Order Detection under Pairwise Measurements

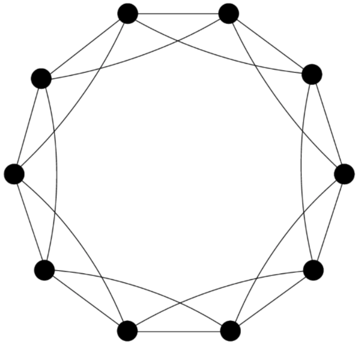
Jiaming Xu

Krannert School of Management
Purdue University

Joint work with
Vivek Bagaria and David Tse (Stanford)
Yihong Wu (Yale)

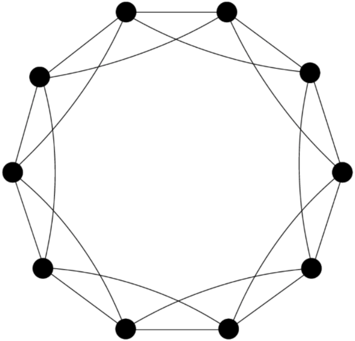
Simons Reunion Workshop, June 8, 2017

Order detection in small-world networks [Cai-Liang-Rakhlin '16]



4-circulant graph

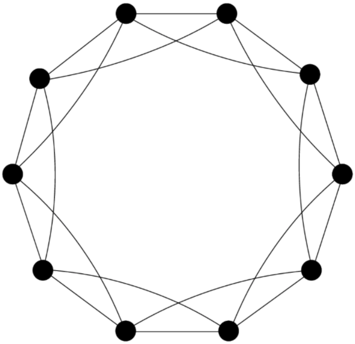
Order detection in small-world networks [Cai-Liang-Rakhlin '16]



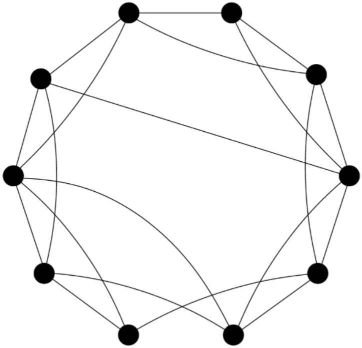
4-circulant graph

- Edge becomes non-edge with probability $1 - p$
- Non-edge becomes edge with probability q

Order detection in small-world networks [Cai-Liang-Rakhlin '16]

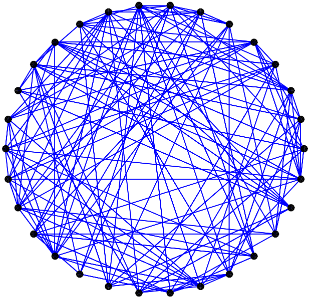


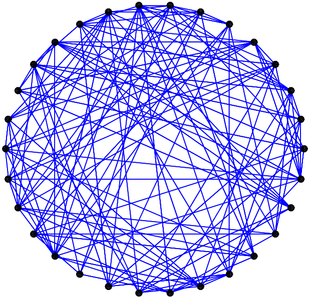
4-circulant graph

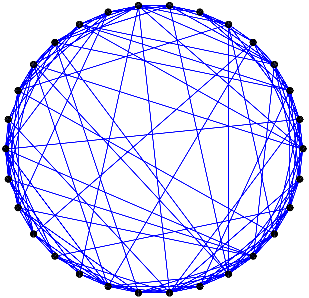


small-world graph

- Edge becomes non-edge with probability $1 - p$
- Non-edge becomes edge with probability q

