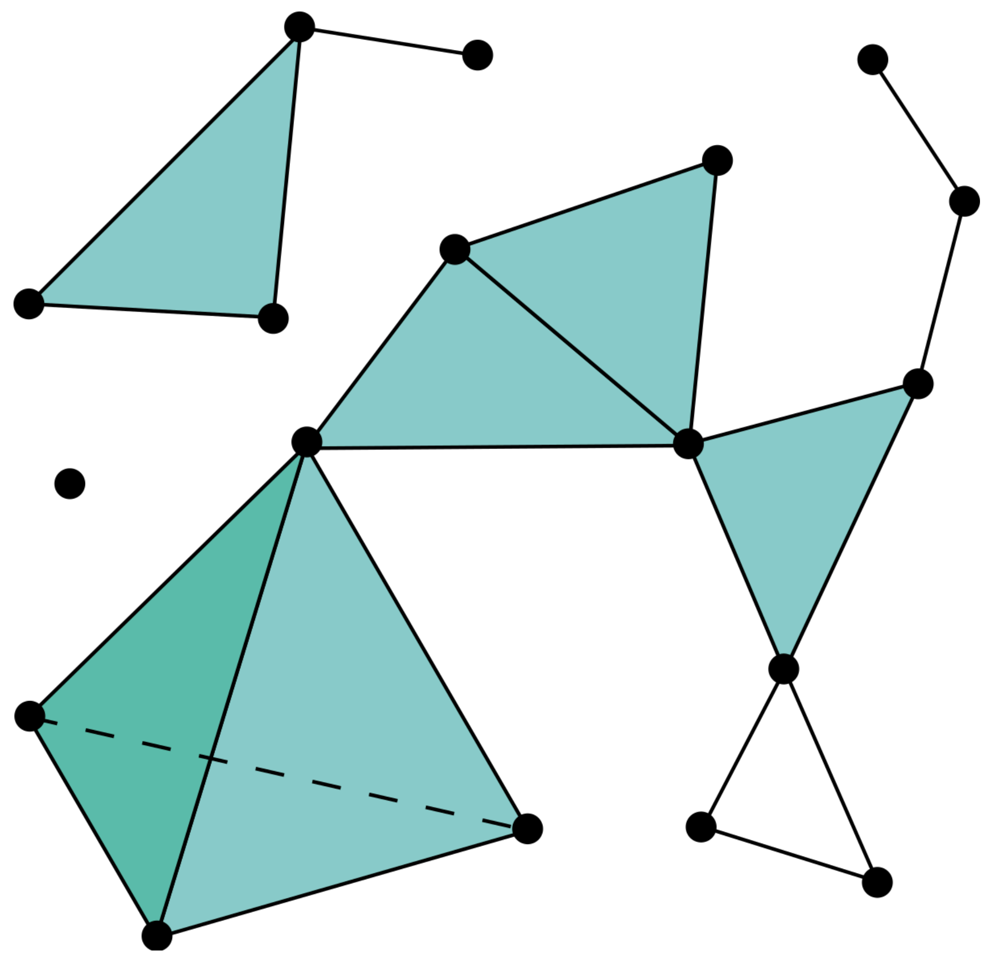
Representing higher-order networks

A tough decision 🤔

- Hypergraphs are not the only way to encode group interactions
- Bipartite graphs
    - “Graph” representation of a hypergraph… but nodes do not directly interact, and can have different meaning
- Simplicial complexes
    - Very established mathematical framework, lots of tools from algebraic topology… but we need to assume closure with respect to faces



Battiston et al., Networks beyond pairwise interactions: Structure and dynamics, Physics Reports, Volume 874, 2020.



From Wikipedia “Simplicial complexes”

Research opportunities & drawbacks

Oh, life is good, but never easy

- Really hot field!
- There are lots of interesting problems from application domains
- Need for generalization of traditional methods, and for new tools, algorithms and frameworks to handle the expressive power of hypergraphs
- Completely new problems available: for example, recovering group interactions from data that is stored naturally in a pairwise fashion (e.g., bluetooth proximity)
- Lack of theoretical tools and softwares with respect to “classic” network science
- Algorithm design tend to become really hard when dealing with hypergraphs!

Key takeaways

Things to remember about (higher-order) networks

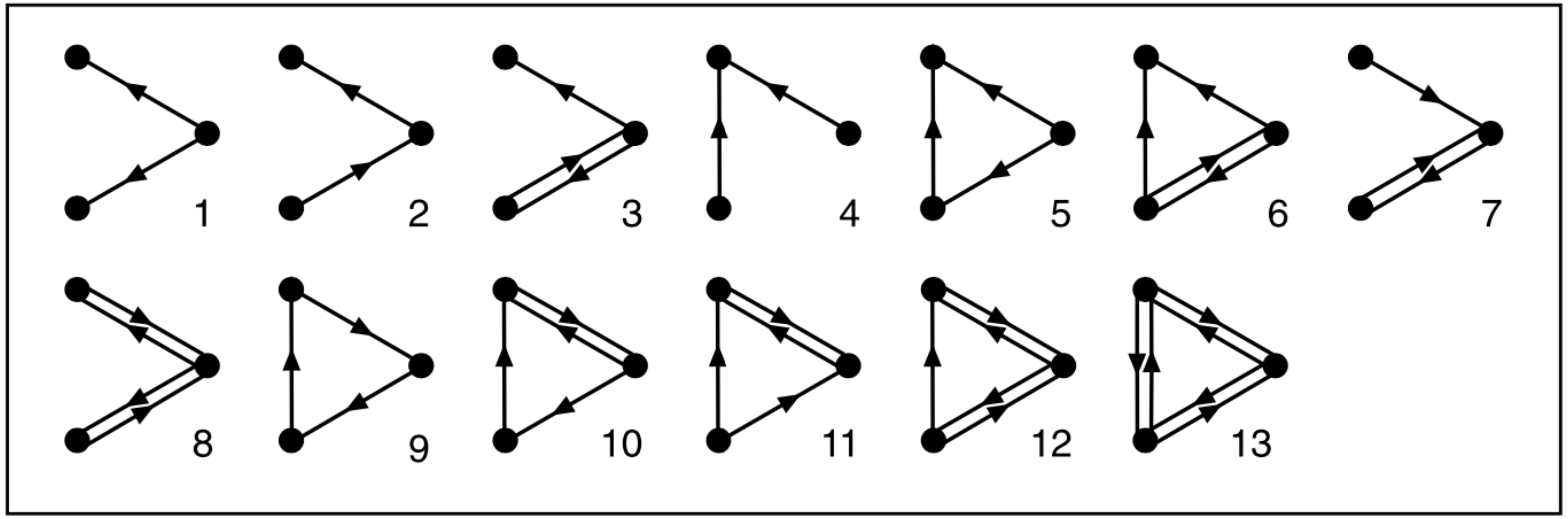
- Networks are a fundamental tool to model and analyze real-world systems of interacting unities
  - Relations are modeled by static dyadic links, and we have countless tools to study them
- But nature is not that simple..
  - Directed, weighted, temporal, multilayer networks have been proposed
- … and it is not even limited to only pairwise interactions
  - Interactions in group are very common: prey / predator networks, co-authorship networks, complexes of proteins, affiliation networks
- We use hypergraphs to naturally model group interactions, we gain expressive power but we need to handle more complexity

So, you were talking about that local structure thing…

Network motifs

The building blocks of networks

- Network motifs are **small patterns** of interactions that **appear** in an **observed network** at a frequency that is **statistically-significantly higher** than in **randomized networks**
- They characterize the local structure of networks

