Lecture 1: The Planted Matching Problem

Jiaming Xu

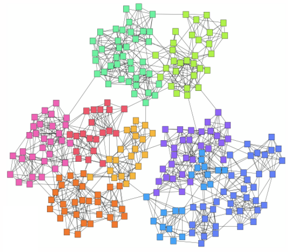
The Fuqua School of Business
Duke University

Joint work with
Mehrdad Moharrami (UIUC) and Cristopher Moore (Santa Fe Institute)
Jian Ding (PKU), Yihong Wu (Yale) and Dana Yang (Cornell)

April 17, 2023
AI4OPT Tutorial Lectures

Statistical model with planted structure

Question: How to recover latent structure from noisy data?



Classical examples

- Recovery of planted clique in Erdős-Rényi graphs
- Community detection under Stochastic Block Model
- Clustering in mixture models

Common theme: low-rank structure

- Underpinning of many phase transitions and algorithms, e.g. spectral method, SDP relaxation, etc

A new zoo of planted problems...

- Planted bipartite matching [Chertkov-Kroc-Krzakala-Vergassola-Zdeborová '10]
- Graph matching (network alignment) [Pedarsani-Grossglauser '11]
- Planted Hamiltonian cycle problem (TSP) [Bagaria-Ding-Tse-W-Xu '18]
- Planted trees [Massoulié-Stephan-Towsley '18]
- Planted k-factors [Sicuro-Zdeborová '20]
- Planted k-nearest-neighbor graph [Ding-Wu-Xu-Yang '19]

A new zoo of planted problems...

- Planted bipartite matching [Chertkov-Kroc-Krzakala-Vergassola-Zdeborová '10]
- Graph matching (network alignment) [Pedarsani-Grossglauser '11]
- Planted Hamiltonian cycle problem (TSP) [Bagaria-Ding-Tse-W-Xu '18]
- Planted trees [Massoulié-Stephan-Towsley '18]
- Planted k-factors [Sicuro-Zdeborová '20]
- Planted k-nearest-neighbor graph [Ding-Wu-Xu-Yang '19]

Common theme: Lack of low-rank structure $\Rightarrow$ new challenges in both statistical analysis and algorithm design

A new zoo of planted problems...

- Planted bipartite matching [Chertkov-Kroc-Krzakala-Vergassola-Zdeborová '10]
- Graph matching (network alignment) [Pedarsani-Grossglauser '11]
- Planted Hamiltonian cycle problem (TSP) [Bagaria-Ding-Tse-W-Xu '18]
- Planted trees [Massoulié-Stephan-Towsley '18]
- Planted k-factors [Sicuro-Zdeborová '20]
- Planted k-nearest-neighbor graph [Ding-Wu-Xu-Yang '19]

Common theme: Lack of low-rank structure $\Rightarrow$ new challenges in both statistical analysis and algorithm design

This tutorial:

- Linear assignment: Planted bipartite matching
- Quadratic assignment: Graph matching (network alignment)

Outline of tutorial

- Lecture 1: Planted matching problem
- Lecture 2: Random graph matching: Information-theoretic limits
- Lecture 3: Random graph matching: Efficient algorithms
- Lecture 4: Random graph matching: Low-degree polynomials and limits of local algorithms

Assignment problems were introduced for facilities location problem by [Koopmans-Beckmann Econometrica '57]

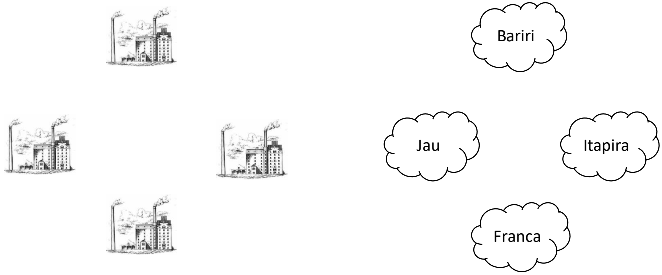
ASSIGNMENT PROBLEMS AND THE LOCATION OF ECONOMIC ACTIVITIES[1]

