S&DS 684 Lecture 11: Hidden Hamiltonian cycle problem and Linear Programming

Yihong Wu

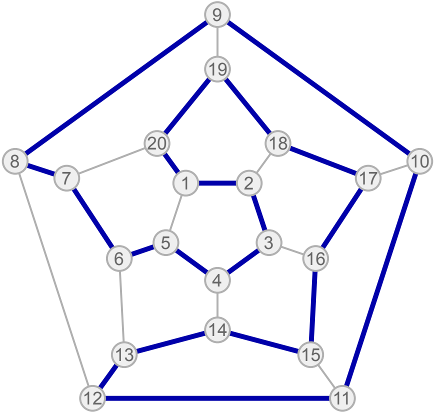
Department of Statistics and Data Science
Yale University

Based on: Vivek Bagaria, Jian Ding, David Tse, Yihong Wu, and Jiaming Xu, "Hidden Hamiltonian Cycle Recovery via Linear Programming", *Operations Research*, vol. 68, no. 1, 2020, https://arxiv.org/abs/1804.05436.

Apr 11, 2023

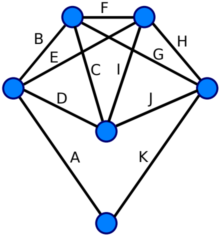
Some elements from graph theory

A Hamiltonian cycle is a cycle that visits each vertex exactly once.



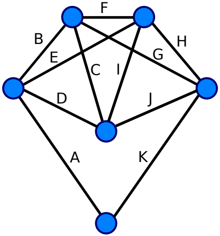
Some elements from graph theory

An Eulerian circuit (or walk, tour) is is a circuit that visits each edge exactly once.

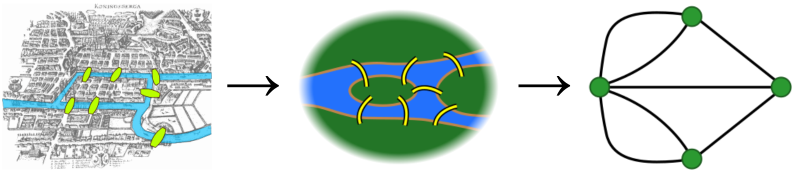


Some elements from graph theory

An Eulerian circuit (or walk, tour) is is a circuit that visits each edge exactly once.



Euler (Seven bridges of Könisberg): Every connected graph with even degrees has an Eulerian circuit.

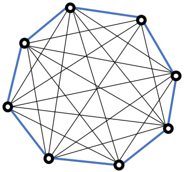


Hidden Hamiltonian cycle model

- Observation: a weighted undirected complete graph on n vertices with weighted adjacency matrix W
- Latent: a Hamiltonian cycle
- Edge weight

$$W_e \overset{\text{ind.}}{\sim} \begin{cases} P & e \in C^* \\ Q & e \notin C^* \end{cases}$$

e.g. two Gaussians/Poissons with different means

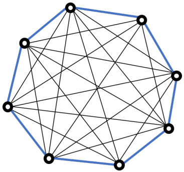


Hidden Hamiltonian cycle model

- Observation: a weighted undirected complete graph on n vertices with weighted adjacency matrix W
- Latent: a Hamiltonian cycle
- Edge weight

$$W_e \overset{\text{ind.}}{\sim} \begin{cases} P & e \in C^* \\ Q & e \notin C^* \end{cases}$$

e.g. two Gaussians/Poissons with different means

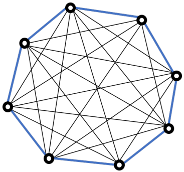


- Goal: observe W, recover C^* with high probability

Hidden Hamiltonian cycle model

- Observation: a weighted undirected complete graph on n vertices with weighted adjacency matrix W
- Latent: a Hamiltonian cycle
- Edge weight

$$W_e \overset{\text{ind.}}{\sim} \begin{cases} P & e \in C^* \\ Q & e \notin C^* \end{cases}$$



e.g. two Gaussians/Poissons with different means

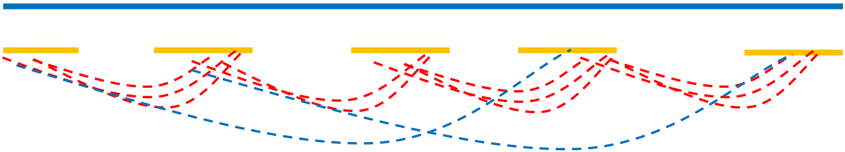
- Goal: observe W, recover C^* with high probability

Remarks:

- P, Q depends on the graph size n
- Hidden Hamiltonian cycle planted in Erdös-Rényi graph [Broder-Frieze-Shamir '94]

Link information in Chicago datasets (Hi-C reads)

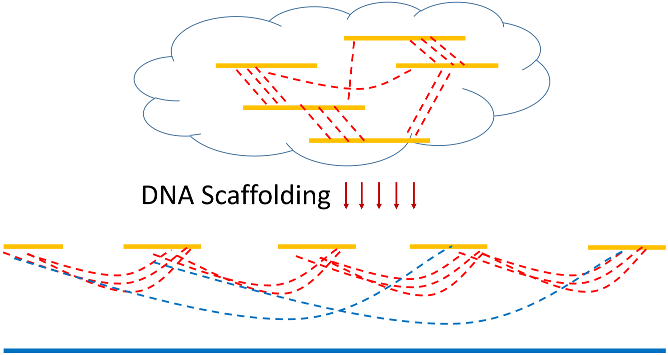
1. Reconstitute chromatin in vitro upon naked DNA
2. Produce cross-links by fixing chromatin with formaldehyde



Chicago datasets generate cross-links among contigs [Putnam et al. '16]

On average more cross-links exist between adjacent contigs

Ordering DNA contigs with Chicago cross-links

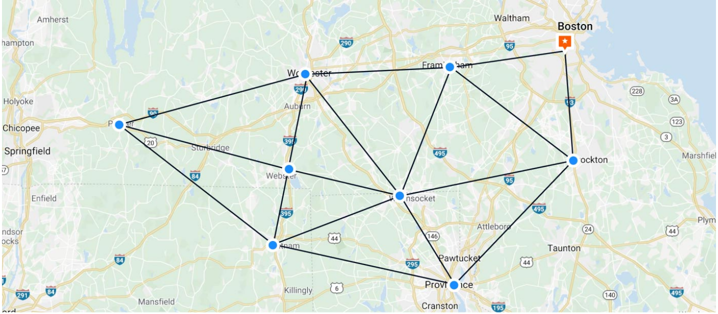


Reduces to traveling salesman problem (TSP)

Find a path (tour) that visits every contig exactly once with the maximum number of cross-links

Traveling salesman problem

Given a weighted graph, find the Hamiltonian cycle (path) with maximum or minimum total weight



Mathematically,

$$\max_{\pi \in S_n} \sum_{i=1}^{n} W_{\pi(i)\pi(i+1)}$$

where each Hamiltonian cycle is represented as $\pi(1), \pi(2), \ldots, \pi(n)$

Key challenges

- Computational: TSP is NP-hard in the worst-case
- Statistical: spurious cross-links between contigs that are far apart

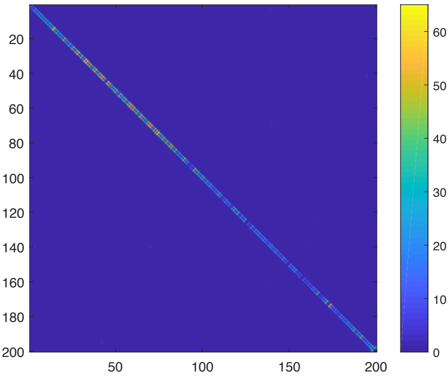
Key challenges

- Computational: TSP is NP-hard in the worst-case
- Statistical: spurious cross-links between contigs that are far apart

Key questions:

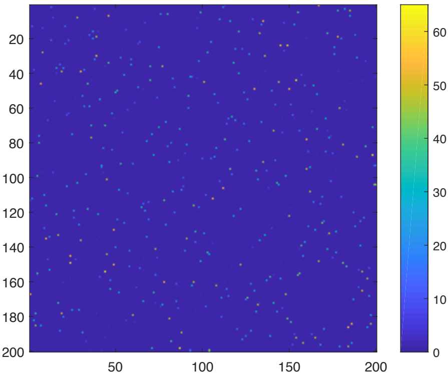
- How to efficiently order hundreds of thousands of contigs?
- How much noise can be tolerated for accurate DNA scaffolding?