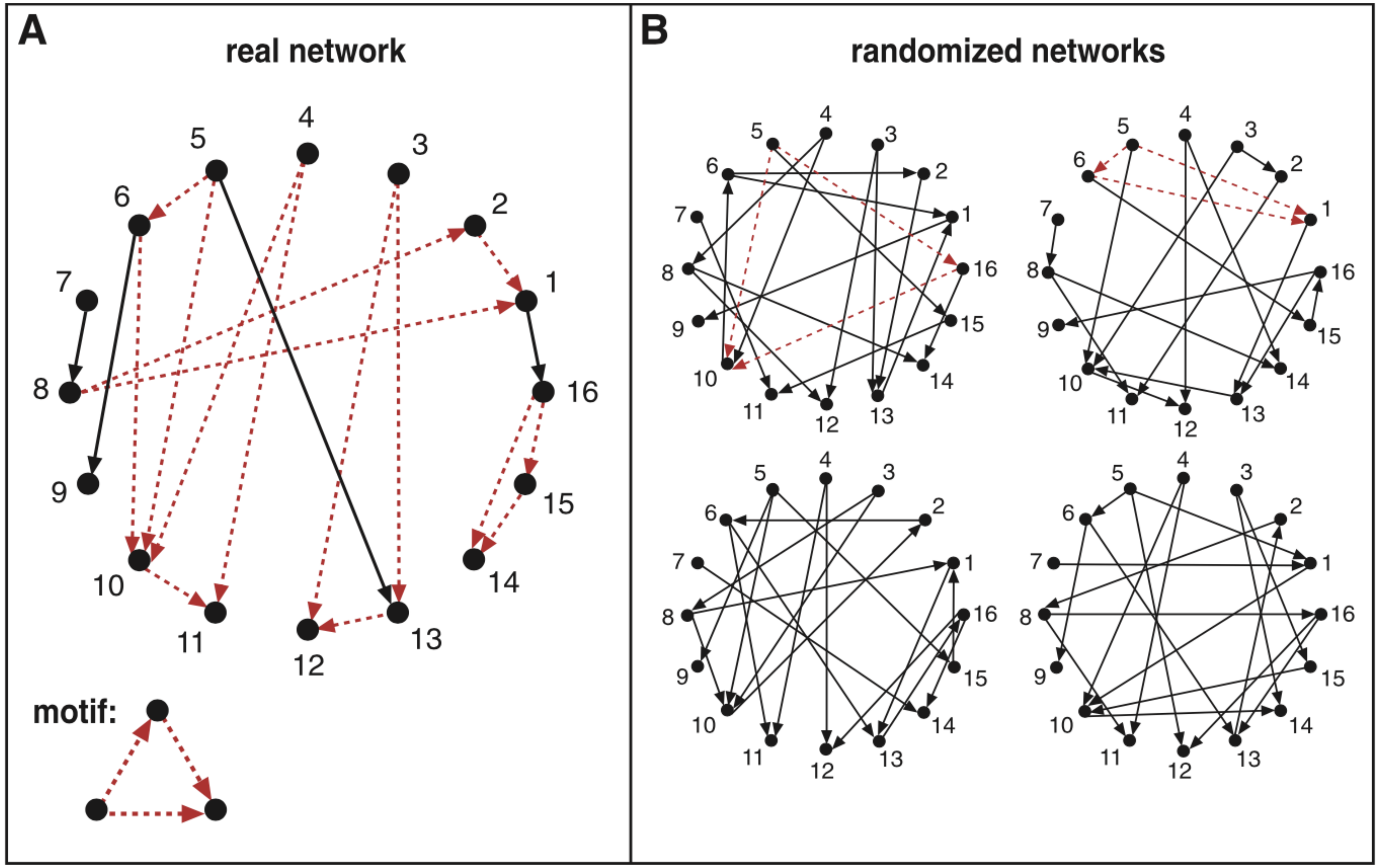


Milo et al., Network motifs: simple building blocks of complex networks. Science. 2002 Oct 25;298(5594):824-7.

Network motifs

The building blocks of networks

- How do we compute frequencies?
- How do we generate random networks?
- How do we perform statistical evaluation?



Milo et al., Network motifs: simple building blocks of complex networks. Science. 2002 Oct 25;298(5594):824-7.

Network motifs

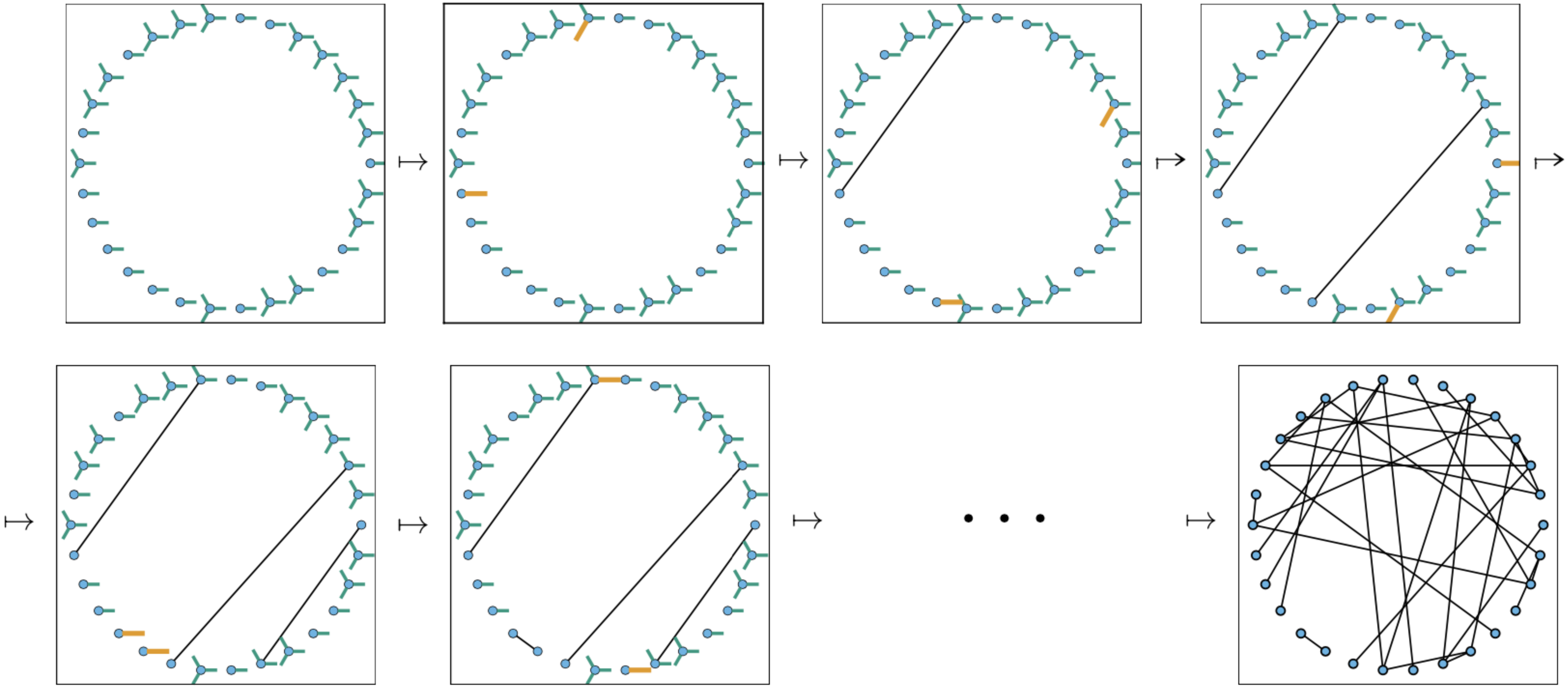
How do we compute frequencies?

- We need algorithms to count the frequency of each motif
- Difficult problem
  ➡ Equivalent to subgraph isomorphism (NP-complete)
- Huge history of proposals in literature
  - Exact counting algorithms
  - Sampling algorithms

Classic trade-off in algorithm design!

Network motifs

How do we generate random networks?



Miller, Joel & Volz, Erik. Incorporating Disease and Population Structure into Models of SIR Disease in Contact Networks. PloS one. 8. e69162. (2013).

Network motifs

How do we perform statistical evaluation?