BY TJALLING C. KOOPMANS AND MARTIN BECKMANN

Two problems in the allocation of indivisible resources are discussed. Both can be interpreted as problems of assigning plants to locations. The first problem, in which cost of transportation between plants is ignored, is found to be a linear programming problem, with which is associated a system of rents that sustains an optimal assignment. The recognition of cost of interplant transportation in the second problem introduces complications which call for more laborious and largely unexplored computations and which also appear to defeat the price system as a means of sustaining an optimal assignment.

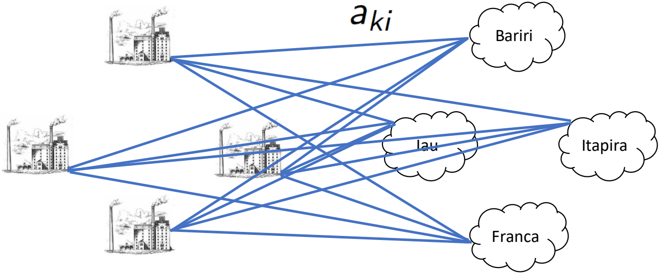
Assignment problems in operations research

Assignment problems were introduced for facilities location problem by [Koopmans-Beckmann Econometrica '57]



Assignment problems in operations research

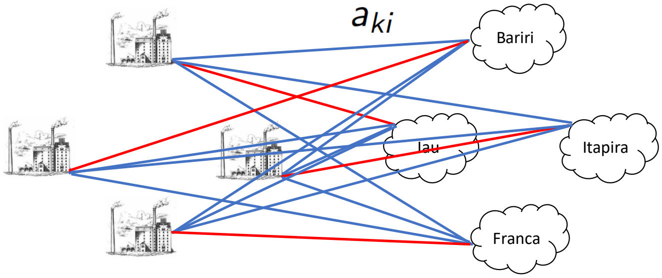
Assignment problems were introduced for facilities location problem by [Koopmans-Beckmann Econometrica '57]



$$\min_{\pi \in \mathcal{S}_n} \sum_{k,i} a_{ki} \pi_{ki}$$

Assignment problems in operations research

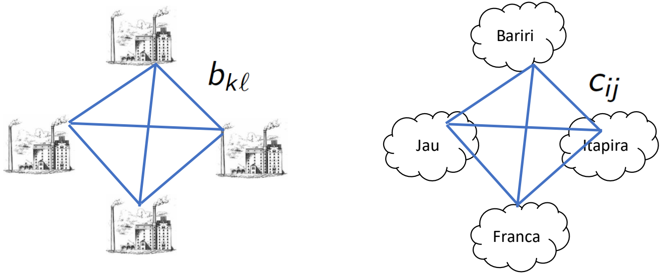
Assignment problems were introduced for facilities location problem by [Koopmans-Beckmann Econometrica '57]



$$\min_{\pi \in \mathcal{S}_n} \sum_{k,i} a_{ki} \pi_{ki}$$

Assignment problems in operations research

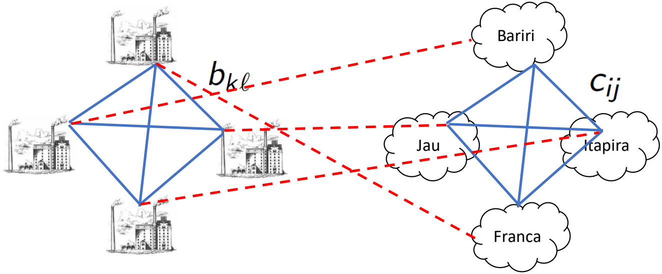
Assignment problems were introduced for facilities location problem by [Koopmans-Beckmann Econometrica '57]



$$\min_{\pi \in \mathcal{S}_n} \sum_{k,\ell} \sum_{i,j} b_{k\ell} \pi_{ki} c_{ij} \pi_{\ell j}$$

