Limiting eigenvalue distribution for the non-backtracking matrix of an Erdős-Rényi random graph

Ke Wang

Department of Mathematics
Hong Kong University of Science and Technology

May 22, 2018

Joint work with Philip Matchett Wood (University of Wisconsin-Madison).

Non-backtracking walks on graphs

Consider a graph G. The non-backtracking random walk traverses edges, with the constraint that most recently traversed edge may not be again traversed in the opposite direction.

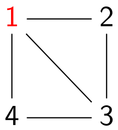
Non-backtracking walks on graphs

Consider a graph G. The non-backtracking random walk traverses edges, with the constraint that most recently traversed edge may not be again traversed in the opposite direction.

Consider a graph G. The non-backtracking random walk traverses edges, with the constraint that most recently traversed edge may not be again traversed in the opposite direction.

Example. Key: red is current position, blue is last position.

Original graph



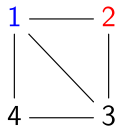
Edge graph (connects to next allowed edge)

Non-backtracking walks on graphs

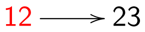
Consider a graph G. The non-backtracking random walk traverses edges, with the constraint that most recently traversed edge may not be again traversed in the opposite direction.

Example. Key: red is current position, blue is last position.

Original graph



Edge graph (connects to next allowed edge)

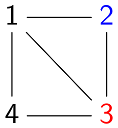


Non-backtracking walks on graphs

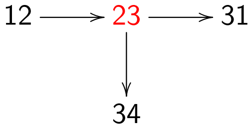
Consider a graph G. The non-backtracking random walk traverses edges, with the constraint that most recently traversed edge may not be again traversed in the opposite direction.

Example. Key: red is current position, blue is last position.

Original graph



Edge graph (connects to next allowed edge)

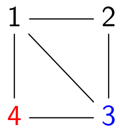


Non-backtracking walks on graphs

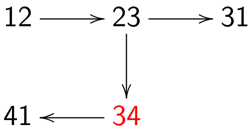
Consider a graph G. The non-backtracking random walk traverses edges, with the constraint that most recently traversed edge may not be again traversed in the opposite direction.

Example. Key: red is current position, blue is last position.

Original graph



Edge graph (connects to next allowed edge)

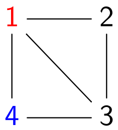


Non-backtracking walks on graphs

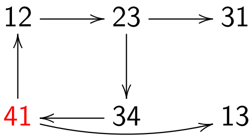
Consider a graph G. The non-backtracking random walk traverses edges, with the constraint that most recently traversed edge may not be again traversed in the opposite direction.

Example. Key: red is current position, blue is last position.

Original graph



Edge graph (connects to next allowed edge)

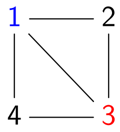


Non-backtracking walks on graphs

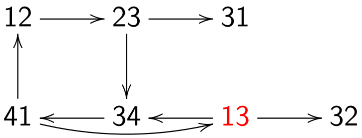
Consider a graph G. The non-backtracking random walk traverses edges, with the constraint that most recently traversed edge may not be again traversed in the opposite direction.

Example. Key: red is current position, blue is last position.

Original graph



Edge graph (connects to next allowed edge)

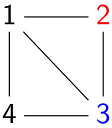


Non-backtracking walks on graphs

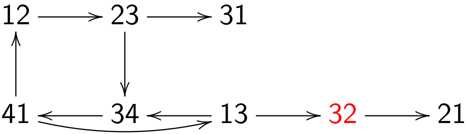
Consider a graph G. The non-backtracking random walk traverses edges, with the constraint that most recently traversed edge may not be again traversed in the opposite direction.

Example. Key: red is current position, blue is last position.

Original graph



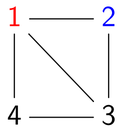
Edge graph (connects to next allowed edge)



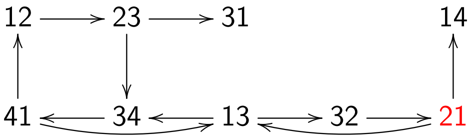
Consider a graph G. The non-backtracking random walk traverses edges, with the constraint that most recently traversed edge may not be again traversed in the opposite direction.

