Spectral gap in random bipartite biregular graphs and applications

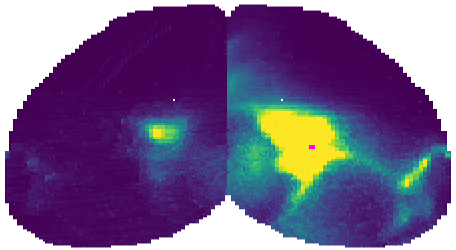
Kameron Decker Harris

Gerandy Brito (Georgia Tech Math)

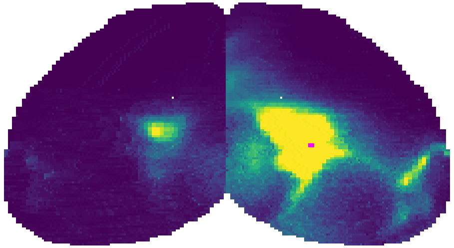
Ioana Dumitriu (UW Math)

Support: NIH Big Data for Genomics & Neuroscience, WRF Postdoctoral fellowship

A little bit about me...

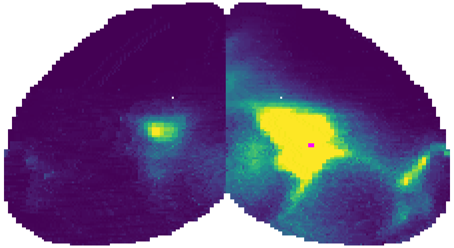


A little bit about me...



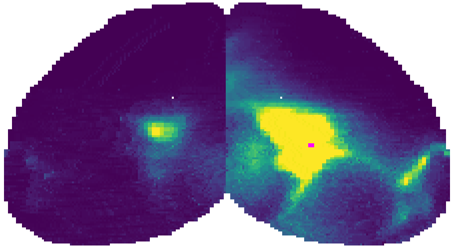
- Ph.D. in Applied Math from UW, December 2017

A little bit about me...



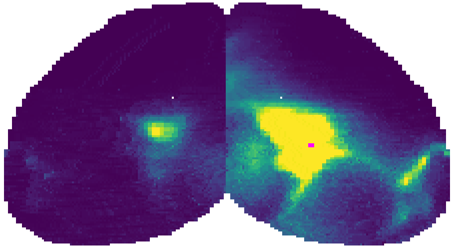
- Ph.D. in Applied Math from UW, December 2017
- Background studying networks, dynamics, & neuroscience

A little bit about me...



- Ph.D. in Applied Math from UW, December 2017
- Background studying networks, dynamics, & neuroscience
- Employed by Raj Rao & Bing Brunton (Biology)
  - Developing methods for clustering human brain recordings
  - Dynamical systems-based timeseries methods in general

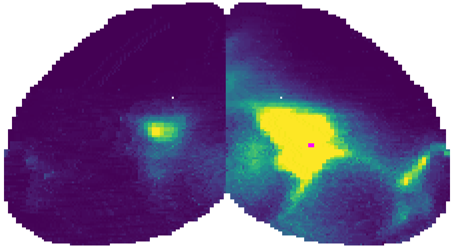
A little bit about me...



- Ph.D. in Applied Math from UW, December 2017
- Background studying networks, dynamics, & neuroscience
- Employed by Raj Rao & Bing Brunton (Biology)
  - Developing methods for clustering human brain recordings
  - Dynamical systems-based timeseries methods in general

- My hope for postdoc:
  - Get to know YOU!   … CS

A little bit about me...

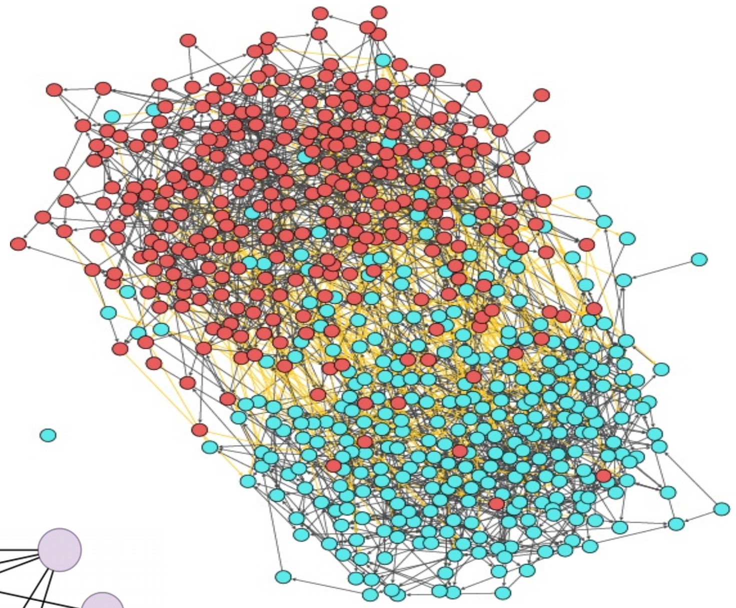


- Ph.D. in Applied Math from UW, December 2017
- Background studying networks, dynamics, & neuroscience
- Employed by Raj Rao & Bing Brunton (Biology)
  - Developing methods for clustering human brain recordings
  - Dynamical systems-based timeseries methods in general

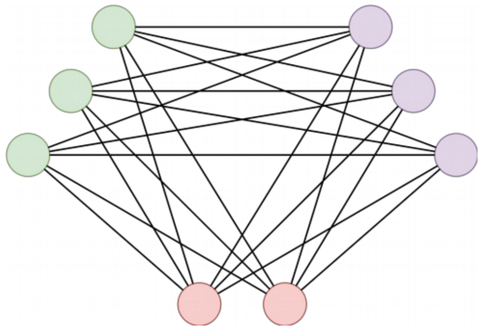
- My hope for postdoc:
  - Get to know YOU!   … CS
  - Strengthen theoretical machine learning understanding, esp. RKHS
  - Teaching, advising, etc.