Mathematical model for DNA scaffolding



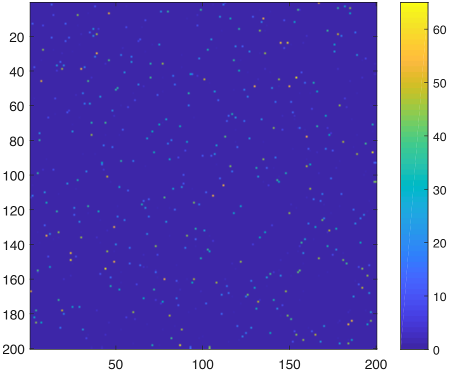
Chicago dataset [Putnam et al. '16]

Mathematical model for DNA scaffolding

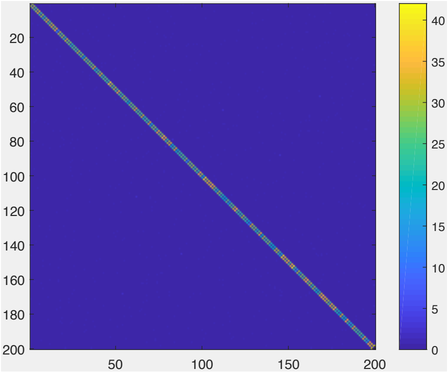


Chicago dataset [Putnam et al. '16]

Mathematical model for DNA scaffolding

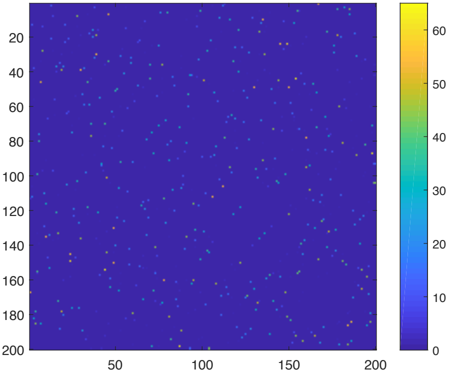


Chicago dataset [Putnam et al. '16]

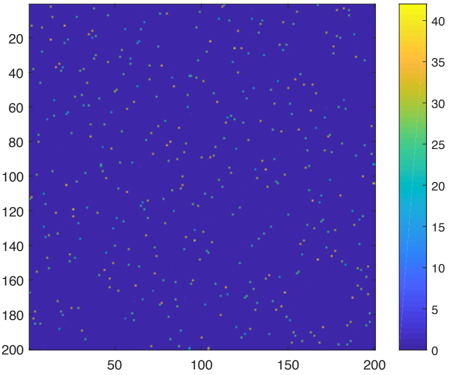


Simulated Poisson data

Mathematical model for DNA scaffolding



Chicago dataset [Putnam et al. '16]



Simulated Poisson data

What is known information-theoretically

Maximum likelihood estimator reduces to solving TSP

$$\widehat{X}_{\mathrm{TSP}} = \arg\max_X \ \langle L, X \rangle$$
$$\text{s.t.} \quad X \text{ is the adjacency matrix of some Hamiltonian cycle}$$

where L is the log likelihood ratio matrix $L_{ij} = \log \frac{dP}{dQ}(W_{ij})$.

- For Gaussian or Poisson (with bigger mean under P), can take $L = W$.
- For simplicity, consider the Gaussian model throughout the lecture:

$$P = N(\mu, 1), \quad Q = N(0, 1).$$

What is known information-theoretically

Maximum likelihood estimator reduces to solving TSP

$$\widehat{X}_{\mathrm{TSP}} = \arg\max_X \; \langle L, X \rangle$$

$$\text{s.t.} \quad X \text{ is the adjacency matrix of some Hamiltonian cycle}$$

where L is the log likelihood ratio matrix $L_{ij} = \log \frac{dP}{dQ}(W_{ij})$.

- For Gaussian or Poisson (with bigger mean under P), can take $L = W$.
- For simplicity, consider the Gaussian model throughout the lecture:

$$P = N(\mu, 1), \quad Q = N(0, 1).$$

Theorem (Sharp threshold)

If $\mu^2 < 4 \log n$, exact recovery is information-theoretically impossible;
If $\mu^2 > 4 \log n$, MLE succeeds in exact recovery.

What is known algorithmically

Spectral method fails (1)

$$W = \mu \cdot \underbrace{X^*}_{\text{“signal”}} + \underbrace{Z}_{\text{noise (GOE)}}$$

where X^* is adj matrix of C^* and can be written as $X^* = \Pi_* X^\dagger \Pi_*^\top$,

- Π_*: permutation matrix corresponding to C^*
- $X^\dagger$ is a circulant matrix: $X_{ij}^\dagger = \mathbf{1}_{\{i-j=\pm 1 \mod n\}}$

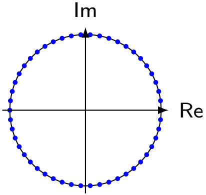
Spectral method fails (1)

$$W = \mu \cdot \underbrace{X^*}_{\text{"signal"}} + \underbrace{Z}_{\text{noise (GOE)}}$$

where X^* is adj matrix of C^* and can be written as $X^* = \Pi_* X^\dagger \Pi_*^\top$,

- Π_*: permutation matrix corresponding to C^*
- $X^\dagger$ is a circulant matrix: $X^\dagger_{ij} = \mathbf{1}_{\{i-j=\pm 1 \mod n\}}$
  - eigenvalues $\lambda_k = 2\cos\frac{2k\pi}{n}$, $k = 0, \dots, n-1$
  - eigenvectors $v_k = (1, e^{i\frac{2k\pi}{n}}, e^{i\frac{4k\pi}{n}}, \dots, e^{i\frac{2(n-1)k\pi}{n}})$ (Fourier basis)

In the noiseless case, second eigenvector recovers the Hamiltonian cycle perfectly:



Spectral method fails (2)

$$W = \mu \cdot X^* + Z : \quad \text{full-rank signal + noise}$$

In the noisy case:

- Spectral gap of cycle: $2 - 2\cos\frac{2\pi}{n} \asymp \frac{1}{n^2}$ vs noise spectrum: $\sqrt{n}$
- So we need $\mu \gg n^{2.5}$ (we will see that simple thresholding requires only $\mu \asymp \sqrt{\log n}$)

Spectral method fails (2)

$$W = \mu \cdot X^* + Z : \quad \text{full-rank signal + noise}$$

In the noisy case:

- Spectral gap of cycle: $2 - 2\cos\frac{2\pi}{n} \asymp \frac{1}{n^2}$ vs noise spectrum: $\sqrt{n}$
- So we need $\mu \gg n^{2.5}$ (we will see that simple thresholding requires only $\mu \asymp \sqrt{\log n}$)

Lesson

Without "low-rank signal + noise", one needs to be careful with spectral method (will revisit this point for the graph matching problem in Lec 14).

Thresholding

- Simple thresholding ("nearest neighbor"): for each vertex, keep the two edges with the largest weights
  - $\mu > \sqrt{8 \log n}$

Thresholding

- Simple thresholding ("nearest neighbor"): for each vertex, keep the two edges with the largest weights
  - $\mu > \sqrt{8 \log n}$
  - Why:
    - For each vertex, two planted edges $N(\mu, 1)$ and $n - 2$ null edges $N(0, 1)$ (maximum fluctuation of the latter $\approx \sqrt{2 \log n}$)
    - There are n vertices in total.
    - So we need $\mathbb{P}\left\{N(\mu, 1) < \sqrt{2 \log n}\right\} = o(1/n)$

Thresholding

- Simple thresholding ("nearest neighbor"): for each vertex, keep the two edges with the largest weights
  - $\mu > \sqrt{8 \log n}$
  - Why:
    - For each vertex, two planted edges $N(\mu, 1)$ and $n-2$ null edges $N(0,1)$ (maximum fluctuation of the latter $\approx \sqrt{2 \log n}$)
    - There are n vertices in total.
    - So we need $\mathbb{P}\left\{N(\mu, 1) < \sqrt{2 \log n}\right\} = o(1/n)$
- Greedy merging [Motahari-Bresler-Tse '13]:
  - $\mu > \sqrt{6 \log n}$

