Reconstruction of hypergraphs from their noisy pairwise observations

NetSci 2022 — Network inference

Simon Lizotte, Jean-Gabriel Young and Antoine Allard

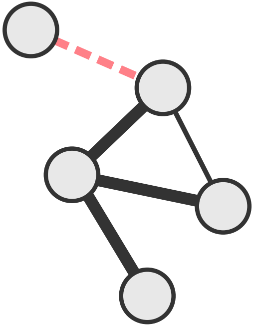
July 28, 2022

simon.lizotte.1@ulaval.ca

Empirical data are *noisy*...

Empirical data are *noisy*...

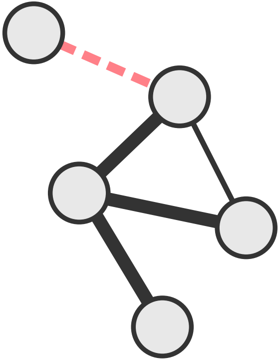
and this applies to *network data*.



It applies to the vast majority of datasets

Whether measurements are

they are *noisy* and *uncertain*.

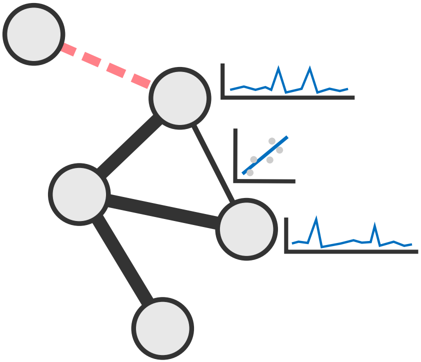


It applies to the vast majority of datasets

Whether measurements are

- vertex times series correlations,

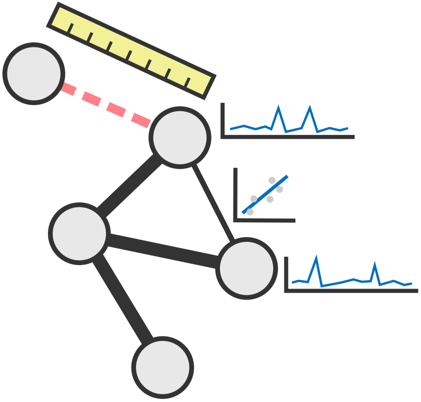
they are *noisy* and *uncertain*.



Whether measurements are

- vertex times series correlations,
- counts of proximity detectors,

they are *noisy* and *uncertain*.

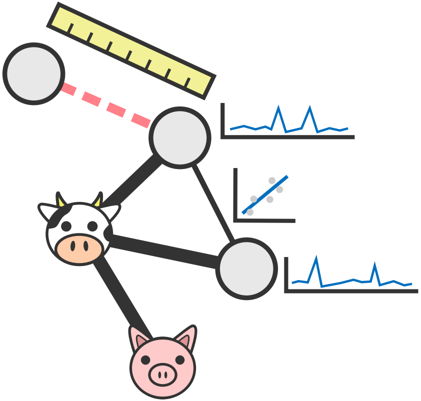


It applies to the vast majority of datasets

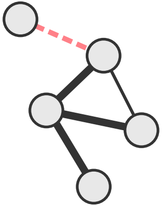
Whether measurements are

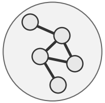
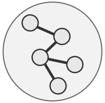
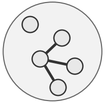
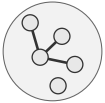
- vertex times series correlations,
- counts of proximity detectors,
- counts of animal interactions,

they are *noisy* and *uncertain*.



Bayesian approach



	Hypothesis (Prior)	**Educated guess (Posterior)**
	0.25	0.20
	0.25	0.60
	0.25	0.19
	0.25	0.01

J.-G. Young, G. T. Cantwell, and M. E. J. Newman, "*Bayesian inference of network structure from unreliable data*", J. Complex Netw. **8**, cnaa046 (2021).

Works in network reconstruction

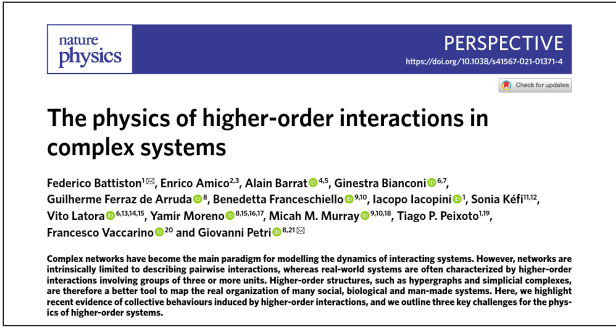
Bayesian approaches:

- C. T. Butts, "*Network inference, error, and informant (in)accuracy: a Bayesian approach*", Soc Networks **25**, 103–140 (2003).
- T. P. Peixoto, "*Network Reconstruction and Community Detection from Dynamics*", Phys. Rev. Lett. **123**, 128301 (2019).

Other approaches:

- V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, and P. Geurts, "*Inferring Regulatory Networks from Expression Data Using Tree-Based Methods*", PLoS One **5**, e12776 (2010).
- A. T. Specht and J. Li, "*LEAP: constructing gene co-expression networks for single-cell RNA-sequencing data using pseudotime ordering*", Bioinformatics **33**, 764–766, (2017).
- H. Matsumoto et al., "*SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation*", Bioinformatics **33**, 2314–2321 (2017).

