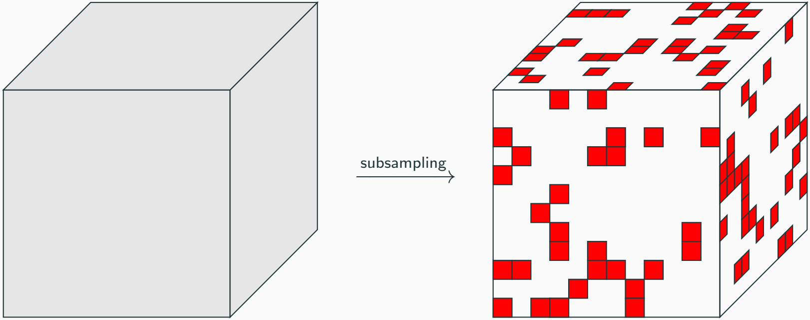
A non-backtracking method for long matrix and tensor completion

Ludovic Stephan

ENSAI - CREST

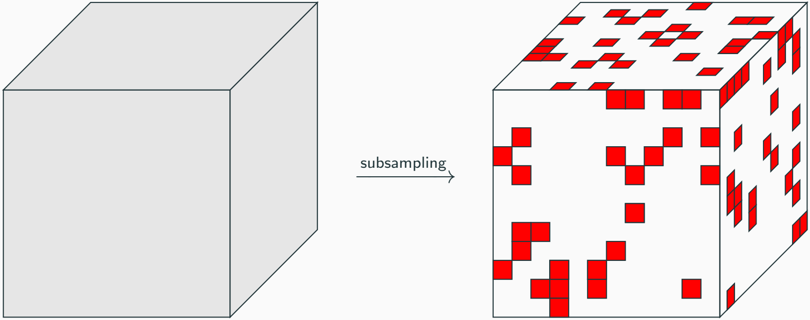
Joint work with Yizhe Zhu (USC)

What is tensor completion ?



Tensor T Observed tensor $\tilde{T}$

What is tensor completion ?



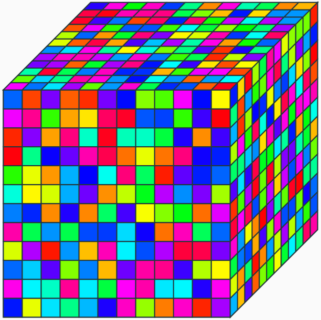
Tensor T Observed tensor $\tilde{T}$

- T is an order-k tensor of size $n \times \cdots \times n$
- The observed tensor $\tilde{T}$ is defined as

$$\tilde{T}_{i_1,\ldots,i_k} = \begin{cases} T_{i_1,\ldots,i_k} & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

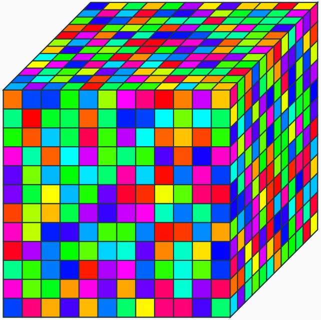
- Goal: Exactly/approximately recover T from $\tilde{T}$ with very few samples (with an efficient algorithm)

Model assumptions

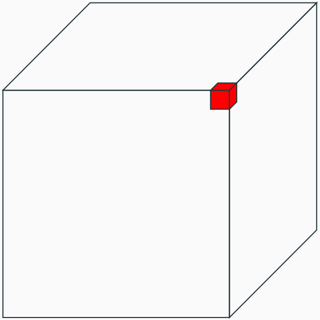


Too many degrees of freedom!

Model assumptions

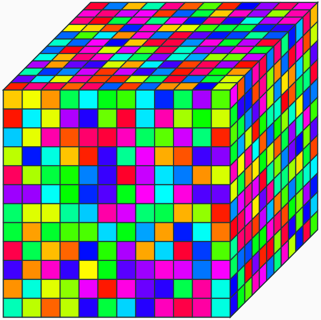


Too many degrees of freedom!

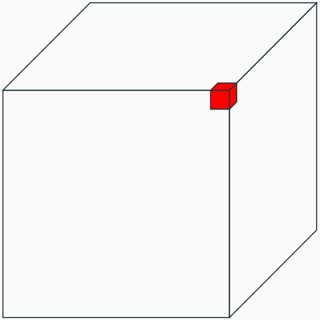


Too localized!

Model assumptions



Too many degrees of freedom!



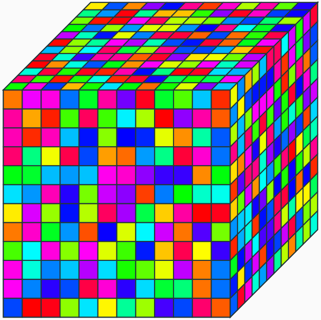
Too localized!

- T has low CP-rank:

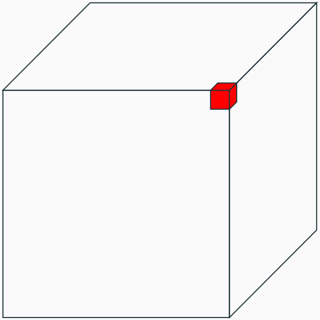
$$T = \sum_{i=1}^{r} \lambda_i \left(w_i^{(1)} \otimes \cdots \otimes w_i^{(k)} \right)$$

$\Rightarrow r \times kn$ degrees of freedom

Model assumptions



Too many degrees of freedom!



Too localized!

- T has low CP-rank:

$$T = \sum_{i=1}^{r} \lambda_i \left(w_i^{(1)} \otimes \cdots \otimes w_i^{(k)} \right)$$

$\Rightarrow r \times kn$ degrees of freedom

- T is delocalized:

$$\|w_i^{(j)}\|_\infty \simeq n^{-1/2}$$

Matrix completion

- Low-rank matrix completion [Candes-Recht '09, Candes-Tao '10, Keshavan-Montanari-Oh '10,...]. When $r = O(1)$, with high probability, uniformly sampling $O(n \log(n))$ entries with convex/ non-convex optimization is sufficient to exactly recover M.

