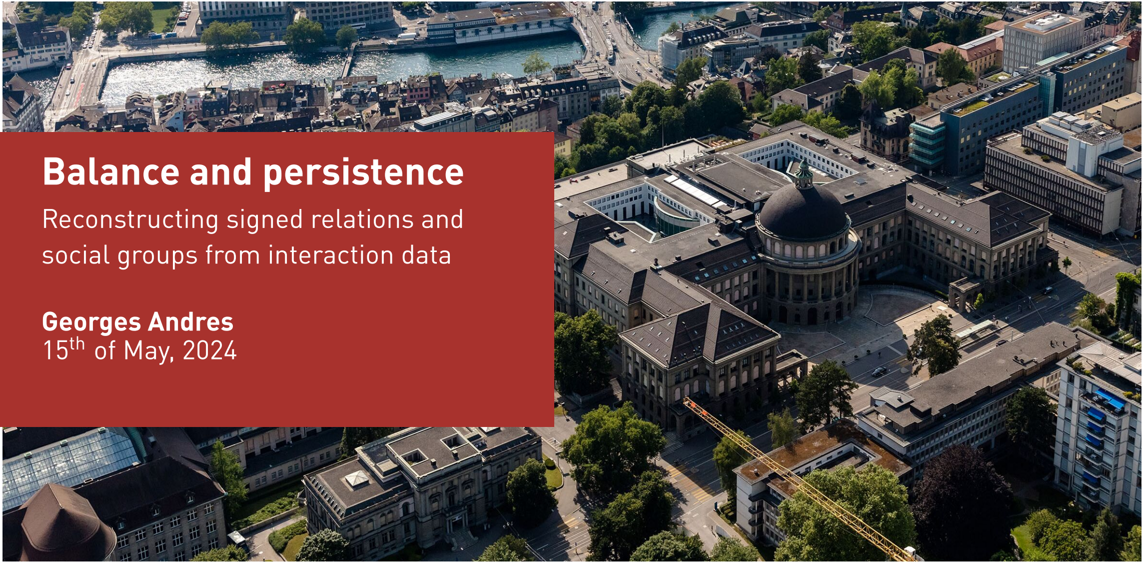
ETH zürich

Chair of Systems Design
www.sg.ethz.ch



Balance and persistence

Reconstructing signed relations and social groups from interaction data

Georges Andres

15th of May, 2024

The persistence of social groups

*"[...] at every moment **destructive forces** attack the life both from within and from without, and, if these alone operated, the unity would soon be resolved into its elements or transformed into other combinations. But opposed to these destructive forces there are **preservative influences**[...]"*

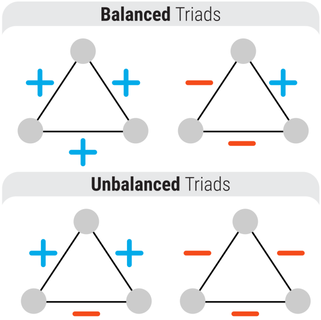
(Georg Simmel, 1898)

The theory of structural balance (SB) (Heider, 1946)

- Balanced vs unbalanced (Cartwright and Harary, 1956)
- **Degree of balance** $\Rightarrow$ destructive vs preservative

Structural balance and persistence

- Resolving unbalance by **leaving the group**
- Balance $\sim$ **low maintenance cost** (Morrissette et al., 1967)
- (+++) triad $\sim$ **cohesion** (Davis, 1963)



Hypothesis

The **higher** the **degree of balance** within a group is, the **more likely** the group is **to persist**.

From interaction data to signed relations and social groups

The data we require is rare

- (i) **temporal**, (ii) **signed relations** (SR), (iii) **groups**
- Surveys or online likes/follows?