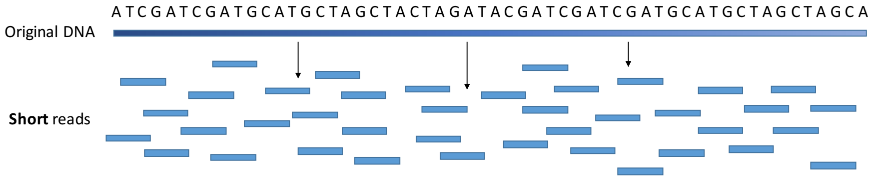




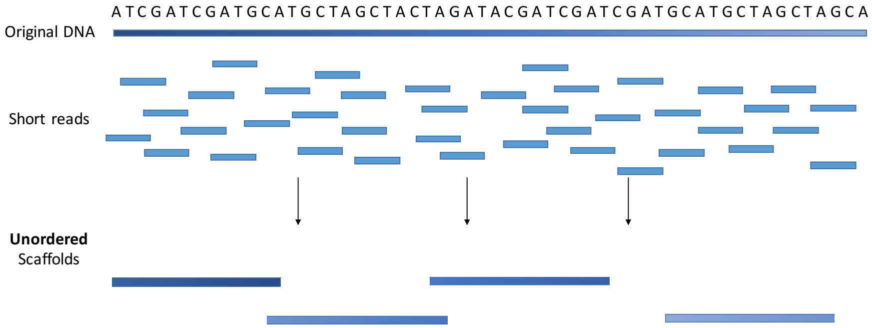


Goal: recover the underlying vertex ordering from observed graph

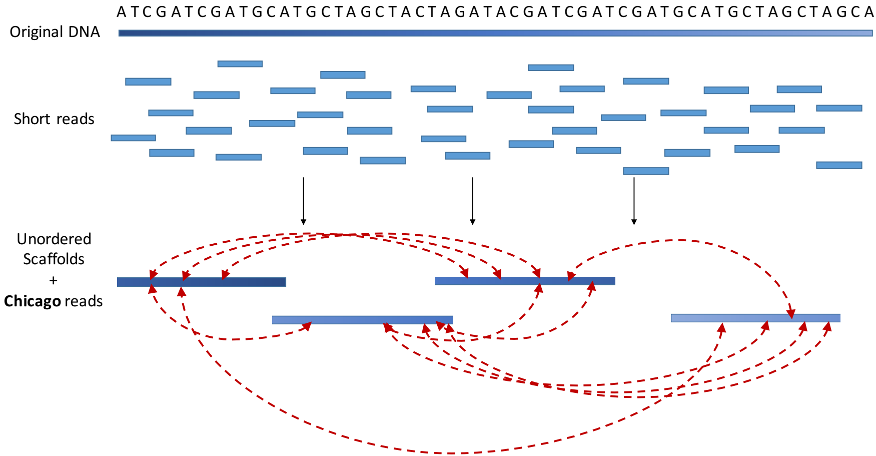
Ordering DNA scaffolds with Chicago reads



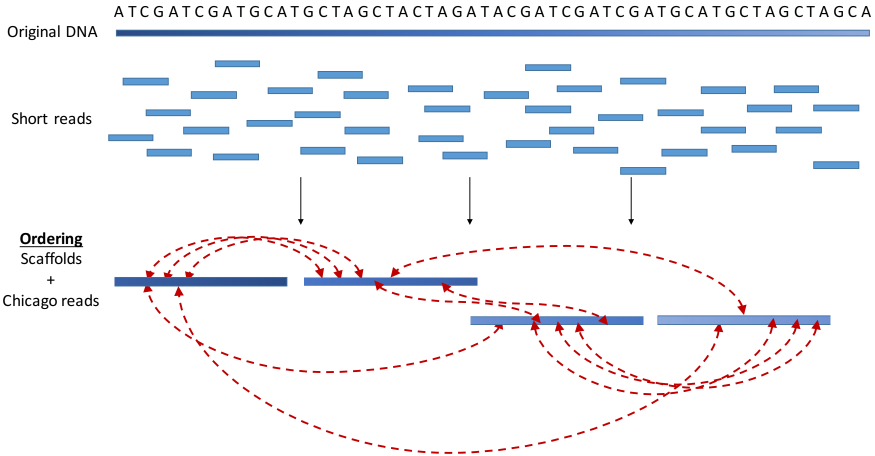
Ordering DNA scaffolds with Chicago reads



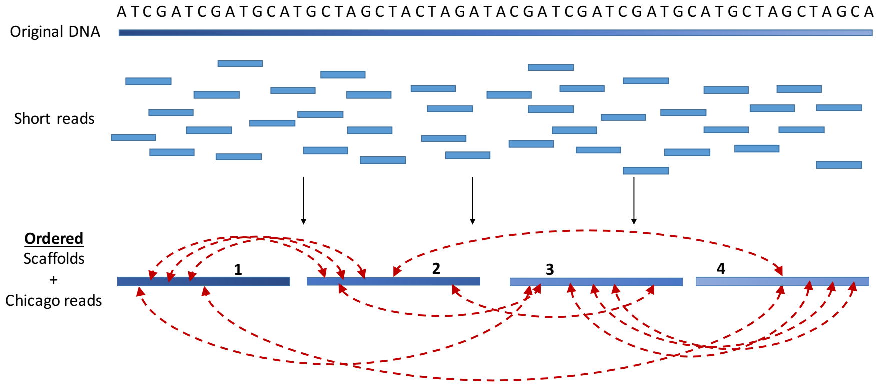
Ordering DNA scaffolds with Chicago reads



