Nonbacktracking Eigenvector Method for Hypergraph Community Detection

Graphs and More Complex Structures for Learning and Reasoning
AAAI 2022 Workshop
February 28th, 2022

Dr. Jamie Haddock
Department of Mathematics
Harvey Mudd College

Background and **graph community detection**.

The **nonbacktracking operator** and eigenvector methods for hypergraphs.

Detectability thresholds and open questions.

Phil Chodrow
UCLA
@PhilChodrow

Nicole Eikmeier
Grinnell
@NicoleEikmeier

Philip S. Chodrow, Nicole Eikmeier, and **JH** (2022). Nonbacktracking spectral clustering of nonuniform hypergraphs. *In preparation*.

Background and graph community detection.

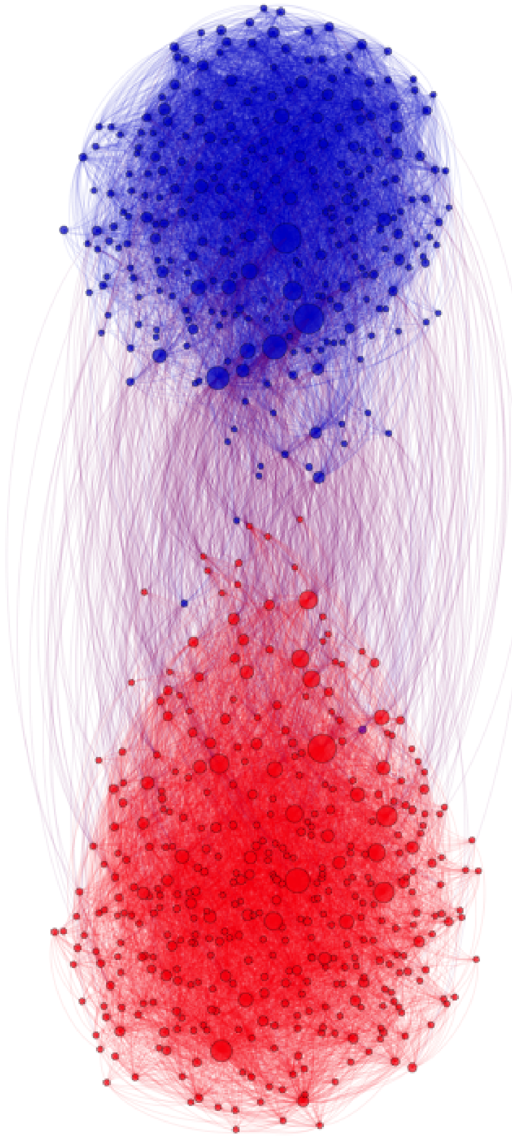
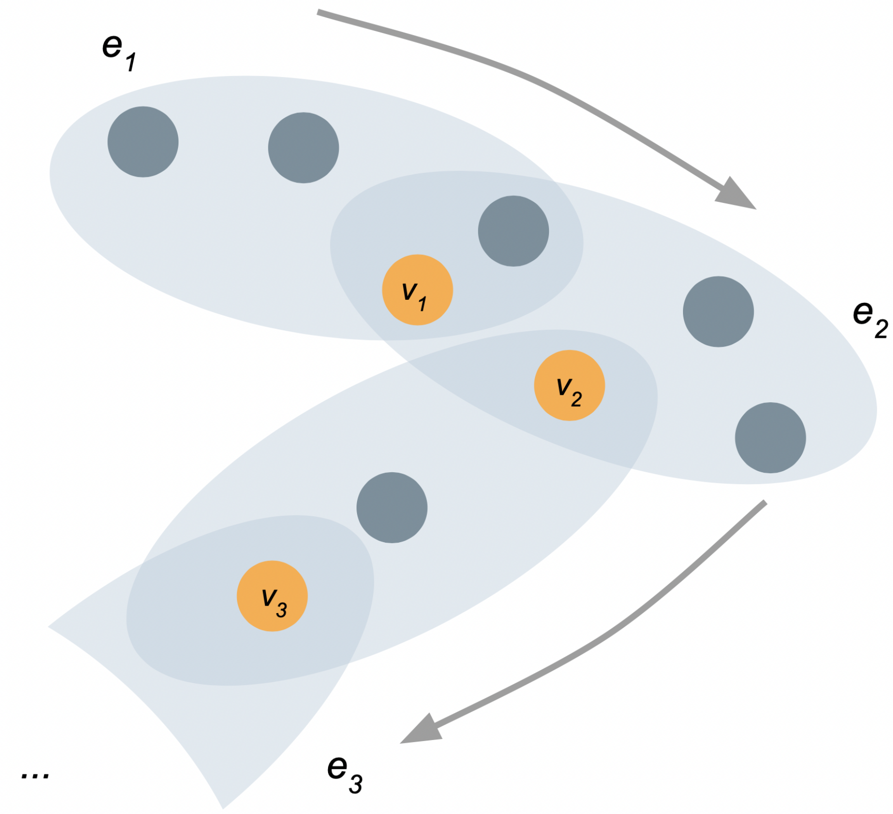


Image from All Things Graphed

Background and **graph community detection**.

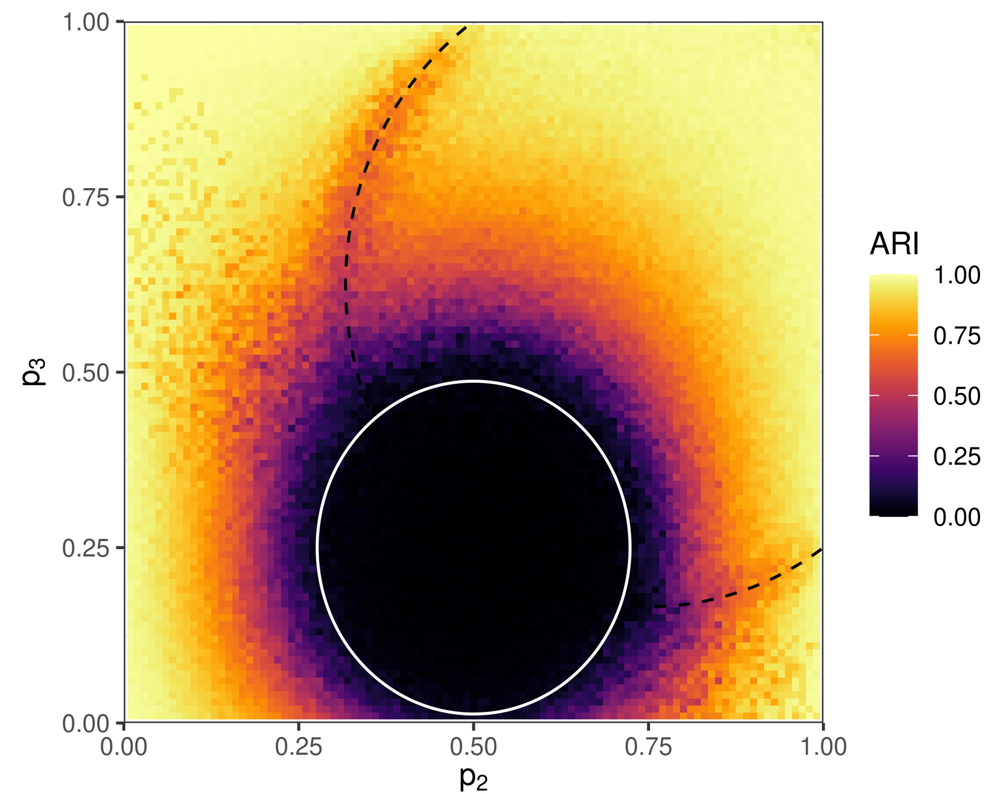
The **nonbacktracking operator** and eigenvector methods for hypergraphs.



Background and **graph community detection**.

The **nonbacktracking operator** and eigenvector methods for hypergraphs.

Detectability thresholds and open questions.



Background and **graph community detection**.

The **nonbacktracking operator** and eigenvector methods for hypergraphs.

Detectability thresholds and open questions.

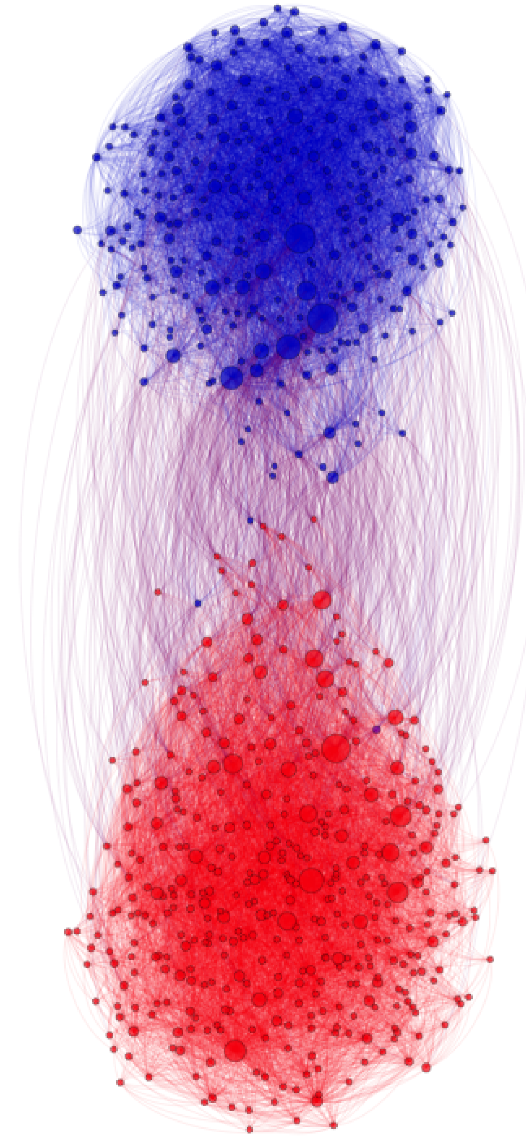
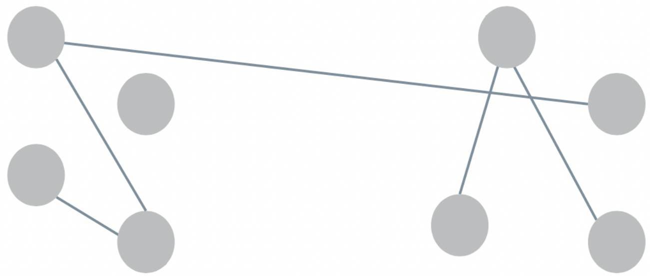


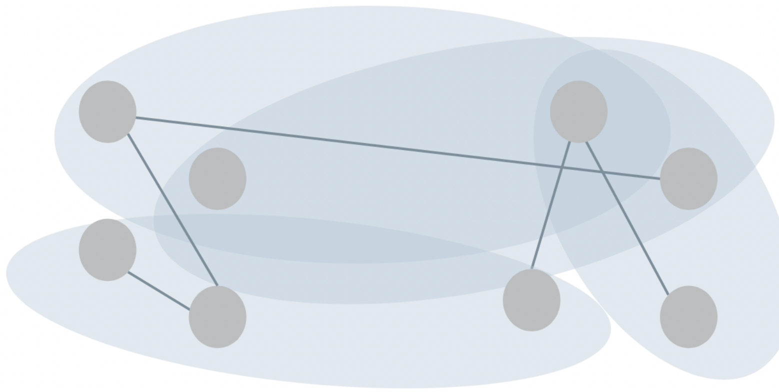
Image from All Things Graphed

Graphs and Hypergraphs

A **graph** consists of a set of nodes $\mathcal{N}$ and a set of edges $\mathcal{E}$. Each edge in $\mathcal{E}$ is a set of two nodes.



In **hypergraphs**, edges in $\mathcal{E}$ can contain *any number* of nodes.



Hypergraph Data

Hypergraph Data

- **Interaction**: nodes are agents, edges are interaction events (socializing in groups, attending events).

Hypergraph Data

- Interaction: nodes are agents, edges are interaction events (socializing in groups, attending events).
- Collaboration: nodes are collaborators, edges are projects or teams (scholarly papers, legislation, etc).