Interaction ID	Individual ID 1	Individual ID 2	timestamp
I_1	A	B	1
I_2	C	B	2
$\mathbf{I}_3$	A	B	3
$\mathbf{I}_3$	A	C	3
$\mathbf{I}_3$	A	D	3
I_4	A	B	4

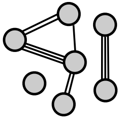
Interaction data more abundant

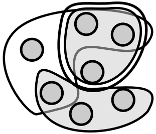
- E.g., email exchanges, phone-calls, FtF contacts etc.
- Short-lived and repeated

  Problem: missing signed relations or groups

Dyadic vs polyadic interactions

- **Networks**: edges between node pairs
- **Hypergraphs**: hyperedges between node groups





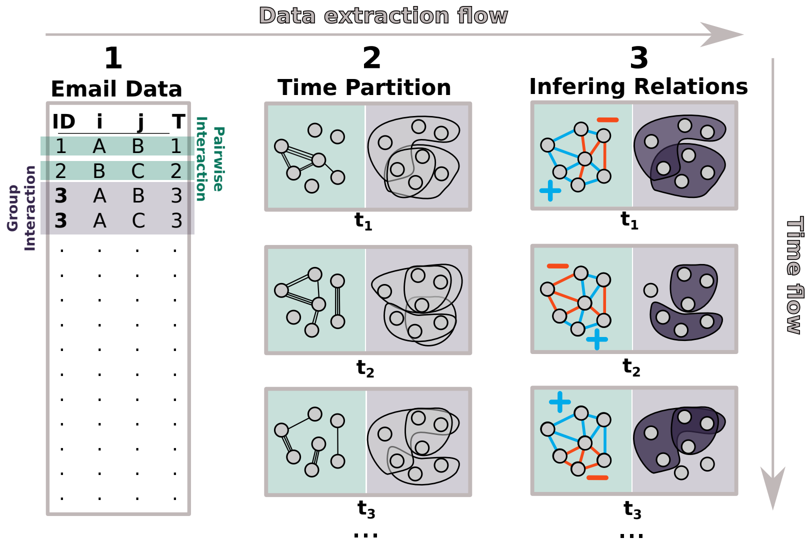
Data challenge

Inferring (i) time-decomposition, (ii) signed relations and (iii) groups from interaction data

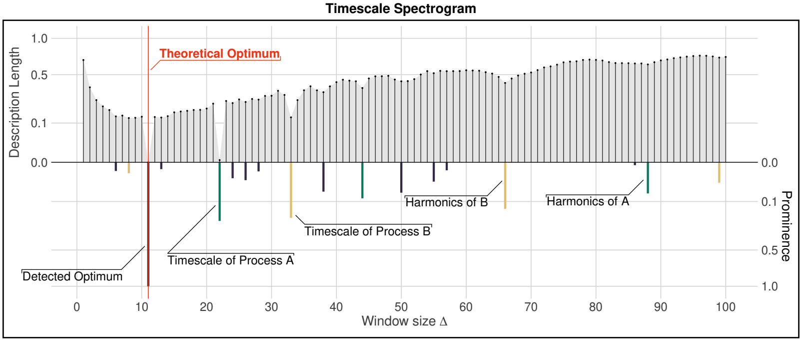
Overview of the data challenge

Data requirement

- **time dependence**
- **signed relations**
- **groups**



Disentangling the spectrum of timescales



Timescale detection*

Bayesian framework for timescale detection in temporal network models

- Employing the **description length** to infer the **spectrum of timescales** in the network

Optimal timescales reveal different system reactions to major shocks

- Tracking the **slowdown** before and **speedup** after shocks in OSS projects

*Casiraghi and Andres, *Disentangling the Timescales of a Complex System*, arXiv (2024)

Inferring SR from dyadic interactions

Working Assumption

more interactions ⇒ + signed relation/links
less interactions ⇒ − signed relation/links

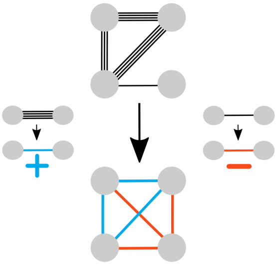
Comparing to random expectation

- Require **null-model for interactions**
- $a_{ij} \in \mathbb{N}$, *observed* interactions between i and j
- $A_{ij} \in \mathbb{N}$, *random* variable for the interactions