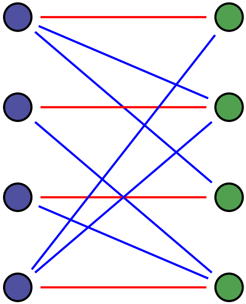
Assignment problems in operations research

Assignment problems were introduced for facilities location problem by [Koopmans-Beckmann Econometrica '57]



$$\min_{\pi \in \mathcal{S}_n} \sum_{k,\ell} \sum_{i,j} b_{k\ell} \pi_{ki} c_{ij} \pi_{\ell j}$$

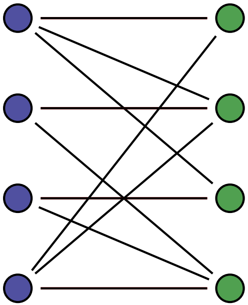
The planted matching model



- A weighted bipartite graph G
- A hidden perfect matching M^*
- All $n(n-1)$ pairs not in M^* are connected w.p. $\frac{d}{n}$
- Edge weight

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

The planted matching model

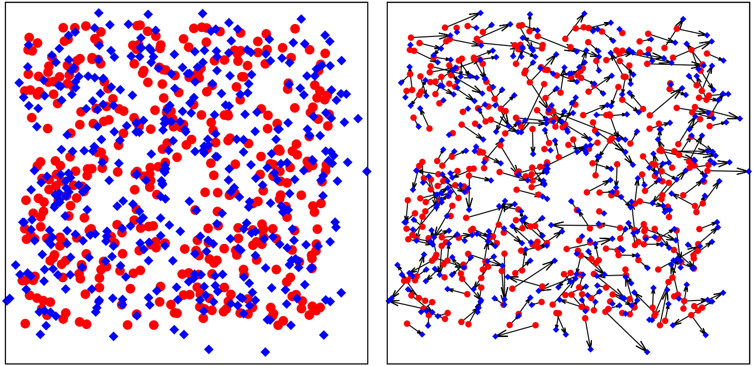


- A weighted bipartite graph G
- A hidden perfect matching M^*
- All $n(n-1)$ pairs not in M^* are connected w.p. $\frac{d}{n}$
- Edge weight

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

- Goal: recover M^* from G

Motivating application: particle tracking



[Chertkov-Kroc-Krzakala-Vergassola-Zdeborová PNAS'10]

- Tracking particles advected by turbulent fluid flow
- Goal: recover the latent correspondence between particles
- $d = n$, $P = |\mathcal{N}(0, \kappa)|$ and $Q = \text{Uniform}[0, n]$

Motivating application: particle tracking

Our replica calculations also show that the "random distance" model presents an interesting phase transition at the diffusivity $\kappa_c \approx 0.174$. For $\kappa^* < \kappa_c$, the MPA π_{MPA} is identical to the special one π^* with high probability, whereas for $\kappa^* > \kappa_c$ the overlap (defined via the Hamming distance) between the most likely assignment π_{MPA} and the special one π^* is extensive, i.e., $O(N)$. The comparison with the finite-dimensional case is discussed in the Results section.

[Chertkov-Kroc-Krzakala-Vergassola-Zdeborová PNAS'10]

- Tracking particles advected by turbulent fluid flow
- Goal: recover latent correspondence between particles based on pairwise distances
- $d = n$, $P = |\mathcal{N}(0, \kappa)|$ and $Q = \text{Uniform}[0, n]$.

Motivating application: particle tracking

Our replica calculations also show that the "random distance" model presents an interesting phase transition at the diffusivity $\kappa_c \approx 0.174$. For $\kappa^* < \kappa_c$, the MPA π_{MPA} is identical to the special one π^* with high probability, whereas for $\kappa^* > \kappa_c$ the overlap (defined via the Hamming distance) between the most likely assignment π_{MPA} and the special one π^* is extensive, i.e., $O(N)$. The comparison with the finite-dimensional case is discussed in the Results section.