Hypergraph Data

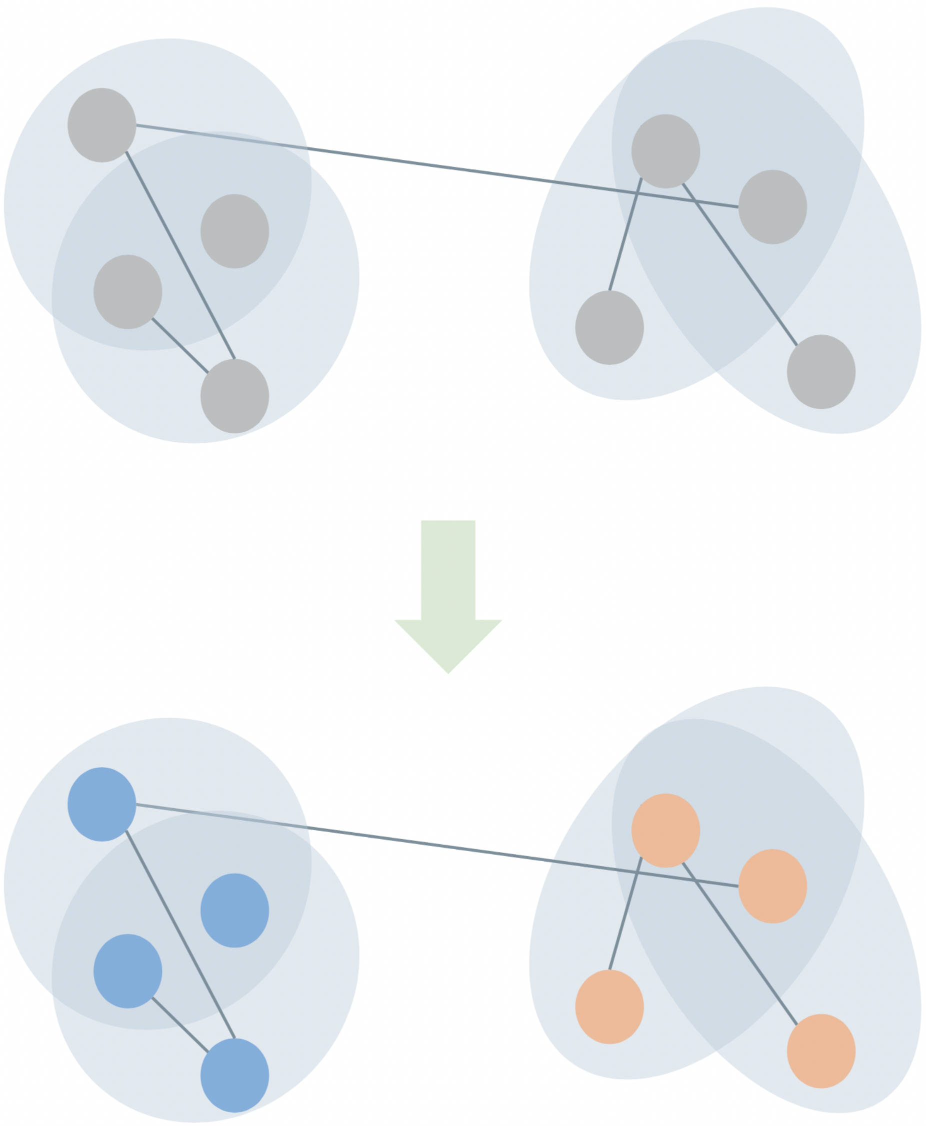
- **Interaction**: nodes are agents, edges are interaction events (socializing in groups, attending events).
- **Collaboration**: nodes are collaborators, edges are projects or teams (scholarly papers, legislation, etc).
- **Co-presence**: nodes are ingredients, edges are recipes formed from those ingredients.

The Hypergraph Community Detection Problem

Given some hypergraph data, assign each node to a *community* (or *cluster*) of "related" nodes.

"*Related*": often interpreted as "*densely interconnected.*"

Applications in social network analysis, drug discovery, image processing, data visualization...



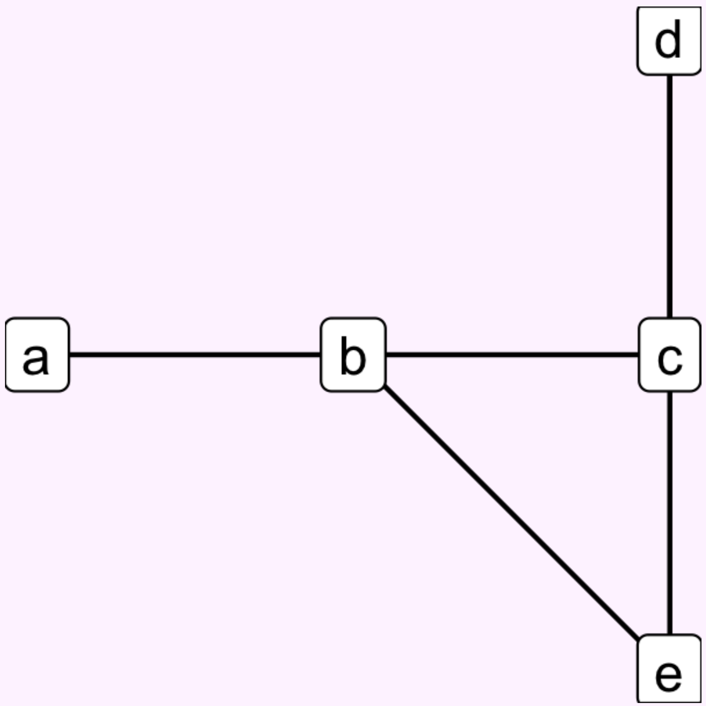
One review in:
P. S. Chodrow, N. Veldt, A. R. Benson (2021). Generative hypergraph clustering: from blockmodels to modularity, *Science Advances*, 7:eabh1303

Reminder

Vector $\mathbf{v} \in \mathbb{R}^n$ is an **eigenvector** of matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with **eigenvalue** $\lambda \in \mathbb{R}$ iff

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v} \, .$$

The graph...



...the adjacency matrix

$$\mathbf{A} = \begin{bmatrix} 0 & \mathbf{1} & 0 & 0 & 0 \\ \mathbf{1} & 0 & \mathbf{1} & 0 & \mathbf{1} \\ 0 & \mathbf{1} & 0 & \mathbf{1} & \mathbf{1} \\ 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & \mathbf{1} & \mathbf{1} & 0 & 0 \end{bmatrix}$$

We have

$$a_{ij} = \begin{cases} 1 & (i,j) \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases}$$

Modeling Graphs with Communities

Take n nodes and divide them into two group a and b.

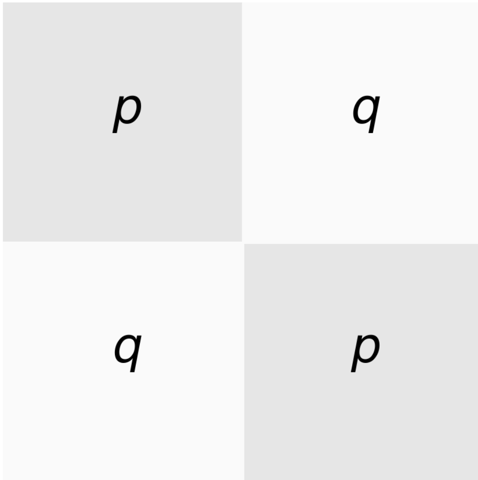
For each pair of nodes i and j, draw an edge with probability

$$p_{ij} = \begin{cases} p & i, j \text{ are in the same group} \\ q & i, j \text{ are in different group} \end{cases}$$

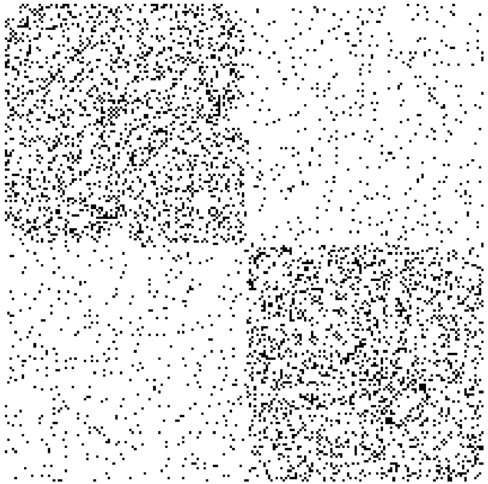
This gives us an adjacency matrix $\mathbf{A}$ with **noisy block structure**.

This is called a **stochastic blockmodel (SBM)**.

$\mathbf{P} =$



$\mathbf{A} =$



Communities and Eigenvectors

Leading eigenpairs of $\mathbf{P}$:

$$\lambda_1 = \frac{n}{2}(p+q)\ , \quad \mathbf{v}_1 = \underbrace{(1,\ldots,1,1)}_{n \text{ copies}}{}^T$$

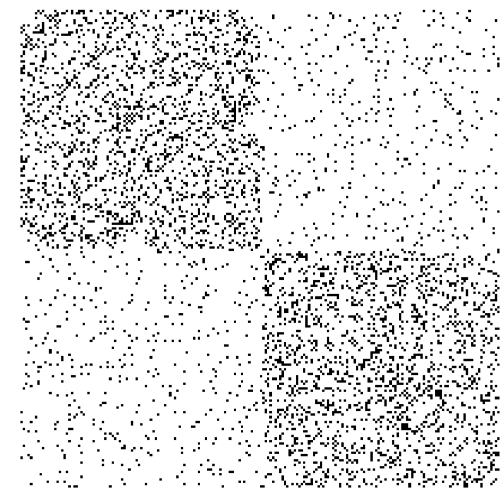
$$\lambda_2 = \frac{n}{2}(p-q)\ , \quad \mathbf{v}_2 = (\underbrace{1,\ldots,1}_{n/2 \text{ copies}}, \underbrace{-1,\ldots,-1}_{n/2 \text{ copies}})^T$$

Fact (from random matrix theory):
Eigenpairs of $\mathbf{A}$ are close to these with high probability as n grows large.

Nadakuditi and Newman (2012). Graph spectra and the detectability of community structure in networks, *Physical Review Letters*

$$\mathbf{P} = \begin{pmatrix} p & q \\ q & p \end{pmatrix}$$

$\mathbf{A} =$

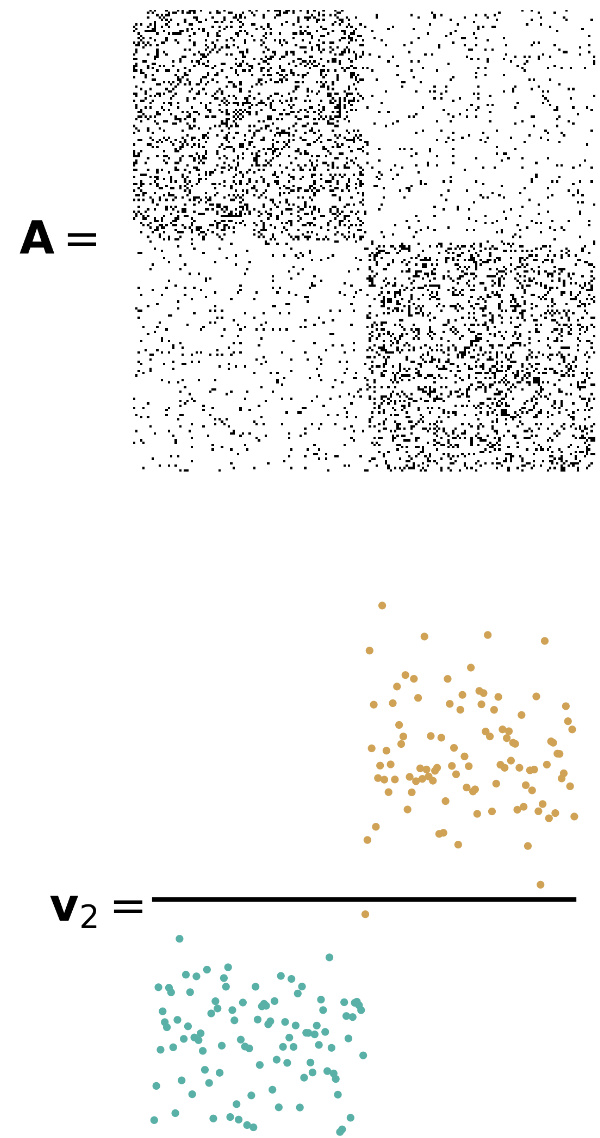


A Graph Community Detection Algorithm

1. Compute the second-largest eigenvector $\mathbf{v}_2$ of $\mathbf{A}$.
2. If $v_{2i} > 0$, guess that node i is in group a, otherwise in group b.