Matrix completion

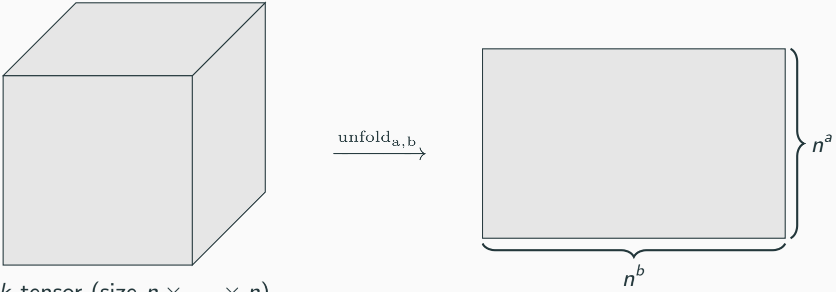
- Low-rank matrix completion [Candes-Recht '09, Candes-Tao '10, Keshavan-Montanari-Oh '10,...]. When $r = O(1)$, with high probability, uniformly sampling $O(n \log(n))$ entries with convex/ non-convex optimization is sufficient to exactly recover M.
- Information threshold $O(rn \log n)$ [Candes-Tao '10]. Best rank dependence: $O(r \log r \cdot n \log n)$ [Ding-Chen '20].

Computational hardness

Computational complexity problem: most tensor problems are hard [Hillar-Lim '09]

- spectral norm
- eigenvalues/singular values
- low-rank approximations

Unfolding



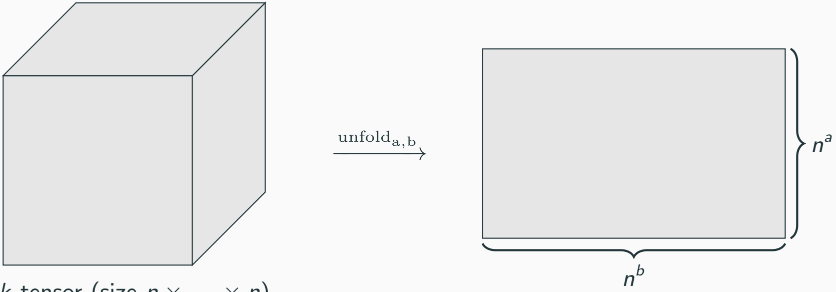
k-tensor (size $n \times \cdots \times n$)

Unfolding matrix (size $n^a \times n^b$)

"Grouping" indices:

$$M_{(i_1,\ldots,i_a),(i_{a+1},\ldots,i_k)} = T_{i_1,\ldots,i_k}$$

Unfolding



k-tensor (size $n \times \cdots \times n$)

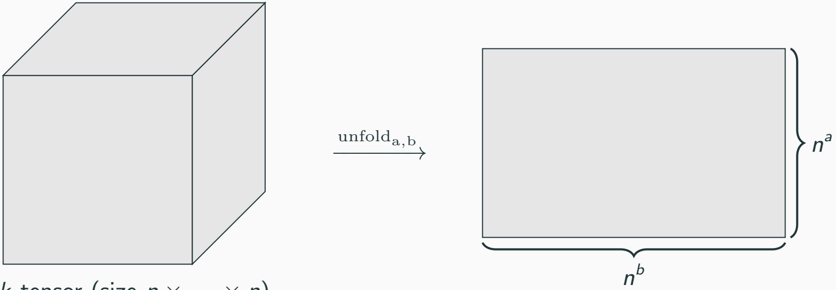
Unfolding matrix (size $n^a \times n^b$)

"Grouping" indices:

$$M_{(i_1,\ldots,i_a),(i_{a+1},\ldots,i_k)} = T_{i_1,\ldots,i_k}$$

Tensor completion on $T \Leftarrow$ Matrix completion on M

Unfolding



k-tensor (size $n \times \cdots \times n$)

Unfolding matrix (size $n^a \times n^b$)

"Grouping" indices:

$$M_{(i_1,\ldots,i_a),(i_{a+1},\ldots,i_k)} = T_{i_1,\ldots,i_k}$$

Tensor completion on $T \Leftarrow$ Matrix completion on M

If k is even: square matrix of size $n^{k/2} \implies \tilde{O}(n^{k/2})$ samples suffice
If k is odd: matrix of size $n^{\lfloor k/2 \rfloor} \times n^{\lceil k/2 \rceil}$

Statistical-computational gap for random tensors

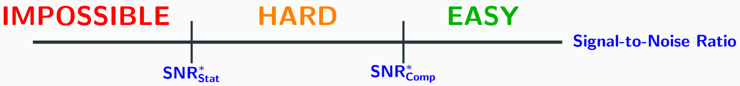
- NP-hard algorithms [Yuan-Zhang '16, Ghadermarzy et al '19, Harris-Zhu '21]: tensor-based norm minimization methods without unfolding
  $\rightarrow$ works with $\tilde{O}(n)$ samples

Statistical-computational gap for random tensors

- NP-hard algorithms [Yuan-Zhang '16, Ghadermarzy et al '19, Harris-Zhu '21]: tensor-based norm minimization methods without unfolding
  $\rightarrow$ works with $\tilde{O}(n)$ samples
- Unfolding-based algorithms with spectral initialization [Montanari and Sun '16, Liu and Moitra '20, Cai et al. '21...]
  $\rightarrow$ works with $\tilde{O}(n^{k/2})$ samples

Statistical-computational gap for random tensors

- NP-hard algorithms [Yuan-Zhang '16, Ghadermarzy et al '19, Harris-Zhu '21]: tensor-based norm minimization methods without unfolding
  $\rightarrow$ works with $\tilde{O}(n)$ samples
- Unfolding-based algorithms with spectral initialization [Montanari and Sun '16, Liu and Moitra '20, Cai et al. '21...]
  $\rightarrow$ works with $\tilde{O}(n^{k/2})$ samples
- Similar gaps in the spiked tensor model $T = \lambda v^{\otimes q} + Z$ [Montanari-Richard '14, Ben Arous-Mei-Montanari-Nica '17, Chen '18, Ben Arous-Gheissari-Jagannath '18, Wein-Alaoui-Moore '19, Perry-Wein-Bandeira '20. . .]