[Chertkov-Kroc-Krzakala-Vergassola-Zdeborová PNAS'10]

- Tracking particles advected by turbulent fluid flow
- Goal: recover latent correspondence between particles based on pairwise distances
- $d = n$, $P = |\mathcal{N}(0, \kappa)|$ and $Q = \text{Uniform}[0, n]$.
- Optimal κ turns out to be $\frac{1}{2\pi} \approx 0.159$

Maximum likelihood estimation as linear assignment

Maximum likelihood estimation reduces to max-weighted matching:

$$\widehat{M}_{\rm ML} = \arg\max_{M \in \mathcal{M}} \sum_{e \in M} \log \frac{P}{Q}(W_e)$$

- Linear assignment: computable in polynomial time
- For certain distributions e.g. exponentials, further reduce to min-weighted matching in terms of W_e

Maximum likelihood estimation as linear assignment

Maximum likelihood estimation reduces to max-weighted matching:

$$\widehat{M}_{\mathrm{ML}} = \arg\max_{M \in \mathcal{M}} \sum_{e \in M} \log \frac{P}{Q}(W_e)$$

- Linear assignment: computable in polynomial time
- For certain distributions e.g. exponentials, further reduce to min-weighted matching in terms of W_e
- How much does $\widehat{M}_{\mathrm{ML}}$ have in common with M^*?

$$\mathrm{overlap}(\widehat{M}_{\mathrm{ML}}, M^*) \triangleq \frac{1}{n}\mathbb{E}\left|\widehat{M}_{\mathrm{ML}} \cap M^*\right| = 1 - \frac{1}{2n}\mathbb{E}\left|\widehat{M}_{\mathrm{ML}} \triangle M^*\right|$$

Maximum likelihood estimation as linear assignment

Maximum likelihood estimation reduces to max-weighted matching:

$$\widehat{M}_{\mathrm{ML}} = \arg\max_{M \in \mathcal{M}} \sum_{e \in M} \log \frac{P}{Q}(W_e)$$

- Linear assignment: computable in polynomial time
- For certain distributions e.g. exponentials, further reduce to min-weighted matching in terms of W_e
- How much does $\widehat{M}_{\mathrm{ML}}$ have in common with M^*?

$$\mathrm{overlap}(\widehat{M}_{\mathrm{ML}}, M^*) \triangleq \frac{1}{n}\mathbb{E}\left|\widehat{M}_{\mathrm{ML}} \cap M^*\right| = 1 - \frac{1}{2n}\mathbb{E}\left|\widehat{M}_{\mathrm{ML}} \triangle M^*\right|$$

- Information-theoretic limit for reconstruction, in terms of
  1. Average degree d
  2. Similarity between P and Q

Maximum likelihood estimation as linear assignment

Maximum likelihood estimation reduces to max-weighted matching:

$$\widehat{M}_{\mathrm{ML}} = \arg\max_{M \in \mathcal{M}} \sum_{e \in M} \log \frac{P}{Q}(W_e)$$

- Linear assignment: computable in polynomial time
- For certain distributions e.g. exponentials, further reduce to min-weighted matching in terms of W_e
- How much does $\widehat{M}_{\mathrm{ML}}$ have in common with M^*?

$$\mathrm{overlap}(\widehat{M}_{\mathrm{ML}}, M^*) \triangleq \frac{1}{n}\mathbb{E}\left|\widehat{M}_{\mathrm{ML}} \cap M^*\right| = 1 - \frac{1}{2n}\mathbb{E}\left|\widehat{M}_{\mathrm{ML}} \triangle M^*\right|$$

- Information-theoretic limit for reconstruction, in terms of
  1. Average degree d
  2. Similarity between P and Q
     - Bhattacharyya coefficient (Hellinger affinity) $B(P,Q) \triangleq \int \sqrt{\mathrm{d}P\mathrm{d}Q}$