Main result

Theorem

Linear programming (LP) *relaxation achieves sharp threshold*

$$\boxed{\begin{array}{ll} \dfrac{\mu^2}{\log n} > 4 : & \textit{LP succeeds} \\ \dfrac{\mu^2}{\log n} < 4 : & \textit{Everything fails} \end{array}}$$

In general

Threshold are determined by Rényi divergence of order $\rho > 0$ from P to Q:

$$D_\rho(P\|Q) \triangleq \frac{1}{\rho - 1} \log \int (dP)^\rho (dQ)^{1-\rho}.$$

- LP works when

  $$\boxed{D_{1/2}(P\|Q) - \log n \to \infty}$$

  optimal under mild assumptions

- Thresholding works when

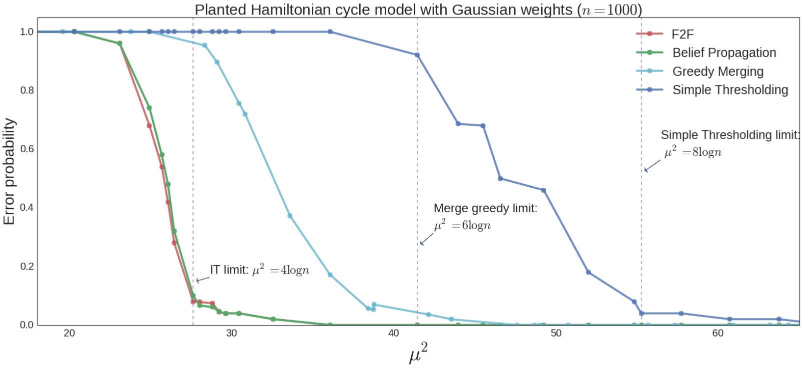
  $$D_{1/2}(P\|Q) - 2\log n \to \infty$$

- Greedy works when

  $$D_{1/3}(Q\|P) - \log n \to \infty$$

Experiments

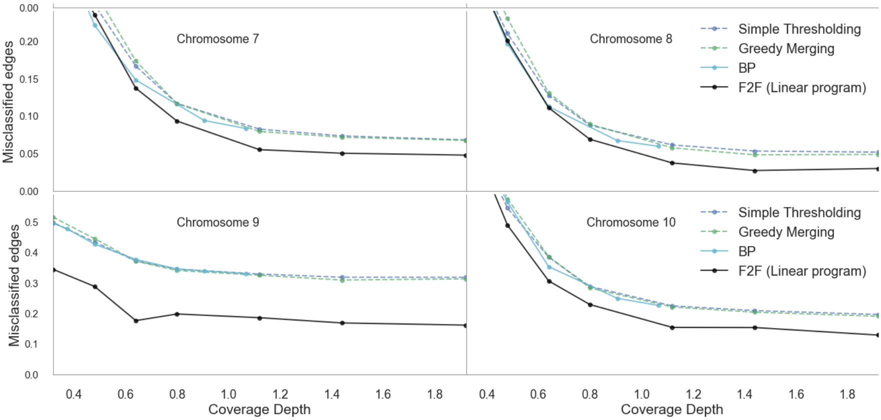
Synthetic data experiment



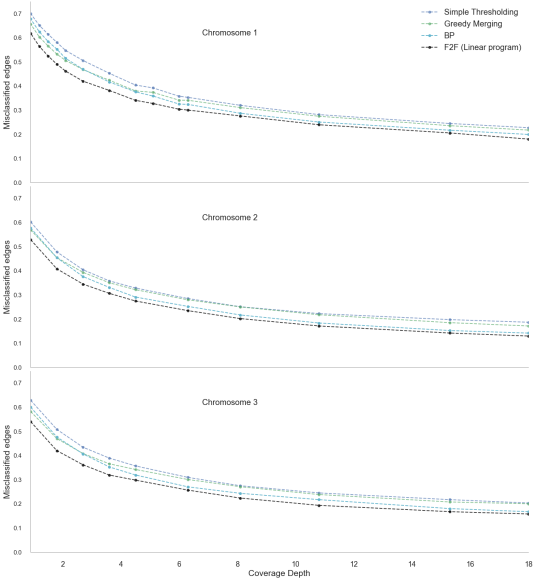
Real-data experiment

- 1000 DNA contigs of size 100 kbps
- 0.45 million Chicago cross-links
- Edge weights $=$ raw number of HiC reads between each pair of contigs
- Ground truth obtained by other (expensive) sequencing technologies

Homosapiens [Putnam et al 16, Genome Research]



Aedes Aegypti (zika mosquito) [Dudchenko et al '16, Science]



Convex relaxations of TSP

Integer Linear Programming reformulation of TSP

$$\begin{aligned}\widehat{X}_{\mathrm{TSP}} = \arg\max_{X} \ & \langle W, X \rangle \\ \text{s.t.} \ & \sum_j X_{ij} = 2, \ \forall i \\ & X_{ij} \in \{0, 1\} \\ & \sum_{i \in I, j \notin I} X_{ij} \geq 2, \ \forall \emptyset \neq I \subset [n]\end{aligned}$$

Integer Linear Programming reformulation of TSP

$$
\begin{aligned}
\widehat{X}_{\mathrm{TSP}} = \arg\max_{X} \ & \langle W, X \rangle \\
\text{s.t.} \ & \sum_{j} X_{ij} = 2, \ \forall i \\
& X_{ij} \in \{0, 1\} \\
& \sum_{i \in I, j \notin I} X_{ij} \geq 2, \ \forall \emptyset \neq I \subset [n]
\end{aligned}
$$

- The last constraint: subtour elimination

Subtour LP

$$
\begin{aligned}
\widehat{X}_{\mathrm{SUB}} = \arg\max_{X} \ & \langle W, X \rangle \\
\text{s.t.} \ & \sum_j X_{ij} = 2, \quad \forall i \\
& X_{ij} \in [0,1] \\
& \sum_{i \in I, j \notin I} X_{ij} \geq 2, \ \forall \emptyset \neq I \subset [n]
\end{aligned}
$$

Subtour LP

$$
\begin{aligned}
\widehat{X}_{\mathrm{SUB}} = \arg\max_{X} \ & \langle W, X \rangle \\
\text{s.t.} \ & \sum_{j} X_{ij} = 2, \quad \forall i \\
& X_{ij} \in [0,1] \\
& \sum_{i \in I, j \notin I} X_{ij} \geq 2, \ \forall \emptyset \neq I \subset [n]
\end{aligned}
$$

- Replacing the integrality constraint with box constraint: SUBTOUR LP relaxation [Dantzig-Fulkerson-Johnson '54, Held-Karp '70]
- Exponentially many linear constraints, nevertheless solvable using interior point method

F2F LP

$$
\begin{aligned}
\widehat{X}_{\mathrm{F2F}} = \arg\max_{X} \ & \langle W, X \rangle \\
\text{s.t.} \ & \sum_j X_{ij} = 2, \quad \forall i \\
& X_{ij} \in [0, 1]
\end{aligned}
$$

- Further dropping subtour elimination constraints $\Longrightarrow$ Fractional 2-factor (F2F) LP

F2F LP

$$\widehat{X}_{\mathrm{F2F}} = \arg\max_{X} \ \langle W, X \rangle$$
$$\text{s.t.} \quad \sum_j X_{ij} = 2, \quad \forall i$$
$$X_{ij} \in [0, 1]$$

- Further dropping subtour elimination constraints $\Longrightarrow$ Fractional 2-factor (F2F) LP
- Extensively studied in worst case [Boyd-Carr '99,Schalekamp-Williamson-van Zuylen '14]
  - The integrality gap $\frac{\mathrm{2F}}{\mathrm{F2F}} \leq \frac{4}{3}$ for metric TSP (min formulation)

F2F LP

$$
\begin{aligned}
\widehat{X}_{\mathrm{F2F}} = \arg\max_{X} \ & \langle W, X \rangle \\
\text{s.t.} \ & \sum_j X_{ij} = 2, \quad \forall i \\
& X_{ij} \in [0,1]
\end{aligned}
$$

- Further dropping subtour elimination constraints $\Longrightarrow$ Fractional 2-factor (F2F) LP
- Extensively studied in worst case [Boyd-Carr '99,Schalekamp-Williamson-van Zuylen '14]
  - The integrality gap $\frac{\mathrm{2F}}{\mathrm{F2F}} \leq \frac{4}{3}$ for metric TSP (min formulation)
- What is the integrality gap whp in our random instance?