Basic unfolding algorithm

Commonly poly-time algorithms: unfolding-based [Montanari and Sun '16, Liu and Moitra '20, Cai et al. '21...]

- Unfold $\tilde{T}$ into $A \in \mathbb{R}^{n \times n^2}$. If $T = \sum_{i=1}^{r} x_i \otimes y_i \otimes z_i$, unfold $\tilde{T}$ in 3 different ways.
- Take the SVD of the hollowed matrix $h(AA^\top) = AA^\top - \mathrm{diag}(AA^T)$ (spectral initialization) + postprocessing
- Diagonal removal improved the performance [Cai et al. '21]

$\rightarrow$ works until $p = O(n^{-k/2} \times \text{polylog}(n))$

Basic unfolding algorithm

Commonly poly-time algorithms: unfolding-based [Montanari and Sun '16, Liu and Moitra '20, Cai et al. '21...]

- Unfold $\tilde{T}$ into $A \in \mathbb{R}^{n \times n^2}$. If $T = \sum_{i=1}^{r} x_i \otimes y_i \otimes z_i$, unfold $\tilde{T}$ in 3 different ways.
- Take the SVD of the hollowed matrix $h(AA^\top) = AA^\top - \mathrm{diag}(AA^T)$ (spectral initialization) + postprocessing
- Diagonal removal improved the performance [Cai et al. '21]

$\rightarrow$ works until $p = O(n^{-k/2} \times \text{polylog}(n))$

What happens if $p \propto n^{-k/2}$?

Not a trivial challenge

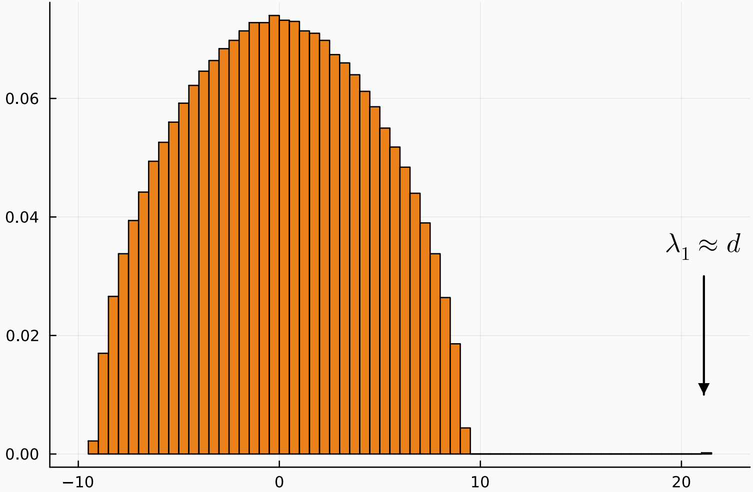


Figure: $T = v \otimes v \otimes v, AA^\top - \mathrm{diag}(AA^\top), p = 20n^{-3/2}$

Not a trivial challenge

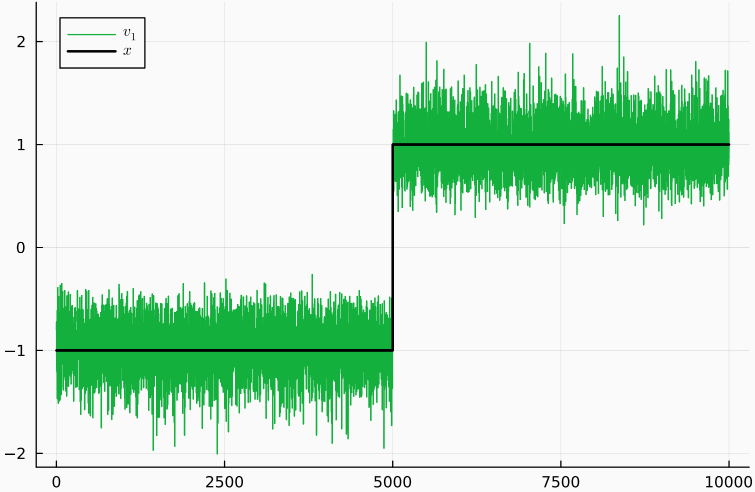


Figure: $T = v \otimes v \otimes v, AA^\top - \mathrm{diag}(AA^\top), p = 20n^{-3/2}$