Main result: phase transition threshold

Theorem (Ding-Wu-X.-Yang '21)

- *If* $\sqrt{d}\, B(P, Q) \leq 1$, *then* $\text{overlap}(\widehat{M}_{\mathrm{ML}}, M^*) \to 1$

Main result: phase transition threshold

Theorem (Ding-Wu-X.-Yang '21)

- *If* $\sqrt{d}\ B(P,Q) \le 1$, *then* $\text{overlap}(\widehat{M}_{\mathrm{ML}}, M^*) \to 1$
- *If* $\sqrt{d}\ B(P,Q) \ge 1+\epsilon$, *then for all* $\widehat{M}$ *and some* $c = c(\epsilon)$

$$\text{overlap}(\widehat{M}, M^*) \le 1 - c$$

for both

- *sparse model:* d, P, Q *fixed; and*
- *dense model:* $d \to \infty$ *and* $Q(x) = \frac{1}{d}\rho(\frac{x}{d})$.

- Resolve the conjecture in [Semerjian-Sicuro-Zdeborová '20]

Some interesting special cases

- Unweighted model: $P = Q$,

$$\boxed{\text{Sharp threshold } d = 1}$$

Coincide with the threshold for emergence of giant component

- Particle tracking: $d = n$, $P = |\mathcal{N}(0, \kappa)|$, $Q = \text{Uniform}[0, n]$:

$$\boxed{\text{Sharp threshold } \kappa = \frac{1}{2\pi}}$$

- Exponential model: $d = n$, $P = \exp(\lambda)$, $Q = \exp(1/n)$:

$$\boxed{\text{Sharp threshold } \lambda = 4}$$

Infinite-order phase transition under exponential model

Theorem (Ding-Wu-X.-Yang '21)

Assume $\lambda = 4 - \epsilon$. There exist absolute constants c_1, c_2:

$$\text{overlap}(\widehat{M}_{\mathrm{ML}}, M^*) \geq 1 - e^{-\frac{c_1}{\sqrt{\epsilon}}};$$

Conversely, for all $\widehat{M}$,

$$\text{overlap}(\widehat{M}, M^*) \leq 1 - e^{-\frac{c_2}{\sqrt{\epsilon}}}.$$

Infinite-order phase transition under exponential model

Theorem (Ding-Wu-X.-Yang '21)

Assume $\lambda = 4 - \epsilon$. There exist absolute constants c_1, c_2:

$$\text{overlap}(\widehat{M}_{\text{ML}}, M^*) \geq 1 - e^{-\frac{c_1}{\sqrt{\epsilon}}};$$

Conversely, for all $\widehat{M}$,

$$\text{overlap}(\widehat{M}, M^*) \leq 1 - e^{-\frac{c_2}{\sqrt{\epsilon}}}.$$

- Optimal reconstruction error is $\exp\left(-\Theta(1/\sqrt{\epsilon})\right)$
- Resolve the ∞-order phase transition conjecture [Semerjian-Sicuro-Zdeborová '20]

Overlap of MLE under exponential model

Theorem (Moharrami-Moore-X. '19)

$$\lim_{n\to\infty} \text{overlap}(\widehat{M}_{\mathrm{ML}}, M^*) = \alpha(\lambda), \quad \textit{if } 0 < \lambda < 4,$$

where $\alpha(\lambda) = 1 - 2\int_0^\infty (1 - F(x))\,(1 - G(x))\,V(x)W(x)\,dx,$

Overlap of MLE under exponential model

Theorem (Moharrami-Moore-X. '19)

$$\lim_{n\to\infty} \text{overlap}(\widehat{M}_{\mathrm{ML}}, M^*) = \alpha(\lambda), \quad \text{if } 0 < \lambda < 4,$$

where $\alpha(\lambda) = 1 - 2\int_0^\infty (1 - F(x))\,(1 - G(x))\,V(x)W(x)\,dx,$

and F, G, V, W is the unique solution to the ODE system

$$\dot{F} = (1-F)(1-G)V$$
$$\dot{G} = -(1-F)(1-G)W$$
$$\dot{V} = \lambda(V - F)$$
$$\dot{W} = -\lambda(W - G)$$