- Explosive phase transitions;



nature physics

PERSPECTIVE

https://doi.org/10.1038/s41567-021-01371-4

Check for updates

The physics of higher-order interactions in complex systems

Federico Battiston[1]✉, Enrico Amico[2,3], Alain Barrat[4,5], Ginestra Bianconi[6,7], Guilherme Ferraz de Arruda[8], Benedetta Franceschiello[9,10], Iacopo Iacopini[1], Sonia Kéfi[11,12], Vito Latora[6,13,14,15], Yamir Moreno[8,15,16,17], Micah M. Murray[9,10,18], Tiago P. Peixoto[1,19], Francesco Vaccarino[20] and Giovanni Petri[8,21]✉

Complex networks have become the main paradigm for modelling the dynamics of interacting systems. However, networks are intrinsically limited to describing pairwise interactions, whereas real-world systems are often characterized by higher-order interactions involving groups of three or more units. Higher-order structures, such as hypergraphs and simplicial complexes, are therefore a better tool to map the real organization of many social, biological and man-made systems. Here, we highlight recent evidence of collective behaviours induced by higher-order interactions, and we outline three key challenges for the physics of higher-order systems.

Importance of higher order interactions

- Explosive phase transitions;
- Social coordination;

nature physics
PERSPECTIVE
https://doi.org/10.1038/s41567-021-01371-4

Check for updates

The physics of higher-order interactions in complex systems

Federico Battiston[1]✉, Enrico Amico[2,3], Alain Barrat[4,5], Ginestra Bianconi[6,7], Guilherme Ferraz de Arruda[8], Benedetta Franceschiello[9,10], Iacopo Iacopini[1], Sonia Kéfi[11,12], Vito Latora[6,13,14,15], Yamir Moreno[8,15,16,17], Micah M. Murray[9,10,18], Tiago P. Peixoto[1,19], Francesco Vaccarino[20] and Giovanni Petri[8,21]✉

RESEARCH

SOCIAL SCIENCE

Experimental evidence for tipping points in social convention

Damon Centola[1,2*], Joshua Becker[1], Devon Brackbill[1], Andrea Baronchelli[3]

Theoretical models of critical mass have shown how minority groups can initiate social change dynamics in the emergence of new social conventions. Here, we study an artificial system of social conventions in which human subjects interact to establish a new coordination equilibrium. The findings provide direct empirical demonstration of the existence of a tipping point in the dynamics of changing social conventions. When minority groups reached the critical mass—that is, the critical group size for initiating social change—they were consistently able to overturn the established behavior. The size of the required critical mass is expected to vary based on theoretically identifiable features of a social setting. Our results show that the theoretically predicted dynamics of critical mass do in fact emerge as expected within an empirical system of social coordination.

broad practical (*18*, *19*) and scientific (*1*, *12*) importance of understanding the dynamics of critical mass in collective behavior, it has not been possible to identify whether there are in fact tipping points in empirical systems because such a test requires the ability to independently vary the size of minority groups within an evolving system of social coordination.

We addressed this problem by adopting an experimental approach to studying tipping-point dynamics within an artificially created system of evolving social conventions. Following the literature on social conventions (*9*, *20*, *21*), we study a system of coordination in which a minority group of actors attempt to disrupt an established equilibrium behavior. In both our theoretical framework and the empirical setting, we adopt the canonical approach of using coordination on a naming convention as a general model for conventional behavior (*21*–*24*). Our experimental approach is designed to test a

Importance of higher order interactions

- Explosive phase transitions;
- Social coordination;
- Multiple species interactions;

nature physics

PERSPECTIVE

https://doi.org/10.1038/s41567-021-01371-4

Check for updates

The physics of higher-order interactions in complex systems

Federico Battiston[1], Enrico Amico[2,3], Alain Barrat[4,5], Ginestra Bianconi[6,7], Guilherme Ferraz de Arruda[8], Benedetta Franceschiello[9,10], Iacopo Iacopini[1], Sonia Kéfi[11,12], Vito Latora[6,13,14,15], Yamir Moreno[8,15,16,17], Micah M. Murray[9,10,18], Tiago P. Peixoto[1,19], Francesco Vaccarino[20] and Giovanni Petri[8,21]

RESEARCH

SOCIAL SCIENCE

Experimental evidence for tipping points in social convention

Damon Centola[1,2*], Joshua Becker[1], Devon Brackbill[1], Andrea Baronchelli[3]

broad practical (18, 19) and scientific (1, 12) importance of understanding the dynamics of critical mass in collective behavior, it has not been possible to identify whether there are in fact tipping points in empirical systems because such a test requires the ability to independently vary the size of minority groups within an evolving system of social coordination.

We addressed this problem by adopting an

nature ecology & evolution

ARTICLES

PUBLISHED: 17 FEBRUARY 2017 | VOLUME: 1 | ARTICLE NUMBER: 0062

Higher-order interactions capture unexplained complexity in diverse communities

Margaret M. Mayfield[1*] and Daniel B. Stouffer[2*]

Natural communities are well known to be maintained by many complex processes. Despite this, the practical aspects of studying them often require some simplification, such as the widespread assumption that direct, additive competition captures the important details about how interactions between species impact community diversity. More complex non-additive 'higher-order' interactions are assumed to be negligible or absent. Notably, these assumptions are poorly supported and have major consequences for the accuracy with which patterns of natural diversity are modelled and explained. We present a mathematically simple framework for incorporating biologically meaningful complexity into models of diversity by including non-additive

Importance of higher order interactions

- Explosive phase transitions;
- Social coordination;
- Multiple species interactions;
- Brain cortical dynamics;
- and many more...

nature physics
PERSPECTIVE
https://doi.org/10.1038/s41567-021-01371-4
Check for updates