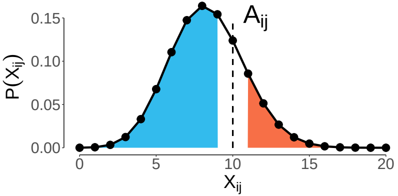
Over- and under-representation of interactions

- $P(X_{ij} < A_{ij})$ large ⇒ $\geq \boldsymbol{A_{ij}}$ **interactions is unlikely**
  Interpret: over-representation and **positive SR**
- $P(X_{ij} > A_{ij})$ large ⇒ $\leq \boldsymbol{A_{ij}}$ **interactions is unlikely**
  Interpret: under-representation and **negative SR**

Interaction Network



Signed Network

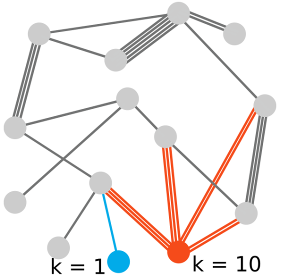


Hypergeometric configuration model

Capturing individuals' heterogeneities

- Individuals exhibit different levels of **fitness** & **activity**
- in-degree $\hat{k}_i^{in} \to$ fitness; out-degree $\hat{k}_i^{out} \to$ activity

  $\Rightarrow$ **Use a network null-model**



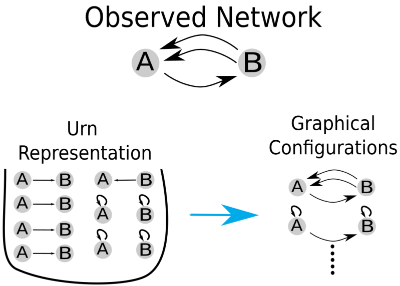
Defining the null-model **(Casiraghi and Nanumyan, 2021)**

- Possible links from i to j: $\Xi_{ij} \equiv \hat{k}_i^{out}\hat{k}_j^{in}$
- Unif. sampling w/o replacement of m links from urn

  $\Rightarrow$ **Multivariate hypergeometric distribution**
- **Degrees preserved in expectation**: $\mathbb{E}[k_i^{out}] = \hat{k}_i^{out}$

Over/under-representation: cumul. marginals

$$P(A_{ij} < a_{ij}) = \sum_{A_{ij}=0}^{a_{ij}-1} \frac{\binom{\Xi_{ij}}{A_{ij}}\binom{m^2-\Xi_{ij}}{m-A_{ij}}}{\binom{m^2}{m}}$$



Inferring singed relations: combine the ingredients!

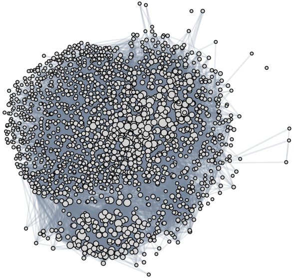
Introducing the Φ-measure

- Combine $P(A_{ij} < a_{ij})$ and $P(A_{ij} > a_{ij})$ in **one index**:

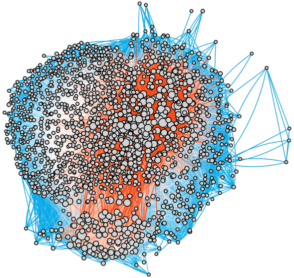
$$\Phi_{ij} = aP(A_{ij} < a_{ij}) + bP(A_{ij} > a_{ij})$$

- $\text{sign}(\Phi_{ij})$ is the **sign** of the relation
- $|\Phi_{ij}|$ is the **weight** on the inferred relation
- a and b can be **learned from known relations**; usually $a > 0$ and $b < 0$
- In absence of additional information, we propose: $1 = a = -b$

Testing the method on real-world data



Interaction network



Inferred signed network

		Inferred Relations: positive	Inferred Relations: negative
Declared Relations	trust	81%	19%
	distrust	26%	74%

Validation

Inference of signed relations*

Providing stochastic network method for inferring signed relations

- **High accuracy** in predicting known relations tested **across 5 datasets**

Unlocking interaction data for the study of signed relations

- Allows for **quantifying structural balance** for a wide variety of systems

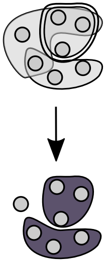
*Andres et al., *Reconstructing signed relations from interaction data*, Scientific Reports (2023)