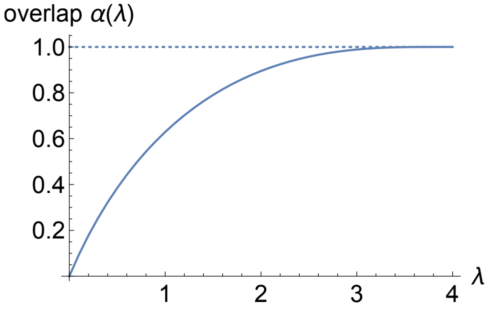
Boundary conditions: $F(x), V(x), G(-x), W(-x) \to \begin{cases} 1 & x \to +\infty \\ 0 & x \to -\infty \end{cases}$

Overlap of MLE under exponential model

Theorem (Moharrami-Moore-X. '19)

$$\lim_{n\to\infty} \text{overlap}(\widehat{M}_{\mathrm{ML}}, M^*) = \alpha(\lambda), \quad \text{if } 0 < \lambda < 4,$$

where $\alpha(\lambda) = 1 - 2\int_0^\infty (1 - F(x))\,(1 - G(x))\,V(x)W(x)\,dx,$

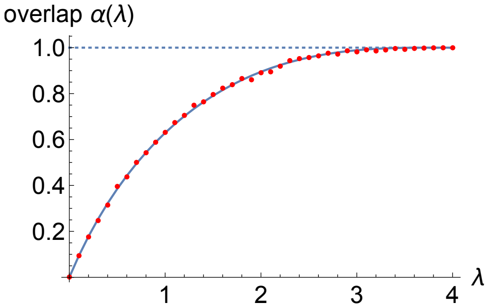


Overlap of MLE under exponential model

Theorem (Moharrami-Moore-X. '19)

$$\lim_{n\to\infty} \text{overlap}(\widehat{M}_{\mathrm{ML}}, M^*) = \alpha(\lambda), \quad \text{if } 0 < \lambda < 4,$$

where $\alpha(\lambda) = 1 - 2\int_0^\infty (1 - F(x))\,(1 - G(x))\,V(x)W(x)\,dx,$

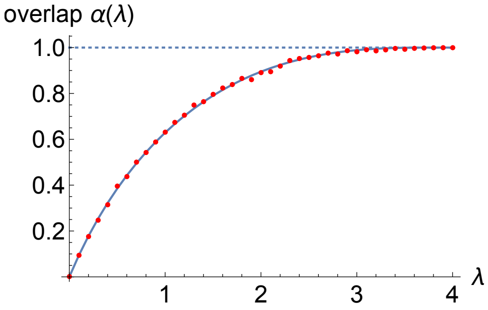


Overlap of MLE under exponential model

Theorem (Moharrami-Moore-X. '19)

$$\lim_{n\to\infty} \text{overlap}(\widehat{M}_{\text{ML}}, M^*) = \alpha(\lambda), \quad \textit{if } 0 < \lambda < 4,$$

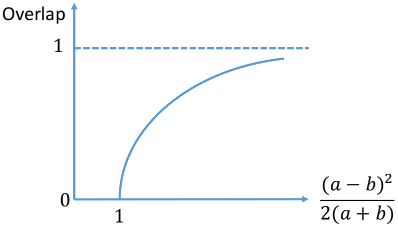
where $\alpha(\lambda) = 1 - 2\int_0^\infty (1 - F(x))\,(1 - G(x))\,V(x)W(x)\,dx,$

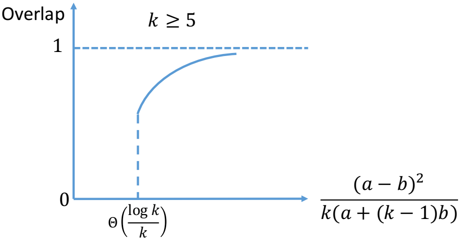


$\alpha(\lambda)$ is infinitely differentiable at threshold $\lambda = 4$!