The physics of higher-order interactions in complex systems

Federico Battiston[1] ✉, Enrico Amico[2,3], Alain Barrat[4,5], Ginestra Bianconi[6,7], Guilherme Ferraz de Arruda[8], Benedetta Franceschiello[9,10], Iacopo Iacopini[1], Sonia Kéfi[11,12], Vito Latora[6,13,14,15], Yamir Moreno[8,15,16,17], Micah M. Murray[9,10,18], Tiago P. Peixoto[1,19], Francesco Vaccarino[20] and Giovanni Petri[8,21] ✉

RESEARCH

SOCIAL SCIENCE

Experimental evidence for tipping points in social convention

Damon Centola[1,2]*, Joshua Becker[1], Devon Brackbill[1], Andrea Baronchelli[3]

broad practical (*18*, *19*) and scientific (*1*, *12*) importance of understanding the dynamics of critical mass in collective behavior, it has not been possible to identify whether there are in fact tipping points in empirical systems because such a test requires the ability to independently vary the size of minority groups within an evolving system of social coordination.
We addressed this problem by adopting an

17514 • The Journal of Neuroscience, November 30, 2011 • 31(48):17514 –17526

Behavioral/Systems/Cognitive

Higher-Order Interactions Characterized in Cortical Activity

Shan Yu (余山),[1] Hongdian Yang (杨鸿典),[1,2] Hiroyuki Nakahara (中原裕之),[3,4] Gustavo S. Santos,[3] Danko Nikolić,[5,6] and Dietmar Plenz[1]
[1]Section on Critical Brain Dynamics, Laboratory of Systems Neuroscience, National Institute of Mental Health, Bethesda, Maryland 20892-9663, [2]Biophysics Program, Institute for Physical Science and Technology, University of Maryland, College Park, Maryland, 20742, [3]Laboratory for Integrated Theoretical Neuroscience, RIKEN Brain Science Institute, Wako City, Saitama 351-0198 Japan, [4]Department of Computational Intelligence and Systems Science, Tokyo

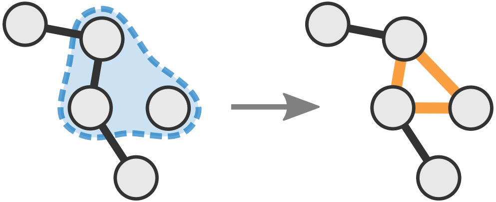
nature ecology & evolution
ARTICLES
PUBLISHED: 17 FEBRUARY 2017 | VOLUME: 1 | ARTICLE NUMBER: 0062

Higher-order interactions capture unexplained complexity in diverse communities

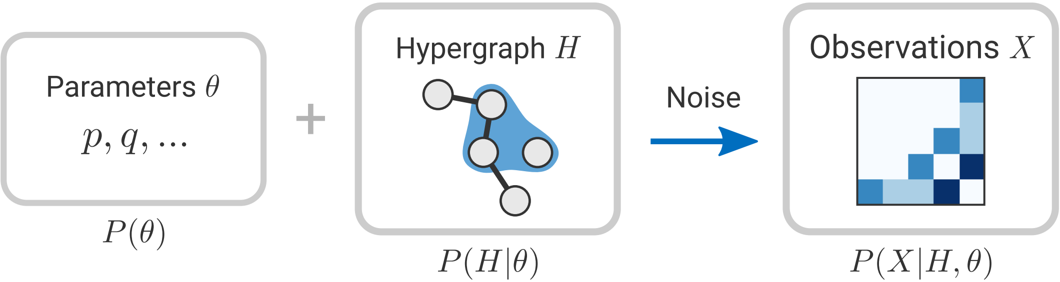
Margaret M. Mayfield[1]* and Daniel B. Stouffer[2]*

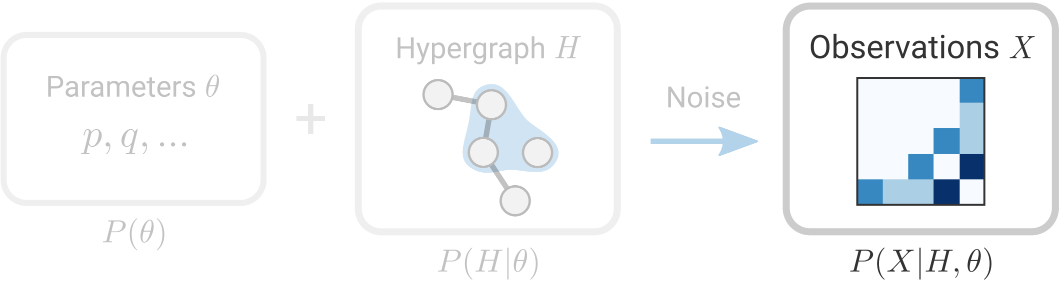
Natural communities are well known to be maintained by many complex processes. Despite this, the practical aspects of studying them often require some simplification, such as the widespread assumption that direct, additive competition captures the important details about how interactions between species impact community diversity. More complex non-additive 'higher-order' interactions are assumed to be negligible or absent. Notably, these assumptions are poorly supported and have major consequences for the accuracy with which patterns of natural diversity are modelled and explained. We present a mathematically simple framework for incorporating biologically meaningful complexity into models of diversity by including non-additive

Explore the importance of *correlations induced* by higher-order interactions in the context of reconstruction.