- Relative abundance $\Delta_i = \dfrac{N_{\text{real}i} - \langle N_{\text{rand}i} \rangle}{N_{\text{real}i} + \langle N_{\text{rand}i} \rangle}$

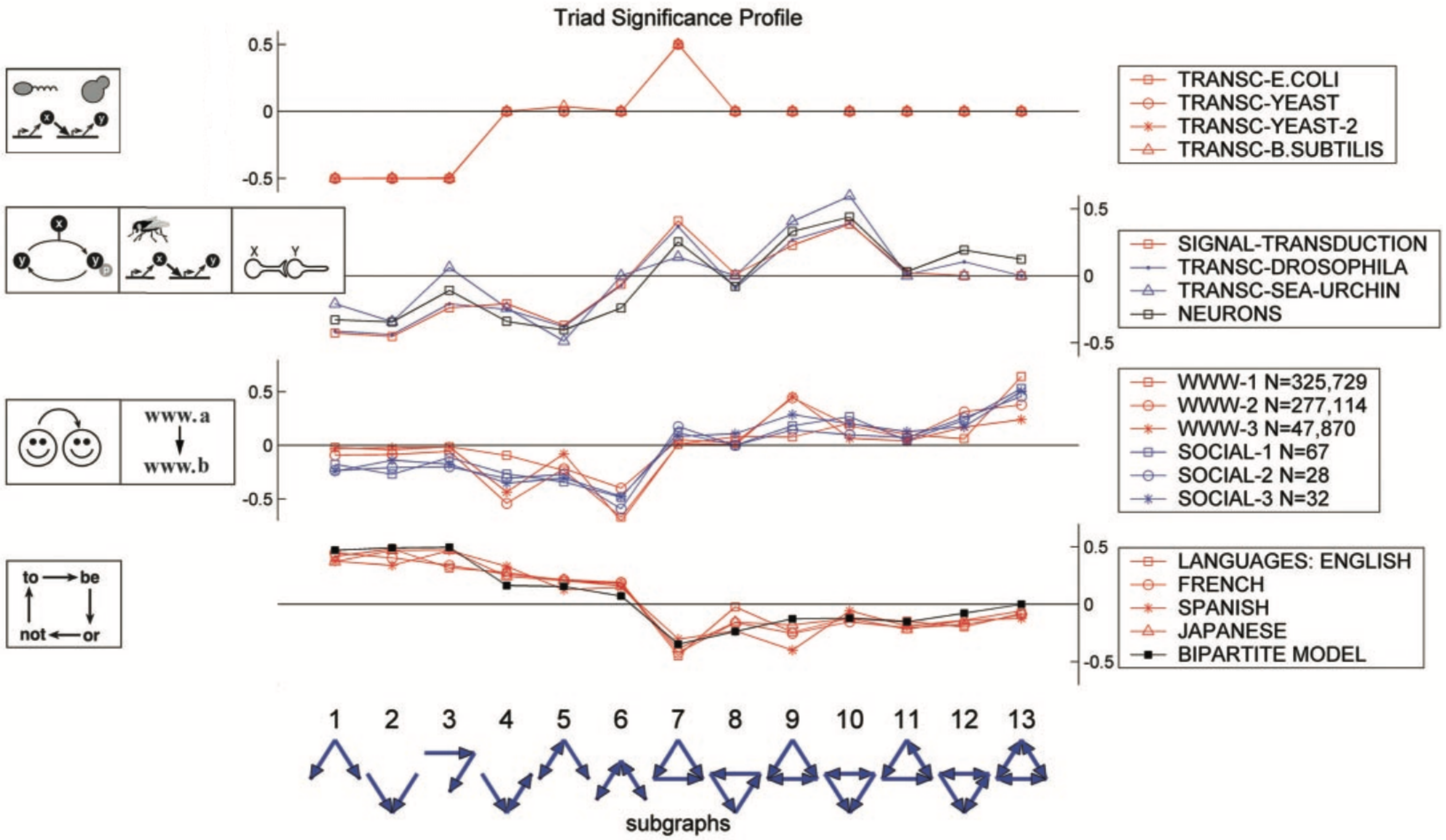
Difference over sum of the observed frequency of motif i wrt the average frequency of motif i in the samples from the configuration model

- Significance profile $\text{SP}_i = \dfrac{\Delta_i}{||\Delta||}$

Normalized vector of the Δ_is

Network motifs

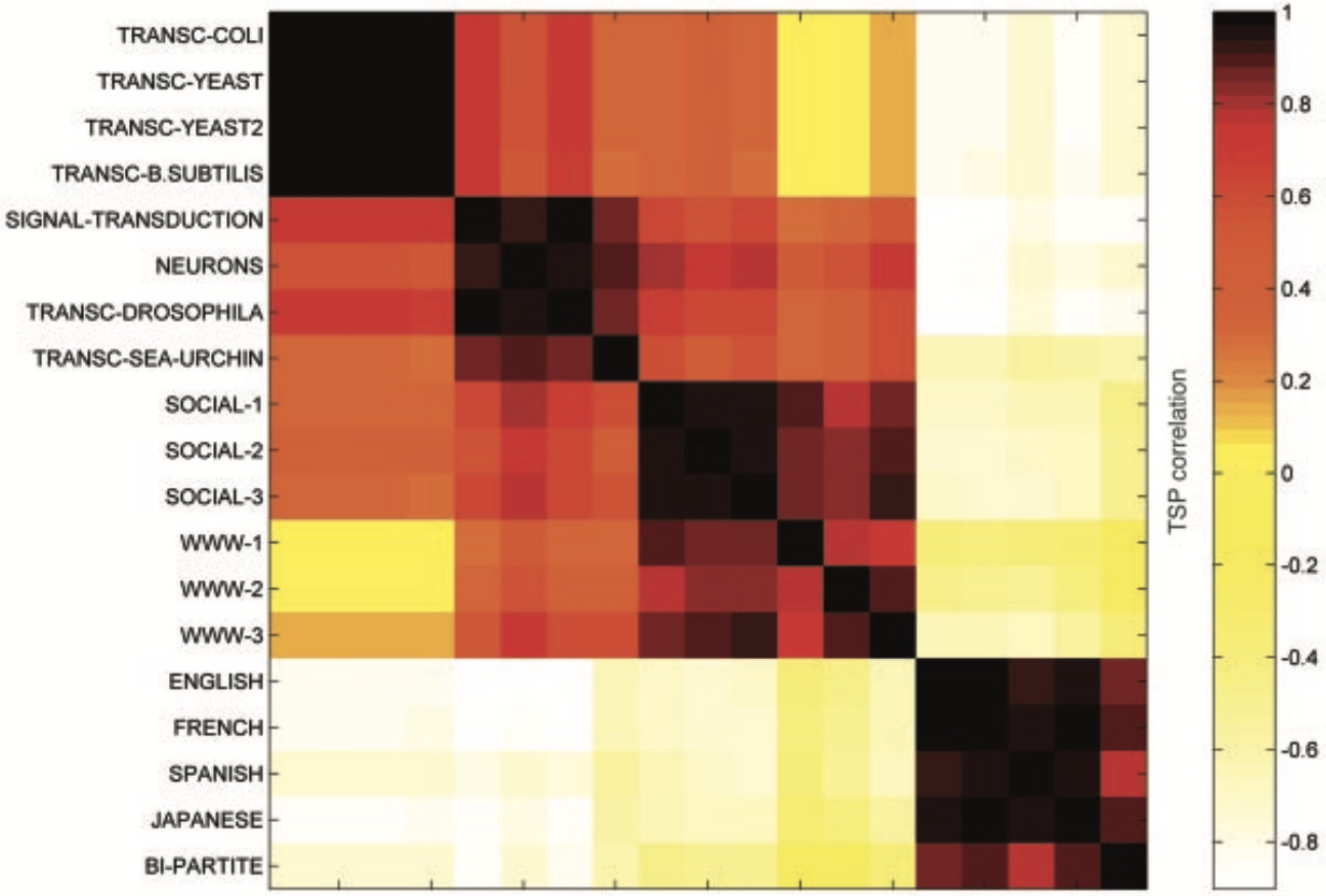
Significance profiles of real-world networks



Milo et al., “Superfamilies of evolved and designed networks.” Science (New York, N.Y.)vol. 303,5663 (2004): 1538-42.

Network motifs

“Superfamilies” of networks



Milo et al., “Superfamilies of evolved and designed networks.” Science (New York, N.Y.)vol. 303,5663 (2004): 1538-42.

Applications of network motifs

So… why should we care about network motifs?

- **Biological networks:** analysis of transcriptomic networks
- **Brain networks:** analysis of the emergence of complex and flexible neural function by composition of fundamental circuits
- **Social networks:** fingerprint of interactions on online social networks
- Change in the distribution of motifs to identify cancer and financial crises

Network motifs on generalized networks

Only static and directed interactions?