Not a trivial challenge

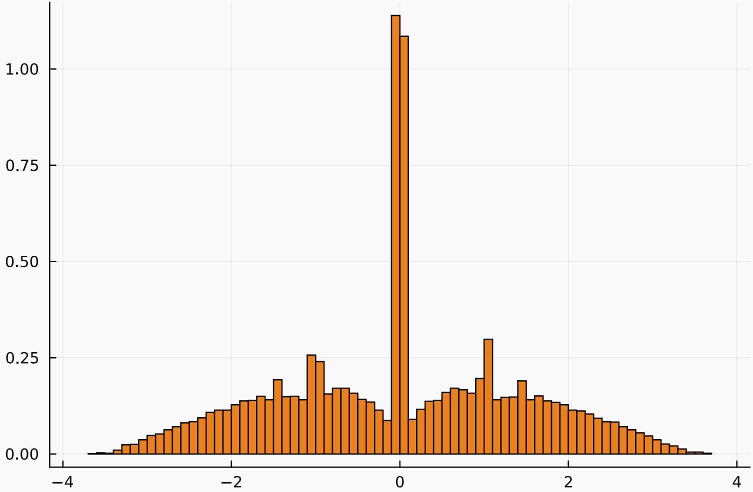


Figure: $AA^\top - \mathrm{diag}(AA^\top), p = 2n^{-3/2}$

Not a trivial challenge

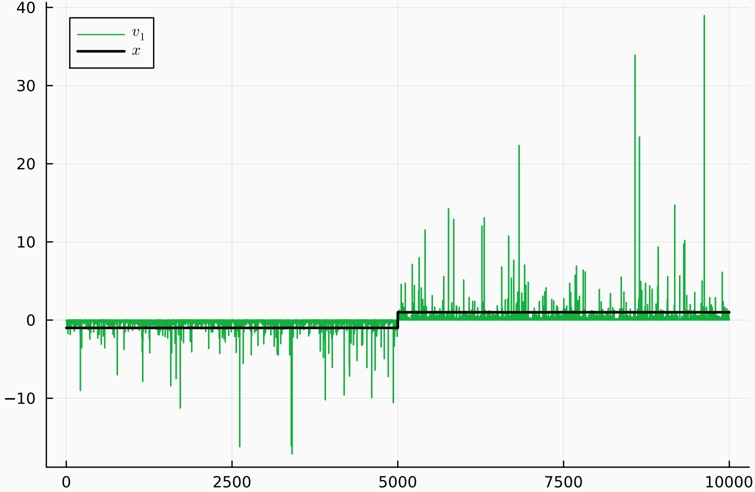
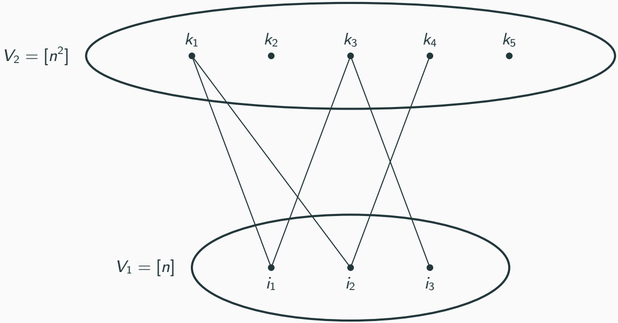


Figure: $AA^\top - \mathrm{diag}(AA^\top), p = 2n^{-3/2}$

A random graph theory explanation

$A \in \mathbb{R}^{n \times n^2}$ corresponds to a (weighted) random bipartite graph with $V_1 = [n], V_2 = [n^2]$.

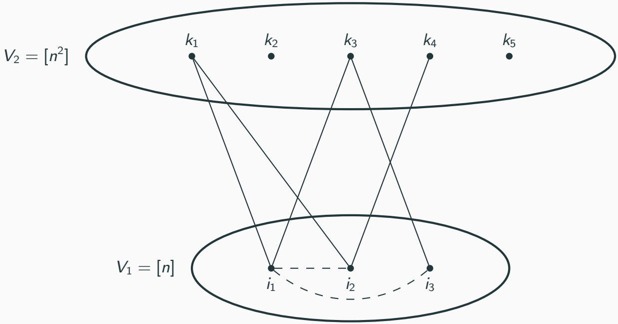


A random graph theory explanation

Hollowed matrix counts walks of length 2, $V_1 \to V_2 \to V_1$:

$$(AA^\top)_{ij} = \sum_k A_{ik} A_{jk}.$$

$h(AA^\top)$ can be seen as the adjacency matrix of a new graph $\tilde{G}$ (dashed edges).



A random graph theory explanation

Fact: $\tilde{G}$ is still sparse (average degree d^2 for $p = dn^{-k/2}$).

In the unweighted (Erdős-Rényi) case:

- if $d^2 \gtrsim \sqrt{\frac{\log(n)}{\log\log(n)}}$: spectrum of $\tilde{G}$ concentrates [Feige-Ofek '05, Benaych-Georges-Bordenave-Knowles '20]
- if $d^2 \ll \sqrt{\frac{\log(n)}{\log\log(n)}}$: no concentration, spectrum dominated by high-degree vertices [Benaych-Georges-Bordenave-Knowles '19]

A random graph theory explanation

Fact: $\tilde{G}$ is still sparse (average degree d^2 for $p = dn^{-k/2}$).

In the unweighted (Erdős-Rényi) case:

- if $d^2 \gtrsim \sqrt{\frac{\log(n)}{\log\log(n)}}$: spectrum of $\tilde{G}$ concentrates [Feige-Ofek '05, Benaych-Georges-Bordenave-Knowles '20]
- if $d^2 \ll \sqrt{\frac{\log(n)}{\log\log(n)}}$: no concentration, spectrum dominated by high-degree vertices [Benaych-Georges-Bordenave-Knowles '19]

$\Rightarrow$ Naive unfolding (probably) doesn't work

Where are we so far?

Recap:

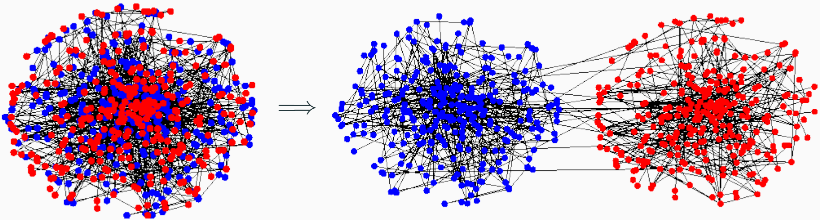
- existing methods do not reach the exact conjectured threshold for tensor completion (no results for "weak recovery").
- It is not a technical but a conceptual issue
- it suffices to solve matrix completion for a rank-r long matrix

Our solution: a new non-backtracking matrix for sparse long matrices

A detour through community detection

Community detection in stochastic block models $\mathcal{G}(n, \frac{a}{n}, \frac{b}{n})$.

- Unknown partition $\sigma \in \{-1, 1\}^n$. Generate a random graph $G = ([n], E)$. i, j is connected with probability $p = \frac{a}{n}$ if $\sigma_i = \sigma_j$ and with probability $q = \frac{b}{n}$ otherwise.
- goal: recover σ from G

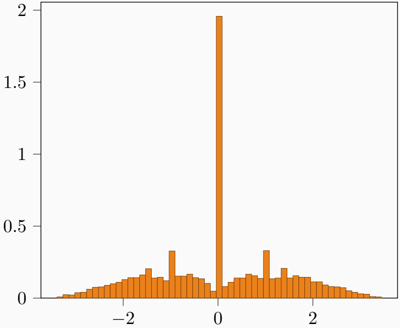


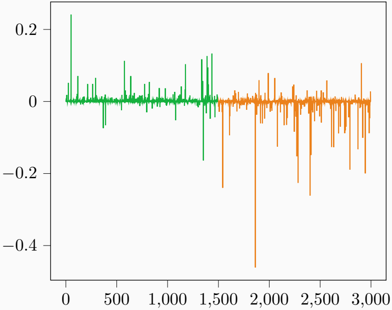
A detour through community detection

$\mathbb{E}[A]$ is low-rank, and $v_2(\mathbb{E}[A]) = \sigma \Rightarrow$ spectral method on A?