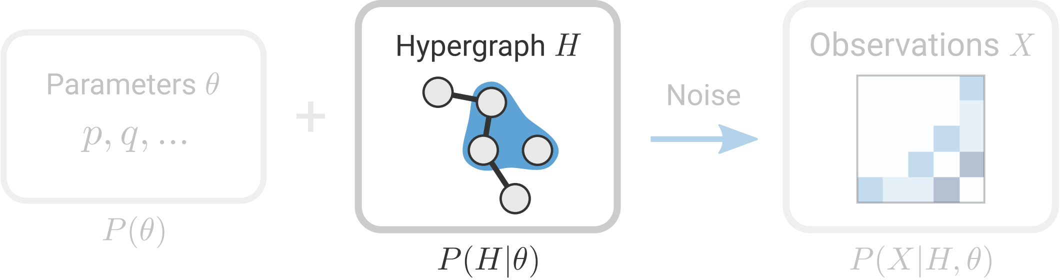


Reconstruction using Bayes formula

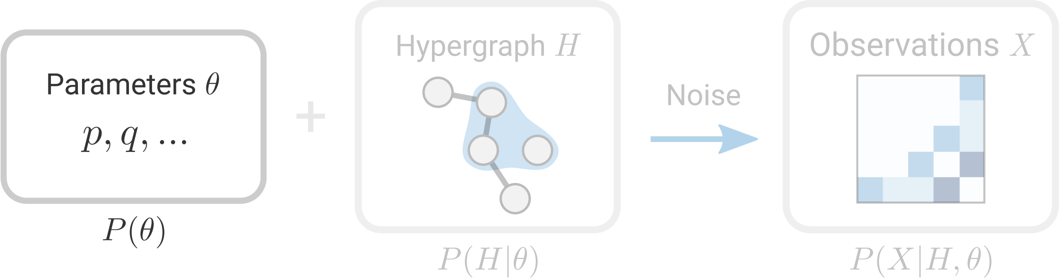


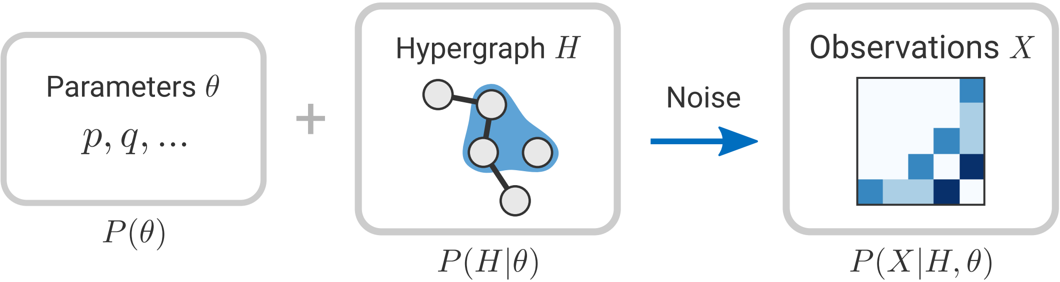

Parameters θ
$p, q, \ldots$
$P(\theta)$
+
Hypergraph H
$P(H|\theta)$
Noise
Observations X
$P(X|H, \theta)$

Reconstruction using Bayes formula



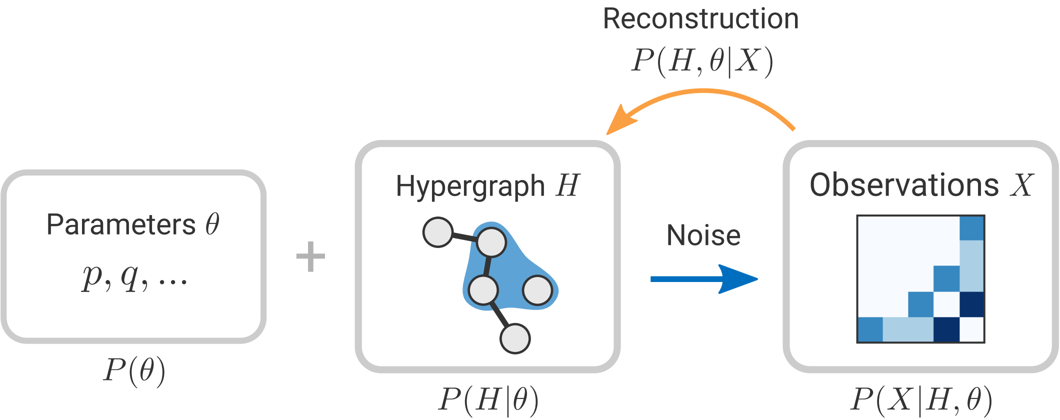
Reconstruction using Bayes formula





Bayes formula:

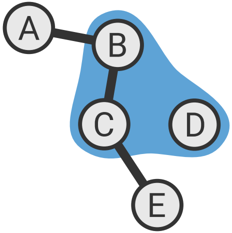
$$\underbrace{P(H,\theta|X)}_{\text{Posterior}} \propto P(X|H,\theta)P(H|\theta)P(\theta)$$



Bayes formula:

$$\underbrace{P(H,\theta|X)}_{\text{Posterior}} \propto P(X|H,\theta)P(H|\theta)P(\theta)$$

Pairwise projection of hypergraphs



We define interaction type ℓ_{ij} of a pair (i, j) as

$$\ell_{ij} = (\text{largest hyperedge size with } i \text{ and } j) - 1.$$

Examples:

- $\ell_{\text{ED}} = 0$
- $\ell_{\text{AB}} = 1$
- $\ell_{\text{BC}} = 2$

Data model $P(X|H, \theta)$

The *size* ($\ell_{ij} + 1$) of the largest hyperedge connecting i and j determines the *measurement rate* ($\mu_{\ell_{ij}}$) of the pair (i, j).

Data model $P(X|H,\theta)$

The *size* ($\ell_{ij}\ +\ 1$) of the largest hyperedge connecting i and j determines the *measurement rate* ($\mu_{\ell_{ij}}$) of the pair (i,j).

The data model is then

$$P(X|H,\theta) = \prod_{i<j} \mathrm{Poi}(x_{ij};\mu_{\ell_{ij}}),$$

where x_{ij} is the count of measurements between i and j.