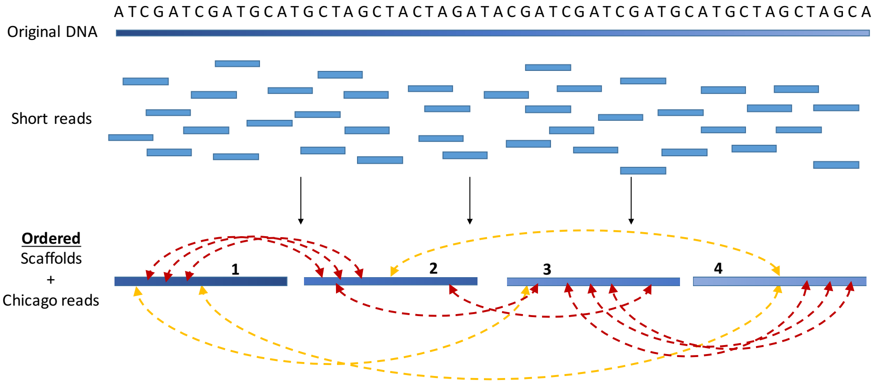
Ordering DNA scaffolds with Chicago reads



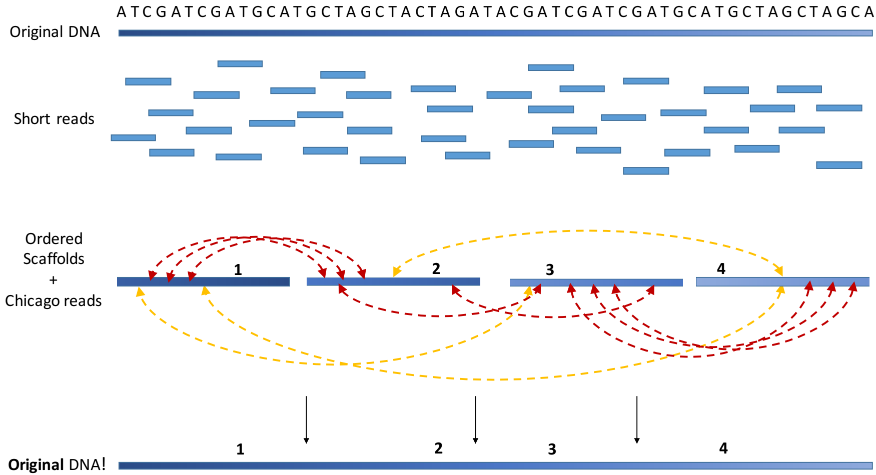
Ordering DNA scaffolds with Chicago reads



Ordering DNA scaffolds with Chicago reads



Ordering DNA scaffolds with Chicago reads



Ordering DNA scaffolds with Chicago reads

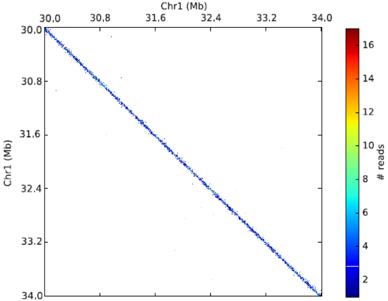


Figure S3 from [Putnam et al. 16]
Chicago reads

Ordering DNA scaffolds with Chicago reads

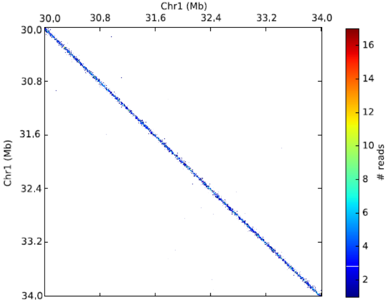
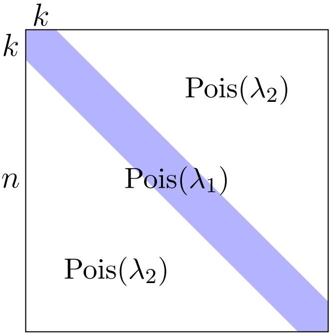


Figure S3 from [Putnam et al. 16]
Chicago reads



Ordering DNA scaffolds with Chicago reads

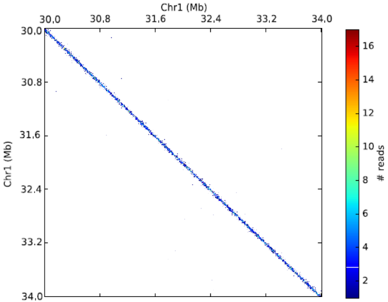
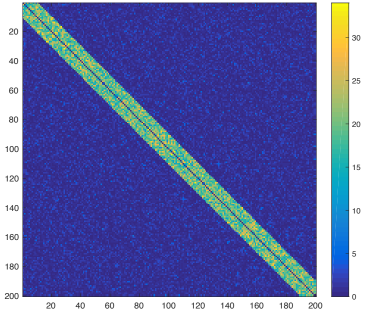