Example. Key: red is current position, blue is last position.

Original graph



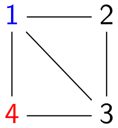
Edge graph (connects to next allowed edge)



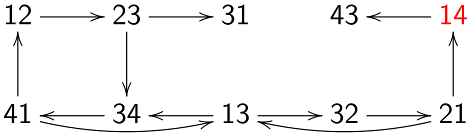
Consider a graph G. The non-backtracking random walk traverses edges, with the constraint that most recently traversed edge may not be again traversed in the opposite direction.

Example. Key: red is current position, blue is last position.

Original graph



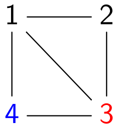
Edge graph (connects to next allowed edge)



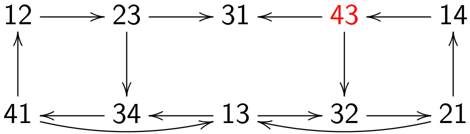
Consider a graph G. The non-backtracking random walk traverses edges, with the constraint that most recently traversed edge may not be again traversed in the opposite direction.

Example. Key: red is current position, blue is last position.

Original graph



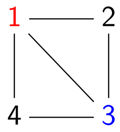
Edge graph (connects to next allowed edge)



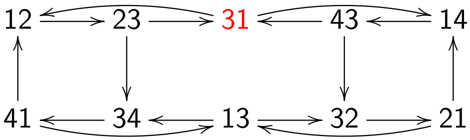
Consider a graph G. The non-backtracking random walk traverses edges, with the constraint that most recently traversed edge may not be again traversed in the opposite direction.

Example. Key: red is current position, blue is last position.

Original graph



Edge graph (connects to next allowed edge)



- Alon, Benjamini, Lubetzky and Sodin (2006): Non-backtracking walks mix faster than simple random walks on regular expander graphs.
- Ben-Hamou, Lubetzky and Peres (2018): Non-backtracking walks mix faster than simple random walks on graphs with minimum degree 3 and degree distribution has exponential tails.
- Cut-off phenomena of non-backtracking walks: Berestycki, Lubetzky, Peres and Sly (2015), Ben-Hamou and Salez (2015) etc.

- Alon, Benjamini, Lubetzky and Sodin (2006): Non-backtracking walks mix faster than simple random walks on regular expander graphs.
- Ben-Hamou, Lubetzky and Peres (2018): Non-backtracking walks mix faster than simple random walks on graphs with minimum degree 3 and degree distribution has exponential tails.
- Cut-off phenomena of non-backtracking walks: Berestycki, Lubetzky, Peres and Sly (2015), Ben-Hamou and Salez (2015) etc.

Non-backtracking matrix B

A simple undirected graph $G = (V, E)$.

- Adjacency matrix A: $A_{ij} = 1$ if $(i, j) \in E$ and 0 otherwise.
- Non-backtracking matrix: for each $(i, j) \in E$, form two directed edges $i \to j$ and $j \to i$. The non-backtracking matrix B is a $2|E| \times 2|E|$ matrix such that

$$B_{i \to j, k \to l} = \begin{cases} 1 & \text{if } j = k \text{ and } i \neq l \\ 0 & \text{otherwise.} \end{cases}$$

Non-backtracking matrix is first introduced by Hashimoto (1989).

- Entries of A^k: the number of walks of length k from one vertex to another.
- Entries of B^k: the number of non-backtracking walks of length k from one directed edge to another.

Non-backtracking matrix B

A simple undirected graph $G = (V, E)$.