Optimality of Fractional 2-Factor LP

Theorem

If $\mu^2 - 4\log n \to \infty$*, then* $\widehat{X}_{\mathrm{F2F}} = X^*$ *with high probability.*

Optimality of Fractional 2-Factor LP

Theorem

If $\mu^2 - 4\log n \to \infty$, then $\widehat{X}_{\mathrm{F2F}} = X^*$ with high probability.

Remarks

- The solution is integral whp.
- This achieves the optimal threshold $\mu^2 = 4\log n$.

Belief propagation

Max-Product Belief Propagation

$$m_{i\to j}(t) = w_{ij} - \operatorname*{2nd\,max}_{\ell\neq j}\{m_{\ell\to i}(t-1)\}$$
$$m_{i\to j}(0) = w_{ij}$$

After T iterations, for each vertex i, keep the two largest incoming messages $m_{\ell\to i}(T)$ and delete the rest.

- BP is exact provided the solution is integral [Bayati-Borgs-Chayes-Zecchina '11]
- It can be shown that $T = O(n^2 \log n)$ whp

SDP relaxations for TSP

Add more constraints to F2F LP

- SDP1 [Cvetković et al '99]: PSD constraint based on second largest eigenvalue of cycle (cf. slide 12)

$$X \preceq \frac{2}{n}\mathbf{J} + 2\cos\frac{2\pi}{n}\left(I - \frac{1}{n}\mathbf{J}\right)$$

SDP relaxations for TSP

Add more constraints to F2F LP

- SDP1 [Cvetković et al '99]: PSD constraint based on second largest eigenvalue of cycle (cf. slide 12)

$$X \preceq \frac{2}{n}\mathbf{J} + 2\cos\frac{2\pi}{n}\left(I - \frac{1}{n}\mathbf{J}\right)$$

  - provably weaker than Subtour LP [Goemans-Rendl '00]

SDP relaxations for TSP

Add more constraints to F2F LP

- SDP1 [Cvetković et al '99]: PSD constraint based on second largest eigenvalue of cycle (cf. slide 12)

$$X \preceq \frac{2}{n}\mathbf{J} + 2\cos\frac{2\pi}{n}\left(I - \frac{1}{n}\mathbf{J}\right)$$

  - provably weaker than Subtour LP [Goemans-Rendl '00]

- SDP2 [Zhao et al '98]: Quadratic Assignment Problem

$$\langle W, X\rangle = \langle W, \Pi \underbrace{X_0}_{\substack{\text{fixed}\\ \text{cycle}}} \Pi^\top\rangle = \left\langle W \otimes X_0, \underbrace{\text{vec}(\Pi)\text{vec}(\Pi)^\top}_{\text{relax..}} \right\rangle$$

SDP relaxations for TSP

Add more constraints to F2F LP

- SDP1 [Cvetković et al '99]: PSD constraint based on second largest eigenvalue of cycle (cf. slide 12)

$$X \preceq \frac{2}{n}\mathbf{J} + 2\cos\frac{2\pi}{n}\left(I - \frac{1}{n}\mathbf{J}\right)$$

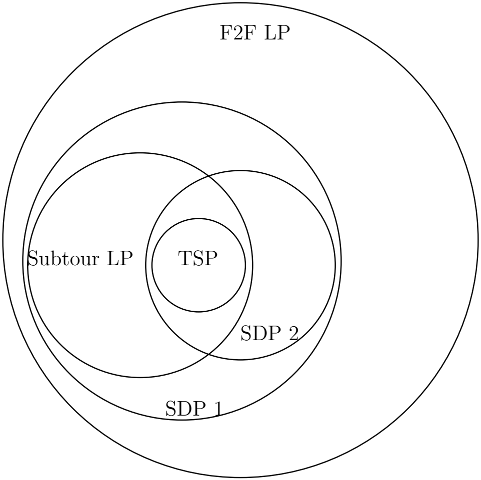
  - provably weaker than Subtour LP [Goemans-Rendl '00]

- SDP2 [Zhao et al '98]: Quadratic Assignment Problem

$$\langle W, X\rangle = \langle W, \Pi \underbrace{X_0}_{\substack{\text{fixed}\\\text{cycle}}} \Pi^\top\rangle = \left\langle W \otimes X_0, \underbrace{\text{vec}(\Pi)\text{vec}(\Pi)^\top}_{\text{relax..}}\right\rangle$$

  - decision variable: $n^2 \times n^2$ matrix
  - provably stronger than SDP1 [de Klerk et al '08]

Different relaxations



F2F LP succeeds $\implies$ all other relaxations succeeed.

Theoretical analysis of convex relaxation

Primal approach vs Dual approach: high level

- Dual argument:
  - Construct dual witness that certifies the ground truth whp (KKT conditions)
  - Successful in proving SDP relaxation attaining sharp threshold for graph partitions: planted clique (Chap. 6), community detection (Chap. 10), etc

Primal approach vs Dual approach: high level

- Dual argument:
  - Construct dual witness that certifies the ground truth whp (KKT conditions)
  - Successful in proving SDP relaxation attaining sharp threshold for graph partitions: planted clique (Chap. 6), community detection (Chap. 10), etc
  - Limitations: construction is ad hoc

Primal approach vs Dual approach: high level

- Dual argument:
  - Construct dual witness that certifies the ground truth whp (KKT conditions)
  - Successful in proving SDP relaxation attaining sharp threshold for graph partitions: planted clique (Chap. 6), community detection (Chap. 10), etc
  - Limitations: construction is ad hoc

- Primal argument:
  - No feasible solution other than the ground truth has a better objective value whp

Primal approach vs Dual approach: high level

- Dual argument:
  - Construct dual witness that certifies the ground truth whp (KKT conditions)
  - Successful in proving SDP relaxation attaining sharp threshold for graph partitions: planted clique (Chap. 6), community detection (Chap. 10), etc
  - Limitations: construction is ad hoc

- Primal argument:
  - No feasible solution other than the ground truth has a better objective value whp
  - Key: for LP, can restrict to **extremal points** (vertices of the feasible polytope)

Dual approach

- KKT conditions (Farkas' lemma): $\widehat{X}_{\mathrm{F2F}} = X^* \iff \exists u \in \mathbb{R}^n$ (dual certificate):

$$
\begin{aligned}
u_i + u_j &\le W_{ij}, \qquad \text{for } i \sim j \text{ in } C^* \\
u_i + u_j &\ge W_{ij}, \qquad \text{for } i \nsim j \text{ in } C^*
\end{aligned}
$$

Dual approach

- KKT conditions (Farkas' lemma): $\widehat{X}_{\mathrm{F2F}} = X^* \iff \exists u \in \mathbb{R}^n$ (dual certificate):

$$\begin{aligned} u_i + u_j &\le W_{ij}, \qquad \text{for } i \sim j \text{ in } C^* \\ u_i + u_j &\ge W_{ij}, \qquad \text{for } i \nsim j \text{ in } C^* \end{aligned}$$

- One feasible choice of dual:

$$u_i = \frac{1}{2} \min\{W_{ij} : j \sim i\}$$

Dual approach

- KKT conditions (Farkas' lemma): $\widehat{X}_{\mathrm{F2F}} = X^* \iff \exists u \in \mathbb{R}^n$ (dual certificate):

$$\begin{aligned} u_i + u_j &\leq W_{ij}, \qquad \text{for } i \sim j \text{ in } C^* \\ u_i + u_j &\geq W_{ij}, \qquad \text{for } i \not\sim j \text{ in } C^* \end{aligned}$$

- One feasible choice of dual:

$$u_i = \frac{1}{2} \min\{W_{ij} : j \sim i\}$$

- This certificate shows correctness if $\mu^2 > 6 \log n$ (same as greedy merging)