Figure S3 from [Putnam et al. 16]
Chicago reads



$n = 200,\ k = 10,\ \lambda_1 = 20,\ \lambda_2 = 1$

Ordering DNA scaffolds with Chicago reads

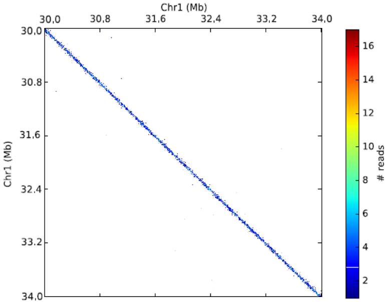
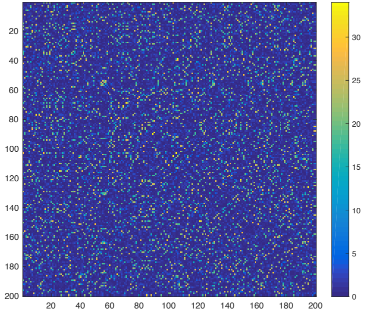


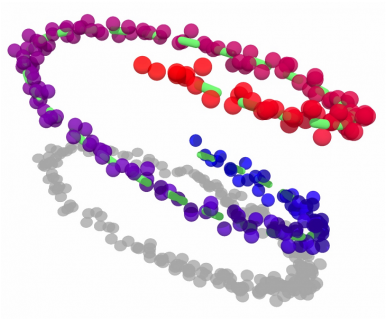
Figure S3 from [Putnam et al. 16]
Chicago reads



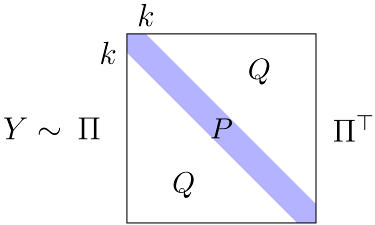
Goal: recover hidden permutation

Data seriation (stringing) [Kendall 71']

- Given a similarity matrix Y for n objects
- Ordering the n objects so that similar objects are near each other

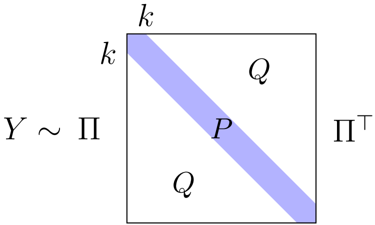


A Planted Ordering Model



- Π is the permutation matrix corresponding to ordering π

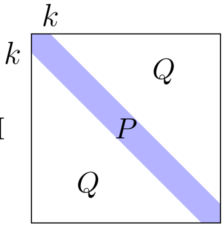
A Planted Ordering Model



- Π is the permutation matrix corresponding to ordering π
- $Y_{ii} = 0$ and for $i \neq j$:

$$Y_{ij} \sim \begin{cases} P & \text{if } |\pi(i) - \pi(j)| \leq k \\ Q & \text{otherwise} \end{cases}$$

A Planted Ordering Model

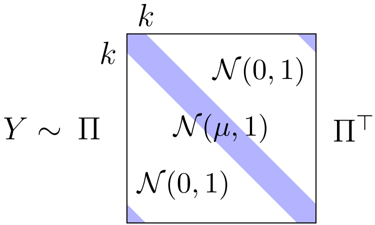
$Y \sim \Pi$  $\Pi^\top$

- Π is the permutation matrix corresponding to ordering π
- $Y_{ii} = 0$ and for $i \neq j$:

$$Y_{ij} \sim \begin{cases} P & \text{if } |\pi(i) - \pi(j)| \leq k \\ Q & \text{otherwise} \end{cases}$$

- Goal: Learn π from observation of Y

A Planted Ordering Model

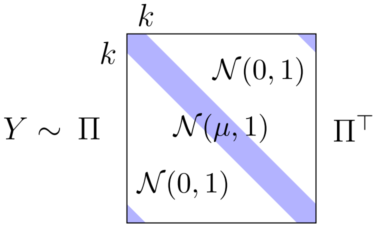


- Π is the permutation matrix corresponding to ordering π
- $Y_{ii} = 0$ and for $i \neq j$:

$$Y_{ij} \sim \begin{cases} P & \text{if } |\pi(i) - \pi(j)| \leq k \\ Q & \text{otherwise} \end{cases}$$

- Goal: Learn π from observation of Y

A Planted Ordering Model



- Π is the permutation matrix corresponding to ordering π
- $Y_{ii} = 0$ and for $i \neq j$:

$$Y_{ij} \sim \begin{cases} P & \text{if } |\pi(i) - \pi(j)| \le k \\ Q & \text{otherwise} \end{cases}$$

- Goal: Learn π from observation of Y
- When $k = 1$, reduces to hidden Hamiltonian cycle model [Broder-Frieze-Shamir 06]