Comparison of phase transition orders

Drastically different from the other well-known planted models such as stochastic block model (conjecture, not fully proven yet)





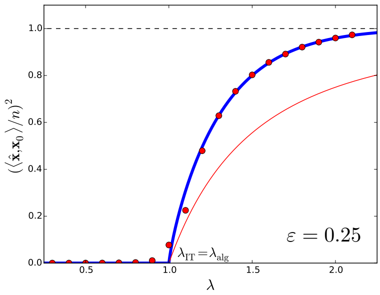
Second-order phase transition
with two groups
[DKMZ'11, MNS'12 13, Massoulié'13]

First-order phase transition
with five or more groups
[DKMZ'11, BMNN '16, AS'16]

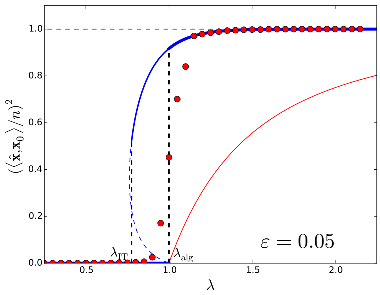
Aside: A phase transition is of pth order if the $(p-2)$th derivative of the average overlap is continuous

Comparison of phase transition orders

Spiked Wigner model: $Y = x_0 x_0^\top + Z$, x_0 is ϵn-sparse [Deshpande-Montanari '14, Krzakala-X.-Zdeborová '16, Barbier et al '16, Lelarge-Miolane '16, Montanari-Venkataramanan '17]



Second-order phase transition



First-order phase transition

Figure from [Montanari-Venkataramanan '17]

Finite-order phase transition for unweighted model

- For unweighted model with $d = 1 + \epsilon$,

$$\epsilon^8 \lesssim \inf_{\widehat{M}} \frac{1}{n} \mathbb{E} \left| \widehat{M} \triangle M^* \right| \lesssim \epsilon.$$

  Thus the phase transition is continuous, and is of finite order.
- Determining the exact order of the phase transition for unweighted model is an open problem

Analysis

- Proof of positive result via maximum likelihood
- Proof of negative result via analyzing posterior distribution
- Proof of tight error lower bound under exponential model
- Proof of overlap of MLE under exponential model

Proof of positive result via maximum likelihood

- At most $\binom{n}{t} t!$ matchings M with $|M \triangle M^*| = 2t$
- Probability that M has higher likelihood than M^* is

$$\mathbb{P}\left\{ \sum_{e \in M \setminus M^*} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e) \geq \sum_{e \in M^* \setminus M} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e) \right\} \leq \left(\frac{d}{n} B^2(\mathcal{P}, \mathcal{Q}) \right)^t$$

Proof of positive result via maximum likelihood

- At most $\binom{n}{t} t!$ matchings M with $|M \triangle M^*| = 2t$
- Probability that M has higher likelihood than M^* is

$$\mathbb{P}\left\{ \sum_{e \in M \setminus M^*} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e) \geq \sum_{e \in M^* \setminus M} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e) \right\} \leq \left(\frac{d}{n} B^2(\mathcal{P}, \mathcal{Q}) \right)^t$$

- Taking union bound $\Rightarrow$

$$\begin{aligned} & \mathbb{P}\left[\exists\ M \text{ with } |M \triangle M^*| \geq 2\beta n \text{ has higher likelihood than } M^*\right] \\ & \leq \sum_{t \geq \beta n} \binom{n}{t} t! \left(\frac{d}{n} B^2(\mathcal{P}, \mathcal{Q}) \right)^t \\ & \to 0 \text{ for some } \beta = o(1), \text{ if } \sqrt{d}\, B(\mathcal{P}, \mathcal{Q}) \leq 1 \end{aligned}$$