- Adjacency matrix A: $A_{ij} = 1$ if $(i, j) \in E$ and 0 otherwise.
- Non-backtracking matrix: for each $(i, j) \in E$, form two directed edges $i \to j$ and $j \to i$. The non-backtracking matrix B is a $2|E| \times 2|E|$ matrix such that

$$B_{i \to j, k \to l} = \begin{cases} 1 & \text{if } j = k \text{ and } i \neq l \\ 0 & \text{otherwise.} \end{cases}$$

Non-backtracking matrix is first introduced by Hashimoto (1989).

- Entries of A^k: the number of walks of length k from one vertex to another.
- Entries of B^k: the number of non-backtracking walks of length k from one directed edge to another.

Goal: Understand the eigenvalues of B.

Theorem (Ihara's formula)

$$\det(I - uB) = (1 - u^2)^{|E|-|V|} \det(I - uA + u^2(D - I)).$$

Here D is the diagonal degree matrix, D_{ii} = degree of vertex i.

- A simple relationship between spectrum of A and B when G is regular.

Spectrum of the non-backtracking matrix B

<u>Goal:</u> Understand the eigenvalues of B.

Theorem (Ihara's formula)

$$\det(I - uB) = (1 - u^2)^{|E|-|V|} \det(I - uA + u^2(D - I)).$$

Here D is the diagonal degree matrix, $D_{ii} =$ degree of vertex i.

- A simple relationship between spectrum of A and B when G is regular.

Spectrum of the non-backtracking matrix B

Goal: Understand the eigenvalues of B.

Theorem (Ihara's formula)

$$\det(I - uB) = (1-u^2)^{|E|-|V|}\det(I - uA + u^2(D-I)).$$

Here D is the diagonal degree matrix, D_{ii} = degree of vertex i.

- A simple relationship between spectrum of A and B when G is regular.

Previous results on the spectrum of B

- Angel, Friedman and Hoory (2015): spectrum of B on tree cover of a finite graph.
- Bordenave, Lelarge and Massoulié (2015): top eigenvalues of B on (very sparse) stochastic block model and Erdős-Rényi $G(n, \frac{c}{n})$.
- Gulikers, Lelarge and Massoulié (2016): top eigenvalues of B on generalized stochastic block models.
- Benaych-Georges, Bordenave and Knowles (2017): spectral radius of B on inhomogeneous Erdős-Rényi random graph.

Ihara's formula: $\det(I - uB) = (1 - u^2)^{|E|-|V|} \det(I - uA + u^2(D - I))$.

$$\text{eigenvalues of } B = \{\pm 1\} \cup \{ \text{ eigenvalues of } H := \begin{pmatrix} A & I - D \\ I & 0 \end{pmatrix} \}.$$

We will call H the *non-backtracking spectrum operator* for the graph.

Non-backtracking matrix of Erdős-Rényi random graphs

Erdős-Rényi random graph $G(n, p)$: edges are drawn independently with probability p. Adjacency matrix A and non-backtracking spectrum operator

$$H = \begin{pmatrix} A & I - D \\ I & 0 \end{pmatrix}.$$

Partial derandomization: replace D by its average, rest of H unchanged. Let $\alpha = (n-1)p - 1$, and define

$$H_0 = \begin{pmatrix} A & I - \mathbb{E}D \\ I & 0 \end{pmatrix} = \begin{pmatrix} A & -\alpha I \\ I & 0 \end{pmatrix}.$$

Heuristic: if $p \gg \log n/n$, an Erdős-Rényi graph $\approx$ a regular graph.

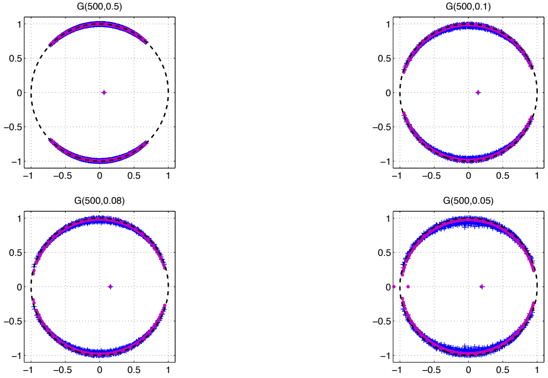


Figure: Spectrum of H_0 (red $\cdot$) and H (blue $+$) for $G(n, p)$.