Motivation: random graphs can model real world

- Example statistics:
  - Average # connections (degree)
  - Number of cycles
  - Other subgraph counts

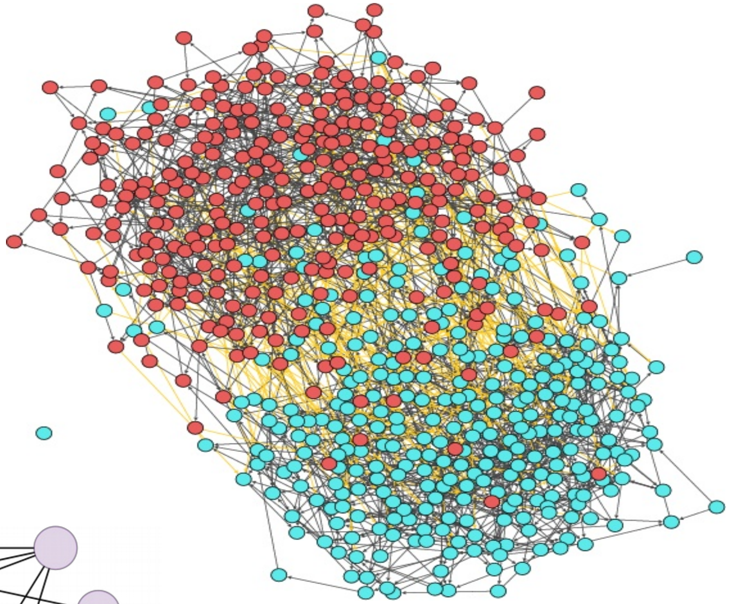


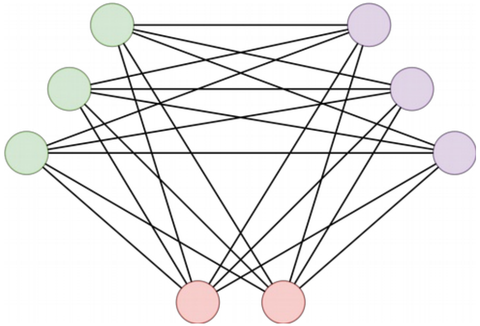
Joshua Mendoza



Motivation: random graphs can model real world

- Example statistics:
  - Average # connections (degree)
  - Number of cycles
  - Other subgraph counts
- Community structure
  - Block models
  - Multi-partite graphs





Joshua Mendoza

Why study the gap?

Why study the gap?

- Spectral “gap” is large → good **expansion**

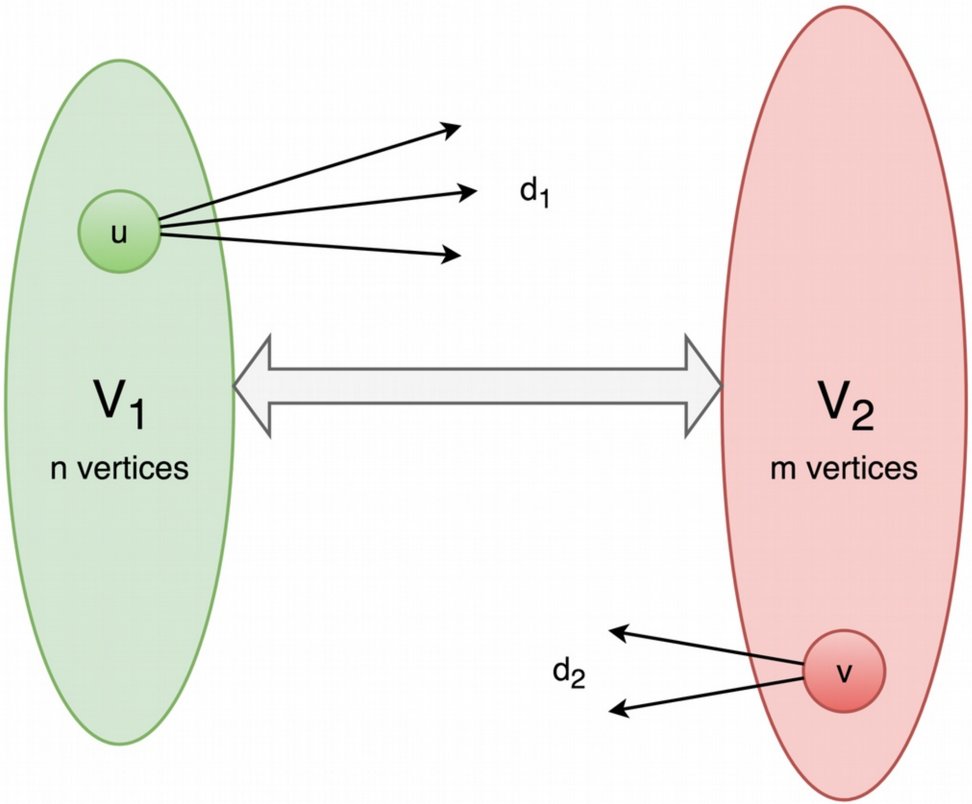
Why study the gap?

- Spectral “gap” is large → good **expansion**
- Many applications:
  - Mixing rates of Markov chains
  - Nonlinear dynamics on networks, e.g. synchronization conditions
  - Community detection, spectral clustering
  - Error correcting codes constructed from graphs
  - Matrix completion

Why study the gap?

- Spectral “gap” is large → good **expansion**
- Many applications:
  - Mixing rates of Markov chains
  - Nonlinear dynamics on networks, e.g. synchronization conditions
  - Community detection, spectral clustering
  - Error correcting codes constructed from graphs
  - Matrix completion
- What’s your favorite, and what did I miss?