Works well if $p - q$ is sufficiently large, variations work for other graph matrices.

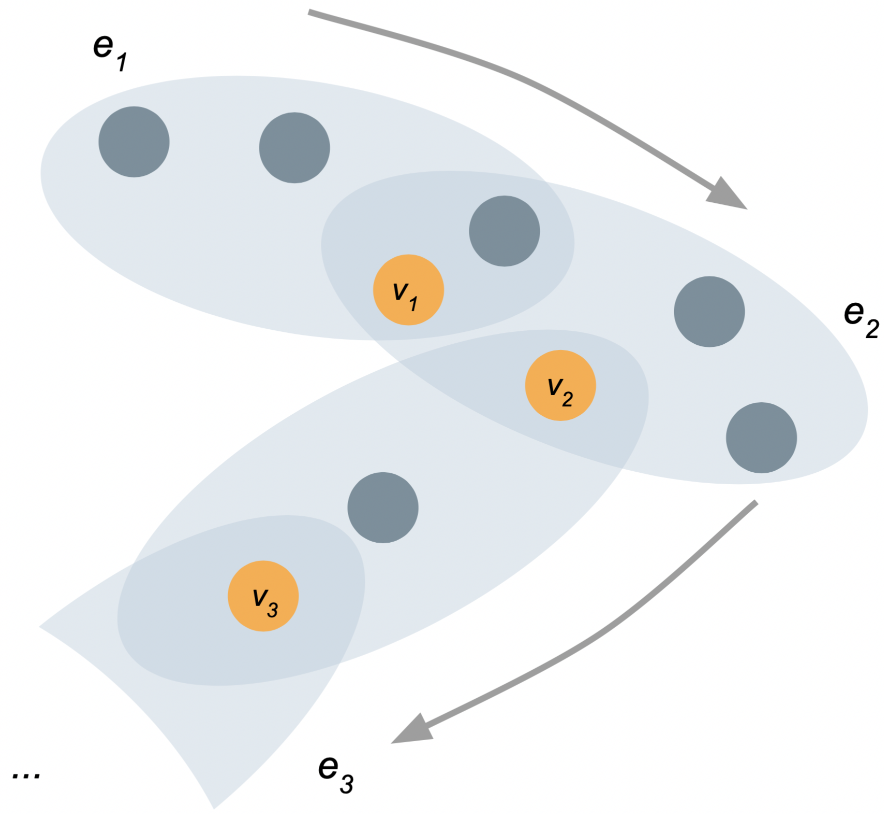
- Decelle et al. (2011) Inference and phase transitions in the detection of modules in sparse networks, *Phys. Rev. Let.* 107.6: 065701
- Krzakala et al. (2013) Spectral redemption in clustering sparse networks, *PNAS* 110 (52) 20935-20940



Background and **graph community detection**.

The **nonbacktracking operator** and eigenvector methods for hypergraphs.

Detectability thresholds and open questions.



Matrices for Hypergraphs?

We could transform the hypergraph into a graph.

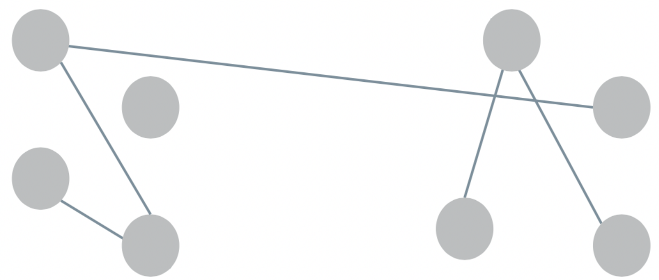
- **Problem**: loses multi-way information.

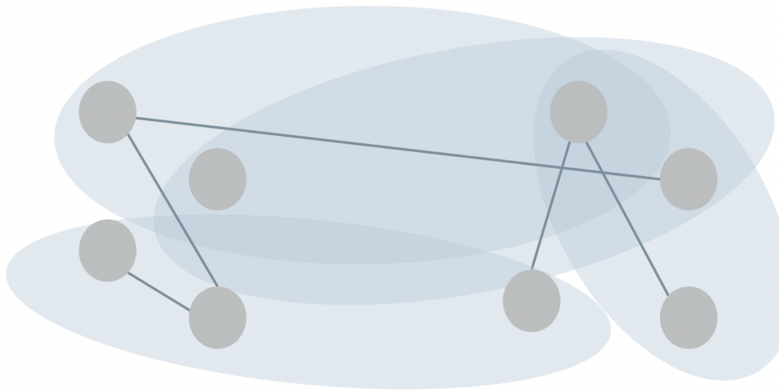
We could construct a set of adjacency tensors $\mathbf{A}^{(2)}$, $\mathbf{A}^{(3)}$, $\mathbf{A}^{(4)}$...

$$a^{(3)}_{ijk} = \begin{cases} 1 & (i,j,k) \in \mathcal{E} \\ 0 & \text{otherwise...} \end{cases}$$

- **Problem**: we know eigenvectors of tensors, but not ***sets*** of tensors.

So, what should we do?....





The Nonbacktracking Operator

The adjacency matrix is $n \times n$ and operates on nodes.

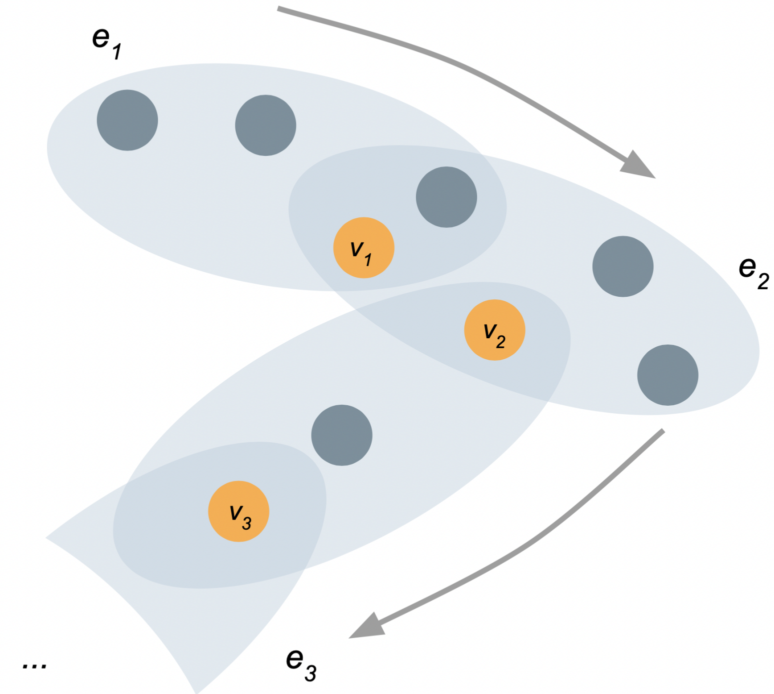
The **nonbacktracking operator** $\mathbf{B}$ is a matrix that operates on *edge-node pairs*. It is often referred to as the **Hashimoto operator**.

Define relation $(e_1, v_1) \rightarrow (e_2, v_2)$:

- $v_1 \in e_1$ and $v_2 \in e_2$
- $v_1 \in e_2 \setminus v_2$
- $e_1 \neq e_2$

Then,

$$\mathbf{B}[(e_1, v_1), (e_2, v_2)] = \begin{cases} 1 & (e_1, v_1) \rightarrow (e_2, v_2) \\ 0 & \text{otherwise.} \end{cases}$$



"I can get to $v_2 \in e_2$ from e_1 by passing through v_1. I can get to $v_3 \in e_3$ from e_2 by passing through v_2..."

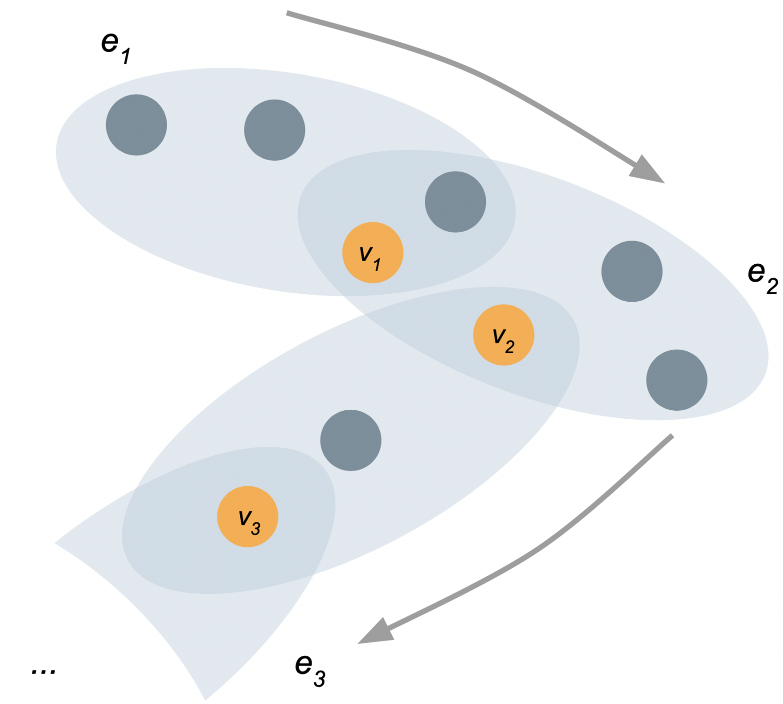
The Nonbacktracking Operator

Popularized (for graphs) by Hashimoto, K. (1990), *Int. J. Math.*

Important theorem for computation by Bass, H. (1992), *Int. J. Math.*

Formulated for hypergraphs by Storm, C. K. (2006). *The Electronic Journal of Combinatorics*.

"Rediscovered" for hypergraphs by Angelini, M. C. et al. (2015), *Allerton Conference*.



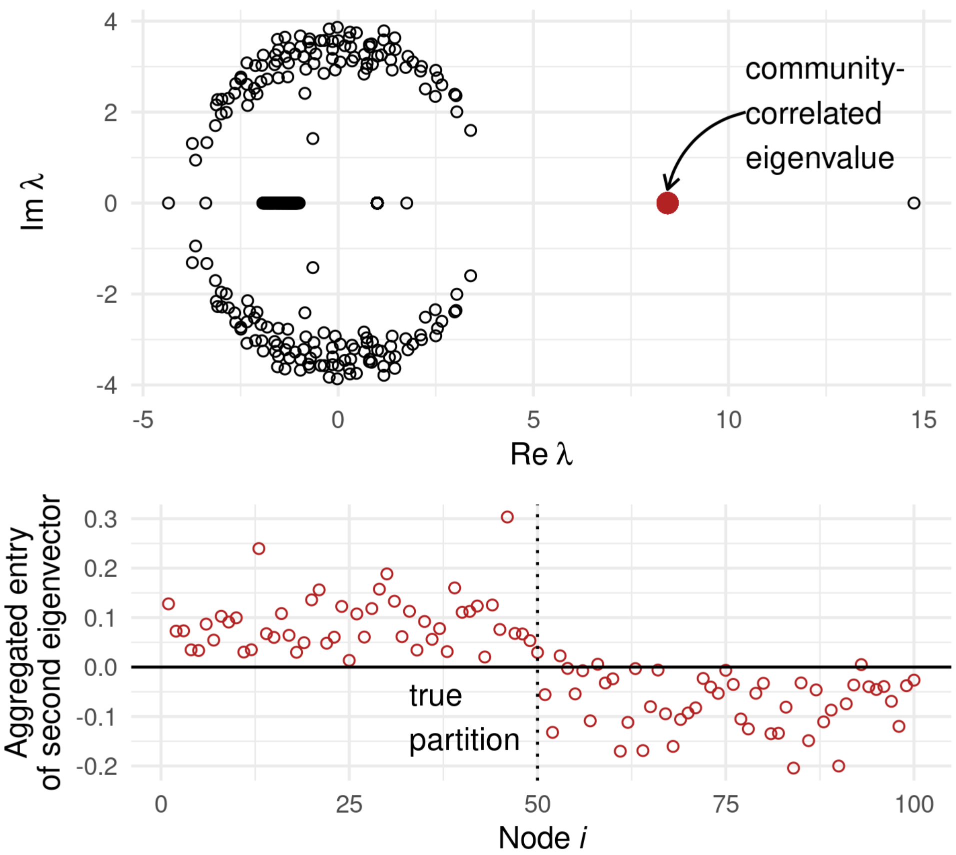
"I can get to $v_2 \in e_2$ from e_1 by passing through v_1. I can get to $v_3 \in e_3$ from e_2 by passing through v_2..."