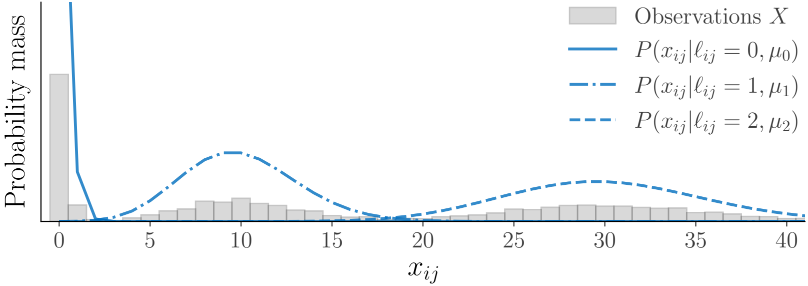
Data model $P(X|H,\theta)$

The *size* ($\ell_{ij} + 1$) of the largest hyperedge connecting i and j determines the *measurement rate* ($\mu_{\ell_{ij}}$) of the pair (i,j).

The data model is then

$$P(X|H,\theta) = \prod_{i<j} \mathrm{Poi}(x_{ij}; \mu_{\ell_{ij}}),$$

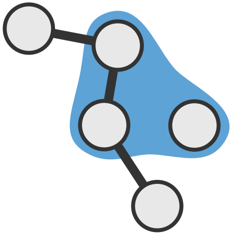
where x_{ij} is the count of measurements between i and j.



Hypergraph prior $P(H|\theta)$

We use a direct generalization of the $G(n, p)$ model:

- 2-edges independent with probability q;
- 3-edges independent with probability p;



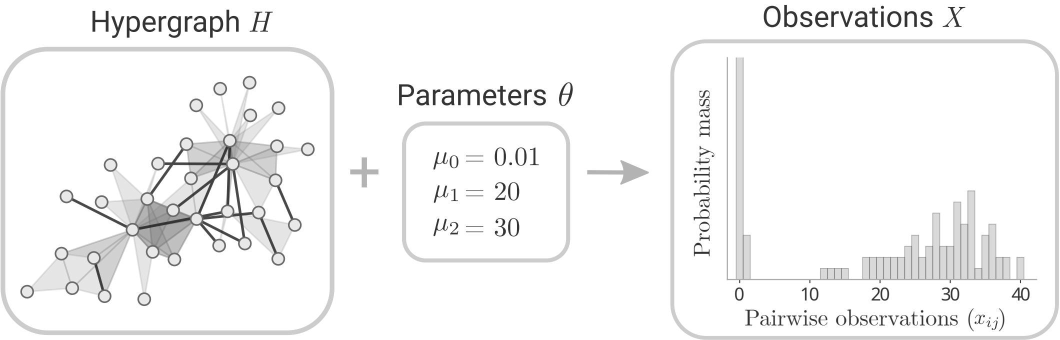
which leads to

$$P(H|\theta) = q^m (1-q)^{\binom{n}{2}-m} p^{\Delta} (1-p)^{\binom{n}{3}-\Delta},$$

where

- m is the number of 2-edges and
- Δ is the number of 3-edges.

Real-world network reconstruction



μ_k: measurement rate for hyperedges of size k,

x_{ij}: count of pairwise measurements for (i, j).

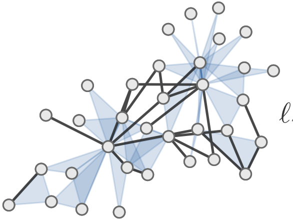
J.-G. Young, G. Petri, and T. P. Peixoto, "*Hypergraph reconstruction from network data*", Commun. Phys. **4**, 135 (2021).

Real-world network reconstruction

Using *MCMC*, we check the *posterior maximum* and the *confusion matrix*.

Real-world network reconstruction

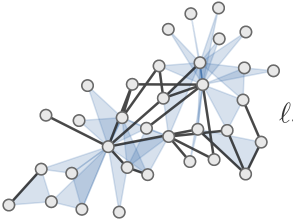
Using *MCMC*, we check the *posterior maximum* and the *confusion matrix*.



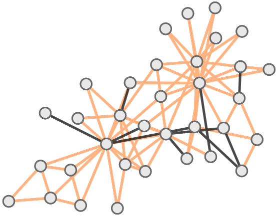
	$\hat{\ell}_{ij}$ = 0	1	2
ℓ_{ij} = 0	483	0	0
1	0	11	0
2	0	5	62

$\ell_{ij} + 1$: largest hyperedge connecting i and j.
$\hat{\ell}_{ij} + 1$: predicted largest hyperedge connecting i and j.

Using *MCMC*, we check the *posterior maximum* and the *confusion matrix*.



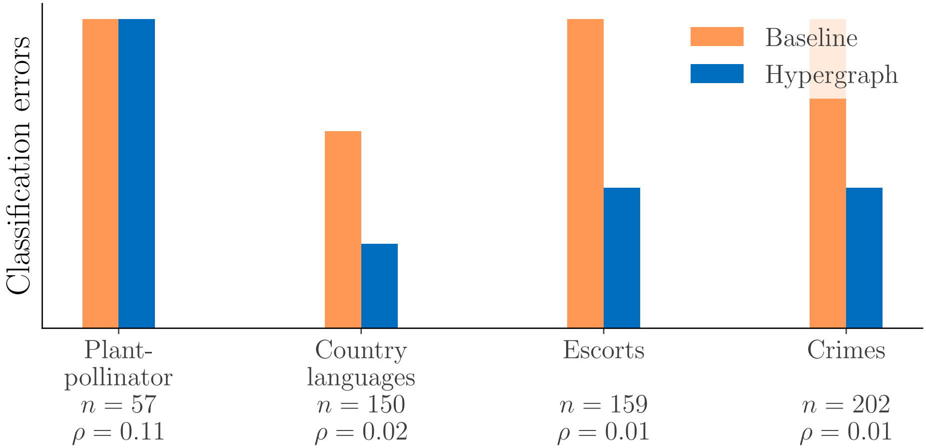
ℓ_{ij} \ $\hat{\ell}_{ij}$	0	1	2
0	483	0	0
1	0	11	0
2	0	5	62