- J. Onnela et al., *"Intensity and coherence of motifs in weighted complex networks"* Physical Review E 71.6 (2005)
- L. Kovanen et al., *"Temporal motifs in time-dependent networks"* Journal of Statistical Mechanics: Theory and Experiment 2011.11 (2011)
- A. Paranjape et al., *"Motifs in temporal networks"* Proceedings of the tenth ACM International Conference on Web Search and Data Mining (2017)
- F. Battiston et al., *"Multilayer motif analysis of brain networks"*, Chaos: An Interdisciplinary Journal of Nonlinear Science 27 (2017)

➡ **Q. Lotito** et al., *"Higher-order motif analysis in hypergraphs"*, in press at Communications Physics (2022) **[Joint work with F. Musciotto, A. Montresor, F. Battiston]**

From traditional to higher-order motifs

Going beyond pairwise interactions

Network motifs: small patterns of interactions that appear in an observed network at a frequency that is statistically-significantly higher than in a randomized network.

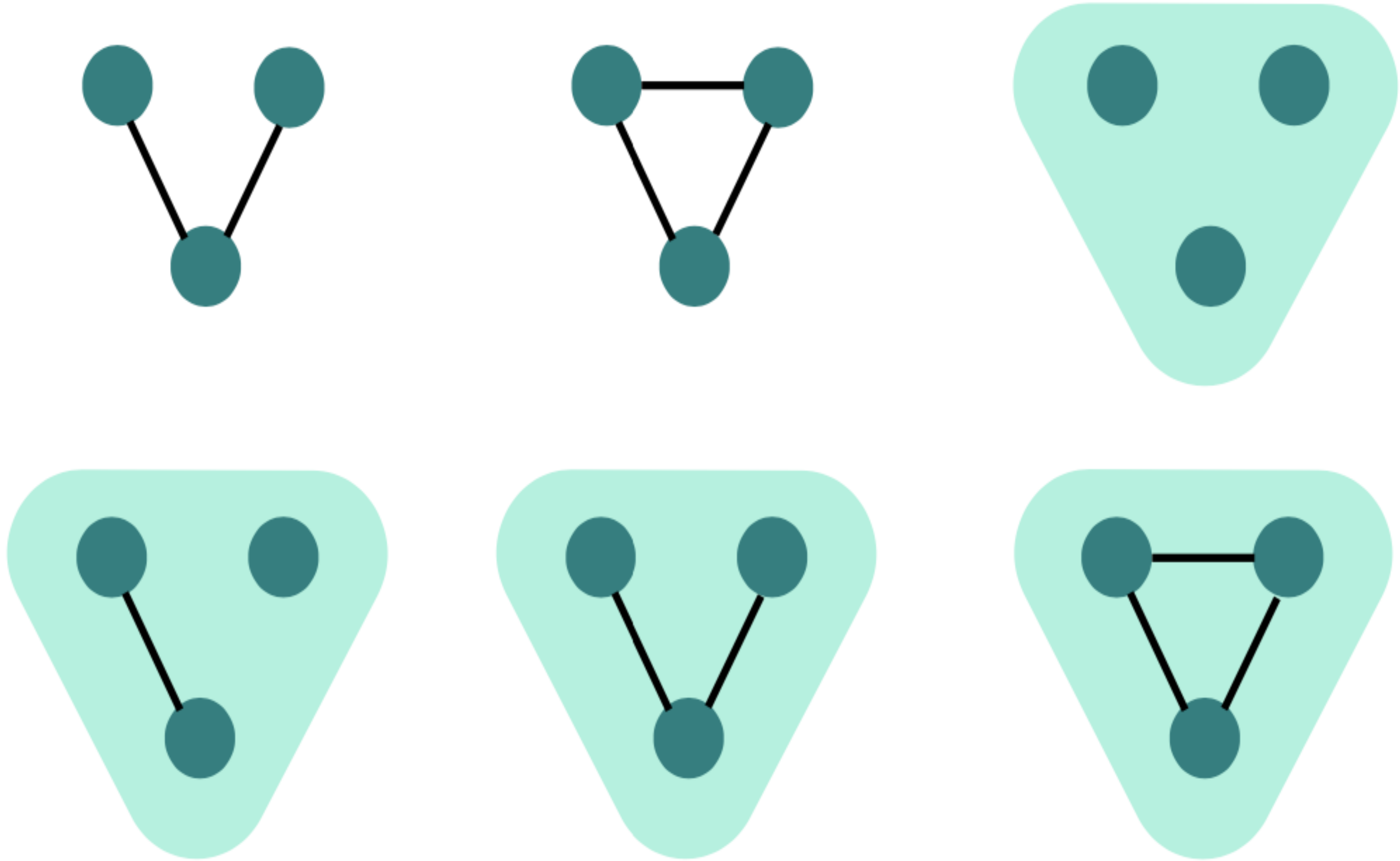
↓

Higher-order network motifs: small patterns of <u>higher-order interactions</u> that appear in an observed <u>hypergraph</u> at a frequency that is statistically-significantly higher than in a <u>randomized hypergraph.</u>

Combinatorics of higher-order motifs

Let's meet these new mathematical objects

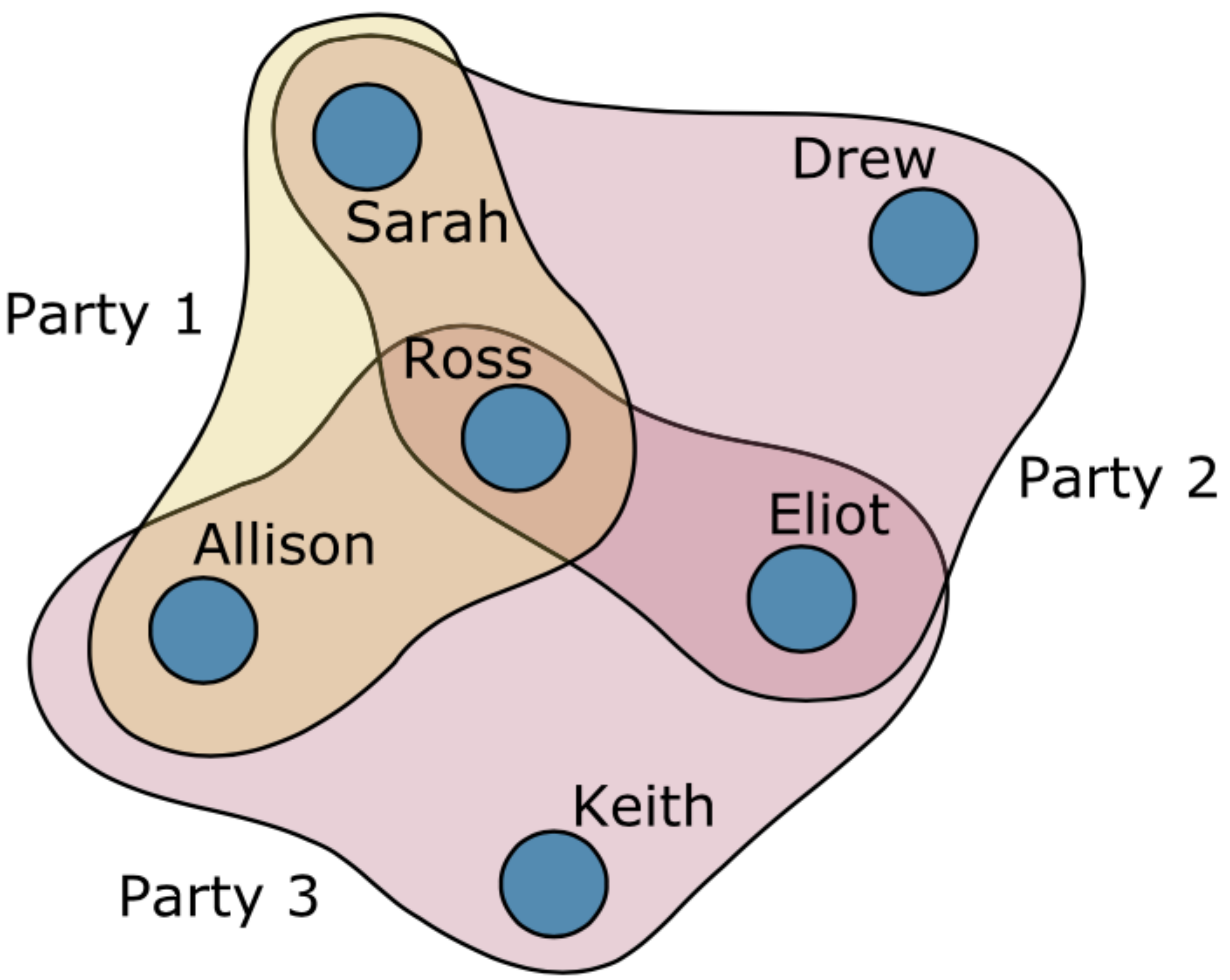
- There are 6 patterns of higher-order interactions involving 3 nodes.
- How many patterns involve 4 nodes?
  - A lot: 171.
- How many patterns involve 9 nodes?
  - More than the atoms in the observable universe