Primal approach

- Show whp for all extremal points $X \neq X^*$:

$$\langle W, X \rangle < \langle W, X^* \rangle$$

- F2F polytope:

$$\left\{ X \in [0,1]^{n \times n} : \sum_{j=1}^{n} X_{ij} = 2 \right\}$$

- The proof heavily exploits the characterization of extremal points

Primal approach

- Show whp for all extremal points $X \neq X^*$:

$$\langle W, X \rangle < \langle W, X^* \rangle$$

- F2F polytope:

$$\left\{ X \in [0,1]^{n \times n} : \sum_{j=1}^{n} X_{ij} = 2 \right\}$$

- The proof heavily exploits the characterization of extremal points
  - F2F polytope is not integral: fractional vertices exist

Primal approach

- Show whp for all extremal points $X \neq X^*$:

$$\langle W, X\rangle < \langle W, X^*\rangle$$

- F2F polytope:

$$\left\{X \in [0,1]^{n\times n} : \sum_{j=1}^{n} X_{ij} = 2\right\}$$

- The proof heavily exploits the characterization of extremal points
  - F2F polytope is not integral: fractional vertices exist
  - Characterization [Balinski '65]: for any vertex X of F2F polytope
    - Half integrality

$$X_{ij} \in \{0, 1/2, 1\}$$

Primal approach

- Show whp for all extremal points $X \neq X^*$:

$$\langle W, X \rangle < \langle W, X^* \rangle$$

- F2F polytope:

$$\left\{ X \in [0,1]^{n \times n} : \sum_{j=1}^{n} X_{ij} = 2 \right\}$$

- The proof heavily exploits the characterization of extremal points
  - F2F polytope is not integral: fractional vertices exist
  - Characterization [Balinski '65]: for any vertex X of F2F polytope
    - Half integrality

$$X_{ij} \in \{0, 1/2, 1\}$$

    - $1/2$'s form disjoint odd cycle connected by path of 1's.

Primal approach

- Show whp for all extremal points $X \neq X^*$:

$$\langle W, X \rangle < \langle W, X^* \rangle$$

- F2F polytope:

$$\left\{ X \in [0,1]^{n \times n} : \sum_{j=1}^{n} X_{ij} = 2 \right\}$$

- The proof heavily exploits the characterization of extremal points
  - F2F polytope is not integral: fractional vertices exist
  - Characterization [Balinski '65]: for any vertex X of F2F polytope
    - Half integrality

$$X_{ij} \in \{0, 1/2, 1\}$$

    - $1/2$'s form disjoint odd cycle connected by path of 1's.

Why half integral?

Usual proofs:

- combinatorial proof [Lovasz-Plummer '86, Schrijver '04]
- linear-algebraic proof
  - F2F polytope (in adjacency vector):

$$\{x \in \mathbb{R}^{\binom{n}{[2]}} : Ax = 2\mathbf{1}\}$$

  - A is $n \times \binom{n}{2}$ zero-one matrix: $A_{ie} = \mathbf{1}_{\{i \in e\}}$
  - Each column of A has exactly two 1's

Why half integral?

Extremal point (basic feasible solution) x is of the following form

$$x = (\underbrace{x_S}_{\text{fractional}}, \underbrace{x_{S^c}}_{\text{integral}})$$

for some $S \subset \binom{n}{[2]}$ of size n, where

- x_S is the solution to the following linear system:

$$A_S x_S = b'$$

Why half integral?

Extremal point (basic feasible solution) x is of the following form

$$x = (\underbrace{x_S}_{\text{fractional}} , \underbrace{x_{S^c}}_{\text{integral}})$$

for some $S \subset \binom{n}{[2]}$ of size n, where

- x_S is the solution to the following linear system:

$$A_S x_S = b'$$

- Cramer's rule:

$$(x_S)_i = \frac{\det(A_S^{(i)})}{\det(A_S)}$$

  - $A_S^{(i)}$ is obtained by substituting the ith column by b', hence $\det(A_S^{(i)}) \in \mathbb{Z}$.
  - Each column of A_S has two 1's $\implies \det(A_S) \in \{0, \pm 1, \pm 2\}$ [Balinski '65]

Balinski's theorem

Any square irreducible zero-one matrix with at most two 1's in each column has determinant in $\{0, \pm 1, \pm 2\}$.

Proof (induction on the matrix size n).

Base case $n = 2$: Direct verification.

Balinski's theorem

Any square irreducible zero-one matrix with at most two 1's in each column has determinant in $\{0, \pm 1, \pm 2\}$.

Proof (induction on the matrix size n).

Base case $n = 2$: Direct verification. Induction from $n - 1$ to n: Fix such an $A \in \{0, 1\}^{n \times n}$.

- Suppose some row (or column) contains zero or one 1's. Then the theorem follows from the induction hypothesis for $n - 1$.

Balinski's theorem

Any square irreducible zero-one matrix with at most two 1's in each column has determinant in $\{0, \pm 1, \pm 2\}$.

Proof (induction on the matrix size n).

Base case $n = 2$: Direct verification. Induction from $n - 1$ to n: Fix such an $A \in \{0, 1\}^{n \times n}$.

- Suppose some row (or column) contains zero or one 1's. Then the theorem follows from the induction hypothesis for $n - 1$.
- Suppose every row and every column contains exactly two 1's. Then A is the adjacency matrix of a 2-regular bipartite graph G. We apply two facts:
  - Every 2-regular graph is a disjoint union of even cycles.
  - G is connected by assumption of irreducibility of A.

  So A is the adjacency matrix of a $2n$-cycle. We can compute such a $\det(A)$ easily:

$\square$

Balinski's theorem

(Proof continued).

That is, $A = \Pi A_0 \Pi'$, where Π, Π' are permutation matrices and

$$A_0 = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \dots & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Direct calculation gives

$$\det(A_0) = \begin{cases} 0 & n \text{ even} \\ 2 & n \text{ odd} \end{cases}$$

So $\det(A) = 0$ or ± 2. $\square$

Warmup: proof of correctness for 2F ILP

2F Integer LP (ILP)

$$\widehat{X}_{2\mathrm{F}} = \arg\max_{X} \; \langle W, X \rangle$$
$$\text{s.t.} \quad \sum_j X_{ij} = 2$$
$$X_{ij} \in \{0,1\}$$

- Solvable using blossom algorithm $O(n^4)$ time [Letchford-Reinelt-Theis '08] but in practice challenging to implement
- Any feasible solution corresponds to a 2-factor (disjoint union of cycles)
- Goal: true Hamiltonian cycle has maximal weight w.h.p.

$$\langle W, X \rangle < \langle W, X^* \rangle \,, \quad \forall \text{ 2-factor } X \neq X^*$$

Encode the solution with difference graph

G and H: simple graphs on the same vertex set with adjacency matrix A and B

- **difference graph** $G - H$: a bicolored graph with signed adjacency matrix $A - B$, with red edge for $-$ and blue edge for $+$
  - Red edge in $G - H \iff$ edge in H but not in G – "Type-II error"
  - Blue edge in $G - H \iff$ edge in G but not in H – "Type-I error"

Encode the solution with difference graph

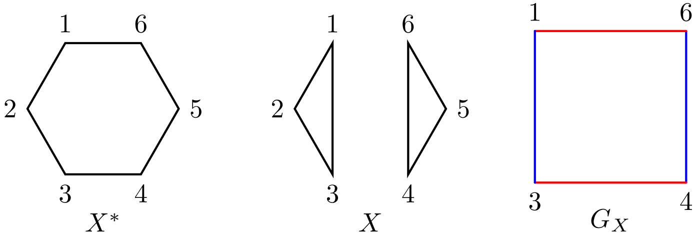
G and H: simple graphs on the same vertex set with adjacency matrix A and B

- **difference graph** $G - H$: a bicolored graph with signed adjacency matrix $A - B$, with red edge for $-$ and blue edge for $+$
  - Red edge in $G - H \iff$ edge in H but not in G – "Type-II error"
  - Blue edge in $G - H \iff$ edge in G but not in H – "Type-I error"
- Fact: difference graph of two k-regular graphs is balanced (for each vertex, red degree = blue degree).

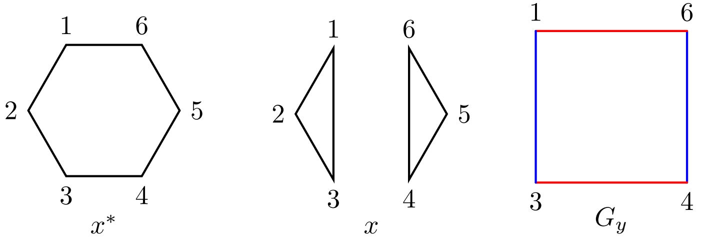
Encode the solution with difference graph

- X^*: the true Hamiltonian cycle; X an arbitrary 2-factor.
- The **difference graph** G_X is a balanced bicolored simple graph encoding $X - X^*$.