Bipartite, biregular random graph model



Fixed d’s (as n grows) mean these graphs are **very sparse**

Previous work: mostly d-regular graphs

$$\lambda_1 = d, \qquad \lambda_2 = 2\sqrt{d-1} \pm \epsilon_n$$

Previous work: mostly d-regular graphs

$$\lambda_1 = d, \qquad \lambda_2 = 2\sqrt{d-1} \pm \epsilon_n$$

- Friedman (2003, 2004) – first gap proof using *selective trace*

Previous work: mostly d-regular graphs

$$\lambda_1 = d, \qquad \lambda_2 = 2\sqrt{d-1} \pm \epsilon_n$$

- Friedman (2003, 2004) – first gap proof using *selective trace*
- Alon (1986) – “gap conjecture” from lower bound

Previous work: mostly d-regular graphs

$$\lambda_1 = d, \qquad \lambda_2 = 2\sqrt{d-1} \pm \epsilon_n$$

- Friedman (2003, 2004) – first gap proof using *selective trace*
- Alon (1986) – “gap conjecture” from lower bound
- Bordenave (2015) – simplified proof using *tangle-free walks*

Previous work: mostly d-regular graphs

$$\lambda_1 = d, \qquad \lambda_2 = 2\sqrt{d-1} \pm \epsilon_n$$

- Friedman (2003, 2004) – first gap proof using *selective trace*
- Alon (1986) – “gap conjecture” from lower bound
- Bordenave (2015) – simplified proof using *tangle-free walks*
- Angel, Friedman, Hoory (2015) – lift model

Previous work: mostly d-regular graphs

$$\lambda_1 = d, \qquad \lambda_2 = 2\sqrt{d-1} \pm \epsilon_n$$

- Friedman (2003, 2004) – first gap proof using *selective trace*
- Alon (1986) – "gap conjecture" from lower bound
- Bordenave (2015) – simplified proof using *tangle-free walks*
- Angel, Friedman, Hoory (2015) – lift model
- Marcus, Spielman, Srivastava (2013) – gap of *bipartite* lifts

Previous work: mostly d-regular graphs

$$\lambda_1 = d, \qquad \lambda_2 = 2\sqrt{d-1} \pm \epsilon_n$$

- Friedman (2003, 2004) – first gap proof using *selective trace*
- Alon (1986) – "gap conjecture" from lower bound
- Bordenave (2015) – simplified proof using *tangle-free walks*
- Angel, Friedman, Hoory (2015) – lift model
- Marcus, Spielman, Srivastava (2013) – gap of *bipartite* lifts

$$\lambda_1 = \sqrt{d_1 d_2}, \quad \lambda_2 \geq \sqrt{d_1 - 1} + \sqrt{d_2 - 2} - \epsilon$$

Feng & Li, Li & Solé (1996)

Our main result

Recall: $\lambda_1 = \sqrt{d_1 d_2}, \quad \lambda_2 \geq \sqrt{d_1 - 1} + \sqrt{d_2 - 2} - \epsilon$

Theorem (Spectral gap). *Let* $A = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}$ *be the adjacency matrix of a bipartite, biregular random graph* $G \sim \mathcal{G}(n, m, d_1, d_2)$. *Without loss of generality, assume* $d_1 \geq d_2$ *or, equivalently,* $n \leq m$. *Then:*

(i) Its second largest eigenvalue $\eta = \lambda_2(A)$ *satisfies*

$$\eta \leq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} + \epsilon'_n$$

asymptotically almost surely, with $\epsilon'_n \to 0$ *as* $n \to \infty$.

(ii) Its smallest positive eigenvalue $\eta^+_{\text{min}} = \min(\{\lambda \in \sigma(A) : \lambda > 0\})$ *satisfies*

$$\eta^+_{\text{min}} \geq \sqrt{d_1 - 1} - \sqrt{d_2 - 1} - \epsilon''_n$$

asymptotically almost surely, with $\epsilon''_n \to 0$ *as* $n \to \infty$. *(Note that this will be almost surely positive if* $d_1 > d_2$*; no further information is gained if* $d_1 = d_2$*.)*

(iii) If $d_1 \neq d_2$*, the rank of* X *is* n *with high probability.*

Our main result

Recall: $\lambda_1 = \sqrt{d_1 d_2}, \quad \lambda_2 \geq \sqrt{d_1 - 1} + \sqrt{d_2 - 2} - \epsilon$

Theorem (Spectral gap). *Let* $A = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}$ *be the adjacency matrix of a bipartite, biregular random graph* $G \sim \mathcal{G}(n, m, d_1, d_2)$. *Without loss of generality, assume* $d_1 \geq d_2$ *or, equivalently,* $n \leq m$. *Then:*

(i) Its second largest eigenvalue $\eta = \lambda_2(A)$ *satisfies*

$$\boxed{\eta \leq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} + \epsilon_n'}$$

Attains Alon-Boppana

asymptotically almost surely, with $\epsilon_n' \to 0$ *as* $n \to \infty$.

(ii) Its smallest positive eigenvalue $\eta_{\min}^+ = \min(\{\lambda \in \sigma(A) : \lambda > 0\})$ *satisfies*

$$\eta_{\min}^+ \geq \sqrt{d_1 - 1} - \sqrt{d_2 - 1} - \epsilon_n''$$

asymptotically almost surely, with $\epsilon_n'' \to 0$ *as* $n \to \infty$. *(Note that this will be almost surely positive if* $d_1 > d_2$*; no further information is gained if* $d_1 = d_2$*.)*

(iii) If $d_1 \neq d_2$*, the rank of* X *is* n *with high probability.*

Our main result

Recall: $\lambda_1 = \sqrt{d_1 d_2}, \quad \lambda_2 \geq \sqrt{d_1 - 1} + \sqrt{d_2 - 2} - \epsilon$

Theorem (Spectral gap). *Let* $A = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}$ *be the adjacency matrix of a bipartite, biregular random graph* $G \sim \mathcal{G}(n, m, d_1, d_2)$. *Without loss of generality, assume* $d_1 \geq d_2$ *or, equivalently,* $n \leq m$. *Then:*

(i) Its second largest eigenvalue $\eta = \lambda_2(A)$ *satisfies*

$$\eta \leq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} + \epsilon'_n$$

Attains Alon-Boppana

asymptotically almost surely, with $\epsilon'_n \to 0$ *as* $n \to \infty$.