Statistical tasks

- Exact recovery:

$$\mathbb{P}\left\{\hat{\pi} = \pi\right\} \xrightarrow{n\to\infty} 1$$

Statistical tasks

- Exact recovery:

$$\mathbb{P}\{\hat{\pi} = \pi\} \xrightarrow{n \to \infty} 1$$

- Detection:

$$\mathcal{H}_0 : \mu = 0 \quad \text{v.s.} \quad \mathcal{H}_1 : \mu > 0$$

Type-I + Type-II error probabilities $\to 0$

Statistical tasks

- Exact recovery:

$$\mathbb{P}\{\hat{\pi} = \pi\} \xrightarrow{n\to\infty} 1$$

- Detection:

$$\mathcal{H}_0 : \mu = 0 \quad \text{v.s.} \quad \mathcal{H}_1 : \mu > 0$$

Type-I + Type-II error probabilities $\to 0$

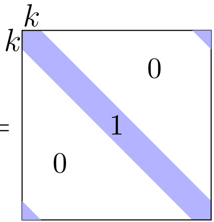
Main Questions

- When is recovery or detection informationally possible?
- Is IT-limit achievable in polynomial-time?

Outline of the remainder

1. Exact recovery
2. Detection
3. Weak recovery
4. Summary and concluding remarks

Exact recovery: maximum likelihood estimation

$A =$ 

$$\begin{aligned} \max \quad & \langle Y, \Pi A \Pi^\top \rangle \\ \text{s.t.} \quad & \Pi \in S_n \end{aligned}$$

- S_n: set of $n \times n$ permutation matrices
- When $k = 1$, maximum weighted Hamiltonian cycle problem

Exact recovery: necessary condition

Theorem (Necessary condition)

Exact recovery *is information-theoretically impossible if*

$$\mu^2 < 2\log n$$

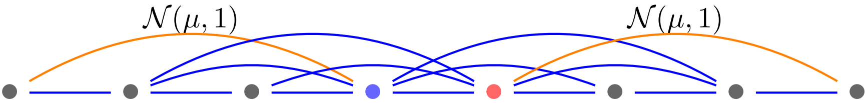
Exact recovery: necessary condition

Theorem (Necessary condition)

Exact recovery *is information-theoretically impossible if*

$$\mu^2 < 2 \log n$$

Independent of bandwidth k



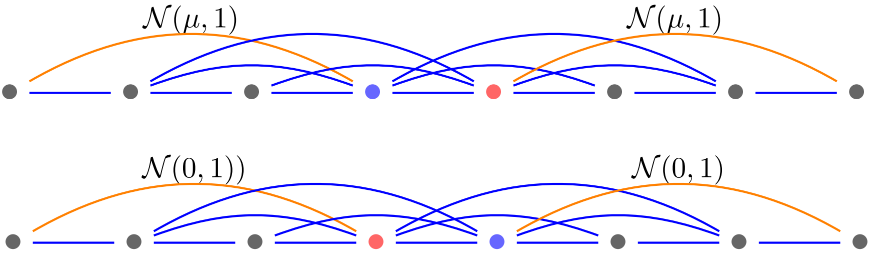
Exact recovery: necessary condition

Theorem (Necessary condition)

Exact recovery *is information-theoretically impossible if*

$$\mu^2 < 2\log n$$

Independent of bandwidth k



Exact recovery: necessary condition

Theorem (Necessary condition $k = 1$)

When $k = 1$, exact recovery is information-theoretically impossible if

$$\mu^2 < 4 \log n$$

Exact recovery: necessary condition

Theorem (Necessary condition $k = 1$)

When $k = 1$, exact recovery is information-theoretically impossible if

$$\mu^2 < 4 \log n$$

Remarks

- MLE fails on the event

$$\mathcal{F} \triangleq \cup_{j>i}\{Y_{i-1,j} + Y_{i,j+1} > Y_{i-1,i} + Y_{j,j+1}\}$$

- $|\{i : Y_{i,i+1} \approx \mu/2\}| \approx ne^{-\mu^2/8}$
- $\mathbb{P}\{Y_{i-1,j} + Y_{i,j+1} > \mu\} \approx e^{-\mu^2/4}$

Exact recovery: necessary condition

Theorem (Necessary condition $k = 1$)

When $k = 1$, exact recovery is information-theoretically impossible if

$$\mu^2 < 4 \log n$$

Remarks

- MLE fails on the event

$$\mathcal{F} \triangleq \cup_{j>i}\{Y_{i-1,j} + Y_{i,j+1} > Y_{i-1,i} + Y_{j,j+1}\}$$

- $|\{i : Y_{i,i+1} \approx \mu/2\}| \approx n e^{-\mu^2/8}$
- $\mathbb{P}\{Y_{i-1,j} + Y_{i,j+1} > \mu\} \approx e^{-\mu^2/4}$
- The necessary condition is tight