  Example:



Encode the solution with difference graph



- Weight of a bicolored graph:

$$w(B) \triangleq \sum_{\text{blue } e \in E(B)} w_e - \sum_{\text{red } e \in E(B)} w_e \,.$$

  Then $w(G_X) = \langle W, X - X^* \rangle$.

- Decomposition into connected components:

$$w(G_X) = \sum w(B_i)$$

  where each component B_i is a connected balanced bicolored graph.

Union bound

Let B be a connected balanced graph with ℓ edges

- $\ell/2$ red and $\ell/2$ blue.
- $w(B) \sim N(-\mu\ell/2, \ell)$
- $\mathbb{P}\{w(B) \geq 0\} \leq \exp(-\mu^2\ell/8)$

Union bound

Let B be a connected balanced graph with ℓ edges

- $\ell/2$ red and $\ell/2$ blue.
- $w(B) \sim N(-\mu\ell/2, \ell)$
- $\mathbb{P}\{w(B) \geq 0\} \leq \exp(-\mu^2\ell/8)$
- The number of distinct B with ℓ edges $\leq (2n)^{\ell/2}$, so that

$$\sum_{\ell \geq 2} (2n)^{\ell/2} e^{-\mu^2\ell/8} = \sum_{\ell \geq 2} (2ne^{-\mu^2/4})^{\ell/2} \xrightarrow{\mu^2=(4+\epsilon)\log n} 0$$

Union bound

Let B be a connected balanced graph with ℓ edges

- $\ell/2$ red and $\ell/2$ blue.
- $w(B) \sim N(-\mu\ell/2, \ell)$
- $\mathbb{P}\{w(B) \geq 0\} \leq \exp(-\mu^2\ell/8)$
- The number of distinct B with ℓ edges $\leq (2n)^{\ell/2}$, so that

$$\sum_{\ell \geq 2} (2n)^{\ell/2} e^{-\mu^2\ell/8} = \sum_{\ell \geq 2} (2ne^{-\mu^2/4})^{\ell/2} \xrightarrow{\mu^2=(4+\epsilon)\log n} 0$$

- [Euler (Seven bridges of Könisberg)]: Every connected graph with even degrees has an Eulerian circuit.

Union bound

Let B be a connected balanced graph with ℓ edges

- $\ell/2$ red and $\ell/2$ blue.
- $w(B) \sim N(-\mu\ell/2, \ell)$
- $\mathbb{P}\{w(B) \geq 0\} \leq \exp(-\mu^2\ell/8)$
- The number of distinct B with ℓ edges $\leq (2n)^{\ell/2}$, so that

$$\sum_{\ell \geq 2} (2n)^{\ell/2} e^{-\mu^2 \ell/8} = \sum_{\ell \geq 2} (2ne^{-\mu^2/4})^{\ell/2} \xrightarrow{\mu^2 = (4+\epsilon)\log n} 0$$

- [Euler (Seven bridges of Könisberg)]: Every connected graph with even degrees has an Eulerian circuit.
- [Kotzig '68]: Every connected balanced bicolored multigraph has an alternating Eulerian circuit.

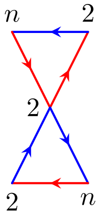
Union bound

Let B be a connected balanced graph with ℓ edges

- $\ell/2$ red and $\ell/2$ blue.
- $w(B) \sim N(-\mu\ell/2, \ell)$
- $\mathbb{P}\{w(B) \geq 0\} \leq \exp(-\mu^2\ell/8)$
- The number of distinct B with ℓ edges $\leq (2n)^{\ell/2}$, so that

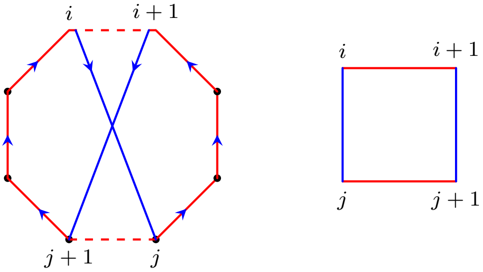
$$\sum_{\ell \geq 2} (2n)^{\ell/2} e^{-\mu^2\ell/8} = \sum_{\ell \geq 2} (2ne^{-\mu^2/4})^{\ell/2} \xrightarrow{\mu^2=(4+\epsilon)\log n} 0$$

- [Euler (Seven bridges of Könisberg)]: Every connected graph with even degrees has an Eulerian circuit.
- [Kotzig '68]: Every connected balanced bicolored multigraph has an alternating Eulerian circuit.
- $\#$ of B's $\leq$ $\#$ of alternating Eulerian circuits



Aside: information-theoretic optimality

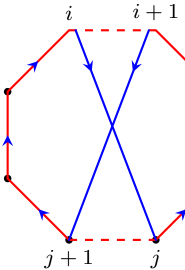
Let's show $\mu \geq 2\sqrt{\log n}$ is necessary. Consider the following cycles:

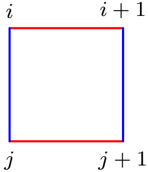


The difference graph B between the true cycle $(1, 2, \ldots, n)$ and the cycle $(1, 2, \ldots, i, j, j-1, \ldots, i+1, j+1, j+2, \ldots, n)$ is a four-cycle.

Aside: information-theoretic optimality

Let's show $\mu \geq 2\sqrt{\log n}$ is necessary. Consider the following cycles:





The difference graph B between the true cycle $(1, 2, \ldots, n)$ and the cycle $(1, 2, \ldots, i, j, j-1, \ldots, i+1, j+1, j+2, \ldots, n)$ is a four-cycle.

- $w(B) \sim N(-2\mu, 4)$.
- Number of such cycles $\sim n^2$.
- Need: $-2\mu + \sqrt{2 \cdot 4 \log(n^2)} < 0$; otherwise, MLE fails (this heuristic can be justified)

Reflection

What we have learned from the previous proof:

- Encode the solution by the difference graph
- Decomposition
- Counting: conditioned on one end of a red edge, the other end has at most 2 choices

Reflection

What we have learned from the previous proof:

- Encode the solution by the difference graph
- Decomposition
- Counting: conditioned on one end of a red edge, the other end has at most 2 choices

Next, for F2F LP

- $X - X^* \in \{0, \pm 1, \pm \frac{1}{2}\}$

Proof of correctness for F2F LP

Proof Outline

1. Encode the solution: for any extremal point X, represent $2(X - X^*)$ as a bicolored multigraph G_X

$$w(G_X) = \langle W, 2(X - X^*) \rangle$$

Proof Outline

1. Encode the solution: for any extremal point X, represent $2(X - X^*)$ as a bicolored multigraph G_X

$$w(G_X) = \langle W, 2(X - X^*) \rangle$$

2. Divide and conquer: decompose G_X as edge-disjoint union of graphs in certain family $\mathcal{F}$

$$w(G_X) = \sum_i w(F_i), \quad F_i \in \mathcal{F}$$

Proof Outline

1. Encode the solution: for any extremal point X, represent $2(X - X^*)$ as a bicolored multigraph G_X

$$w(G_X) = \langle W, 2(X - X^*) \rangle$$

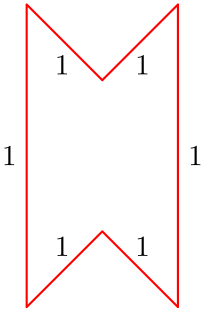
2. Divide and conquer: decompose G_X as edge-disjoint union of graphs in certain family $\mathcal{F}$