Spectrum of partly averaged matrix H_0

Eigenvalues of H_0 can be quantified by the eigenvalues of A.

Theorem (combining FK81,KS03,Vu07, BGBK17)

Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of A. If $p \gg \frac{\log n}{n}$, then the following two facts hold a.a.s.:

$$\lambda_1 = np(1 + o(1));$$

$$\max_{2 \leq i \leq n} |\lambda_i| \leq 2\sqrt{np(1-p)}(1 + o(1)).$$

Theorem (W.-Wood 2018; spectrum of H_0)

If $p \gg \frac{1}{\sqrt{n}}$, then almost surely for large n, the matrix H_0 has 2 real eigenvalues

$$\mu_1 = np(1+o(1)) \quad \textit{and} \quad \mu_2 = 1 + o(1).$$

All other eigenvalues μ are complex with magnitude

$$|\mu| = \sqrt{\alpha} = \sqrt{(n-1)p-1} = \sqrt{np}(1+o(1)),$$

and have real parts distributed according to the semi-circular law.

Explicit diagonalization: $\det(xI - H_0) = \prod_{i=1}^{n}(x^2 - \lambda_i x + \alpha)$. Roots are $\frac{1}{2}\left(\lambda_i \pm \sqrt{\lambda_i^2 - 4\alpha}\right)$. Cases: $|\lambda_i| \geq 2\sqrt{\alpha}$ (real eigenvals) or $|\lambda_i| < 2\sqrt{\alpha}$. Compare $2\sqrt{\alpha} = 2\sqrt{(n-1)p-1}$ with $2\sqrt{np(1-p)}$.

Theorem (W.-Wood 2018; spectrum of H_0)

If $p \gg \frac{1}{\sqrt{n}}$, then almost surely for large n, the matrix H_0 has 2 real eigenvalues

$$\mu_1 = np(1 + o(1)) \quad \textit{and} \quad \mu_2 = 1 + o(1).$$

All other eigenvalues μ are complex with magnitude

$$|\mu| = \sqrt{\alpha} = \sqrt{(n-1)p - 1} = \sqrt{np}(1 + o(1)),$$

and have real parts distributed according to the semi-circular law.

Explicit diagonalization: $\det(xI - H_0) = \prod_{i=1}^{n}(x^2 - \lambda_i x + \alpha)$. Roots are $\frac{1}{2}\left(\lambda_i \pm \sqrt{\lambda_i^2 - 4\alpha}\right)$. Cases: $|\lambda_i| \geq 2\sqrt{\alpha}$ (real eigenvals) or $|\lambda_i| < 2\sqrt{\alpha}$. Compare $2\sqrt{\alpha} = 2\sqrt{(n-1)p - 1}$ with $2\sqrt{np(1-p)}$.

Are spectra of H_0 and H close in bulk?

For a matrix M_n with eigenvalues $\lambda_1, \dots, \lambda_n$, the function

$$\mu_{M_n}(z) = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i}(z)$$

is the *empirical spectral measure*.

Theorem (W.-Wood 2018; bulk convergence)

Assume $np/\log n \to \infty$. *Then*

$$\mu_{\frac{1}{\sqrt{\alpha}}H} - \mu_{\frac{1}{\sqrt{\alpha}}H_0} \to 0 \quad a.s. \quad as\ n \to \infty.$$

Sketch of proof: perturbative approach

View H **as a perturbation of** H_0 where

$$E = \frac{1}{\sqrt{\alpha}} H - \frac{1}{\sqrt{\alpha}} H_0 = \begin{pmatrix} 0 & I + \frac{1}{\alpha}(I - D) \\ 0 & 0 \end{pmatrix}.$$

Warning: eigenvalue perturbation of non-normal matrices can be tricky.