Exact recovery: naïve thresholding $k = 1$

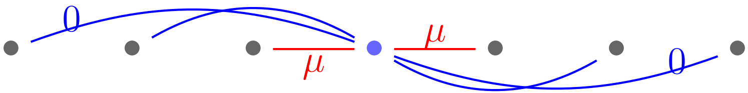
- When $k = 1$, MLE $\Rightarrow$ maximum weighted Hamiltonian cycle

Exact recovery: naïve thresholding $k = 1$

- When $k = 1$, MLE $\Rightarrow$ maximum weighted Hamiltonian cycle
- A naïve thresholding algorithm:
  For every vertex, keep the two edges with the largest weights

Exact recovery: naïve thresholding $k = 1$

- When $k = 1$, MLE $\Rightarrow$ maximum weighted Hamiltonian cycle
- A naïve thresholding algorithm:
  For every vertex, keep the two edges with the largest weights



Theorem (naïve thresholding $k = 1$)

When $k = 1$, the naïve thresholding achieves **exact recovery** *if*

$$\mu^2 > 8 \log n$$

Exact recovery: naïve thresholding for general k

- A naïve thresholding algorithm for general k:
  For every vertex, keep the $2k$ edges with the largest weights

Theorem (naïve thresholding for general k)

When $k = 1$, the naïve thresholding exactly recovers $2k$-NN graph if

$$\mu^2 > 8 \log n + 4 \log k$$

Exact recovery: naïve thresholding for general k

- A naïve thresholding algorithm for general k:
  For every vertex, keep the $2k$ edges with the largest weights

Theorem (naïve thresholding for general k)

When $k = 1$, the naïve thresholding exactly recovers $2k$-NN graph if

$$\mu^2 > 8 \log n + 4 \log k$$

Remarks

When $k = 1$, a factor of 2 gap to the IT limit $\mu^2 = 4 \log n$

Exact recovery: greedy merging $k = 1$

Greedy merging [Motahari-Bresler-Tse '13]

1. Initialize the set of edges to be empty
2. Among all vertices with degree less than 2, connect two vertices i, j with largest Y_{ij}
3. Repeat Step 2

Exact recovery: greedy merging $k = 1$

Greedy merging [Motahari-Bresler-Tse '13]

1. Initialize the set of edges to be empty
2. Among all vertices with degree less than 2, connect two vertices i, j with largest Y_{ij}
3. Repeat Step 2

Theorem (Greedy merging $k = 1$)

When $k = 1$, the greedy merging achieves **exact recovery** *if*

$$\mu^2 > 6 \log n$$

Exact recovery: greedy merging $k = 1$

Greedy merging [Motahari-Bresler-Tse '13]

1. Initialize the set of edges to be empty
2. Among all vertices with degree less than 2, connect two vertices i, j with largest Y_{ij}
3. Repeat Step 2

Theorem (Greedy merging $k = 1$)

When $k = 1$, the greedy merging achieves **exact recovery** *if*

$$\mu^2 > 6 \log n$$

Remarks

i and j will not be connected if

$$Y_{ij} < \min\{Y_{i-1,i}, Y_{i,i+1}\} \quad \text{or} \quad Y_{ij} < \min\{Y_{j-1,j}, Y_{j,j+1}\}$$

Exact recovery: greedy merging for genearl k

Greedy merging for general k

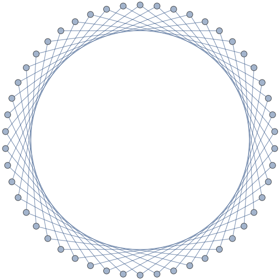
1. Initialize the set of edges to be empty
2. Among all vertices with degree less than $2k$, connect two vertices i, j with largest Y_{ij}
3. Repeat Step 2

Theorem (Greedy merging for general k)

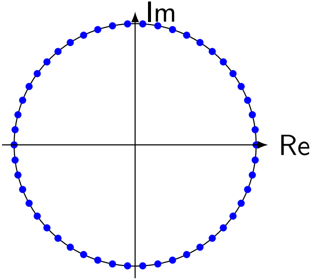
The greedy merging exactly recovers the $2k$-NN graph if