A detour through community detection

$\mathbb{E}[A]$ is low-rank, and $v_2(\mathbb{E}[A]) = \sigma \Rightarrow$ spectral method on A? **No!**





$p = \frac{a}{n}, q = \frac{b}{n}$. High-degree vertices dominate the spectrum. v_2 localized around high-degree vertices.

[Krivelevich-Sudakov '01,Benaych-Georges, Bordenave, Knowles '19, Alt-Ducatez-Knowles '23]

Non-backtracking matrix for graphs

Proposed in [Krzakala et al. '13]

Defined on the oriented edges of G:

$$\vec{E} = \{u \to v : \{u, v\} \in E\}, |\vec{E}| = 2|E|.$$

Non-backtracking matrix for graphs

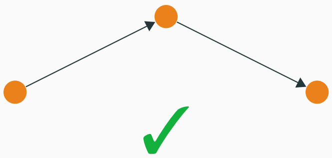
Proposed in [Krzakala et al. '13]

Defined on the oriented edges of G:

$$\vec{E} = \{u \to v : \{u, v\} \in E\}, |\vec{E}| = 2|E|.$$

The non-backtracking matrix B is defined: for $u \to v, x \to y \in \vec{E}$,

$$B_{u \to v, x \to y} = \mathbf{1}_{v=x} \mathbf{1}_{u \neq y}.$$



Non-backtracking matrix for graphs

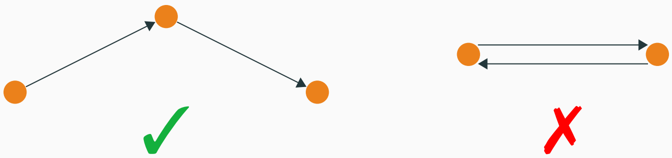
Proposed in [Krzakala et al. '13]

Defined on the oriented edges of G:

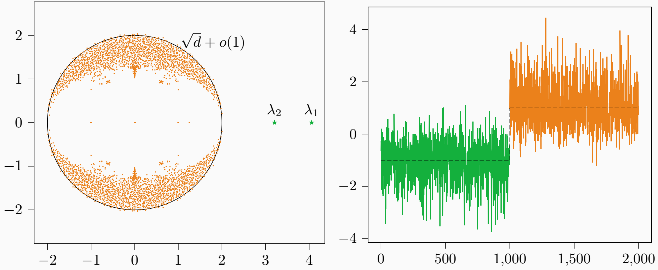
$$\vec{E} = \{u \to v : \{u, v\} \in E\}, |\vec{E}| = 2|E|.$$

The non-backtracking matrix B is defined: for $u \to v, x \to y \in \vec{E}$,

$$B_{u\to v, x\to y} = \mathbf{1}_{v=x}\mathbf{1}_{u\neq y}.$$

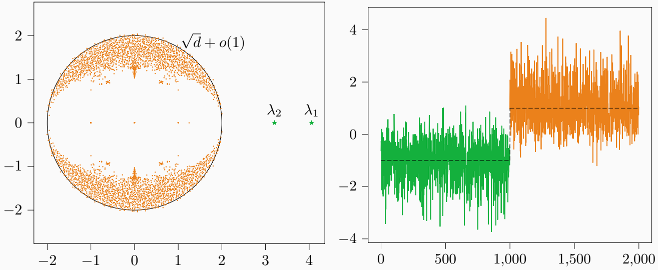


Non-backtracking spectral method



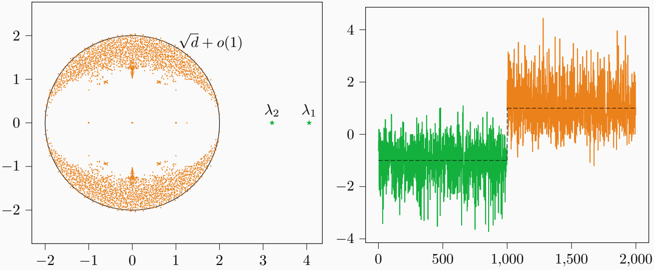
- If $(a-b)^2 > 2(a+b)$, then the second eigenvector of B can be used to detect the community structure. [Bordenave, Lelarge, Massoulié '18]

Non-backtracking spectral method



- If $(a-b)^2 > 2(a+b)$, then the second eigenvector of B can be used to detect the community structure. [Bordenave, Lelarge, Massoulié '18]
- B is non-Hermitian: avoid the localization effect from high degree vertices when G is very sparse.

Non-backtracking spectral method



- If $(a-b)^2 > 2(a+b)$, then the second eigenvector of B can be used to detect the community structure. [Bordenave, Lelarge, Massoulié '18]
- B is non-Hermitian: avoid the localization effect from high degree vertices when G is very sparse.
- Can be generalized for very sparse matrix completion: *estimate a low-rank structure from sparse observations* with $O(n)$ many samples. [Bordenave-Coste-Nadakuditi '23]

Long matrix reconstruction

- Rectangular matrix M of size $n \times m$ ($m \gg n$), with SVD

$$M = \sum_{i=1}^{r} \nu_i \phi_i \psi_i^\top, \quad MM^\top = \sum_{i=1}^{r} \nu_i^2 \phi_i \phi_i^\top$$

- Masking matrix X with $X_{ij} \sim \mathrm{Ber}(p)$, $\quad p = \frac{d}{\sqrt{mn}}$.
- Observed matrix:

$$A = \frac{X \circ M}{p} \quad \text{so that} \quad \mathbb{E}[A] = M$$

Long matrix reconstruction

- Rectangular matrix M of size $n \times m$ ($m \gg n$), with SVD

$$M = \sum_{i=1}^{r} \nu_i \phi_i \psi_i^\top, \quad MM^\top = \sum_{i=1}^{r} \nu_i^2 \phi_i \phi_i^\top$$