(ii) Its smallest positive eigenvalue $\eta^+_{\min} = \min(\{\lambda \in \sigma(A) : \lambda > 0\})$ *satisfies*

$$\eta^+_{\min} \geq \sqrt{d_1 - 1} - \sqrt{d_2 - 1} - \epsilon''_n$$

asymptotically almost surely, with $\epsilon''_n \to 0$ *as* $n \to \infty$. *(Note that this will be almost surely positive if* $d_1 > d_2$*; no further information is gained if* $d_1 = d_2$*.)*

(iii) If $d_1 \neq d_2$*, the rank of* X *is* n *with high probability.*

First result of this kind for rectangular matrices, conjectured to hold for d-regular adjacency w/ d > 2 by Costello & Vu (2008)

Non-backtracking theorem

Theorem *If B is the non-backtracking matrix of a bipartite, biregular random graph $G \sim \mathcal{G}(n, m, d_1, d_2)$, then its second largest eigenvalue*

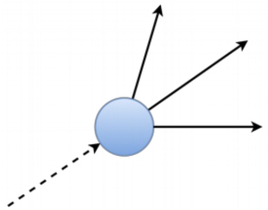
$$|\lambda_2(B)| \leq ((d_1 - 1)(d_2 - 1))^{1/4} + \epsilon_n$$

asymptotically almost surely, with $\epsilon_n \to 0$ as $n \to \infty$. Equivalently, there exists a sequence $\epsilon_n \to 0$ as $n \to \infty$ so that

$$\mathbb{P}\left[|\lambda_2(B)| - ((d_1 - 1)(d_2 - 1))^{1/4} > \epsilon_n\right] \to 0 \quad as \quad n \to \infty \ .$$

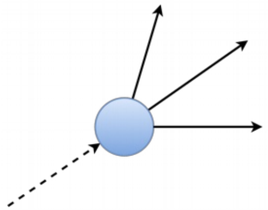
Main result follows from this

Sketch of proof

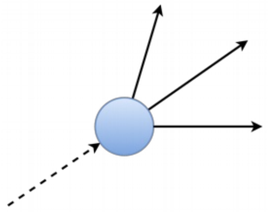


Sketch of proof

1) Look at non-backtracking matrix B

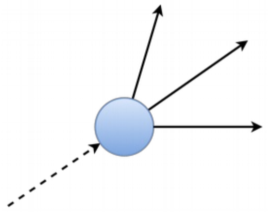


Sketch of proof



1) Look at non-backtracking matrix B
2) Subtract leading eigenspace: $\bar{B} = B - S$

Sketch of proof



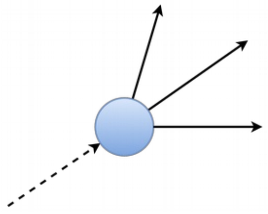
1) Look at non-backtracking matrix B
2) Subtract leading eigenspace: $\bar{B} = B - S$
3) Use moments to bound matrix norm:

$$\mathbb{E}\left(\|\bar{B}^{\ell}\|^{2k}\right) \leq \mathbb{E}\left(\mathrm{Tr}\,\underbrace{\left((\bar{B}^{\ell})(\bar{B}^{\ell})^{*}\right)^{k}}\right)$$

Trace depends on diagonals of this matrix

Diagonals = non-backtracking circuits

Sketch of proof



1) Look at non-backtracking matrix B
2) Subtract leading eigenspace: $\bar{B} = B - S$
3) Use moments to bound matrix norm:

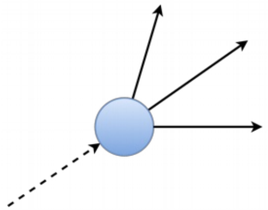
$$\mathbb{E}\left(\|\bar{B}^\ell\|^{2k}\right) \leq \mathbb{E}\left(\mathrm{Tr}\,\underbrace{\left((\bar{B}^\ell)(\bar{B}^\ell)^*\right)^k}\right)$$

Trace depends on diagonals of this matrix

Diagonals = non-backtracking circuits

4) Leading order is from circuits that are **trees**

Sketch of proof



1) Look at non-backtracking matrix B
2) Subtract leading eigenspace: $\bar{B} = B - S$
3) Use moments to bound matrix norm:

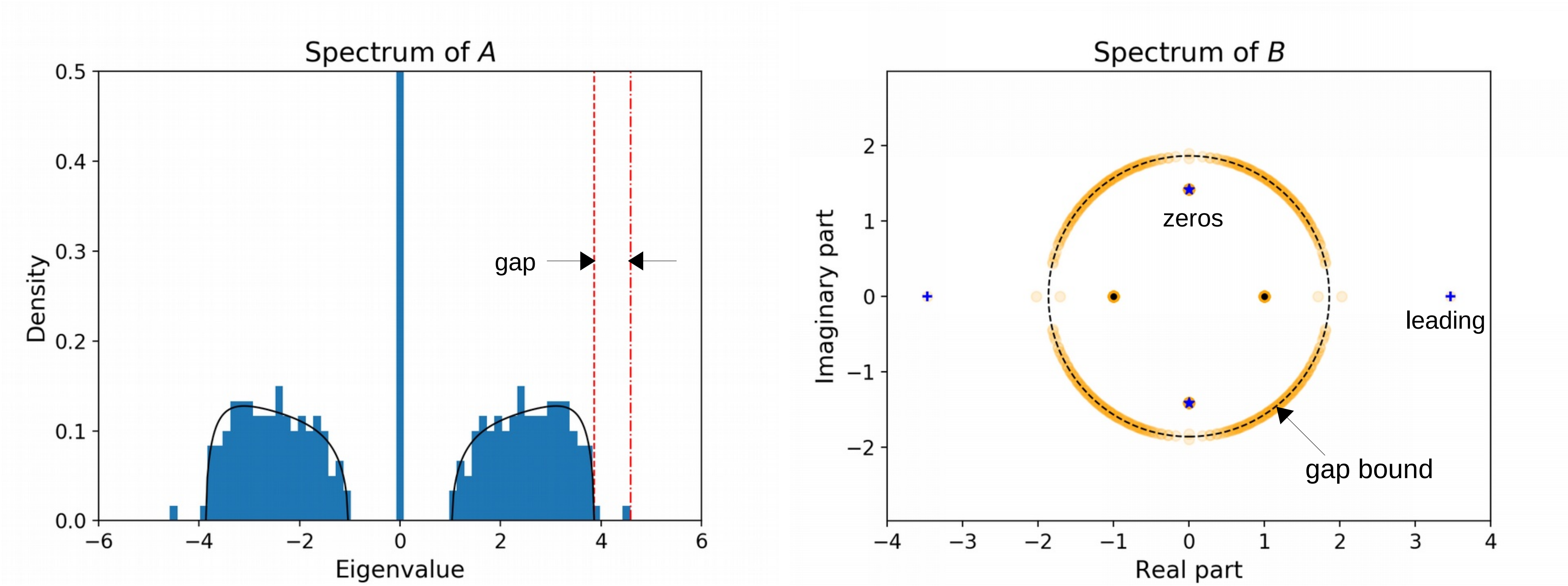
$$\mathbb{E}\left(\|\bar{B}^\ell\|^{2k}\right) \leq \mathbb{E}\left(\mathrm{Tr}\,\underbrace{\left((\bar{B}^\ell)(\bar{B}^\ell)^*\right)^k}\right)$$

Trace depends on diagonals of this matrix

Diagonals = non-backtracking circuits

4) Leading order is from circuits that are **trees**
5) Relate spectra of B and A (Ihara-Bass formula, “zeta” function)

Graphical depiction of the spectra



Ihara-Bass: $\sigma(B) = \{\pm 1\} \bigcup \{\lambda : \ D - \lambda A + \lambda^2 I \text{ is not invertible}\}$

So, wouldn't it be simpler to work with A directly?

- But it doesn't work. Take d-regular example:

Goal: $$\mathrm{Tr}\left(A - \frac{d}{n}11^*\right)^k \leq n\left(2\sqrt{d-1} + o(1)\right)^k$$

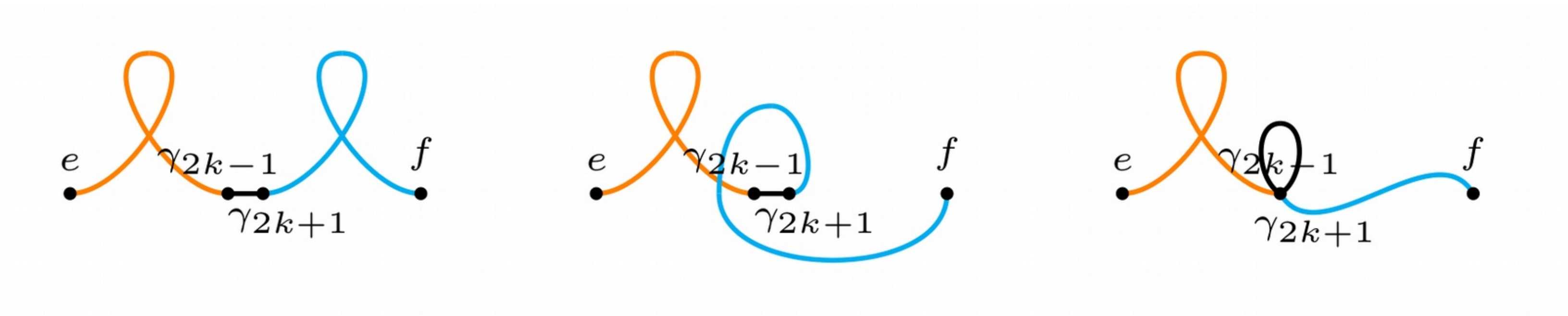
But: $$\mathbb{E}\,\mathrm{Tr}\left(A - \frac{d}{n}11^*\right)^k \geq d^k \mathbb{P}(\Omega) \geq d^k n^{-c}$$

Event = existence of isolated K_{d+1}
occurs with non-negligible prob.