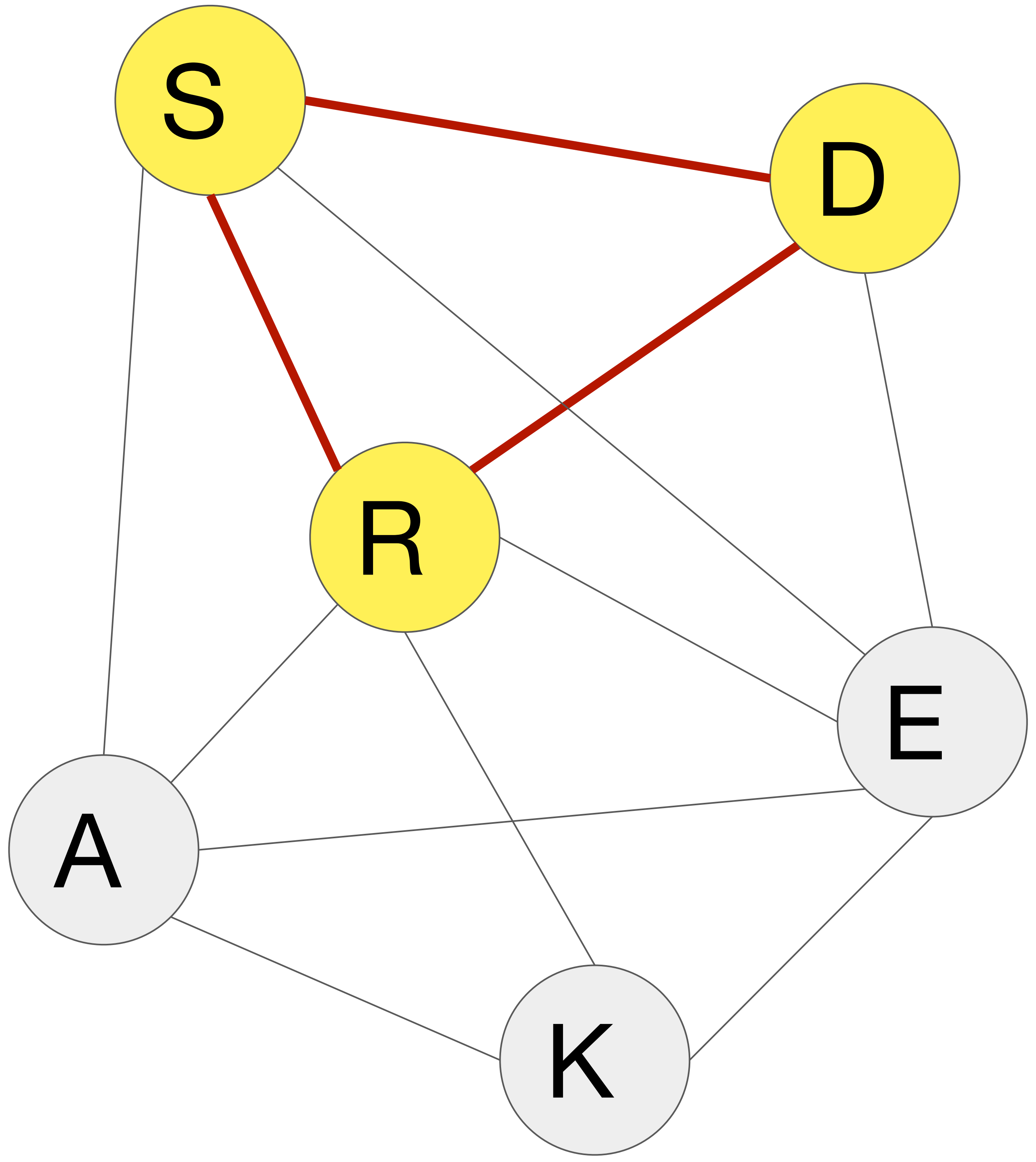
Q. Lotito et al., "Higher-order motif analysis in hypergraphs", in press at Communications Physics (2022)

Projection of a hypergraph

Brief observation 🧐



Projection →



Higher-order motif discovery - exact algorithm

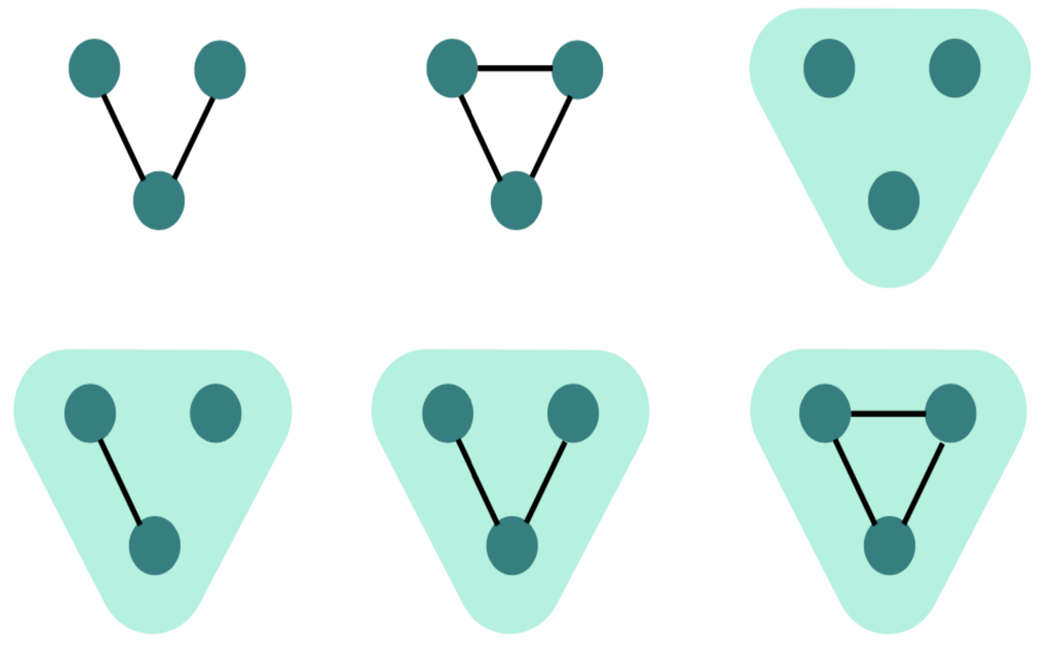
Baseline: project and compute

- Project hyperedges and exploit existing algorithms
- We care only about the vertices of the output subgraphs (candidates)
- In order to validate a candidate, we build the corresponding induced sub-hypergraph and check if it is connected
- We need an efficient way to update the count of the right motif! (Classes of isomorphism) Hint: precomputation, the subgraphs we consider are really small
- Advantages: it correctly performs higher-order motifs discovery, it relies on well-established algorithms
- Disadvantages: creates lots of edges and discards lots of subgraphs