Inferring groups from polyadic interactions

Working assumption

many polyadic **interactions** $\Rightarrow$ individuals part of **social group**

Interaction Hypergraph

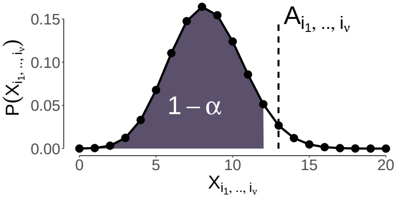


Social groups

Comparing to random expectation

- Which groups interacted **more than expected**?
- Employ **hypergraph null-model** for interactions
- Perform a significance test ($\alpha = 0.05$):

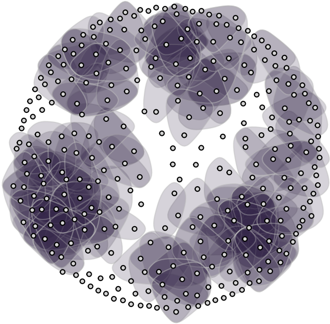
$$\text{Pr}(X_{i_1,..,i_\nu} \leq A_{i_1,..,i_\nu}) > 1 - \alpha$$

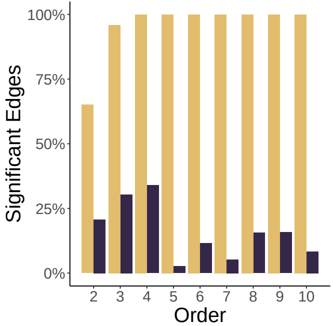


Modeling sparsity

- Real hypergraphs are **extremely sparse**
  $\Rightarrow$ **Zero-inflating** the hypergraph model

Inference of groups despite sparsity





Inference of groups*

Constructing a sparsity-adapted stochastic hypergraph model

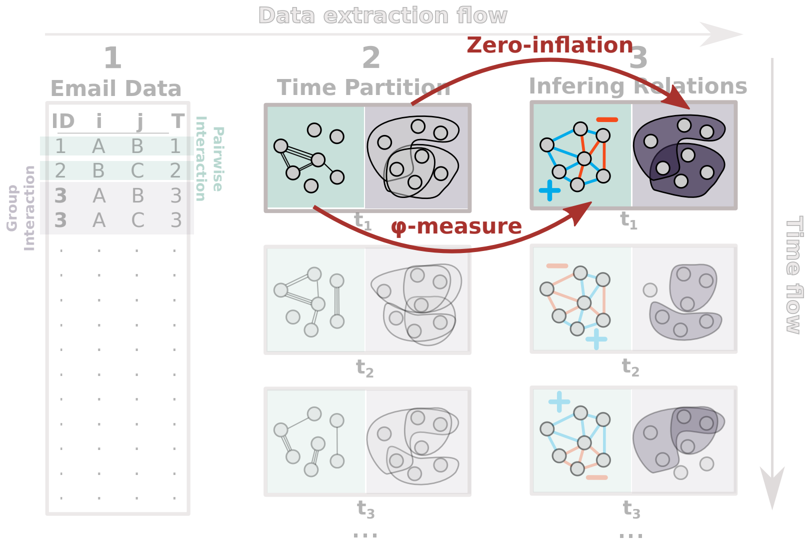
- Derive **analytical formulation** and provide a method for **fitting to data**

Enabling the inference of groups through ensemble perspective

- **Social groups can be differentiated** from spurious interactions

*Andres et al. *Stochastic Modeling of Hypergraphs: Zero-inflation to counteract the curse of dimensionality* (in preparation)

Final data representation



The persistence of social groups

Hypothesis

The **higher** the **degree of balance** within a group is, the **more likely** the group is **to persist**.