Following Bordenave (2015), Friedman

Reason: “Tangled” paths

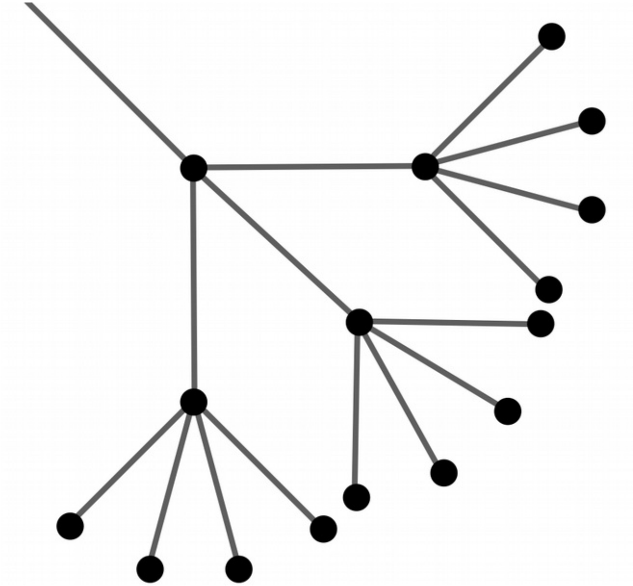
- Elucidated by Friedman, Bordenave



l-tangle-free: all *l*-neighborhoods contain at most one cycle

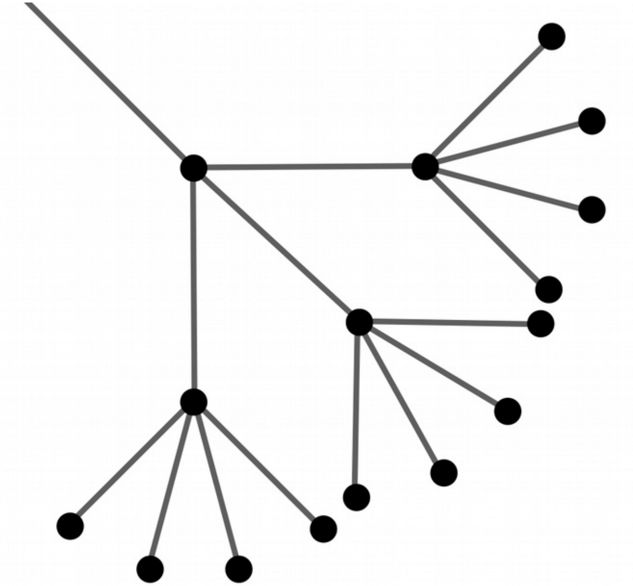
Lemma *Fix a constant $c < 1/8$ and let $\ell = c\log_d(n)$. Let $G \sim \mathcal{G}(n, m, d_1, d_2)$ be a bipartite, biregular random graph. Then G is ℓ-tangle-free with probability at least $1 - n^{-1/2}$.*

“Tree-like property”



Proof following Lubetzky & Sly (2010)

Lemma *Fix a constant $c < 1/8$ and let $\ell = c \log_d(n)$. Let $G \sim \mathcal{G}(n, m, d_1, d_2)$ be a bipartite, biregular random graph. Then G is ℓ-tangle-free with probability at least $1 - n^{-1/2}$.*

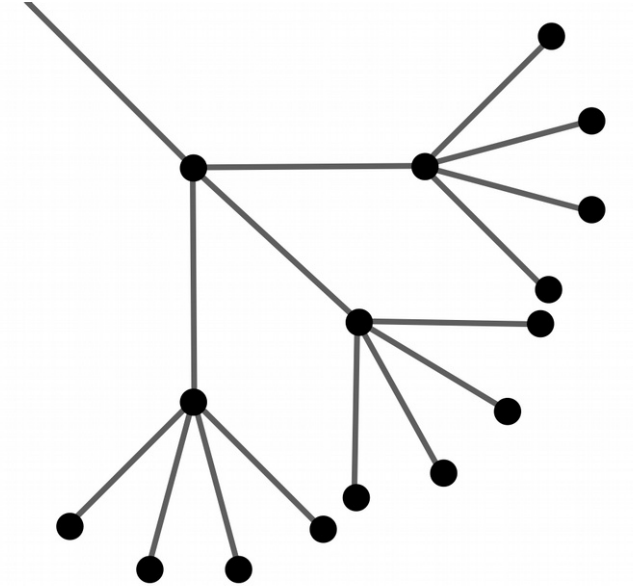


“Tree-like property”

Proof following Lubetzky & Sly (2010)

1) Configuration model: random matching of half-edges

Lemma *Fix a constant $c < 1/8$ and let $\ell = c\log_d(n)$. Let $G \sim \mathcal{G}(n, m, d_1, d_2)$ be a bipartite, biregular random graph. Then G is ℓ-tangle-free with probability at least $1 - n^{-1/2}$.*

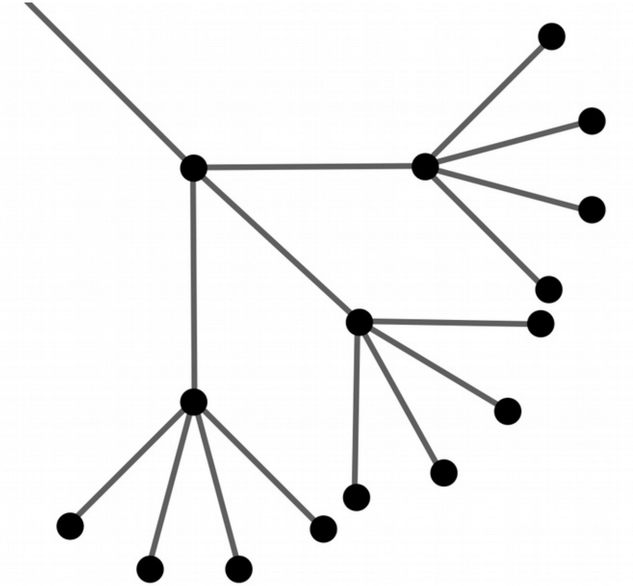


"Tree-like property"

Proof following Lubetzky & Sly (2010)

1) Configuration model: random matching of half-edges
2) Consider depth i exploration of neighborhood

Lemma *Fix a constant $c < 1/8$ and let $\ell = c\log_d(n)$. Let $G \sim \mathcal{G}(n, m, d_1, d_2)$ be a bipartite, biregular random graph. Then G is ℓ-tangle-free with probability at least $1 - n^{-1/2}$.*