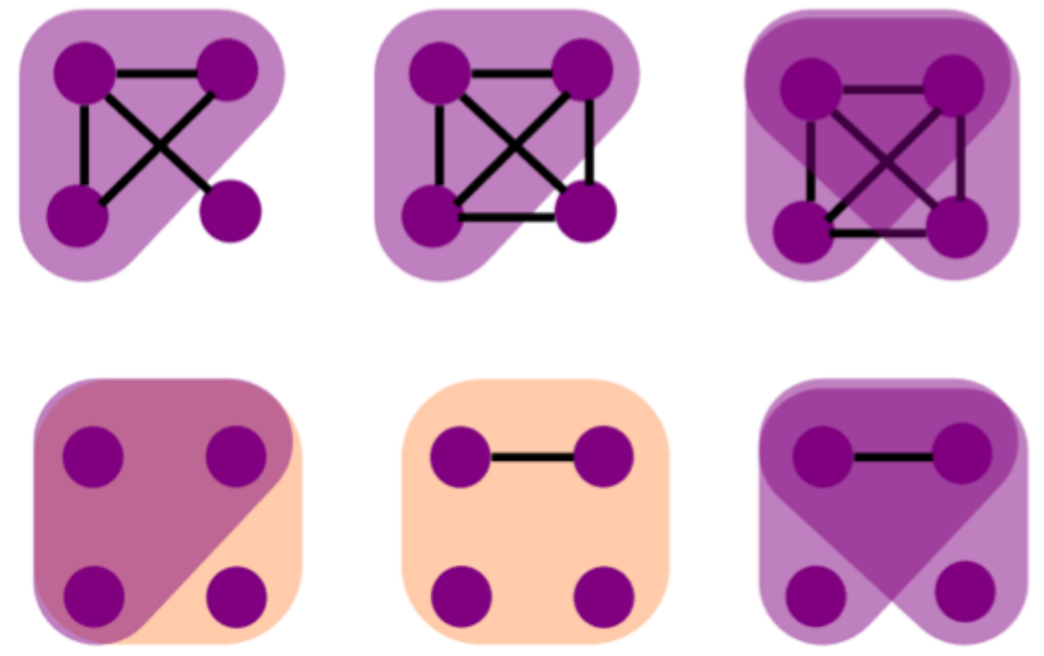
Higher-order motif discovery - exact algorithm

Exploiting higher-order structures to speed up the enumeration 🚀

- Hyperedges already define sets of connected nodes 🧐
- Higher-order motifs of order 3
  ➡ Iterate over all the hyperedges of size 3 and construct their nested pairwise structure (“fill in” the hyperedge), then run a classic algorithm for lower-order motifs (the ones which do not involve a group interaction)



- Higher-order motifs of order 4
  ➡ Iterate over all the hyperedges of size 4 and construct their nested pairwise structure, iterate over all the hyperedges of size 3 and add nodes from the neighborhood of the hyperedges, run classic algorithm for lower-order motifs (the ones which do not involve a group interaction).



Statistical evaluation

Remember the configuration model? Relative abundance?

- A generalization of the notion of the configuration model for hypergraphs has been proposed by P. Chodrow (*"Configuration Models of Random Hypergraphs"*, Journal of Complex Networks, 2019)

- We employ again the notions of relative abundance and significance profile

  - Relative abundance $\Delta_i = \dfrac{N_{\text{real}i} - \langle N_{\text{rand}i} \rangle}{N_{\text{real}i} + \langle N_{\text{rand}i} \rangle}$

  - Significance profile $\text{SP}_i = \dfrac{\Delta_i}{\|\Delta\|}$

Local structure of real-world hypergraphs - datasets

What about real-world networks with group interactions?

Naturally encoded

- Co-authorship data (PACS, DBLP, History, Geology)
- Biological data (gene-disease, NDC-classes, NDC- substances)
- Tech data of e-mails exchange (Enron, EU)

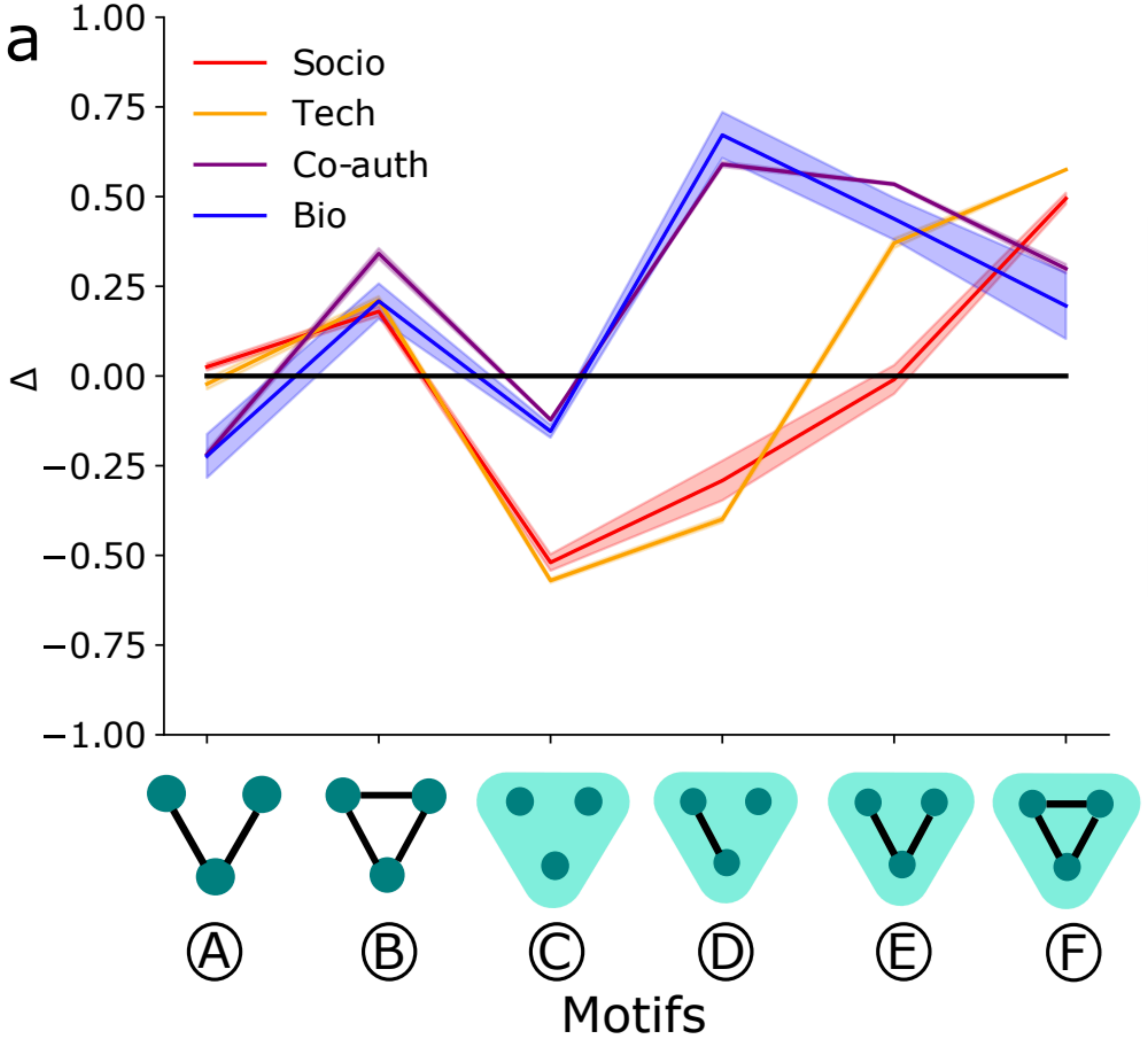
Inferred

- Face-to-face and proximity contacts of individuals and animals from the SocioPatterns repository (Baboons, Schools, Hospital, Conference, etc.)