ℓ_{ij} \ $\hat{\ell}_{ij}$	0	1	2
0	483	0	0
1	0	7	4
2	0	3	64

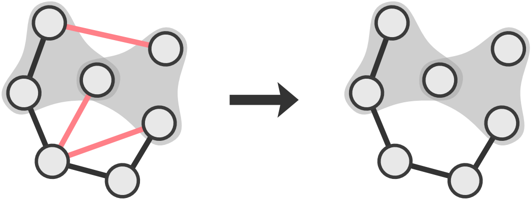
$\ell_{ij} + 1$: largest hyperedge connecting i and j.
$\hat{\ell}_{ij} + 1$: predicted largest hyperedge connecting i and j.

More real-world hypergraphs

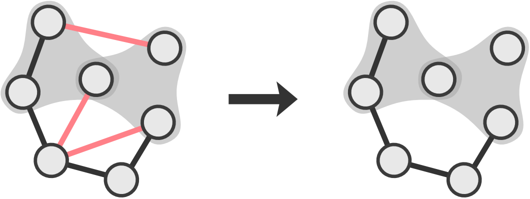


An easy hypergraph

We remove 2-edges that create triangles. As a result, 3-edges can be deduced directly from projected triangles.

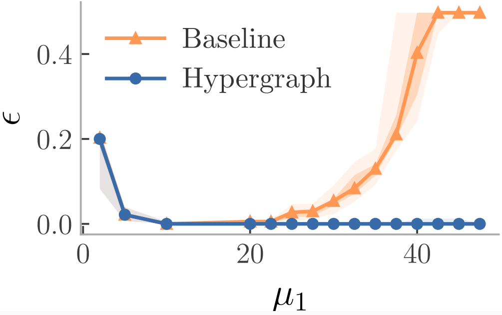


We remove 2-edges that create triangles. As a result, 3-edges can be deduced directly from projected triangles.



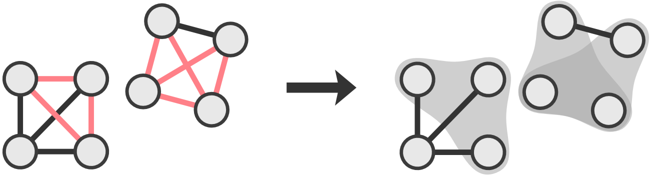
$$\epsilon = \frac{\#\text{ misclassified interactions}}{\#\text{ interactions}}$$

μ_1 : measurement rate of 2-edges.

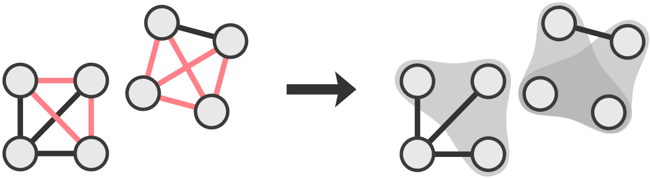


A difficult hypergraph

We create complete subgraphs and randomly promote triangles to 3-edges. As a result, triangles don't distinguish 2-edges from 3-edges.

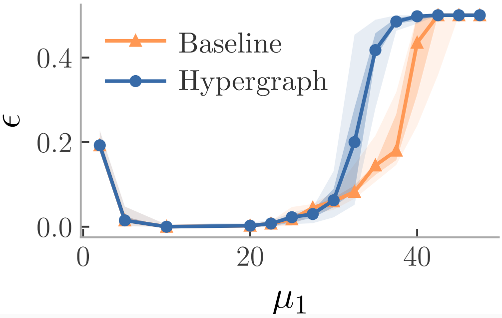


We create complete subgraphs and randomly promote triangles to 3-edges. As a result, triangles don't distinguish 2-edges from 3-edges.

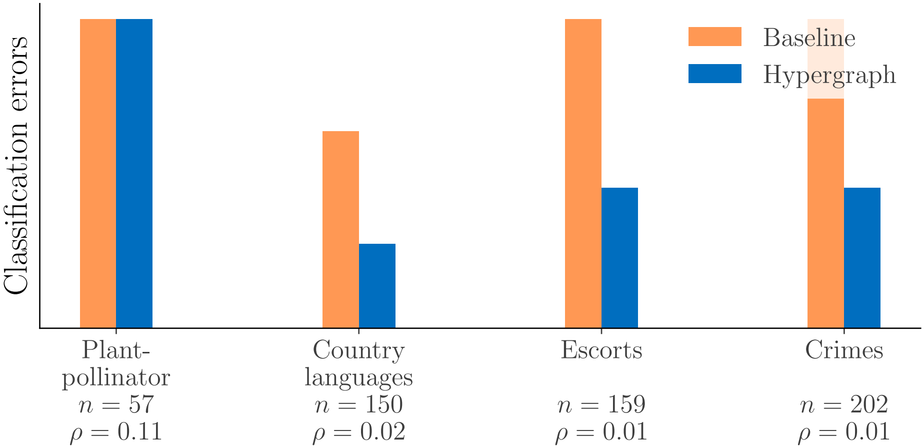


$$\epsilon = \frac{\text{\# misclassified interactions}}{\text{\# interactions}}$$

μ_1 : measurement rate of 2-edges.