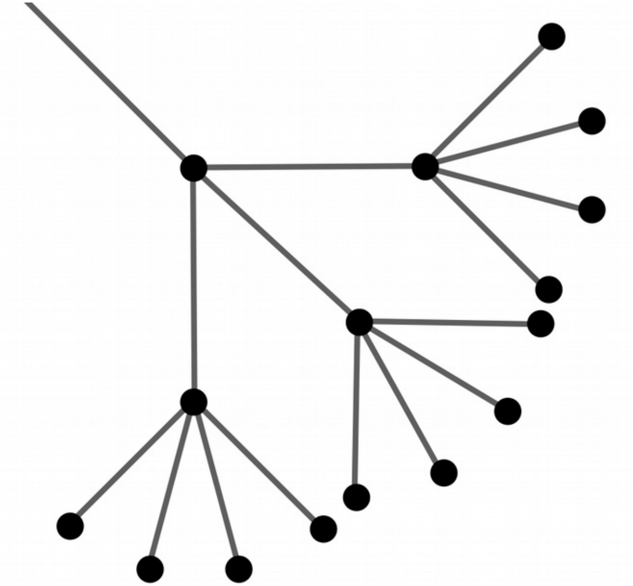


"Tree-like property"

<u>Proof</u> following Lubetzky & Sly (2010)

1) Configuration model: random matching of half-edges
2) Consider depth i exploration of neighborhood
3) At most d^{i+1} half-edges to match here

Lemma *Fix a constant $c < 1/8$ and let $\ell = c \log_d(n)$. Let $G \sim \mathcal{G}(n, m, d_1, d_2)$ be a bipartite, biregular random graph. Then G is ℓ-tangle-free with probability at least $1 - n^{-1/2}$.*

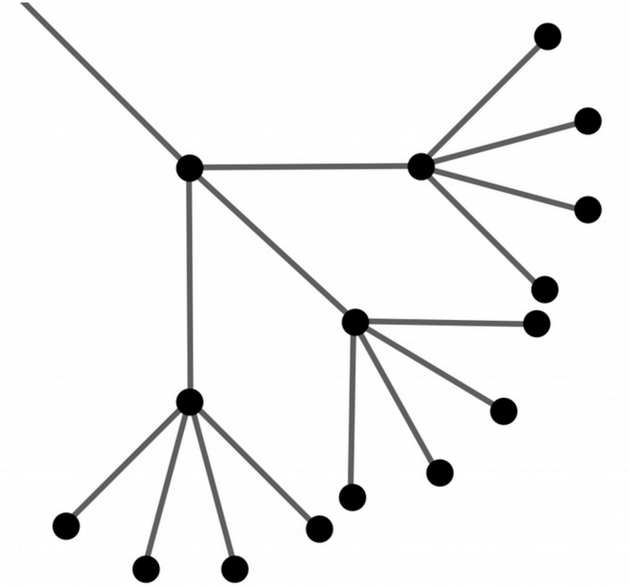


"Tree-like property"

Proof following Lubetzky & Sly (2010)

1) Configuration model: random matching of half-edges
2) Consider depth i exploration of neighborhood
3) At most d^{i+1} half-edges to match here
4) Consider event $A_{i,k}$: k^{th} edge in depth i → cycle

Lemma *Fix a constant* $c < 1/8$ *and let* $\ell = c \log_d(n)$. *Let* $G \sim \mathcal{G}(n, m, d_1, d_2)$ *be a bipartite, biregular random graph. Then* G *is* ℓ*-tangle-free with probability at least* $1 - n^{-1/2}$.



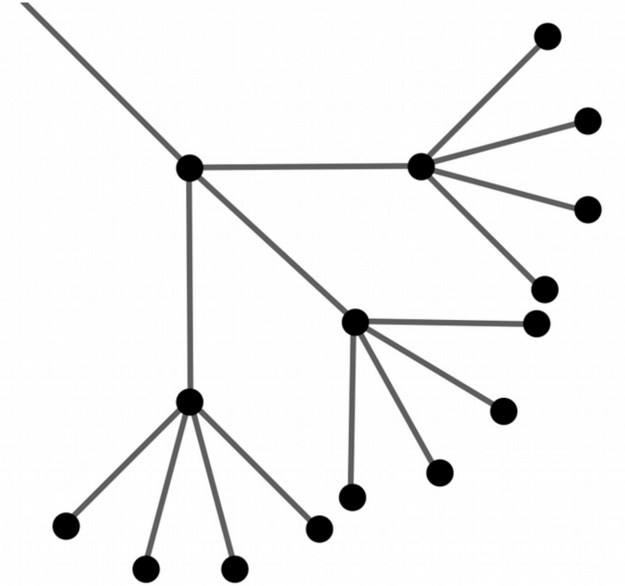
"Tree-like property"

Proof following Lubetzky & Sly (2010)

1) Configuration model: random matching of half-edges
2) Consider depth i exploration of neighborhood
3) At most d^{i+1} half-edges to match here
4) Consider event $A_{i,k}$: k^{th} edge in depth i → cycle

$$\mathbb{P}(A_{i,k}) \leq \frac{\overbrace{(k-1)(d-1)}^{\text{candidates}}}{\underbrace{nd - 2d^{i+1}}_{\text{total unmatched}}} \leq \frac{d^{i+1}(d-1)}{nd - 2d^{i+1}} \leq \frac{d^{i+1}}{n}$$

Lemma *Fix a constant $c < 1/8$ and let $\ell = c\log_d(n)$. Let $G \sim \mathcal{G}(n,m,d_1,d_2)$ be a bipartite, biregular random graph. Then G is ℓ-tangle-free with probability at least $1 - n^{-1/2}$.*



"Tree-like property"

<u>Proof</u> following Lubetzky & Sly (2010)

1) Configuration model: random matching of half-edges
2) Consider depth i exploration of neighborhood
3) At most d^{i+1} half-edges to match here
4) Consider event $A_{i,k}$: k^{th} edge in depth i → cycle

$$\mathbb{P}(A_{i,k}) \leq \frac{\overbrace{(k-1)(d-1)}^{\text{candidates}}}{\underbrace{nd - 2d^{i+1}}_{\text{total unmatched}}} \leq \frac{d^{i+1}(d-1)}{nd - 2d^{i+1}} \leq \frac{d^{i+1}}{n}$$