- Masking matrix X with $X_{ij} \sim \mathrm{Ber}(p)$, $\quad p = \frac{d}{\sqrt{mn}}$.
- Observed matrix:

$$A = \frac{X \circ M}{p} \quad \text{so that} \quad \mathbb{E}[A] = M$$

Assumptions:

$$r, \sqrt{n}\|\phi_i\|_\infty = O(\mathrm{polylog}(n))$$

Goal: estimate singular values and left singular vectors of M: ν_i, ϕ_i, with sample size $O(\sqrt{mn})$

Long matrix reconstruction

- Rectangular matrix M of size $n \times m$ ($m \gg n$), with SVD

$$M = \sum_{i=1}^{r} \nu_i \phi_i \psi_i^\top, \quad MM^\top = \sum_{i=1}^{r} \nu_i^2 \phi_i \phi_i^\top$$

- Masking matrix X with $X_{ij} \sim \mathrm{Ber}(p)$, $\quad p = \frac{d}{\sqrt{mn}}$.
- Observed matrix:

$$A = \frac{X \circ M}{p} \quad \text{so that} \quad \mathbb{E}[A] = M$$

Assumptions:

$$r, \sqrt{n}\|\phi_i\|_\infty = O(\mathrm{polylog}(n))$$

Goal: estimate singular values and left singular vectors of M: ν_i, ϕ_i, with sample size $O(\sqrt{mn})$

Estimating the full SVD of M needs $O(m)$ [Bordenave-Coste-Nadakuditi '23]!

Non-backtracking wedge matrix

First idea: take the non-backtracking matrix of $\tilde{G} \Rightarrow$ doesn't work

Non-backtracking wedge matrix

First idea: take the non-backtracking matrix of $\tilde{G} \Rightarrow$ doesn't work

Better idea: work directly on *oriented wedges* in G

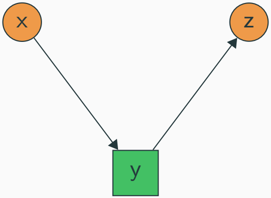
$$\vec{E}_2 = \{(x, y, z) \in V_1 \times V_2 \times V_1, z \neq x\}$$

Non-backtracking wedge matrix

First idea: take the non-backtracking matrix of $\tilde{G}$ $\Rightarrow$ doesn't work

Better idea: work directly on *oriented wedges* in G

$$\vec{E}_2 = \{(x, y, z) \in V_1 \times V_2 \times V_1, z \neq x\}$$

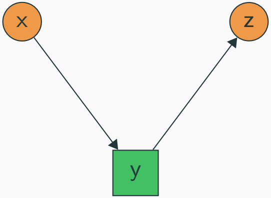


Non-backtracking wedge matrix

First idea: take the non-backtracking matrix of $\tilde{G}$ $\Rightarrow$ doesn't work

Better idea: work directly on *oriented wedges* in G

$$\vec{E}_2 = \{(x, y, z) \in V_1 \times V_2 \times V_1, z \neq x\}$$



$\Rightarrow$ B has size $\sim n^2 m p^2 = d^2 n$: independent from m

Non-backtracking wedge matrix

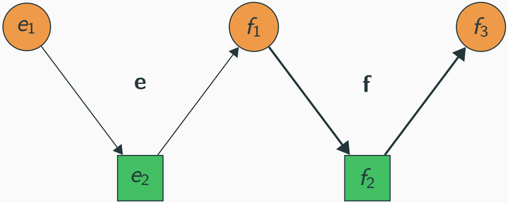
Defined B index by $\vec{E}$ as

$$B_{ef} = \begin{cases} A_{f_1 f_2} A_{f_3 f_2} & \text{if } e_3 = f_1 \text{ and } e_2 \neq f_2 \\ 0 & \text{otherwise} \end{cases}$$

Non-backtracking wedge matrix

Defined B index by $\vec{E}$ as

$$B_{ef} = \begin{cases} A_{f_1 f_2} A_{f_3 f_2} & \text{if } e_3 = f_1 \text{ and } e_2 \neq f_2 \\ 0 & \text{otherwise} \end{cases}$$

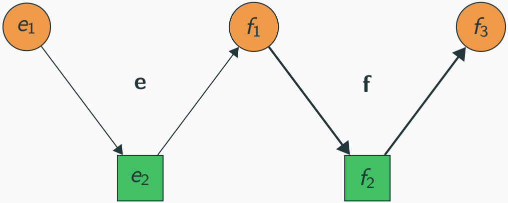


e, f form a non-backtracking walk of length 4, starting from V_1, ending in V_1.

Non-backtracking wedge matrix

Defined B index by $\vec{E}$ as

$$B_{ef} = \begin{cases} A_{f_1 f_2} A_{f_3 f_2} & \text{if } e_3 = f_1 \text{ and } e_2 \neq f_2 \\ 0 & \text{otherwise} \end{cases}$$



e, f form a non-backtracking walk of length 4, starting from V_1, ending in V_1.

Thresholds

$$\rho = \sqrt{mn}\|M \circ M\|, \qquad L = \sqrt{mn} \max_{x\in[n],y\in[m]} |M_{xy}|.$$

Thresholds

$$\rho = \sqrt{mn}\|M \circ M\|, \qquad L = \sqrt{mn} \max_{x \in [n], y \in [m]} |M_{xy}|.$$

Two important thresholds:

$$\vartheta_1 = \sqrt{\rho / d}$$

- decreases as $d^{-1/2}$

Thresholds

$$\rho = \sqrt{mn}\|M \circ M\|, \qquad L = \sqrt{mn} \max_{x\in[n],y\in[m]} |M_{xy}|.$$

Two important thresholds:

$$\vartheta_1 = \sqrt{\rho/d}$$

- decreases as $d^{-1/2}$

$$\vartheta_2 = L/d$$

- decreases as d^{-1}

Thresholds

$$\rho = \sqrt{mn}\|M \circ M\|, \qquad L = \sqrt{mn} \max_{x\in[n], y\in[m]} |M_{xy}|.$$