Idea: apply the Tao-Vu replacement principle as a perturbation result, comparing

$$\frac{1}{\sqrt{\alpha}} H_0 \quad \text{with} \quad \frac{1}{\sqrt{\alpha}} H = \frac{1}{\sqrt{\alpha}} H_0 + E.$$

Sketch of proof: perturbative approach

View H **as a perturbation of** H_0 where

$$E = \frac{1}{\sqrt{\alpha}}H - \frac{1}{\sqrt{\alpha}}H_0 = \begin{pmatrix} 0 & I + \frac{1}{\alpha}(I - D) \\ 0 & 0 \end{pmatrix}.$$

Warning: eigenvalue perturbation of non-normal matrices can be tricky.

Idea: apply the Tao-Vu replacement principle as a perturbation result, comparing

$$\frac{1}{\sqrt{\alpha}}H_0 \quad \text{with} \quad \frac{1}{\sqrt{\alpha}}H = \frac{1}{\sqrt{\alpha}}H_0 + E.$$

Sketch of proof: perturbative approach

View H **as a perturbation of** H_0 where

$$E = \frac{1}{\sqrt{\alpha}}H - \frac{1}{\sqrt{\alpha}}H_0 = \begin{pmatrix} 0 & I + \frac{1}{\alpha}(I - D) \\ 0 & 0 \end{pmatrix}.$$

Warning: eigenvalue perturbation of non-normal matrices can be tricky.

Idea: apply the Tao-Vu replacement principle as a perturbation result, comparing

$$\frac{1}{\sqrt{\alpha}}H_0 \quad \text{with} \quad \frac{1}{\sqrt{\alpha}}H = \frac{1}{\sqrt{\alpha}}H_0 + E.$$

The Tao-Vu replacement principle for perturbations

Theorem (Tao-Vu, with appendix by Krishnapur, 2010)

If $\frac{1}{n}\|M_n\|_F^2 + \frac{1}{n}\|M_n + P_n\|_F^2$ bounded a.s. and if for almost every $z \in \mathbb{C}$

$$\frac{1}{n}\log|\det(M_n - zI)| - \frac{1}{n}\log|\det(M_n + P_n - zI)| \to 0$$

almost surely, **then** $\mu_{M_n} - \mu_{M_n+P_n} \to 0$ *almost surely.*

1) Note that M_n and $M_n + P_n$ can be dependent (internally, too)
2) Determinant is a product of singular values.
3) In fact, the following conditions imply converging log determinants:
There exists $f(z,n) \geq 1$, a function of n and z so that
(i) $f(z,n)\|P_n\| \to 0$ almost surely, and
(ii) $\|(M_n - zI)^{-1}\| \leq f(z,n)$ almost surely, for all suff. large n.

The Tao-Vu replacement principle for perturbations

Theorem (Tao-Vu, with appendix by Krishnapur, 2010)

If $\frac{1}{n}\|M_n\|_F^2 + \frac{1}{n}\|M_n + P_n\|_F^2$ *bounded a.s. and if for almost every* $z \in \mathbb{C}$

$$\frac{1}{n}\log|\det(M_n - zI)| - \frac{1}{n}\log|\det(M_n + P_n - zI)| \to 0$$

almost surely, **then** $\mu_{M_n} - \mu_{M_n+P_n} \to 0$ *almost surely.*

1) Note that M_n and $M_n + P_n$ can be dependent (internally, too)

2) Determinant is a product of singular values.

3) In fact, the following conditions imply converging log determinants:

There exists $f(z,n) \geq 1$, a function of n and z so that

(i) $f(z,n)\|P_n\| \to 0$ almost surely, and

(ii) $\|(M_n - zI)^{-1}\| \leq f(z,n)$ almost surely, for all suff. large n.