Interactions are sparse



Take-aways:

- We introduced a model to reconstruct *hypergraphs* from *noisy pairwise observations*;
- We applied it on *real-world hypergraphs*;
- We showed that it works well because interactions are *sparse*.

Available soon:

- Preprint on arXiv
- Python/C++ implementation on GitHub

Thanks to my advisors Jean-Gabriel-Young and Antoine Allard.

Contact me at simon.lizotte.1@ulaval.ca

We suppose the following process

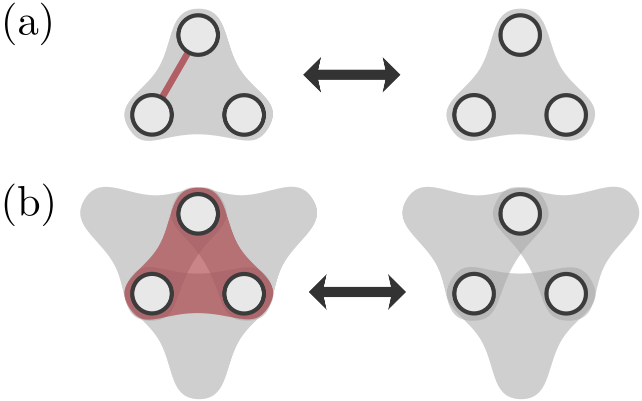
1. Edges of type $\ell_{ij} = 2$ are placed uniformly with probability q_2,
2. Edges of type $\ell_{ij} = 1$ are placed uniformly with probability q_1 among the remaining unconnected pairs,

which results in

$$P(G|\theta) = q_1^{m_1}(1 - q_1)^{\binom{n}{2} - m_1 - m_2} \times q_2^{m_2}(1 - q_2)^{\binom{n}{2} - m_2}, \tag{1}$$

where m_k is the number of edges of type $\ell_{ij} = k$.

Symmetries of the hypergraph model



Prior distributions

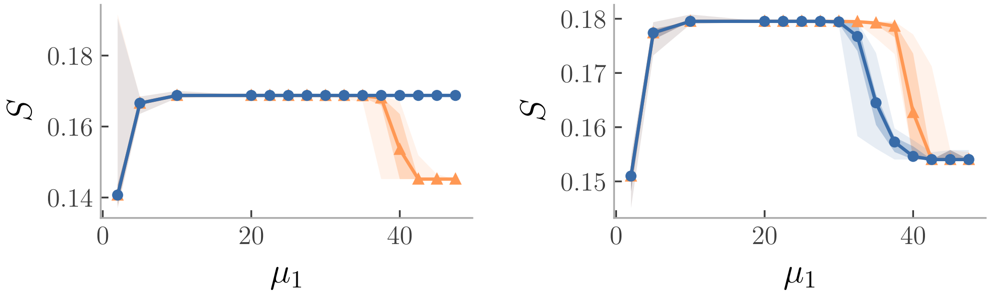
Hypergraph model:

$$p, q \sim \text{Beta}(\xi, \zeta) \tag{2}$$
$$\mu_0 \sim \text{Gamma}(\alpha_0, \beta_0) \tag{3}$$
$$\mu_1|\mu_0 \sim \text{TruncGamma}_{(\mu_0,\infty)}(\alpha_1, \beta_1) \tag{4}$$
$$\mu_2|\mu_0 \sim \text{TruncGamma}_{(\mu_0,\infty)}(\alpha_2, \beta_2). \tag{5}$$

Categorical model:

$$q_1, q_2 \sim \text{Beta}(\xi, \zeta) \tag{6}$$
$$\mu_0 \sim \text{Gamma}(\alpha_0, \beta_0) \tag{7}$$
$$\mu_1|\mu_0 \sim \text{TruncGamma}_{(\mu_0,\infty)}(\alpha_1, \beta_1) \tag{8}$$
$$\mu_2|\mu_1 \sim \text{TruncGamma}_{(\mu_1,\infty)}(\alpha_2, \beta_2). \tag{9}$$

Entropy for different structures and parameters



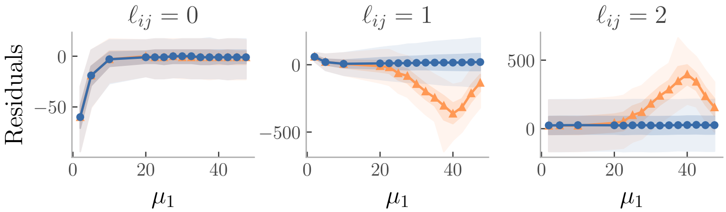
(a) Hypergraph for which every projected triangle is a 3-edge.

(b) Hypergraph for which every 2-edge is part of a projected triangle.

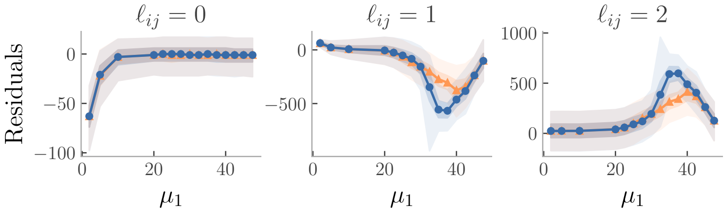
$$S = -\sum_{k=0}^{2} \rho_k \log_3 \rho_k, \tag{10}$$

where ρ_k is the proportion of interactions predicted as $\hat{\ell}_{ij} = k$.

Residuals (posterior predictive check)

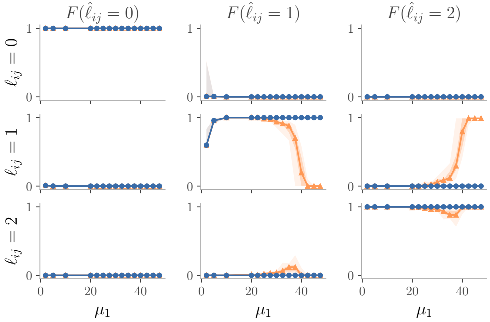


(c) Hypergraph for which every projected triangle is a 3-edge.