$$\mu^2 > 6\log n + 6\log k$$

From $2k$-NN graph to vertex ordering

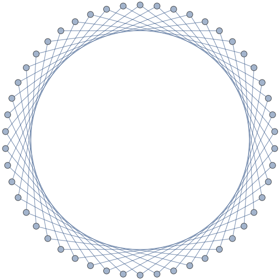


$2k$-NN graph

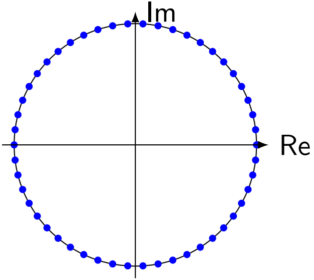


Eigenvector v_2 of circulant graph

From $2k$-NN graph to vertex ordering



$2k$-NN graph



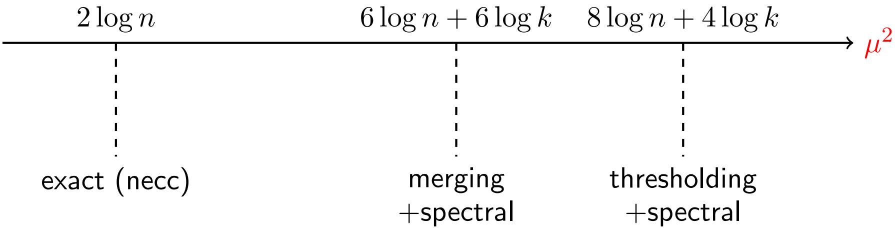
Eigenvector v_2 of circulant graph

$$v_2 = (\omega^{\pi(1)}, \ldots, \omega^{\pi(n)}),$$

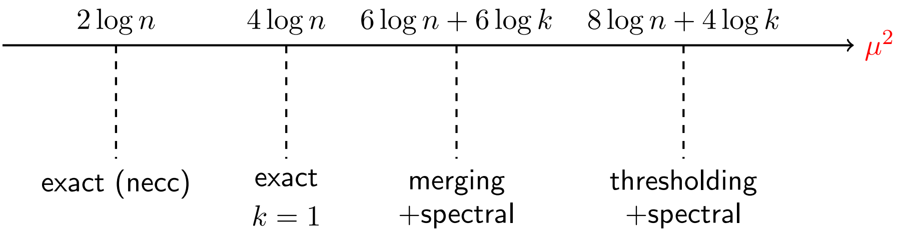
where $\omega = \exp\left(\frac{2\pi i}{n}\right)$ is the n^{th} root of unity

Exact recovery of vertex ordering for genearl k

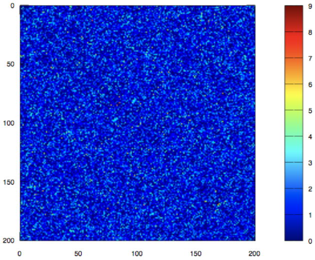
1. Estimate $2k$-NN graph A
2. Let v_2 denote the (complex) eigenvector of A corresponding to the 2nd largest eigenvalue
3. Sort the phase of v_2 and output the ordering


$2 \log n$
$6 \log n + 6 \log k$
$8 \log n + 4 \log k$
μ^2
exact (necc)
merging
+spectral
thresholding
+spectral

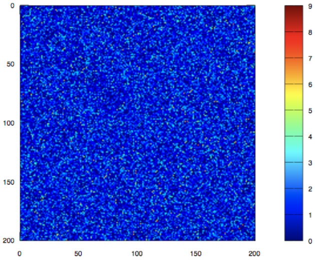
Summary for exact recovery



Detection threshold

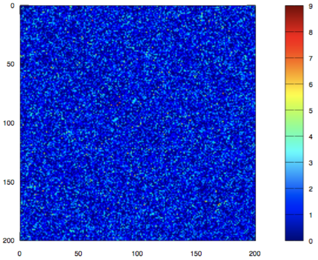


H_0

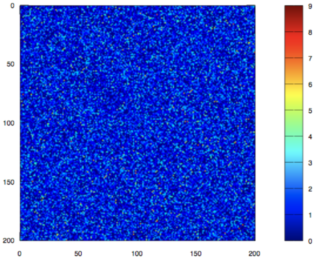


H_1

Detection threshold



H_0



H_1

Theorem

Detection *is possible if and only if*

$$k^2\mu^2 \to \infty$$

Proof of detection threshold

- Upper bound: sum statistic $\sum_{i<j} Y_{ij}$
- Lower bound: bounded second moment

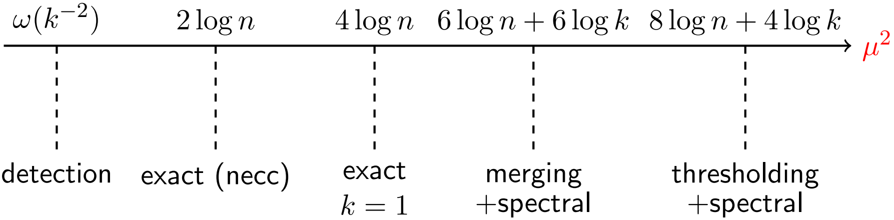
$$\mathbb{E}_{Y\sim\mathbb{Q}}\left[\left(\frac{\mathbb{P}(Y)}{\mathbb{Q}(Y)}\right)^2\right] = \mathbb{E}_{\pi,\pi'} \exp\left(\mu^2 \omega(\pi,\pi')\right),$$

where

$$\omega(\pi,\pi') = \sum_{i<j} \mathbf{1}_{\{|\pi(i)-\pi(j)|\le k,\ |\hat{\pi}(i)-\hat{\pi}(j)|\le k\}}$$