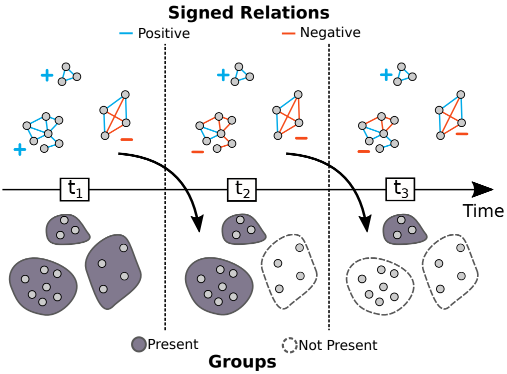
Performing a logistic regression

- Dependent variable: individual i being in a group g_j that persists from $t \to t+1$
- Independent variables: importance of triads $n(g_j^t)$

Quantifying the degree of balance

- Calculated from **inferred signed network**
- **Importance** associated with each **type of triad**:

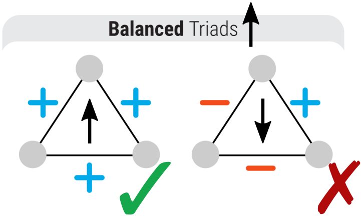
$$n(g_j^t) = (n_{+++}, n_{++-}, n_{+--}, n_{---})$$



Structural balance and group persistence

How different triads affect the persistence of groups

- **(+++)** consistently has a **positive effect** on persistence
  $\Rightarrow$ **cohesion**
- **(+--)** has a **negative effect** when significant
  $\Rightarrow$ **polarization**



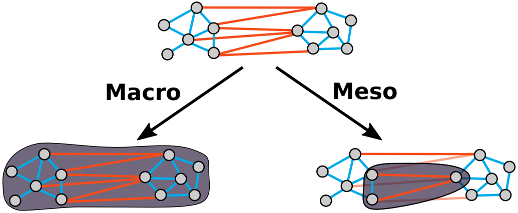
Structural balance vs cohesion

For small groups, only the **cohesive triad (+++)** positively affects persistence

Structural balance and group persistence

Interpretation: scales matter!

- **Macro**: polarization can be **very persistent**
- **Meso**: polarization is **not persistent**

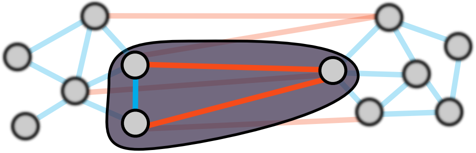


Uncovering the effects of balance

- Signed network $\Rightarrow$ groups

  vs

- **Signed network $\not\Rightarrow$ groups**



Structural balance: beneficial or not?

To answer this, we must consider the **interplay** between **group and signed structures**

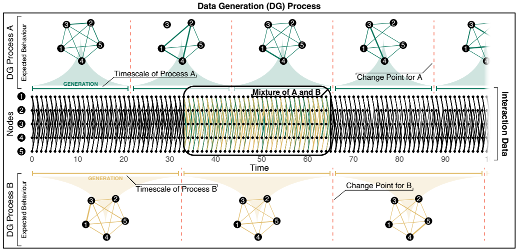
Conclusion

Structural balance and group dynamics

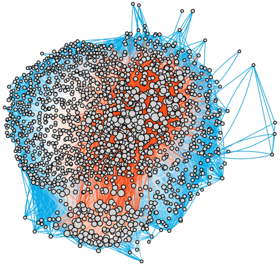
- Small groups: **cohesion vs balance**

Unlocking the potential of interaction data

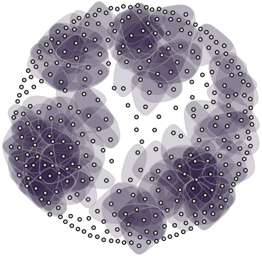
- **Disentangle** a network's **timescales**
- Infer signed **relations with Φ-method**
- Infer social **groups via zero-inflation**



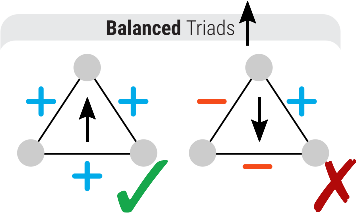
Timescale inference



Signed inference



Group inference



Balance & persistence