The Nonbacktracking Operator

Cool topological connections: prime cycles and zeta functions.

Can represent hyperedges of all sizes in the same matrix!

In our stochastic blockmodel from before, eigenvector $\mathbf{v}_2$ is correlated with communities if λ_2 is real.



Precise control over community-correlated eigenvalues in the graph case:

Bordenave et al. (2018): Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. *Annals of Probability*.

Issue: Computation

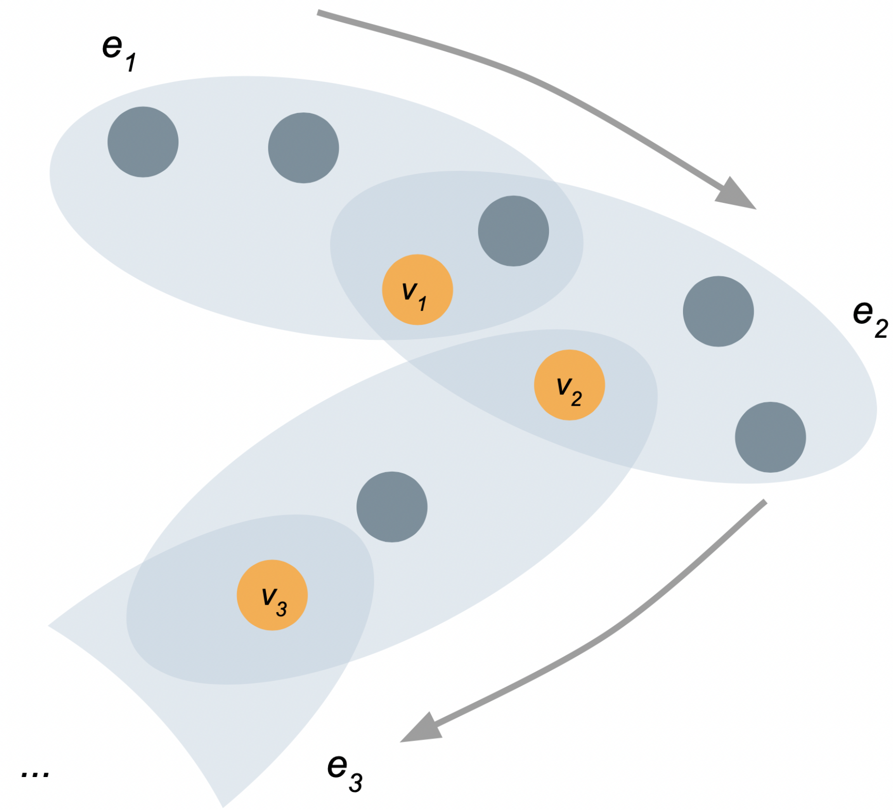
$\mathbf{B}$ is indexed by edge-node pairs.

So, $\mathbf{B}$ is of size $m\langle k\rangle \times m\langle k\rangle$, where m is the number of edges and $\langle k\rangle$ is the average edge size.

A ***small*** data set might have $n = 300$ nodes, $m = 8,000$ edges, and average edge size 2.5.

$m\langle k\rangle = 8,000 \times 2.5 = 20,000$, which is already a pretty big matrix.

Eigenpair computations get expensive fast...



A Generalized Ihara-Bass Theorem

Theorem (PSC, JH, NE '22): Under mild conditions, if λ is an eigenvalue of $\mathbf{B}$, then either:

1. $\lambda \in \{1, -1, -2, \ldots, 1 - \bar{k}\}$ and carries no structural information about the hypergraph, or
2. λ is an eigenvalue of the matrix

$$\mathbf{B}' = \begin{bmatrix} \mathbf{0} & \mathbb{D} - \mathbf{I}_{\bar{k}n} \\ (\mathbf{I}_{\bar{k}} - \mathbf{K}) \otimes \mathbf{I}_n & \mathbb{A} + (2\mathbf{I}_{\bar{k}} - \mathbf{K}) \otimes \mathbf{I}_n \end{bmatrix} \in \mathbb{R}^{2\bar{k}n \times 2\bar{k}n} .$$

- $\bar{k}$ is the number of distinct edge sizes, n is the number of nodes.
- $\mathbb{A} \in \mathbb{R}^{\bar{k}n \times \bar{k}n}$ collects adjacency information for each hyperedge size.
- $\mathbb{D} \in \mathbb{R}^{\bar{k}n \times \bar{k}n}$ collects node degrees for each hyperedge size.
- $\mathbf{K} \in \mathbb{R}^{\bar{k} \times \bar{k}}$ lists possible edge sizes.
- $\mathbf{I}_\ell \in \mathbb{R}^{\ell \times \ell}$ is the matrix identity of size ℓ.
- $\otimes$ is the Kronecker product.

Proof Sketch

1. $\mathbf{B}$ can be written as $\mathbf{ST} - \mathbf{R}$ for suitable operators $\mathbf{S}$, $\mathbf{T}$ and $\mathbf{R}$, which also satisfy handy relations like $\mathbf{TS} = \mathbb{A}$.
2. Consider $\det(\lambda \mathbf{I} - \mathbf{B})$, substitute $\mathbf{B} = \mathbf{ST} - \mathbf{R}$, and use the *push-through identity*:

$$\det(\mathbf{X} + \mathbf{YZ}) = \det(\mathbf{X}) \det(\mathbf{I} + \mathbf{ZX}^{-1}\mathbf{Y})$$

   (*provided all inverses, sums, and products are defined*).
3. Simplify, obtaining

$$\det(\lambda \mathbf{I} - \mathbf{B}) = \det(\lambda \mathbf{I} - \mathbf{B}') \det(\mathbf{uninformative\ part})\,.$$

Approach based on a proof of the the graph Ihara-Bass formula in:
M. C. Kempton (2016). Non-backtracking random walks and a weighted Ihara's theorem. *Open Journal of Discrete Mathematics* 6, 207-226

Issue: Computation

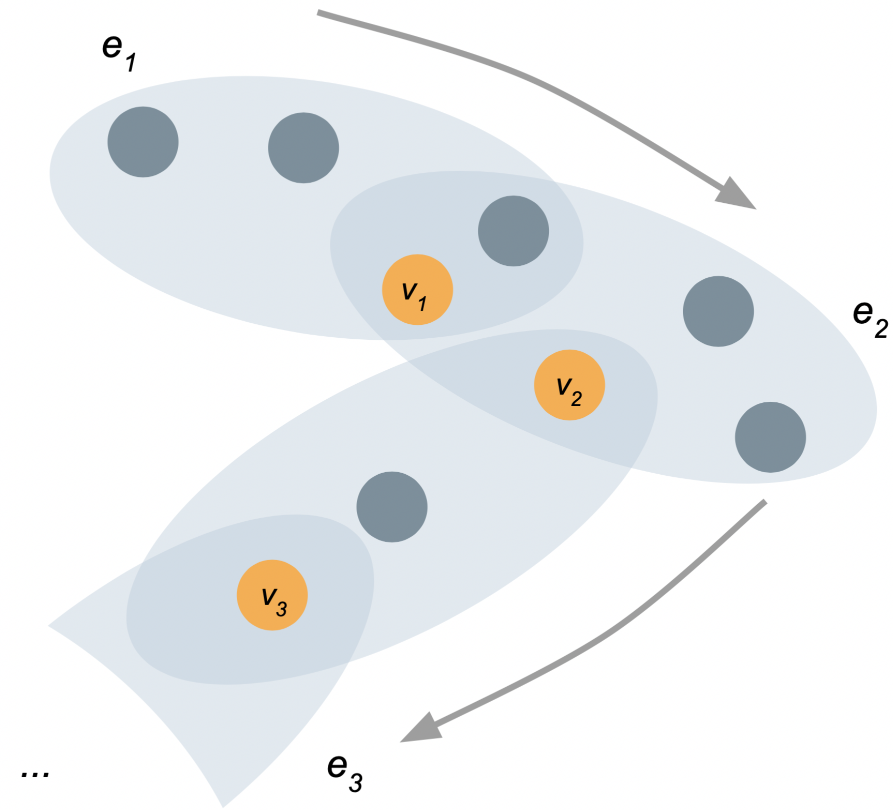
A ***small*** data set might have $n = 300$ nodes, $m = 8,000$ edges, and average edge size 2.5.

If $\bar{k} = 3$, then we can compute eigenvectors in

$$2n\bar{k} = 1,800 \ll 20,000 = m\langle k \rangle$$

dimensions instead.

We can do that 100x-1,000x faster!

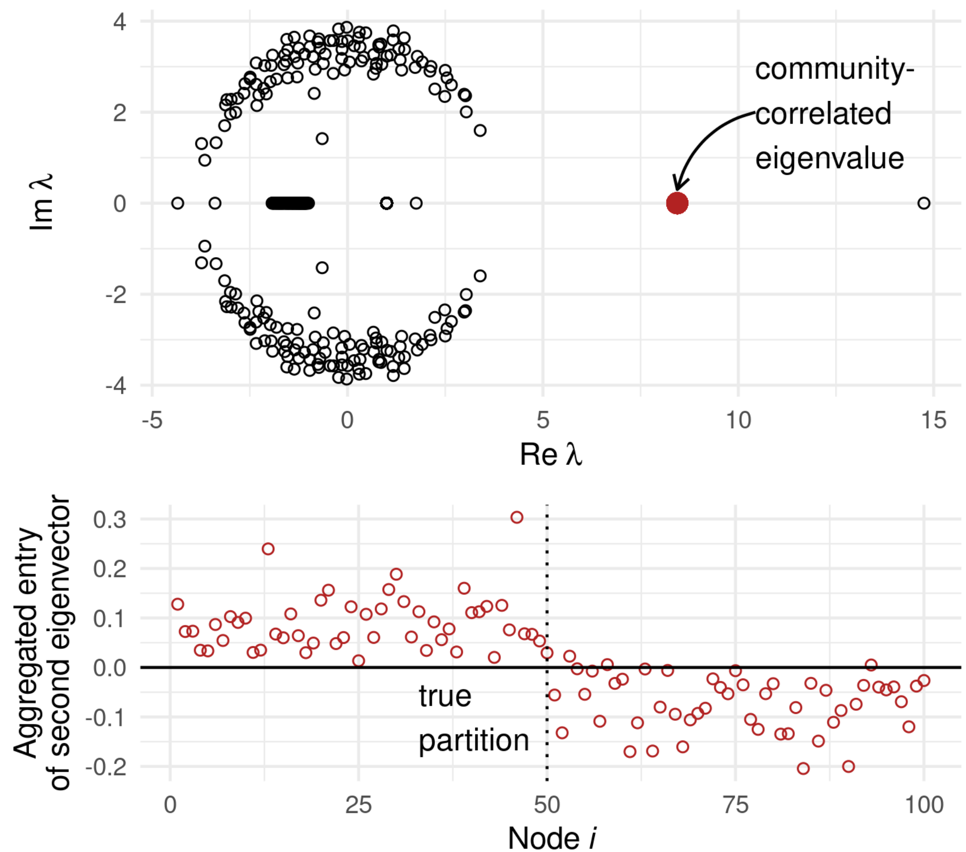


First Algorithm

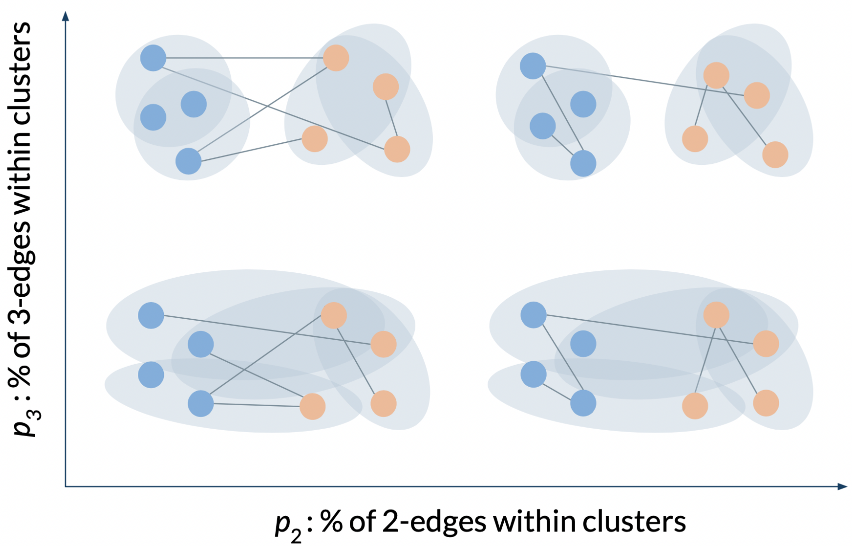
1. Compute the second eigenpair $(\lambda_2, \mathbf{v}_2)$ of $\mathbf{B}'$.
2. If λ_2 is real, separate $\mathbf{v}_2 = (\alpha, \beta)$, with $\alpha, \beta \in \mathbb{R}^{n\bar{k}}$.
3. If