$$w(G_X) = \sum_i w(F_i), \quad F_i \in \mathcal{F}$$

3. Counting: Show that whp $w(F) < 0$ for all $F \in \mathcal{F}$

Step 1: Bicolored multigraph representation

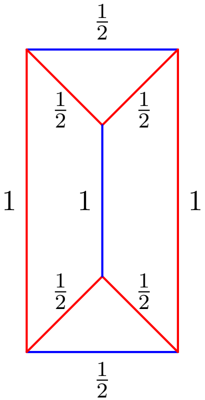
Example 1



X^*: true cycle

Step 1: Bicolored multigraph representation

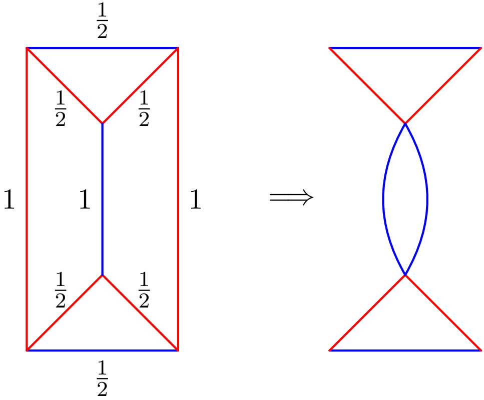
Example 1



X: extremal solution

Step 1: Bicolored multigraph representation

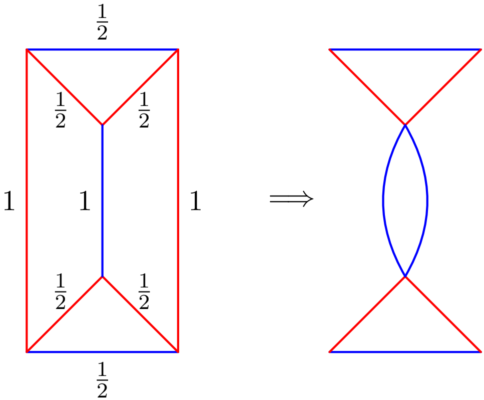
Example 1



X: extremal solution $\quad G_X$ encodes $2(X - X^*)$

Step 1: Bicolored multigraph representation

Example 1

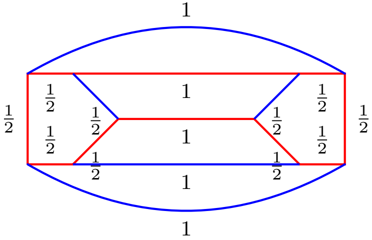


X: extremal solution $\quad G_X$ encodes $2(X - X^*)$

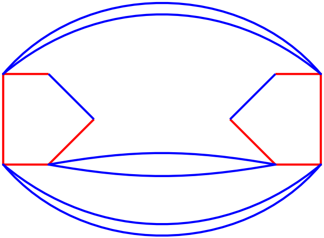
Key observation

G_X is always balanced: red degree $=$ blue degree

Example 2



$\Downarrow$



Step 2: Edge decomposition

Theorem (Kotzig '68)

Every connected balanced bicolored multigraph has an alternating Eulerian circuit.

Step 2: Edge decomposition

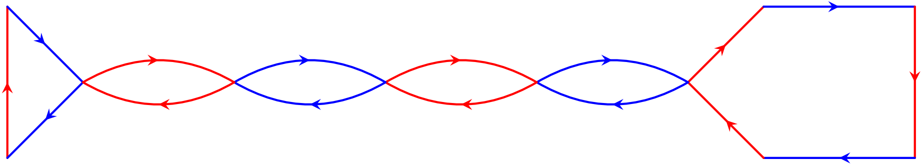
Theorem (Kotzig '68)

Every connected balanced bicolored multigraph has an alternating Eulerian circuit.

Remarks

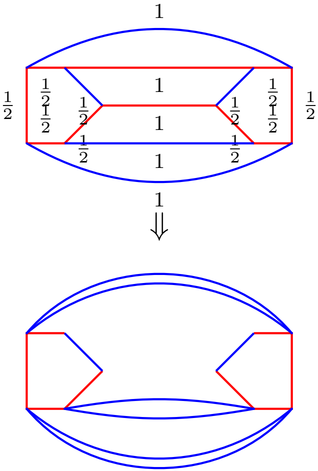
- An Eulerian circuit traverses a double edge twice

Recall: Example 1



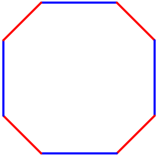
“Dumbbell” structure

Example 2



Decompose as $C_4 + C_4 + C_8$.

Example: even cycles



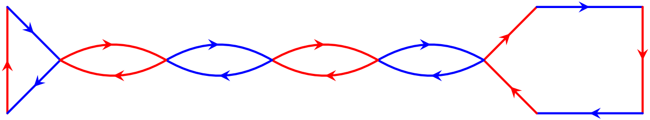
Consider an ℓ-cycle

- $\ell/2$ red and $\ell/2$ blue.
- $w(B) \sim N(-\mu\ell/2, \ell)$.
- Number of alternating ℓ-cyles $\asymp n^{\ell/2}$
- Need

$$\underbrace{\ell\mu/2}_{\text{“signal”}} > \underbrace{\sqrt{2 \cdot \ell \cdot \log(n^{\ell/2})}}_{\text{“noise fluctuation”}} = \ell\sqrt{\log n}$$

which is ensured by $\mu = (2+\epsilon)\sqrt{\log n}$. (This is the same calculation in the proof of correctness for 2F IP – slide 40)

Example: Dumbbell calculation



Consider a dumbbell B with k double edges and ℓ single edges. Then

- $w(B) \sim N(-\mu(k + \ell/2), \mathbf{4}k + \ell)$.
- # of labelings for double edges: $n^{\frac{k+2}{2}}$
- # of labelings for single edges conditioned on double edges: $n^{\frac{\ell}{2}-2}$
- Since $\mu = (2+\epsilon)\sqrt{\log n}$, we have[1]

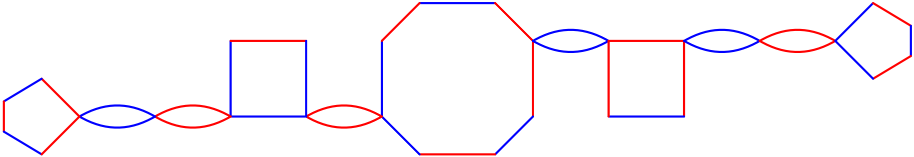
$$\underbrace{(k + \ell/2)\mu}_{\text{“signal”}} > \underbrace{\sqrt{2 \cdot 4k \cdot \log(n^{\frac{k+2}{2}})} + \sqrt{2 \cdot \ell \cdot \log(n^{\frac{\ell}{2}-2})}}_{\text{“noise fluctuation”}}$$

[1] Details: $2k - \sqrt{4k^2 + 8k} + \ell - \sqrt{\ell^2 - 4\ell} = -\frac{4k}{k+\sqrt{k^2+2k}} + \frac{4\ell}{\ell+\sqrt{\ell^2-4\ell}} > 0$.

Step 2: Edge decomposition

$\mathcal{U}$: collection of graphs recursively constructed

1. Start with an even cycle in alternating colors
2. **Blossoming procedure**: At each step, contract an edge in any cycle and attach a **flower** (path of double edges followed by an alternating odd cycle)

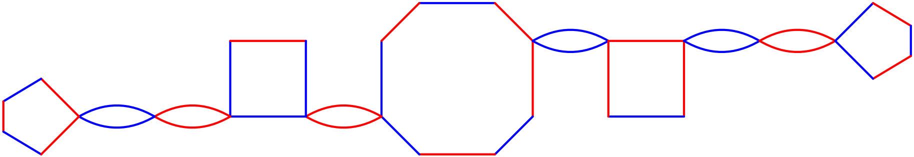


Obtained by starting with an 10-cycle and blossoming 4 times

Step 2: Edge decomposition

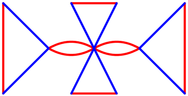
$\mathcal{U}$: collection of graphs recursively constructed

1. Start with an even cycle in alternating colors
2. **Blossoming procedure**: At each step, contract an edge in any cycle and attach a **flower** (path of double edges followed by an alternating odd cycle)

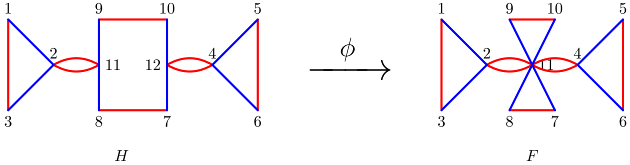


Obtained by starting with an 10-cycle and blossoming 4 times

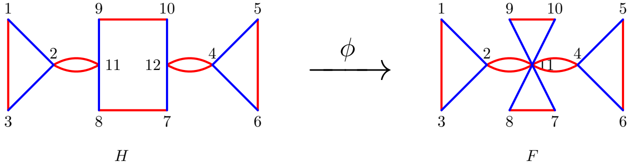
However, not every G_X is of this form...