Finish with following & union bound over vertices

$$\mathbb{P}(B_\ell(v) \text{ is not } \ell\text{-tangle-free}) = \mathbb{P}\left(\sum_{i=1}^{\ell-1}\sum_{k=1}^{m_i} A_{i,k} > 1\right) \leq \mathbb{P}(Z > 1) = O\left(\frac{d^{4\ell+1}}{n^2}\right) = O\left(n^{-3/2}\right)$$

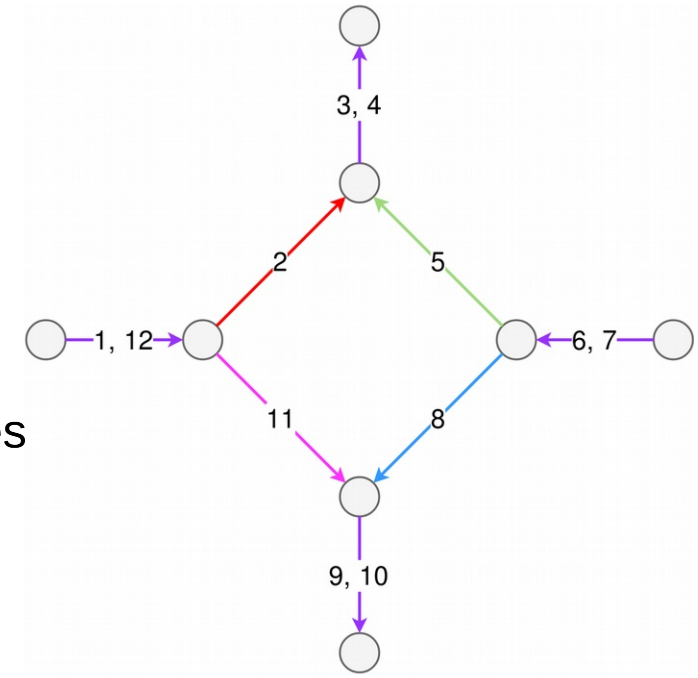
How this appears in the proof:

- We form a bound for this:

$$\mathbb{E}\left(\|\bar{B}^\ell\|^{2k}\right) \leq \mathbb{E}\left(\mathrm{Tr}\,\underbrace{\left((\bar{B}^\ell)(\bar{B}^\ell)^*\right)^k}\right)$$

Count non-backtracking circuits of diff types
Weight by expectation
Use tangle-free property

$$\ell \leq c\log(n), \quad k = \frac{\log(n)}{\log(\log(n))}$$



Example: k = l = 2

OK: on to applications of the gap

OK: on to applications of the gap

- Spectral “gap” is large $\rightarrow$ good expansion

OK: on to applications of the gap

- Spectral “gap” is large → good expansion
- Many applications:

OK: on to applications of the gap

- Spectral “gap” is large → good expansion
- Many applications:
  - Mixing rates of Markov chains

OK: on to applications of the gap

- Spectral “gap” is large → good expansion
- Many applications:
  - Mixing rates of Markov chains
  - Community detection via spectral clustering

OK: on to applications of the gap

- Spectral “gap” is large → good expansion
- Many applications:
  - Mixing rates of Markov chains
  - Community detection via spectral clustering
  - Error correcting codes constructed from graphs

OK: on to applications of the gap

- Spectral “gap” is large → good expansion
- Many applications:
  - Mixing rates of Markov chains
  - Community detection via spectral clustering
  - Error correcting codes constructed from graphs
  - Matrix completion

OK: on to applications of the gap

- Spectral “gap” is large → good expansion
- Many applications:
  - Mixing rates of Markov chains
  - Community detection via spectral clustering
  - Error correcting codes constructed from graphs
  - Matrix completion
- Common tool, the **expander mixing Lemma**:

$$\left|\frac{E(A,B)}{|E|} - \frac{|A||B|}{nm}\right| \leq \underbrace{\boxed{\frac{\lambda_2}{\sqrt{d_1 d_2}}}}_{\text{Gap}} \sqrt{\frac{|A||B||A^c||B^c|}{(nm)^2}}$$

A proof in: De Winter, Schillewaert, Verstraete (2012)

Matrix completion

		-1		
			1	
1	1	-1	1	-1
1				-1
		-1		

Observed

1	1	-1	1	-1
1	1	-1	1	-1
1	1	-1	1	-1
1	1	-1	1	-1
1	1	-1	1	-1

Find

Data points are “edges” in the graph:

$$(i, j) \in E \iff \text{entry } (i, j) \text{ is observed}$$

Expansion is related to complexity

Solve the problem:

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \gamma_2(X) \\ \text{subject to} \quad & X_{ij} = Y_{ij},\ (i,j) \in E, \end{aligned}$$

where $\gamma_2(Y) = \min_{UV^*=Y} \max_{i,j} \|u_i\|_2 \, \|v_j\|_2 \leq \|Y\|_*$

So, if observations from (d_1, d_2)-regular graph:

$$\frac{1}{nm} \|\hat{Y} - Y\|_F^2 \leq 7.2 \, \gamma_2(Y)^2 \frac{\lambda_2}{\sqrt{d_1 d_2}}$$

Extends & improves result of Heiman, Schechtman, Shraibman (2014)

Conclusions

- Proof that bipartite biregular graphs have large gap "Ramanujan"
  - (Hopefully) simpler to understand than in past
- Surprising side result: full rank matrix X in $A = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}$
- Highlighted some nice applications:
  - Mostly showing how knowing gap yields explicit bounds
    - Community detection in general
    - LDPC error correcting codes
    - Matrix completion
  - New result for rectangular matrix completion

Thank you for listening!

Scott Rinckenberger