$$u_i = \sum_{k=1}^{\bar{k}} \alpha_{ik} < 0 \ ,$$

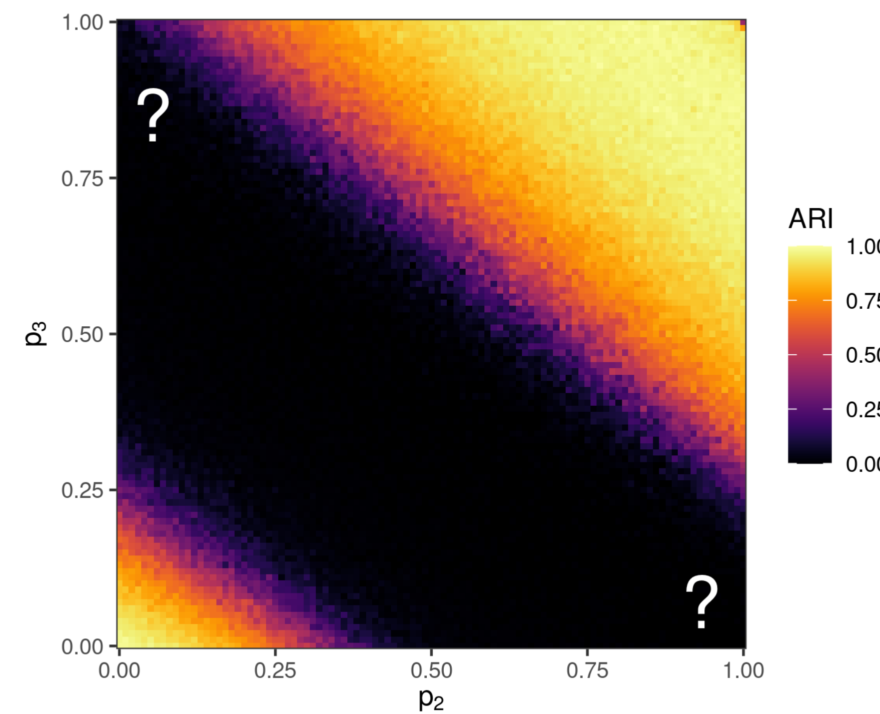
assign i to cluster A, else assign i to cluster B.



Synthetic Testbed



Eigenvector Algorithm



Adjusted Rand Index (ARI):

- ARI = 1: perfect community detection.
- ARI = 0: random noise.

Belief Propagation…

Belief Propagation...

...is the "cavity method" of statistical physics.

Belief Propagation…

…is the "cavity method" of statistical physics.

…is an approximate method for statistical inference.

Belief Propagation...

...is the "cavity method" of statistical physics.

...is an approximate method for statistical inference.

...is a discrete-time dynamical system.

Belief Propagation…

…is the "cavity method" of statistical physics.

…is an approximate method for statistical inference.

…is a discrete-time dynamical system.

Formally, iterate these updates to convergence:

$$\mu_{iR}^{(s)} \leftarrow \frac{1}{Z_{iR}} \prod_{Q \in \binom{[n]}{|R|} \setminus R} \nu_{Qi}^{(s)}$$

$$\nu_{Ri}^{(s)} \leftarrow \frac{1}{Z_{Ri}} \sum_{\mathbf{z} : z_i = s} \mathbb{P}(a_R | \mathbf{z}_R) \prod_{j \in R \setminus i} \mu_{jR}^{(z_j)}$$

$\mu_{iR}^{(s)}$ is "node i's confidence that it belongs to community s based on other nodes in tuple R."

Z_{iR} and Z_{Ri} are normalization constants.

$\mathbb{P}(a_R|\mathbf{z}_R)$ is our stochastic blockmodel: specifies how likely there are to be a_R edges on tuple R given some community labels $\mathbf{z}_R$.

A Linear Approximation

Theorem (PSC, NE, JH '22): Consider a stochastic blockmodel in which:

- Every node has the same expected number of attached edges.
- The expected number of attached edges does not depend on the number of nodes n.

Then:

- BP has an approximate fixed point $\bar{\mathbf{x}}$ that contains no cluster information.
- The Jacobian derivative $\mathcal{J}(\bar{\mathbf{x}})$ of the BP dynamics around $\bar{\mathbf{x}}$ has $O(n^{-1})$ entries, except for a block of the form

$$\mathbf{J} = \sum_{k=1}^{\bar{k}} \mathbf{C}_k \otimes \mathbf{B}_k + O(n^{-1}) \ .$$

- $\mathbf{C}_k$ is a matrix of parameters that depends on the stochastic blockmodel $\mathbb{P}$.
- $\mathbf{B}_k$ is our friend the nonbacktracking operator, restricted to edges of size k.
- $\otimes$ is the Kronecker product.

Result argued heuristically for graphs in:

Krzakala et al. (2013) Spectral redemption in clustering sparse networks, *PNAS* 110 (52) 20935-20940

A Cheat

$\mathbf{J}$ can be a *very* large matrix.

As before, we can use a smaller one:

Theorem (PSC, JH, NE '22): Under mild conditions, if λ is an "interesting" eigenvalue of $\mathbf{J}$, then λ is also an eigenvalue of the $2n\ell\bar{k}$ matrix

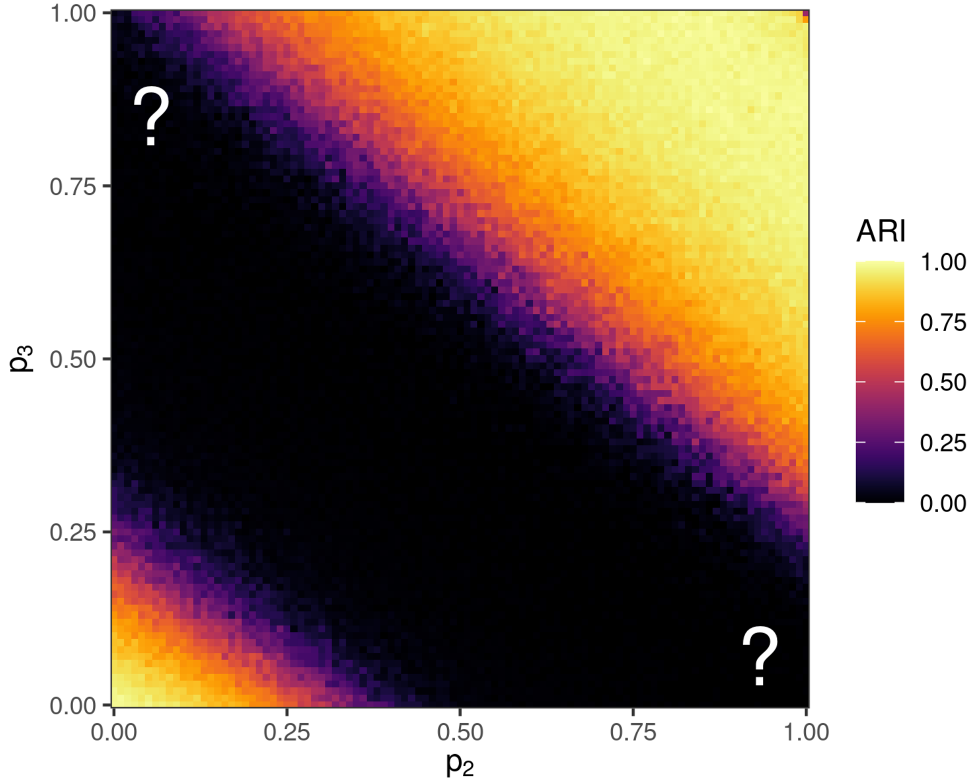
$$\mathbf{J}' = (\mathbf{G} \otimes \mathbf{I}_n) \begin{bmatrix} \mathbf{0} & \mathbf{I}_\ell \otimes \mathbb{D} \\ \mathbf{0} & \mathbf{I}_\ell \otimes \mathbb{A} \end{bmatrix} - \bar{\mathbf{G}} \begin{bmatrix} \mathbf{0} & \mathbf{I}_\ell \otimes \mathbf{I}_{\bar{k}} \\ \mathbf{I}_\ell \otimes (\mathbf{K} - \mathbf{I}_{\bar{k}-1}) & \mathbf{I}_\ell \otimes (\mathbf{K} - 2\mathbf{I}_{\bar{k}-1}) \end{bmatrix} \otimes \mathbf{I}_n$$

where ℓ is the number of communities and $\mathbf{G}$, $\bar{\mathbf{G}}$ hold statistical parameters.

Proof is a little messier this time.

Ok, but does it work?

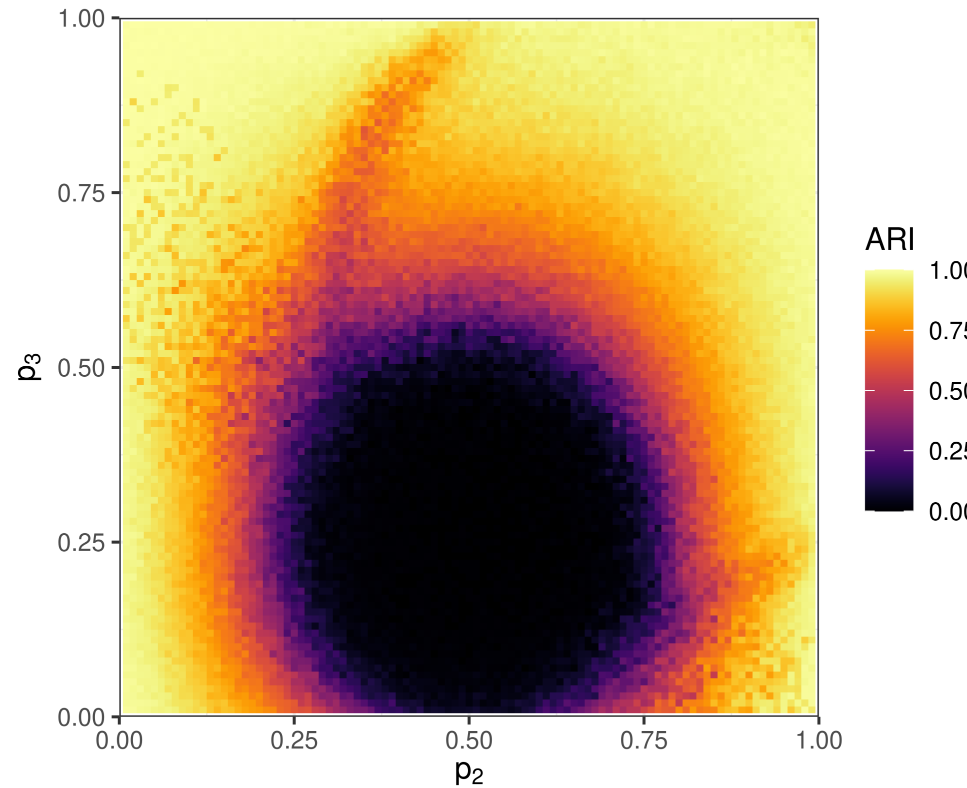
Recall that we were having issues with parameter combinations in which edges of different sizes carried different kinds of information.



Ok, but does it work?

Recall that we were having issues with parameter combinations in which edges of different sizes carried different kinds of information.