- Heuristically, $\omega(\pi,\pi') \sim \mathrm{Pois}(2k^2)$
- Hence, if $k^2\mu^2 = O(1)$, then the second moment is bounded

Summary for exact recovery and detection



Weak recovery

Weak recovery:

$$\underbrace{\frac{1}{nk}\sum_{i<j}\mathbf{1}_{\{|\pi(i)-\pi(j)|\le k,\ |\hat{\pi}(i)-\hat{\pi}(j)|\le k\}}}_{\text{overlap}} \to 1$$

Weak recovery for $k = 1$

Theorem

When $k = 1$, weak recovery is information-theoretically possible if and only if

$$\mu^2 > 2 \log n$$

Weak recovery for $k = 1$

Theorem

When $k = 1$, weak recovery is information-theoretically possible if and only if

$$\mu^2 > 2 \log n$$

Remarks

- Upper bound: analysis of MLE

Weak recovery for $k = 1$

Theorem

When $k = 1$, weak recovery is information-theoretically possible if and only if

$$\mu^2 > 2 \log n$$

Remarks

- Upper bound: analysis of MLE
- Lower bound: rate distortion argument

Proof of lower bound for weak recovery

-

$$\begin{aligned} I(Y;\pi) &\ge I(\hat{\pi};\pi) \\ &\ge \min_{\mathbb{E}[\omega(\tilde{\pi},\pi)]=(1+o(1))n} I(\tilde{\pi};\pi) \\ &\approx H(\pi) \approx n \log n \end{aligned}$$

Proof of lower bound for weak recovery

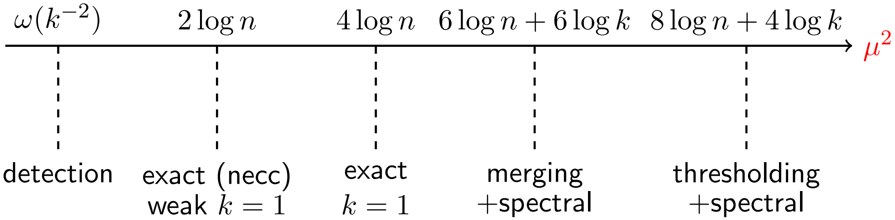
-

$$
\begin{aligned}
I(Y;\pi) &\geq I(\hat{\pi};\pi) \\
&\geq \min_{\mathbb{E}[\omega(\tilde{\pi},\pi)]=(1+o(1))n} I(\tilde{\pi};\pi) \\
&\approx H(\pi) \approx n \log n
\end{aligned}
$$

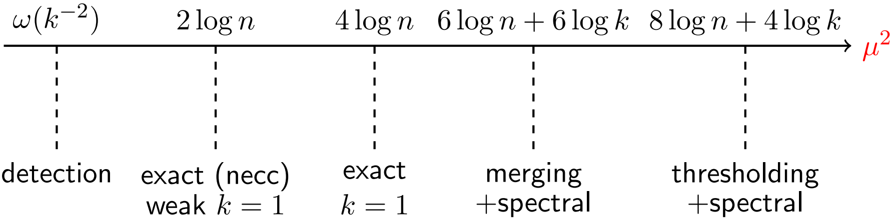
-

$$
\begin{aligned}
I(Y;\pi) &= \min_{\mathbb{Q}} D(\mathbb{P}_{Y|\pi} \| \mathbb{Q} \mid \mathbb{P}_{\pi}) \\
&\leq D(P_{Y|\pi^*} \| \mathcal{N}(0,1)^{\otimes \binom{n}{2}} | \mathbb{P}_{\pi}) \\
&= n\mu^2/2
\end{aligned}
$$

Conclusion and remarks



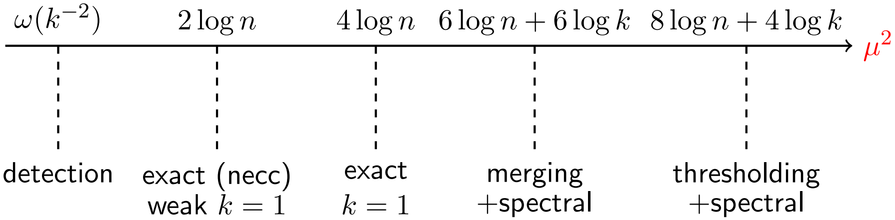
Conclusion and remarks



Future work

- Recovery threshold for general k
- SDP relaxation of MLE

Conclusion and remarks



Future work

- Recovery threshold for general k
- SDP relaxation of MLE
- Real data experiment