- Graph homomorphism $\phi : H \to F$ is a vertex map that preserves edges and edge multiplicity



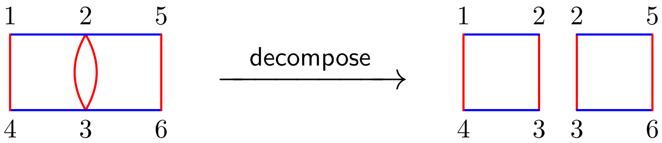
- Graph homomorphism $\phi : H \to F$ is a vertex map that preserves edges and edge multiplicity



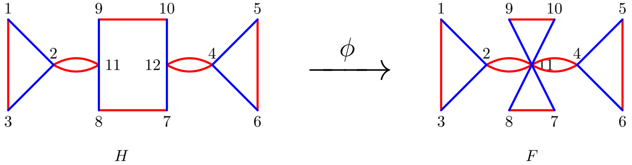
Lemma (Decomposition)

Every balanced bicolored multigraph G with edge multiplicity at most 2 can be decomposed as an union of elements in

$$\mathcal{F} = \{F : V(F) \subset [n], H \to F \text{ for some } H \in \mathcal{U}\}$$



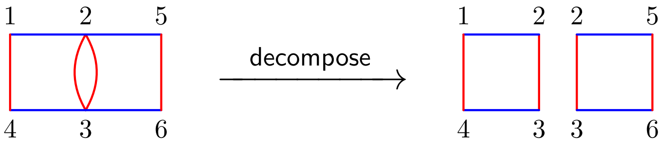
- Graph homomorphism $\phi : H \to F$ is a vertex map that preserves edges and edge multiplicity



Lemma (Decomposition)

Every balanced bicolored multigraph G with edge multiplicity at most 2 can be decomposed as an union of elements in

$$\mathcal{F} = \{F : V(F) \subset [n], H \to F \text{ for some } H \in \mathcal{U}\}$$



- It remains to show $\min_{F \in \mathcal{F}} w(F) < 0$ whp

Step 3: Counting

$$\mathcal{F}_{k,\ell} = \{F \in \mathcal{F} : E(F) \text{ consists of } k \text{ double edges and } \ell \text{ single edges } \}$$

Lemma (Counting isomorphism classes)

The number of distinct $H \in \mathcal{U}$ with k double edges and ℓ single edges is at most $C^{k+\ell}$ for universal constant C.

Lemma (Counting homomorphisms)

For each $H \in \mathcal{U}$, there exists $0 \le r \le \ell/2$

- *Number of labelings for double edges:*

$$\le (Cn)^{k/2+r/2}$$

- *Number of labelings for single edges conditioned on double edges*

$$\le (Cn)^{\ell/2-r}$$

Step 4: Probabilistic arguments

$\mathcal{F}_{k,\ell} = \{F \in \mathcal{F} : E(F)$ consists of k double edges and ℓ single edges $\}$

Lemma

For any $k \geq 0$ and $\ell \geq 3$, with probability at least $1 - n^{-\Theta(k+\ell)}$,

$$\max_{F \in \mathcal{F}_{k,\ell}} \left(w(F) - \mathbb{E}\left[w(F)\right]\right) \leq (1+\epsilon)\,(2k+\ell)\,\sqrt{\log n}$$

Step 4: Probabilistic arguments

$\mathcal{F}_{k,\ell} = \{F \in \mathcal{F} : E(F)$ consists of k double edges and ℓ single edges $\}$

Lemma

For any $k \geq 0$ and $\ell \geq 3$, with probability at least $1 - n^{-\Theta(k+\ell)}$,

$$\max_{F \in \mathcal{F}_{k,\ell}} \left(w(F) - \mathbb{E}\left[w(F)\right]\right) \leq (1+\epsilon)\,(2k+\ell)\,\sqrt{\log n}$$

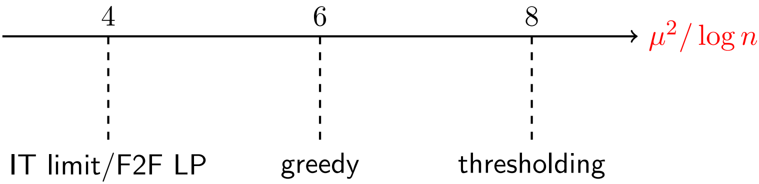
Remarks

- Total: $2k+\ell$ edges, half red half blue. Weights on red edges $\sim N(\mu, 1)$. Weights on blue edges $\sim N(0,1)$.

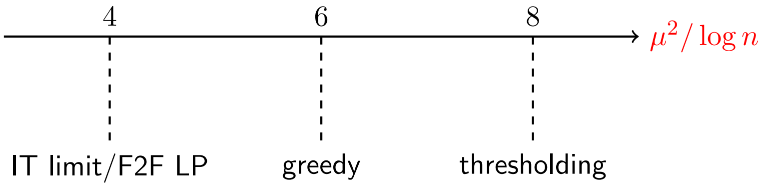
$$w(F) \sim N(-(k+\ell/2)\mu, \mathbf{4}k+\ell)$$

- Proof: Counting $\mathcal{F}_{k,\ell}$ and large deviation bounds

Conclusion and remarks



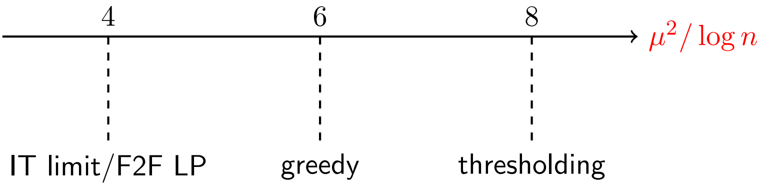
Conclusion and remarks



Extensions/Open problems

- More realistic models: k-NN graph (Watts-Strogatz small-world graph)
  - IT limit becomes $\sqrt{2 \log n}$ for $k \geq 2$ [Ding-W-Xu-Yang, '19]
  - NOT achieved by LP. Efficient algo remains open

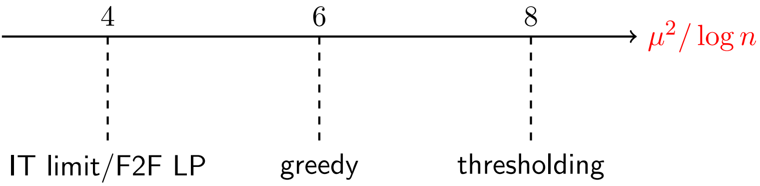
Conclusion and remarks



Extensions/Open problems

- More realistic models: k-NN graph (Watts-Strogatz small-world graph)
  - IT limit becomes $\sqrt{2 \log n}$ for $k \geq 2$ [Ding-W-Xu-Yang, '19]
  - NOT achieved by LP. Efficient algo remains open

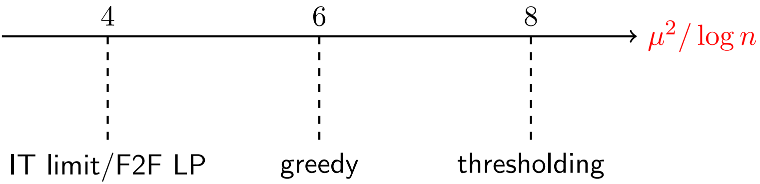
Conclusion and remarks



Extensions/Open problems

- More realistic models: k-NN graph (Watts-Strogatz small-world graph)
  - IT limit becomes $\sqrt{2\log n}$ for $k \geq 2$ [Ding-W-Xu-Yang, '19]
  - NOT achieved by LP. Efficient algo remains open

Conclusion and remarks



Extensions/Open problems

- More realistic models: k-NN graph (Watts-Strogatz small-world graph)
  - IT limit becomes $\sqrt{2\log n}$ for $k \geq 2$ [Ding-W-Xu-Yang, '19]
  - NOT achieved by LP. Efficient algo remains open

References

- Vivek Bagaria, Jian Ding, David Tse, W. & Jiaming Xu (2018). *Hidden Hamiltonian Cycle Recovery via Linear Programming*, https://arxiv.org/abs/1804.05436