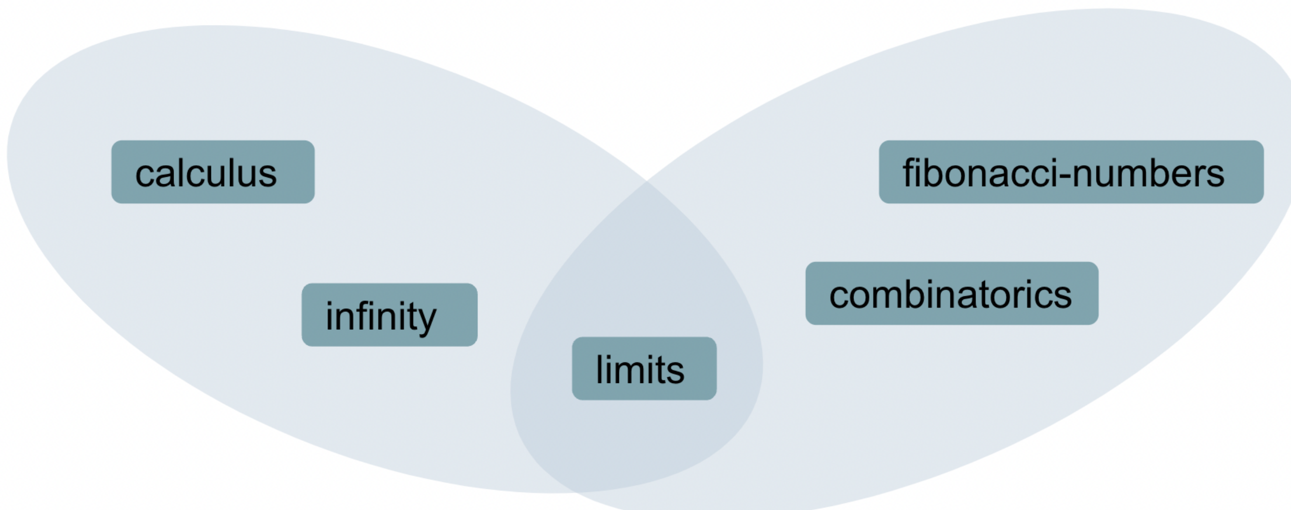
Working with the more complicated matrix $\mathbf{J}$ increases computation time, but also allows us to detect communities for more parameter combinations.



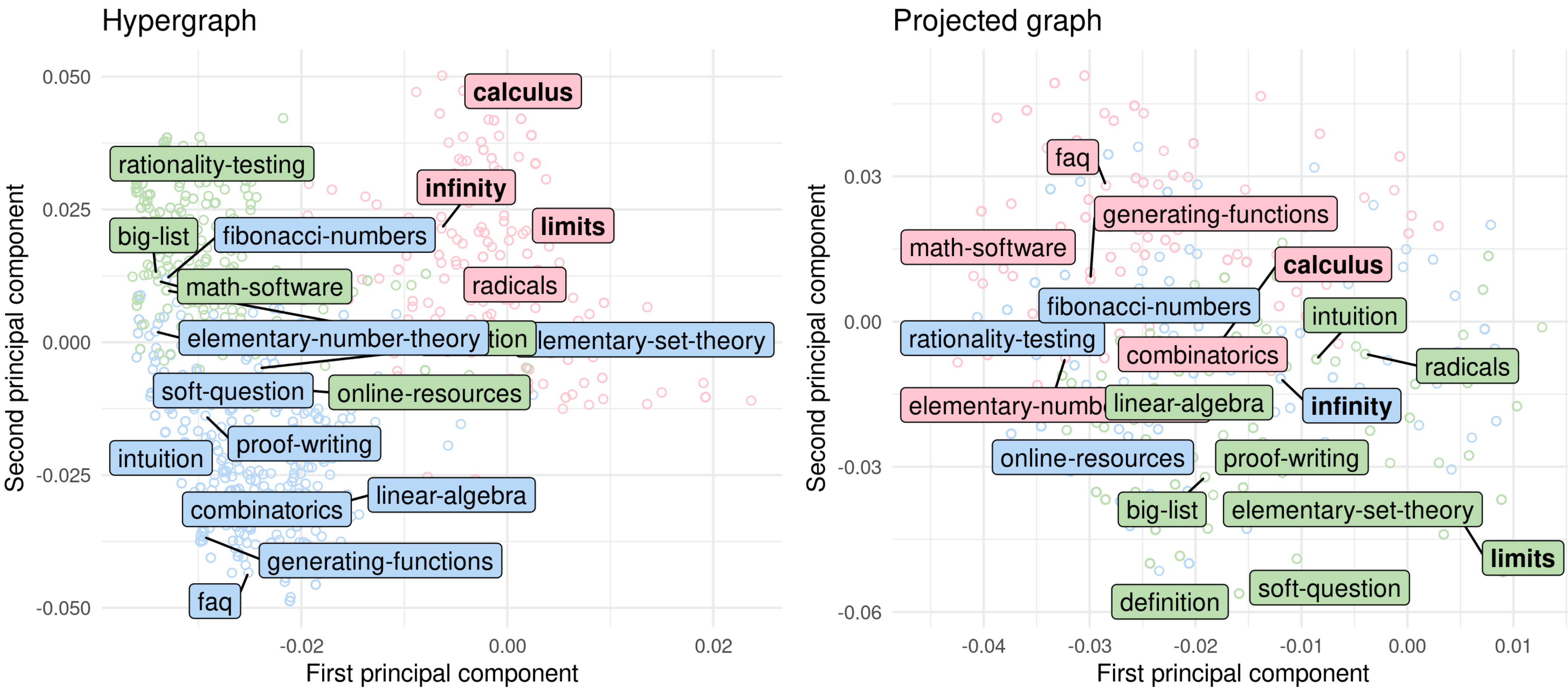
Example Data: Mapping Math with StackExchange Tags

“Does the derivative of a function at a vertical asymptote have to approach infinity?”

“Proof that Fibonacci numbers grow faster than any polynomial?”



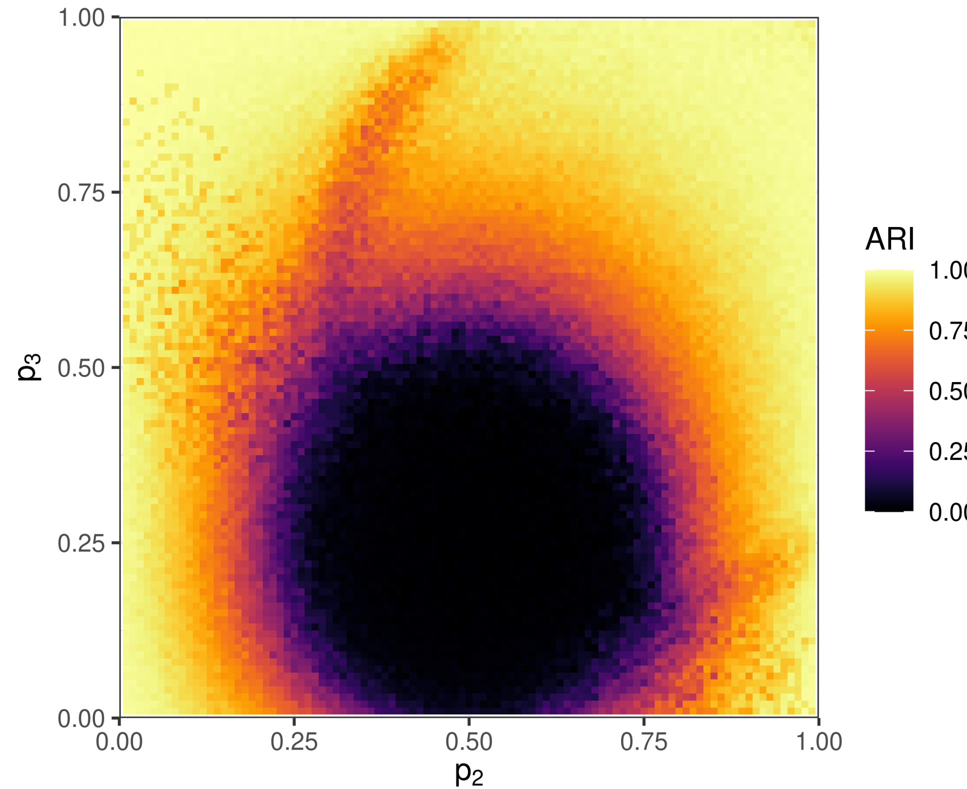
Example Data: Mapping Math with StackExchange Tags



Background and **graph community detection**.

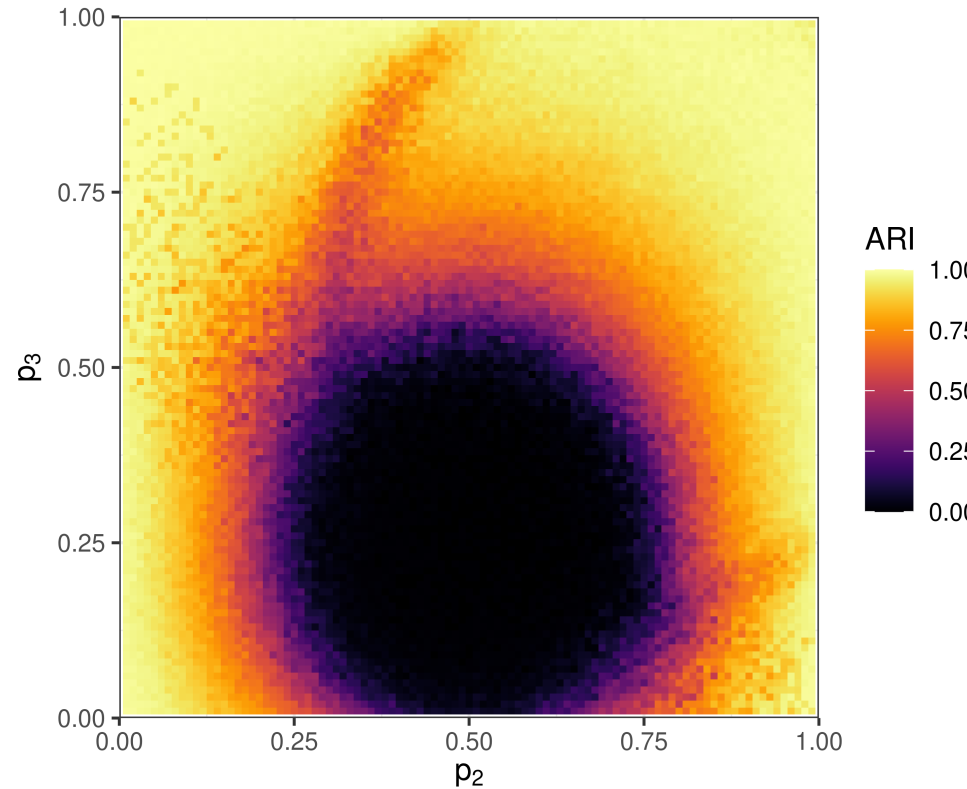
The **nonbacktracking operator** and eigenvector methods for hypergraphs.

Detectability thresholds and open questions.



Algorithmic Thresholds

Recall the ***suspiciously round*** region where our algorithm totally failed to learn any cluster information.

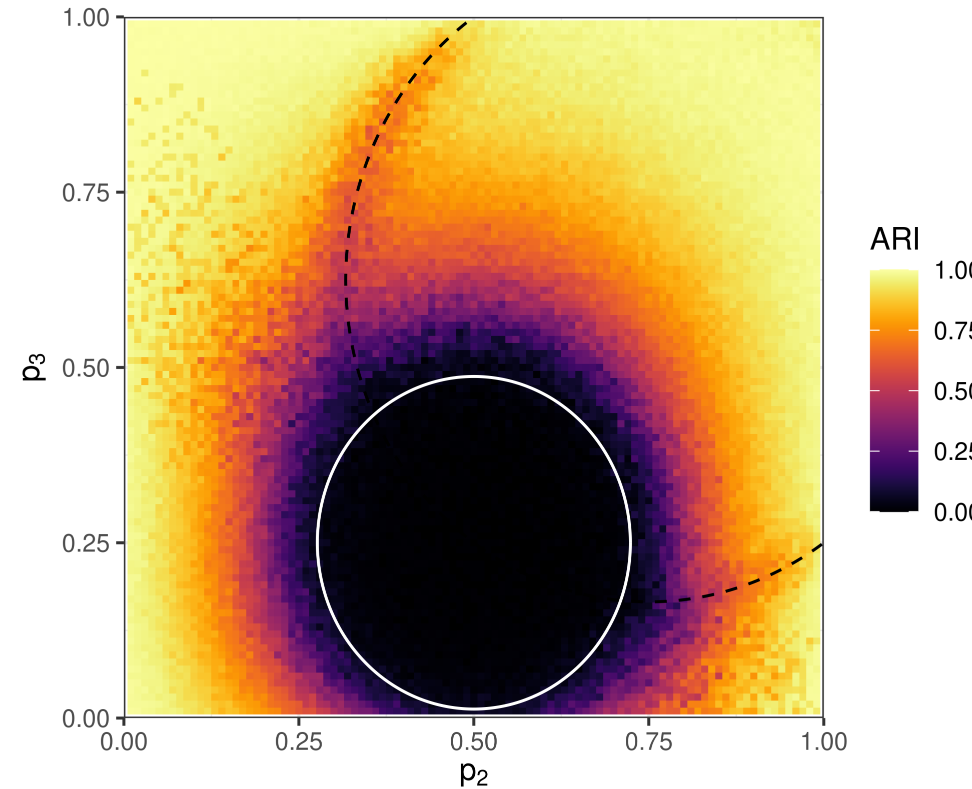


Algorithmic Thresholds

Recall the ***suspiciously round*** region where our algorithm totally failed to learn any cluster information.