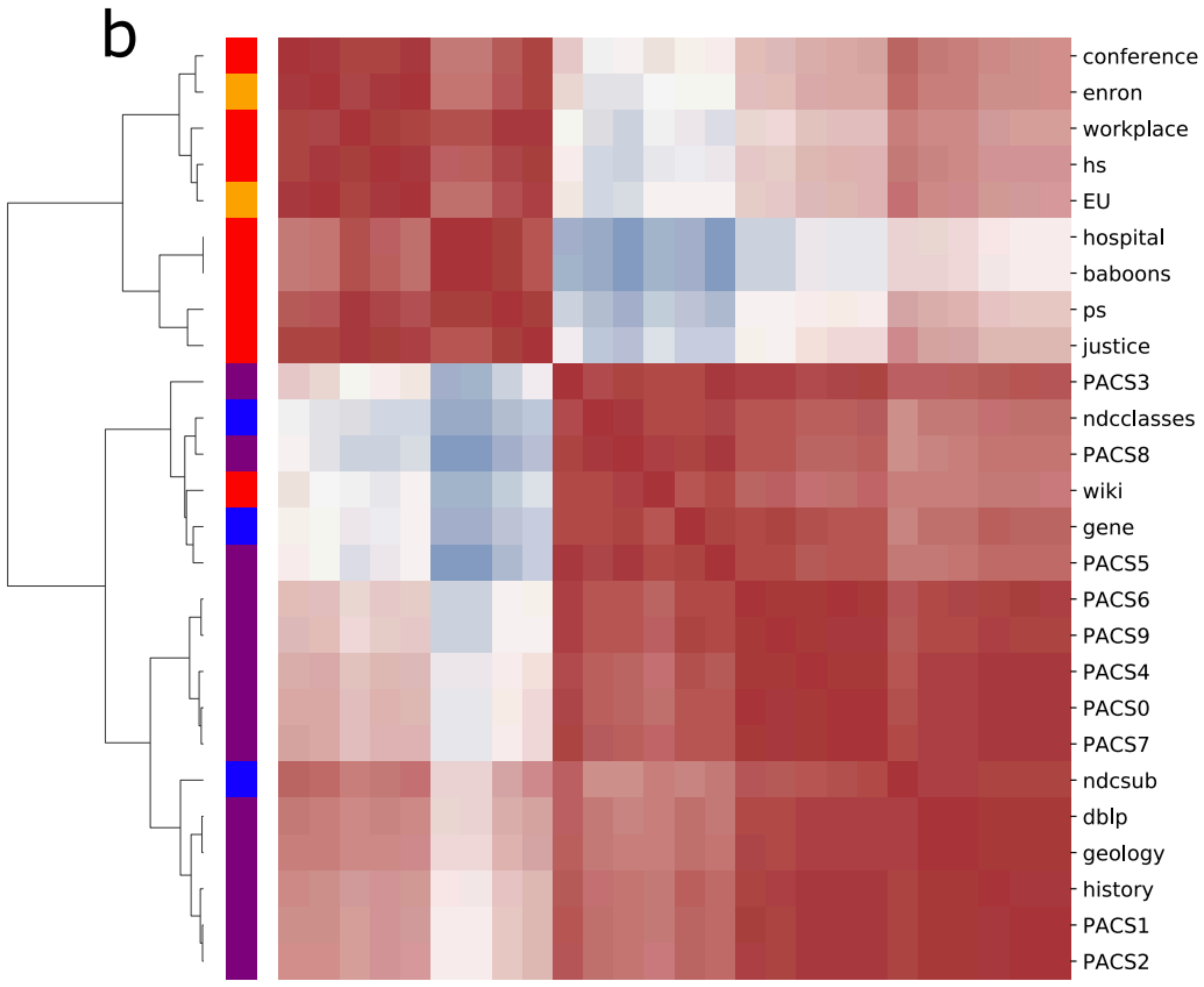
Local structure of real-world hypergraphs