The Tao-Vu replacement principle for perturbations

Theorem (Tao-Vu, with appendix by Krishnapur, 2010)

If $\frac{1}{n}\|M_n\|_F^2 + \frac{1}{n}\|M_n + P_n\|_F^2$ *bounded a.s. and if for almost every* $z \in \mathbb{C}$

$$\frac{1}{n}\log|\det(M_n - zI)| - \frac{1}{n}\log|\det(M_n + P_n - zI)| \to 0$$

almost surely, **then** $\mu_{M_n} - \mu_{M_n+P_n} \to 0$ *almost surely.*

1) Note that M_n and $M_n + P_n$ can be dependent (internally, too)
2) Determinant is a product of singular values.
3) In fact, the following conditions imply converging log determinants:
There exists $f(z, n) \geq 1$, a function of n and z so that
(i) $f(z, n)\|P_n\| \to 0$ almost surely, and
(ii) $\|(M_n - zI)^{-1}\| \leq f(z, n)$ almost surely, for all suff. large n.

The Tao-Vu replacement principle for perturbations

Theorem (Tao-Vu, with appendix by Krishnapur, 2010)

If $\frac{1}{n}\|M_n\|_F^2 + \frac{1}{n}\|M_n + P_n\|_F^2$ *bounded a.s. and if for almost every* $z \in \mathbb{C}$

$$\frac{1}{n}\log|\det(M_n - zI)| - \frac{1}{n}\log|\det(M_n + P_n - zI)| \to 0$$

almost surely, **then** $\mu_{M_n} - \mu_{M_n+P_n} \to 0$ *almost surely.*

1) Note that M_n and $M_n + P_n$ can be dependent (internally, too)

2) Determinant is a product of singular values.

3) In fact, the following conditions imply converging log determinants:

There exists $f(z,n) \geq 1$, a function of n and z so that

(i) $f(z,n)\|P_n\| \to 0$ almost surely, and

(ii) $\|(M_n - zI)^{-1}\| \leq f(z,n)$ almost surely, for all suff. large n.

Sketch of proof

Apply with $M_n = \frac{1}{\sqrt{\alpha}} H_0$ and $P_n = E$. We verify the following conditions.

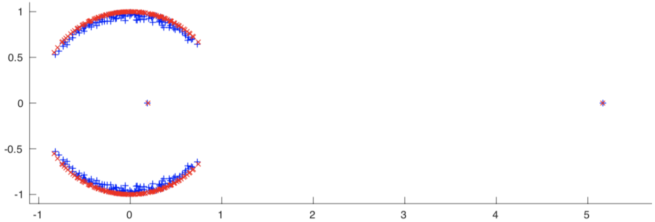
- Both $\frac{1}{2n}\|M_n\|_F^2$ and $\frac{1}{2n}\|M_n + P_n\|_F^2$ are almost surely bounded.
- $\|P_n\| \to 0$ almost surely.
- Construct a constant $C(z) > 0$ such that $\|(M_n - zI)^{-1}\| \leq C(z)$ almost surely.

Distance between eigenvalues of H and H_0

Theorem (Wang-W. 2018)

Let $p \gg \frac{\log^{3/2} n}{n^{1/6}}$. Then with probability $1 - o(1)$, every eigenvalue of $\frac{1}{\sqrt{\alpha}} H$ is within $R = 40\sqrt{\frac{\log n}{np^2}}$ of an eigenvalue of $\frac{1}{\sqrt{\alpha}} H_0$.

Eigenvalues of $\frac{1}{\sqrt{\alpha}} H$ and $\frac{1}{\sqrt{\alpha}} H_0$ when $n = 100$ and $p = 0.3$



Sketch of Proof

Apply the Bauer-Fike theorem.

Theorem (Bauer-Fike theorem)

If H_0 is diagonalizable by the matrix Y, then

$$\max_j \min_i |\mu_j(H_0 + E) - \mu_i(H_0)| \leq \|E\| \cdot \|Y\| \cdot \|Y^{-1}\|.$$

For eigenvalue λ_i of A, μ_{2i-1} and μ_{2i} solutions of $\mu^2 - \lambda_i \mu + \alpha = 0$. Let v_i be the eigenvector of A. Then $Y^{-1} H_0 Y = \text{diag}(\mu_1, \ldots, \mu_{2n})$ where

$$Y = (Y_1, \ldots, Y_n) \quad \text{and} \quad Y_i = \begin{pmatrix} \mu_{2i-1} v_i & \mu_{2i} v_i \\ v_i & v_i \end{pmatrix}.$$

The eigenvalues of $Y^* Y$ are unions of eigenvalues of $Y_i^* Y_i$.