This region can be estimated!

Strategy: ask when $\mathbf{J}$ has an eigenvalue > 1, using approximations analogous to known results for graphs.



Algorithmic Thresholds

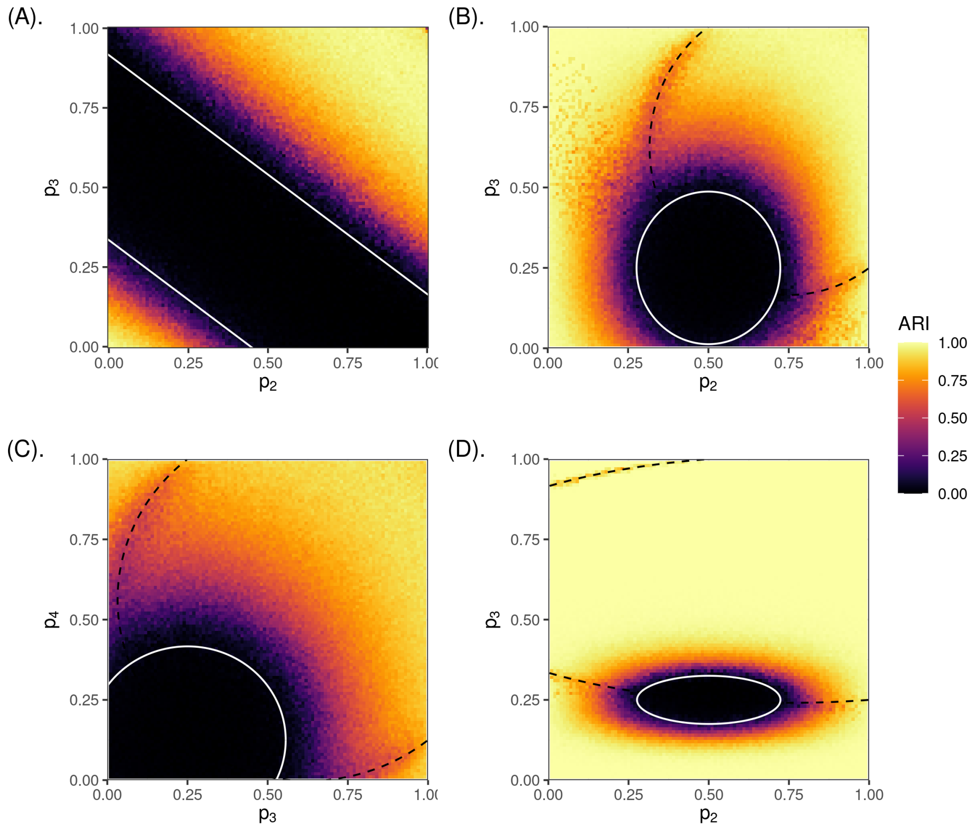
Conjecture: In a 2-group testbed with edge sizes $k_1, k_2, \ldots$ and c_k edges of size k per node, detection is possible outside the ellipsoid with centroid $(x_{k_1}, x_{k_2}, \ldots)$ and radii $(r_{k_1}, r_{k_2} \ldots)$, where:

$$x_k = \frac{1 - a_k}{2 - a_k}$$

$$r_k = \frac{\sqrt{(k-1)c_k}}{2 - a_k}$$

$$a_k = \frac{1 - 2^{2-k}}{1 - 2^{1-k}} .$$

Proof will involve some random matrix theory (future work).



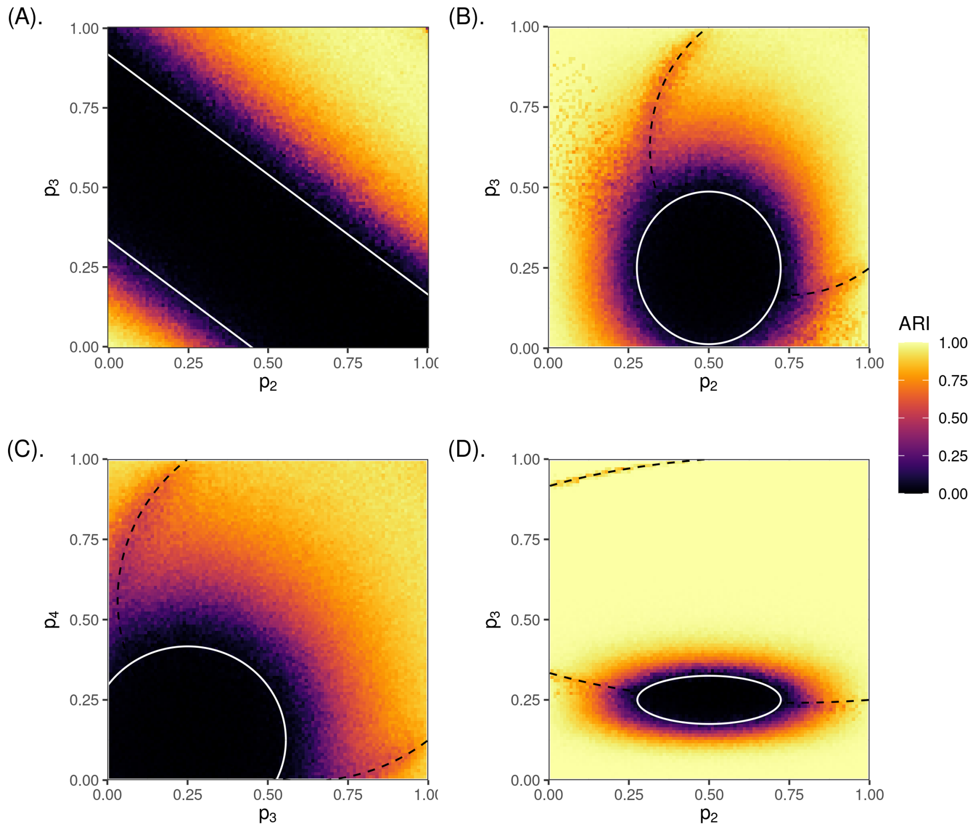
Detectability Thresholds

In graphs, failure of nonbacktracking spectral clustering coincides with an **information-theoretic bound** on the clustering problem.

- **No algorithm** can reliably detect communities.

We conjecture the same thing for hypergraphs: inside that ellipse, the clustering problem is not just difficult but *theoretically* impossible.

Recent proof for graphs:
Mossel et al. (2018) A proof of the blockmodel threshold conjecture, *Combinatorica*.

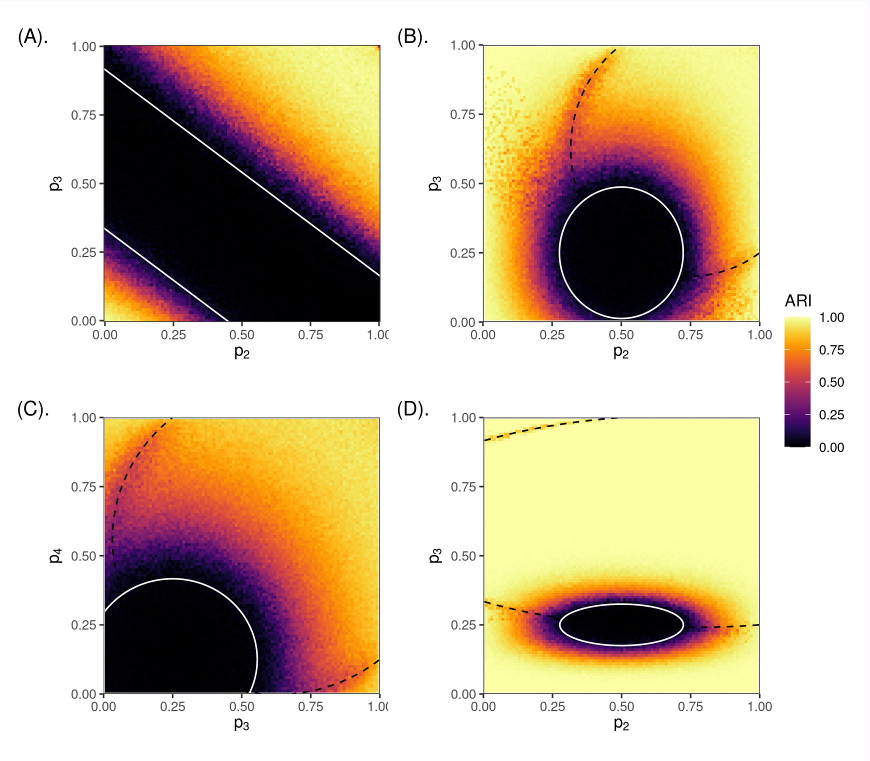


Wrapping Up

Future Directions

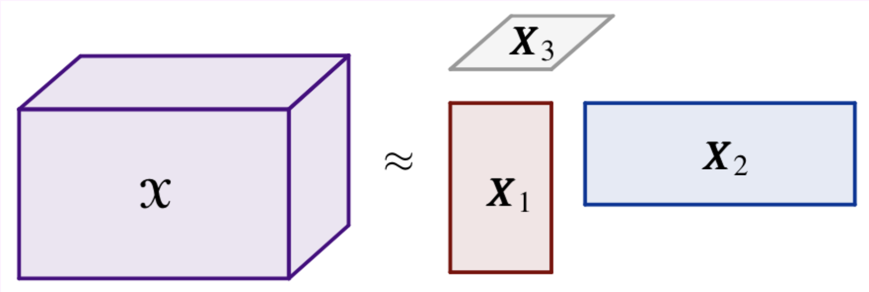
Future Directions

- Prove the conjectured detectability threshold.



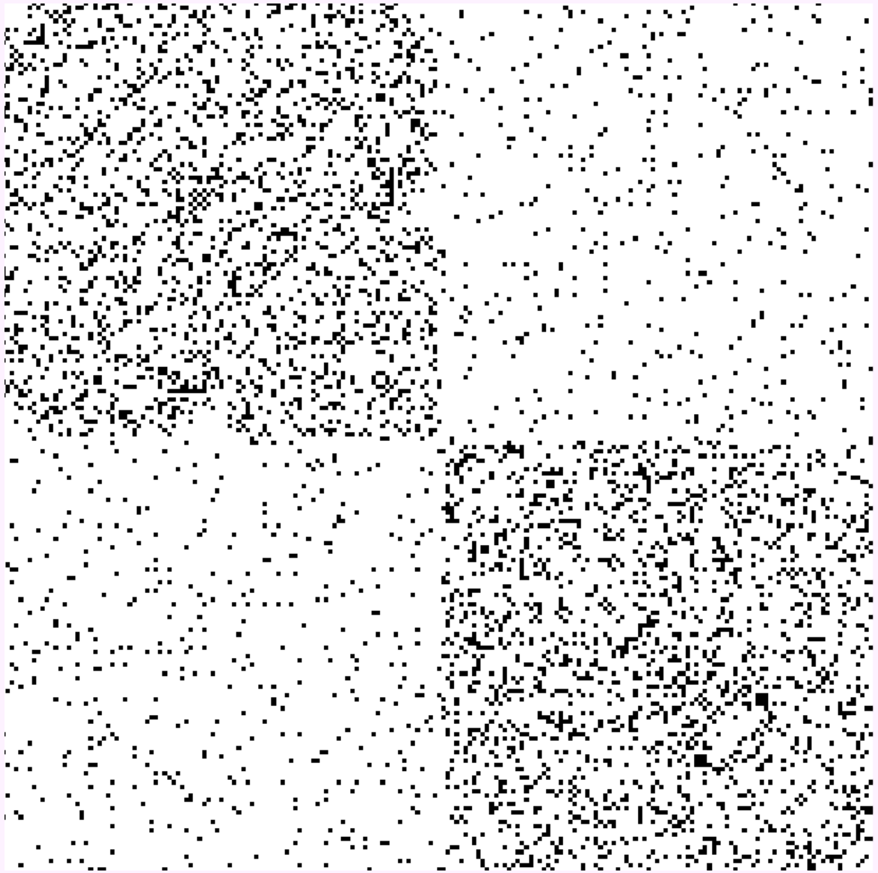
Future Directions

- Prove the conjectured detectability threshold.
- Develop tensor-based hypergraph community detection methods.