Motifs of order 3

Significance profiles grouped by domain



Correlation matrix of the SPs

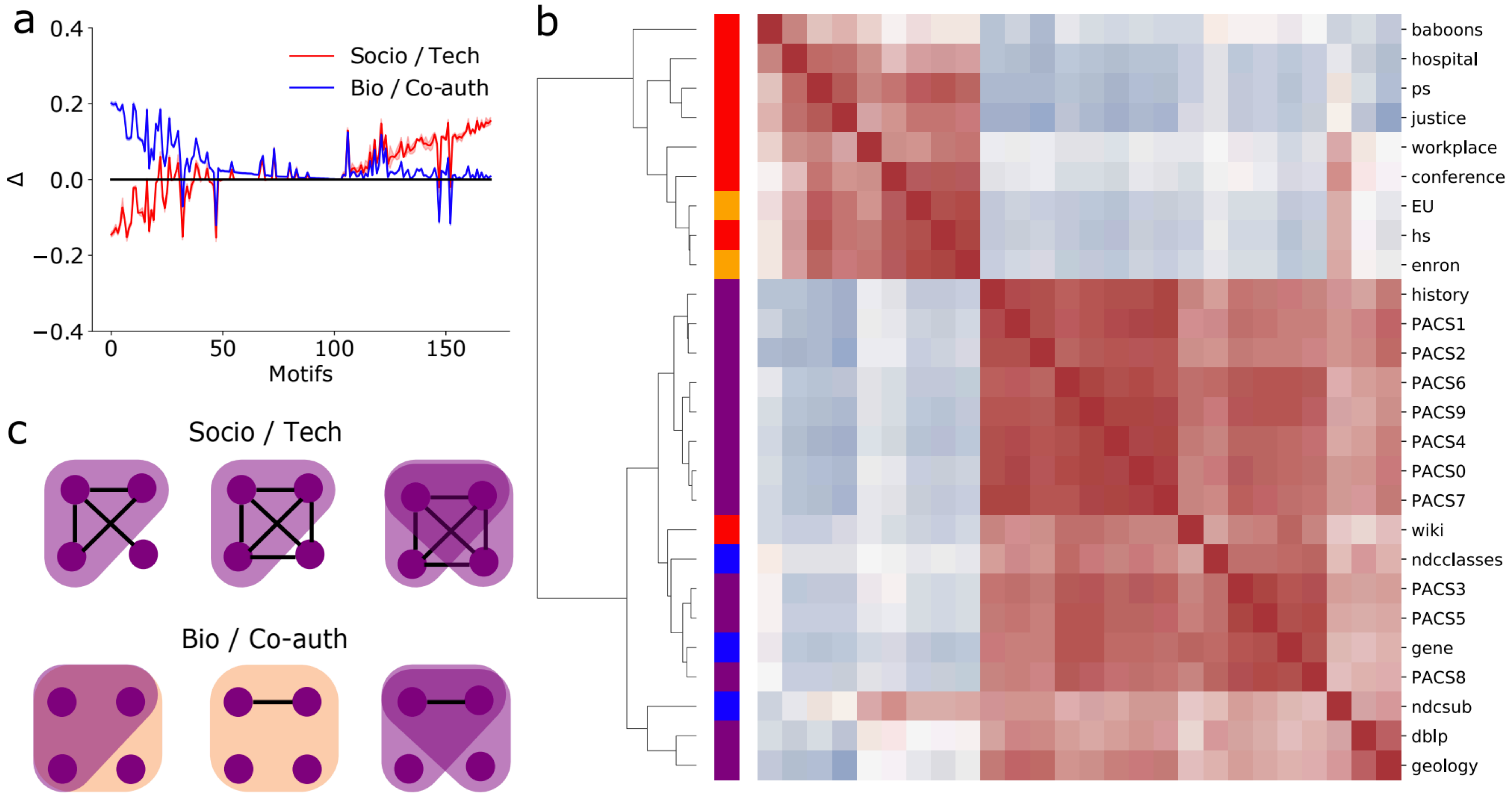


Local structure of real-world hypergraphs

Motifs of order 4

Significance profiles grouped by domain

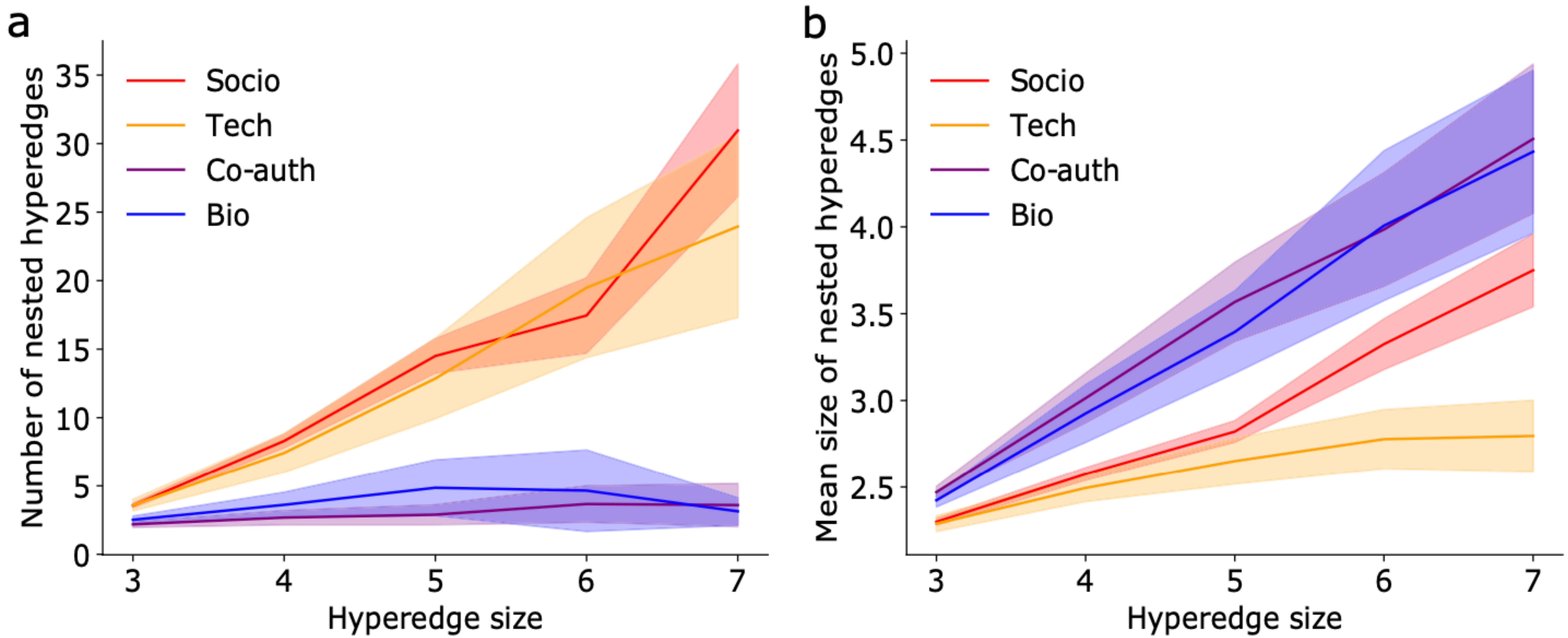
Correlation matrix of the SPs



Most over-expressed motifs of order 4

Nested organization of group interactions

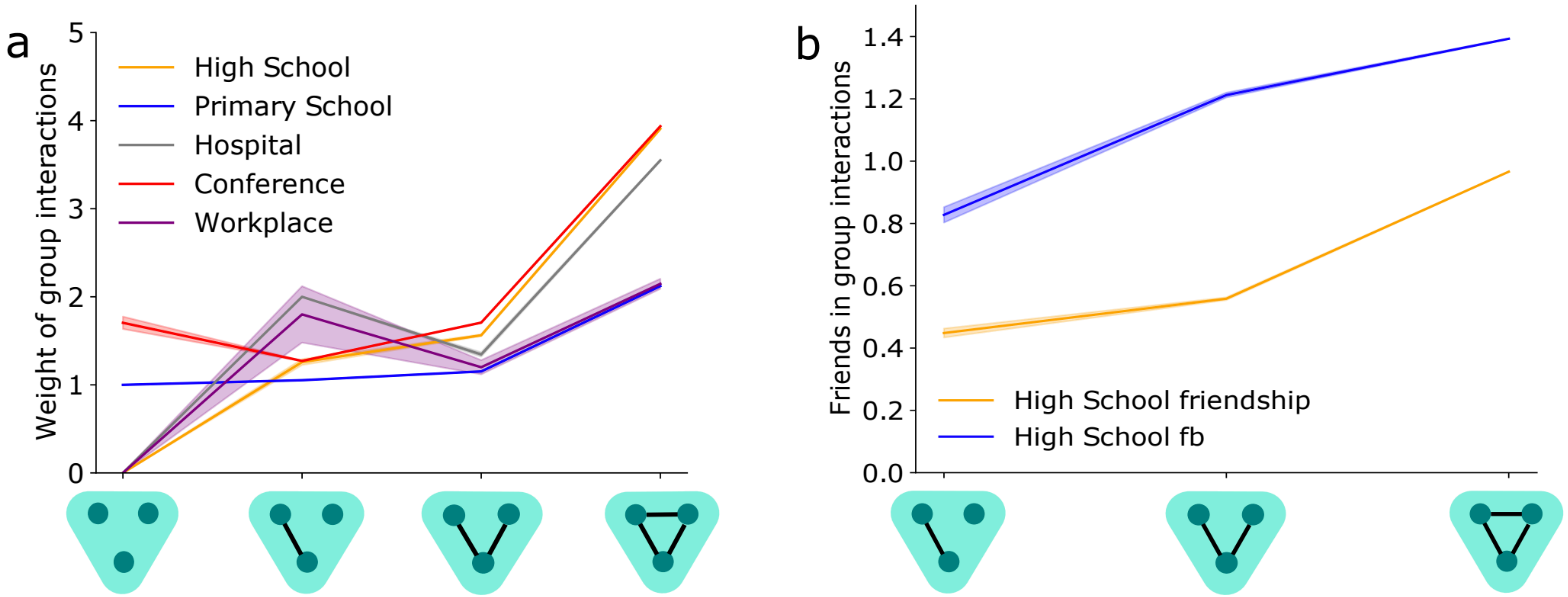
What happens “inside” large hyperedges?



NB: here we do not compute motifs of size bigger than 4 (that is not feasible as of now), we just settle for extracting statistics of the nested structure of hyperedges

Structural reinforcement

A rich supporting nested structure of pairwise links makes group interactions stronger 💪



Key takeaways

Things to remember about (higher-order) network motifs

- Network motifs are small patterns of interactions that are statistically over-expressed with respect to a null model
- Network motifs is an important tool in network science: **it highlights the local structure characteristics of complex networks**
- Traditional network motifs are not able to “see” patterns of group interactions, therefore we have proposed an extension of this notion to hypergraphs!

Key takeaways

Things to remember about (higher-order) network motifs

- We have seen that running algorithms on the projection of a hypergraphs can be misleading and slow
- We have proposed an efficient algorithm for higher-order motif discovery that exploits higher-order interactions
- We extracted fingerprints at the microscale of a variety of real-world systems with higher-order interactions and showed the emergence of families of hypergraphs characterized by similar local structures
- We defined the “nested structure” of hyperedges, showed that real-world hypergraphs from different domains show very different behaviors, and highlighted that richer nested structure of pairwise links make group interactions stronger!

Future work

Spoiler alert

- Pushed by the growing awareness of the relevance of the higher-order nature of interactions in many real-world systems, **higher-order network motifs can pave the way to applications in a number of domains!** (In a way similar to what traditional motifs did)
- Given the possible applications in data-intensive domains, a limitation of our proposed approach is its scalability
  - Sampling and approximated methods will be critical

  ✓ But we are working on it, stay tuned!
- Of course, there is still room for taking into account other features of the interactions and embed them into the higher-order motif framework

References

If you are interested in learning more 📚

- F. Battiston et al., *Networks beyond pairwise interactions: Structure and dynamics*, Physics Reports, Volume 874, 2020, Pages 1-92, ISSN 0370-1573 (2020)
- Milo et al., *Network motifs: simple building blocks of complex networks,* Science. 2002 Oct 25;298(5594):824-7 (2002)
- Milo et al., *Superfamilies of evolved and designed networks*, Science (New York, N.Y.), vol. 303,5663 (2004): 1538-42 (2004)
- Q. Lotito et al., *Higher-order motif analysis in hypergraphs*, in press at Communications Physics (2022)

Thank you for the attention!

Hope you had fun, hope you’ll have even more fun at the aperitivo 🍻

PS: I’m very open for collaborations, if you are interested in my work drop me a line ☺️

Quintino Francesco Lotito

University of Trento, Italy

quintino.lotito@unitn.it

@FraLotito (Twitter and Telegram)