Also we show $\|E\| \leq 20\sqrt{\frac{\log n}{np}}$ w.h.p.

Sketch of Proof

Apply the Bauer-Fike theorem.

Theorem (Bauer-Fike theorem)

If H_0 is diagonalizable by the matrix Y, then

$$\max_j \min_i |\mu_j(H_0 + E) - \mu_i(H_0)| \le \|E\| \cdot \|Y\| \cdot \|Y^{-1}\|.$$

For eigenvalue λ_i of A, μ_{2i-1} and μ_{2i} solutions of $\mu^2 - \lambda_i \mu + \alpha = 0$. Let v_i be the eigenvector of A. Then $Y^{-1} H_0 Y = \text{diag}(\mu_1, \dots, \mu_{2n})$ where

$$Y = (Y_1, \dots, Y_n) \quad \text{and} \quad Y_i = \begin{pmatrix} \mu_{2i-1} v_i & \mu_{2i} v_i \\ v_i & v_i \end{pmatrix}.$$

The eigenvalues of $Y^* Y$ are unions of eigenvalues of $Y_i^* Y_i$.
Also we show $\|E\| \le 20\sqrt{\frac{\log n}{np}}$ w.h.p.

Future questions

- Mixing rate of non-backtracking walks on random graphs

Consider the transition probability matrix P of size $2|E| \times 2|E|$ of the non-backtracking walks.

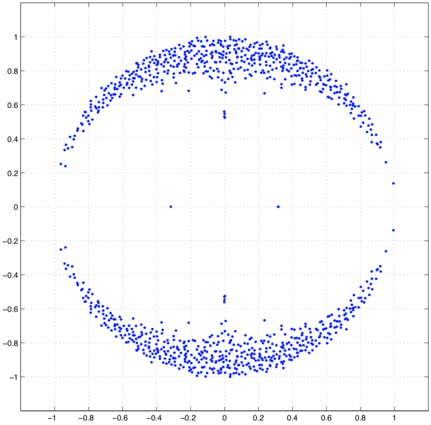
$$P_{i\to j,k\to l} = \begin{cases} \frac{1}{d(j)-1} & \text{if } j = k \text{ and } i \neq l \\ 0 & \text{otherwise.} \end{cases}$$

Top eigenvalues of P contain information of the mixing rate of non-backtracking walks.

Future questions

- Spectrum of B for very sparse random graphs

For $G(n, \frac{c}{n})$, Bordenave, Lelarge and Massoulié (2015): $\lambda_1 = c + o(1)$ and $\max_{i \neq 1} |\lambda_i| \leq \sqrt{c} + o(1)$. Plot of eigenvalues of B of $G(n,p)$ with $n = 500$ and $p = \frac{10}{500}$.



- Properties of eigenvectors of B

Observed: top eigenvectors of B are usually robust against localization.

- Krzakala, Moore, Mossel, Neeman, Sly, Zdeborová, Zhang (2013): "spectral redemption conjecture" for stochastic block model. The second eigenvector of B contains information on the global block structure.

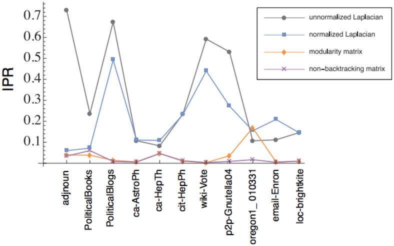


Figure: taken from T. Kawamoto (2016) "Localized eigenvectors of the non-backtracking matrix". IPR$=\frac{\sum_i v_i^4}{(\sum_i v_i^2)^2}$.

THANK YOU!