Future Directions

- Prove the conjectured detectability threshold.
- Develop tensor-based hypergraph community detection methods.
- Consider other hypergraph block models.



Summary

The nonbacktracking operator enables eigenvector techniques for community detection in hypergraphs.

Determinant identities help us speed up computation.

There are open questions around the fundamental limits of hypergraph community detection.

Phil Chodrow
UCLA
@PhilChodrow

Nicole Eikmeier
Grinnell
@NicoleEikmeier

Philip S. Chodrow, Nicole Eikmeier, and **JH** (2022). Nonbacktracking spectral clustering of nonuniform hypergraphs. *In preparation*.

Thanks everyone!

Questions?

Extra slides

High School Social Contacts

$n = \mathbf{327}$ students (nodes) in a French high school.

$m = \mathbf{7,818}$ social contact events (edges) measured by wearable sensors.

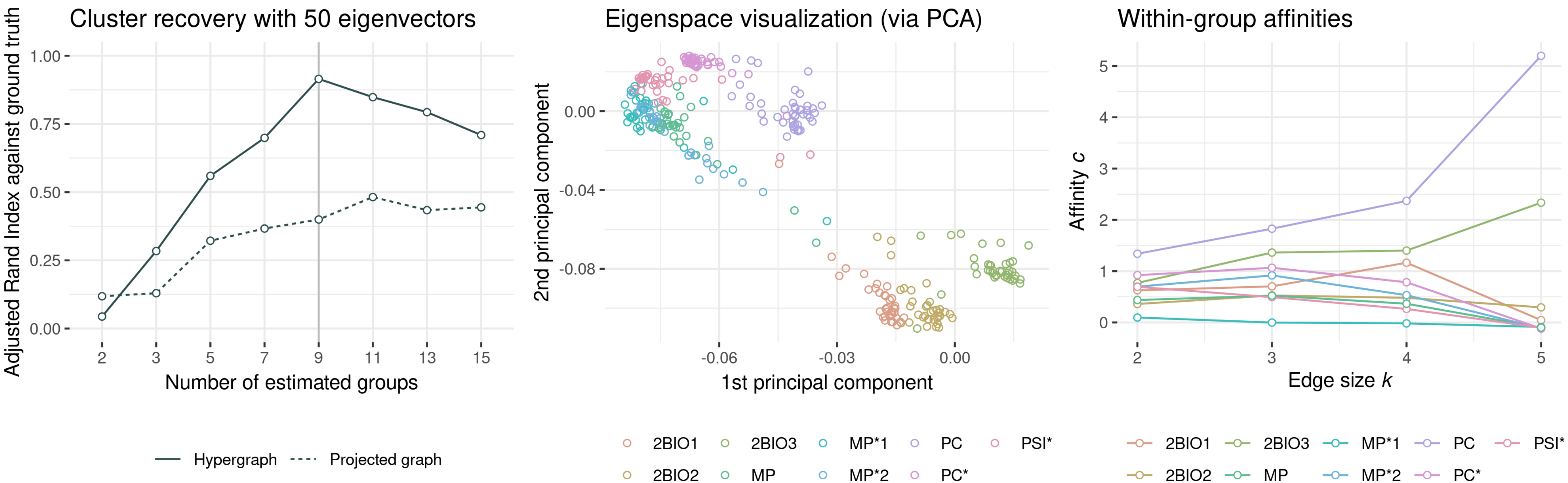
Average number of participants in interaction $\langle \boldsymbol{k} \rangle = \mathbf{2.3}$

Cluster labels are the classes to which students are assigned.

Data originally from:
R. Mastrandrea et al. (2015), Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries, and friendship surveys. *PLoS One* 10:9, e0136497

Prepared by A. R. Benson et al. (2018), Simplicial closure and higher-order link prediction. *Proceedings of the National Academy of Sciences* 10.1073/pnas.1800683115

High School Social Contacts



On the Other Hand...Senate Bills

$n = \mathbf{293}$ U.S. senators (nodes) cosponsoring bills.

$m = \mathbf{20,006}$ bills (edges) in period 1973-2016.

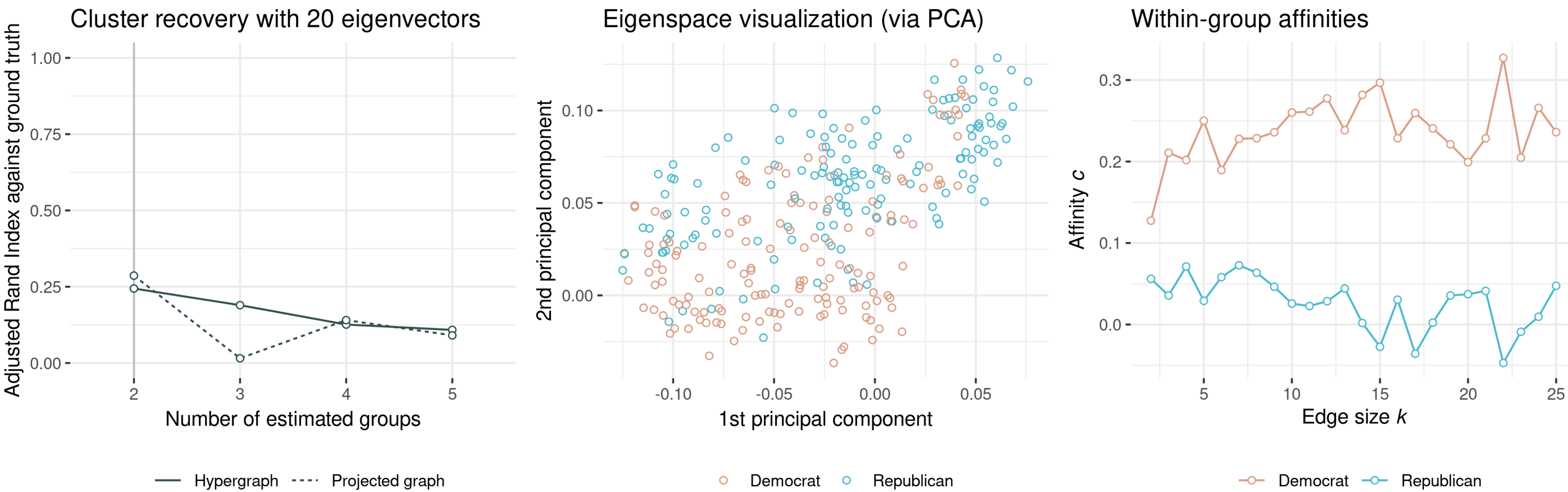
Average number of cosponsors $\langle k \rangle = \mathbf{7.3}$.

Community labels are Democrat/Republican.

Data originally from:
J. Fowler (2006), Legislative cosponsorship networks in the U.S. House and Senate. *Social Networks* 28:4, 454--465

Prepared by A. R. Benson et al. (2018), Simplicial closure and higher-order link prediction. *Proceedings of the National Academy of Sciences* 10.1073/pnas.1800683115

Senate Bills



Big Picture: you want hypergraph methods when edges of different sizes give you different information about the community structure.

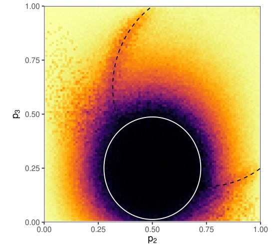
My Papers

Foundations of Network Data Science

What **models** accurately reflect features of network data?

What **algorithms** can we use to learn these models?

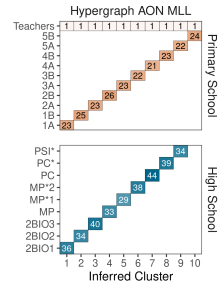
What **mathematical challenges** arise from these questions?



Nonbacktracking specral clustering of nonuniform hypergraphs

PSC, Nicole Eikmeier, and Jamie Haddock

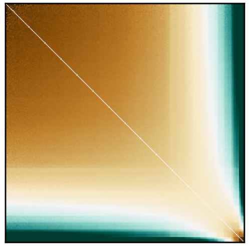
In preparation (2022)



Generative hypergraph clustering: from blockmodels to modularity

PSC, Nate Veldt, and Austin Benson

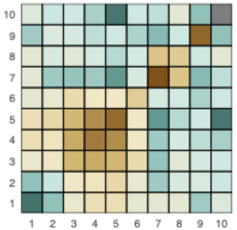
Science Advances (2021)



Moments of uniformly random multigraphs with fixed degree sequences

PSC

SIAM J. Mathematics of Data Science (2020)



Configuration models of random hypergraphs

PSC

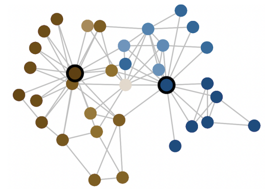
J. Complex Networks (2020)

Models of Biosocial Systems

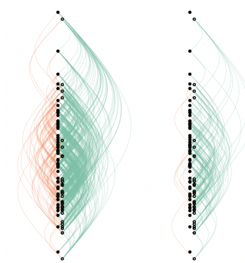
How can **individual decisions** lead to large-scale social division or hierarchy?

What **mechanisms** do math models need to capture these phenomena?

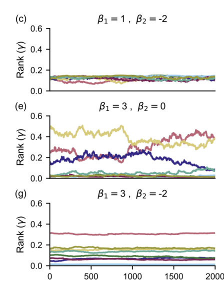
What can we **prove** or **approximate** about the behavior of these models?



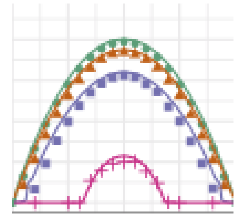
Smoothly nonlinear opinion dynamics
Heather Zinn Brooks, **PSC**, and Mason A. Porter
In preparation (2022)



Model-based approaches to layer aggregation in animal dominance networks
PSC, Kelly Finn, and and Mason A. Porter
In preparation (2022)



Emergence of hierarchy in networked endorsement dynamics
Mari Kawakatsu, **PSC**, Nicole Eikmeier, and Dan Larremore
Proc. National Academy of Sciences (2021)



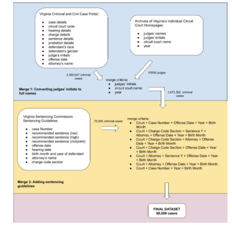
Local symmetry and global structure in adaptive voter models
PSC and Peter Mucha
SIAM J. Applied Math (2020)

Data Science and Social Responsibility

How can mathematics **unify and support** methods in quantitative sociology?

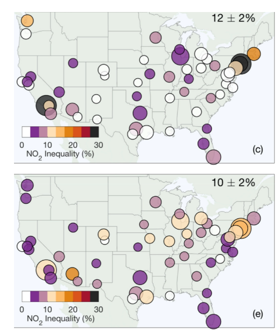
How can we acquire and analyze data sets on **equity** and **justice**?

How can we model the history (and future?) of **representation** in our discipline?



Impact of race on sentencing in a state court
Hinal Jajal (undergraduate mentee), **PSC**
Ongoing work (2022)

Dynamics of gender representation in mathematical subfields
Ben Brill (undergraduate mentee) et al.
Ongoing work (2022)



Space-based observational constraints on NO_2 air pollution inequality from diesel traffic in major U.S. cities
Demetillo et al. *Geophysics Review Letters* (2021)

Structure and information in spatial segregation
PSC
Proc. National Academy of Sciences (2017)