Two important thresholds:

$$\vartheta_1 = \sqrt{\rho/d}$$

- decreases as $d^{-1/2}$

$$\vartheta_2 = L/d$$

- decreases as d^{-1}

Total threshold (Signal-to-noise ratio):

$$\vartheta = \max(\vartheta_1, \vartheta_2)$$

Our results: eigenvalues

Theorem (Stephan-Z. '24)

- *(**Outliers**) For any ν_i satisfying $\nu_i > \vartheta$, there exists an eigenvalue λ_i of B with*

$$|\lambda_i - \nu_i^2| = O(n^{-c})$$

- *(**Bulk**) All other eigenvalues are asymptotically confined in a circle of radius ϑ^2*

Our results: eigenvalues

Theorem (Stephan-Z. '24)

- *(**Outliers**) For any ν_i satisfying $\nu_i > \vartheta$, there exists an eigenvalue λ_i of B with*

$$|\lambda_i - \nu_i^2| = O(n^{-c})$$

- *(**Bulk**) All other eigenvalues are asymptotically confined in a circle of radius ϑ^2*

Similar to the Kesten-Stigum threshold in community detection [Bordenave-Lelarge-Massoulié '18, Mossel-Neeman-Sly '18]

Results: eigenvalues

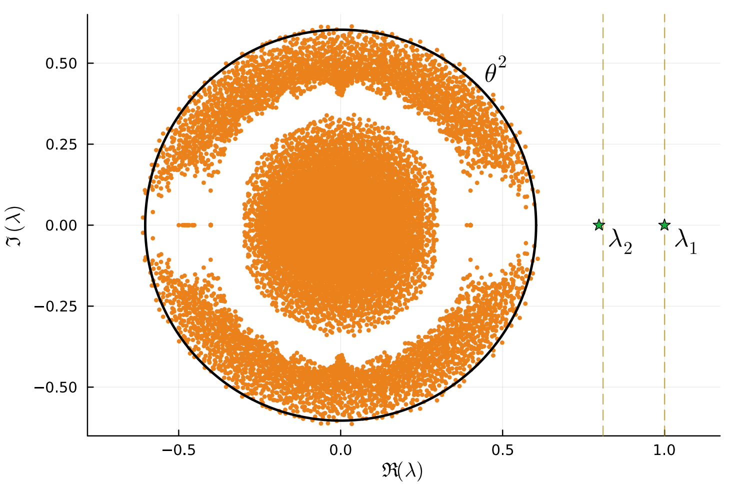


Figure: M is of rank-2, spectrum of $B, d = 3$

Embedding eigenvectors

Need an embedding procedure from $\mathbb{R}^{\vec{E}_2}$ to $\mathbb{R}^n$

Embedding eigenvectors

Need an embedding procedure from $\mathbb{R}^{\vec{E}_2}$ to $\mathbb{R}^n$

- For a right eigenvector ξ^R of B:

$$\zeta^R(x) = \sum_{e:e_1=x} A_{e_1 e_2} A_{e_3 e_2} \xi^R(e), \quad \forall x \in [n].$$

- For a left eigenvector ξ^L:

$$\zeta^L(x) = \sum_{e:e_1=x} \xi^L(e)$$

Our results: eigenvectors

Theorem (Stephan-Z. '24)

Assume that $\nu_i > \vartheta$, and let $\xi_i^{L/R}$ the left/right eigenvectors associated to λ_i. Then, there exists a γ_i such that

$$\gamma_i = 1 - O(d^{-1})$$

and

$$\left| \langle \zeta_i^{L/R}, \phi_i \rangle - \sqrt{\gamma_i} \right| = O(n^{-c})$$

Our results: eigenvectors

Theorem (Stephan-Z. '24)

Assume that $\nu_i > \vartheta$, and let $\xi_i^{L/R}$ the left/right eigenvectors associated to λ_i. Then, there exists a γ_i such that

$$\gamma_i = 1 - O(d^{-1})$$

and

$$\left| \langle \zeta_i^{L/R}, \phi_i \rangle - \sqrt{\gamma_i} \right| = O(n^{-c})$$

Weak recovery when $d \to \infty$. Explicit γ_i when d is fixed.

Our results: eigenvectors

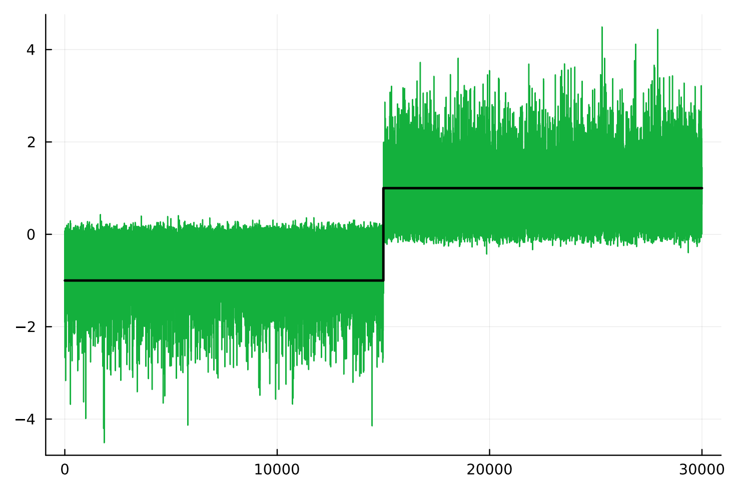


Figure: $B, d = 3$

Application in tensor completion

- $T = x_1 \otimes \cdots \otimes x_k$, $x_i \in \frac{1}{\sqrt{n}}\{\pm 1\}$. Sample with probability $p = \frac{d}{n^{k/2}}$.
- Unfold $\tilde{T}$ in k different ways (the most unbalanced unfolding) [Ben Arous, Huang, Huang '23]. Apply the non-backtracking method to $\mathrm{Unfold}(\tilde{T})$.