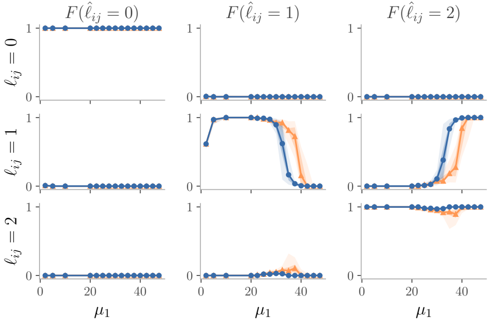


(d) Hypergraph for which every 2-edge is part of a projected triangle.

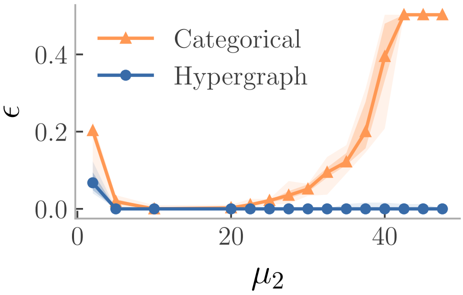
Confusion matrices



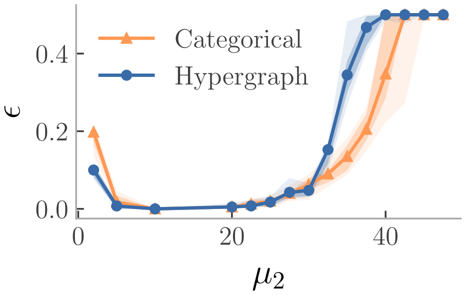
(e) Hypergraph for which every projected triangle is a 3-edge.



(f) Hypergraph for which every 2-edge is part of a projected triangle.

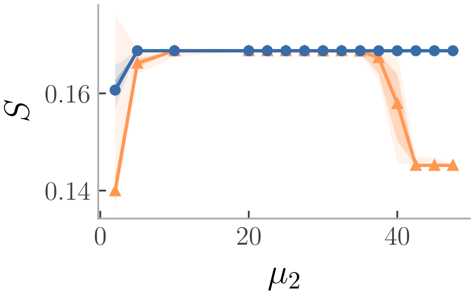


(g) Hypergraph for which every projected triangle is a 3-edge.

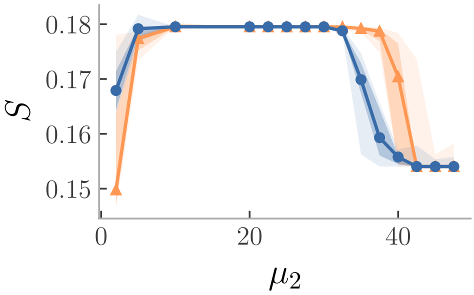


(h) Hypergraph for which every 2-edge is part of a projected triangle.

Entropy when varying μ_2

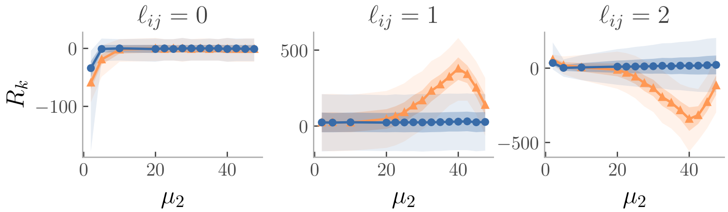


(i) Hypergraph for which every projected triangle is a 3-edge.

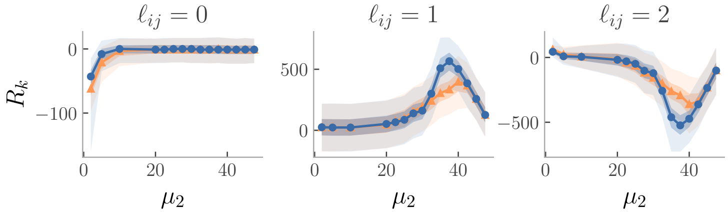


(j) Hypergraph for which every 2-edge is part of a projected triangle.

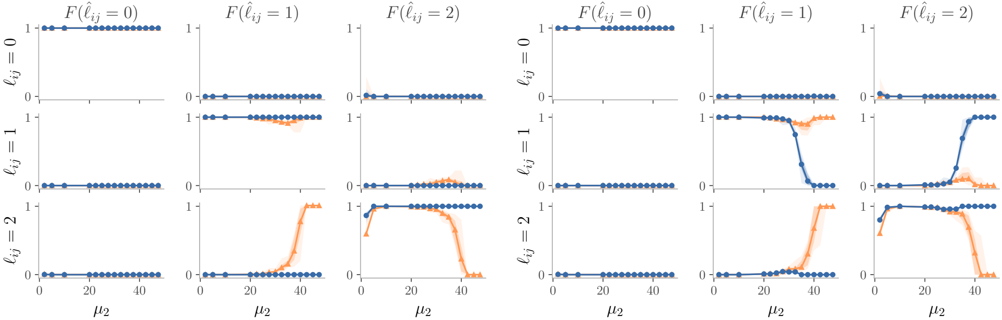
Residuals when varying μ_2



(k) Hypergraph for which every projected triangle is a 3-edge.



(l) Hypergraph for which every 2-edge is part of a projected triangle.



(m) Hypergraph for which every projected triangle is a 3-edge.

(n) Hypergraph for which every 2-edge is part of a projected triangle.