Application in tensor completion

- $T = x_1 \otimes \cdots \otimes x_k$, $x_i \in \frac{1}{\sqrt{n}}\{\pm 1\}$. Sample with probability $p = \frac{d}{n^{k/2}}$.
- Unfold $\tilde{T}$ in k different ways (the most unbalanced unfolding) [Ben Arous, Huang, Huang '23]. Apply the non-backtracking method to $\mathrm{Unfold}(\tilde{T})$.
- When sample size is $\alpha n^{k/2}$ with $\alpha > 1$, one can find unit eigenvectors such that

$$\langle v_i, x_i \rangle = \frac{\sqrt{\alpha^2 - 1}}{\alpha} + O(n^{-c}).$$

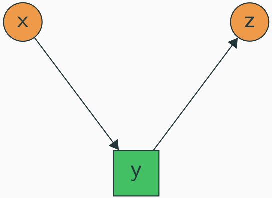
Application in tensor completion

- $T = x_1 \otimes \cdots \otimes x_k$, $x_i \in \frac{1}{\sqrt{n}}\{\pm 1\}$. Sample with probability $p = \frac{d}{n^{k/2}}$.
- Unfold $\tilde{T}$ in k different ways (the most unbalanced unfolding) [Ben Arous, Huang, Huang '23]. Apply the non-backtracking method to $\mathrm{Unfold}(\tilde{T})$.
- When sample size is $\alpha n^{k/2}$ with $\alpha > 1$, one can find unit eigenvectors such that

$$\langle v_i, x_i \rangle = \frac{\sqrt{\alpha^2 - 1}}{\alpha} + O(n^{-c}).$$

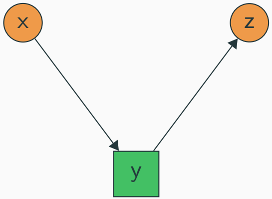
- $T = \sum_{i=1}^{r} \lambda_i \left(w_i^{(1)} \otimes \cdots \otimes w_i^{(k)} \right)$, under the orthonormal condition on $w_1^{(j)}, \ldots, w_r^{(j)}$, the same analysis apply. $O(n^{k/2})$ samples for nontrivial approximation.

Proof idea

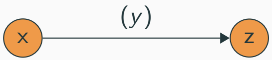


- $\deg(y) = 2$ w.h.p
- "independent" wedges
- Associated weight:
  $W_e = A_{xy}A_{zy}$

Proof idea



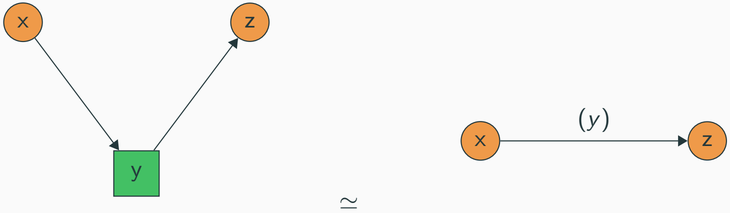
$\simeq$



- $\deg(y) = 2$ w.h.p
- "independent" wedges
- Associated weight: $W_e = A_{xy} A_{zy}$

- Erdős-Rényi graph $G \sim \mathcal{G}(n, d^2/n)$
- Associated weight: $W_e = A_{xY} A_{zY}$, $Y \sim \mathrm{Unif}([m])$

Proof idea

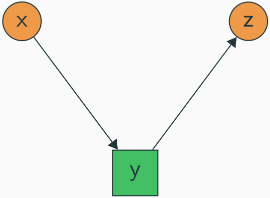


- $\deg(y) = 2$ w.h.p
- "independent" wedges
- Associated weight: $W_e = A_{xy}A_{zy}$

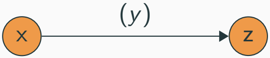
- Erdős-Rényi graph $G \sim \mathcal{G}(n, d^2/n)$
- Associated weight: $W_e = A_{xY}A_{zY}$, $Y \sim \mathrm{Unif}([m])$

- Extract information from a sparse Erdős-Rényi bipartite graph with random edge weights

Proof idea



$\simeq$



- $\deg(y) = 2$ w.h.p
- "independent" wedges
- Associated weight: $W_e = A_{xy}A_{zy}$

- Erdős-Rényi graph $G \sim \mathcal{G}(n, d^2/n)$
- Associated weight: $W_e = A_{xY}A_{zY}$, $Y \sim \mathrm{Unif}([m])$

- Extract information from a sparse Erdős-Rényi bipartite graph with random edge weights
- Bulk eigenvalues: high moment methods on random bipartite graphs.
- Top eigenvalues and eigenvectors: local tree approximation. Galton-Watson tree with random weights [Stephan and Massoulié '20]

Conclusions

- Very sparse and long matrices are not standard in random matrix theory. We define a new non-backtracking matrix tailored for it (random bipartite graphs with more attention to V_1).
- The corresponding spectral method for tensor completion reaches the conjectured threshold in [Barak-Moitra '15].

Conclusions

- Very sparse and long matrices are not standard in random matrix theory. We define a new non-backtracking matrix tailored for it (random bipartite graphs with more attention to V_1).
- The corresponding spectral method for tensor completion reaches the conjectured threshold in [Barak-Moitra '15].
- Does not work for finite aspect ratio ($m = O(n)$) considered in [Bordenave-Coste-Nadakuditi '23]. Is there a unified spectral algorithm for all aspect ratios?

Conclusions

- Very sparse and long matrices are not standard in random matrix theory. We define a new non-backtracking matrix tailored for it (random bipartite graphs with more attention to V_1).
- The corresponding spectral method for tensor completion reaches the conjectured threshold in [Barak-Moitra '15].
- Does not work for finite aspect ratio ($m = O(n)$) considered in [Bordenave-Coste-Nadakuditi '23]. Is there a unified spectral algorithm for all aspect ratios?
- Statistical-computational gap between $O(n)$ and $O(n^{k/2})$ samples:
  - Possible with $O(n)$ samples and polynomial-time algorithms with non-uniform/ adaptive sampling [Haselby-Iwen-Karnik-Wang '24]
  - Rank-1 case is different, can be estimated with $O(n)$ samples [Stephan-Z. '24, Gomez-Leos, López '24] by solving linear systems.
  - can we justify this gap with a hardness proxy ?

Thank you!

L. Stephan, Y. Zhu, A non-backtracking method for long matrix and tensor completion, COLT 2024.