Proof of positive result via maximum likelihood

- At most $\binom{n}{t} t!$ matchings M with $|M \triangle M^*| = 2t$
- Probability that M has higher likelihood than M^* is

$$\mathbb{P}\left\{ \sum_{e \in M \setminus M^*} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e) \geq \sum_{e \in M^* \setminus M} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e) \right\} \leq \left(\frac{d}{n} B^2(\mathcal{P}, \mathcal{Q}) \right)^t$$

- Taking union bound $\Rightarrow$

$$\mathbb{P}\left[\exists\ M \text{ with } |M \triangle M^*| \geq 2\beta n \text{ has higher likelihood than } M^*\right]$$
$$\leq \sum_{t \geq \beta n} \binom{n}{t} t! \left(\frac{d}{n} B^2(\mathcal{P}, \mathcal{Q}) \right)^t$$
$$\to 0 \text{ for some } \beta = o(1), \text{ if } \sqrt{d}\, B(\mathcal{P}, \mathcal{Q}) \leq 1$$

Aside: If $\sqrt{d} B(P, Q) \to 0$, then $M_{\mathrm{ML}} = M^*$ whp

Analysis

- Proof of positive result via maximum likelihood
- Proof of negative result via analyzing posterior distribution
- Proof of tight error lower bound under exponential model
- Proof of overlap of MLE under exponential model

Negative result via analyzing posterior distribution

- ...But MLE does not maximize overlap
- Optimal estimator is maximum posterior marginal
- Need to analyze posterior distribution: Gibbs distribution over perfect matchings

$$\mu_W(m) \propto \exp\left(\sum_{e \in m} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e)\right)$$

Negative result via analyzing posterior distribution

- ...But MLE does not maximize overlap
- Optimal estimator is maximum posterior marginal
- Need to analyze posterior distribution: Gibbs distribution over perfect matchings

$$\mu_W(m) \propto \exp\left(\sum_{e \in m} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e)\right)$$

Crucial observation

Sampling from posterior distribution is optimal within a factor of two.

Negative result via analyzing posterior distribution

- ...But MLE does not maximize overlap
- Optimal estimator is maximum posterior marginal
- Need to analyze posterior distribution: Gibbs distribution over perfect matchings

$$\mu_W(m) \propto \exp\left(\sum_{e\in m} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e)\right)$$

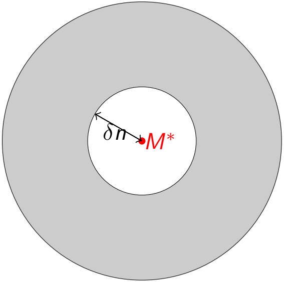
Crucial observation

Sampling from posterior distribution is optimal within a factor of two.

Proof: Let $\widetilde{M}$ be sampled from posterior distribution. Then for any estimator $\widehat{M}$, $(M^*, \widehat{M}) \overset{\text{law}}{=} (\widetilde{M}, \widehat{M})$ and

$$\mathbb{E}|\widetilde{M}\triangle M^*| \le \mathbb{E}|\widetilde{M}\triangle\widehat{M}| + \mathbb{E}|\widehat{M}\triangle M^*| = 2\mathbb{E}|\widehat{M}\triangle M^*|.$$

Analysis of posterior distribution



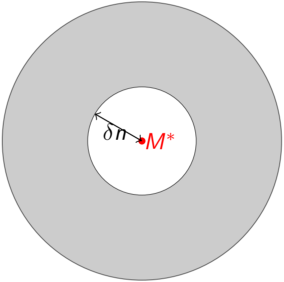
- Upper bound the posterior mass of matchings near M^*:

$$\frac{\mu_W(\mathcal{M}_{\rm near})}{\mu_W(M^*)} \le e^{7\epsilon\delta n} \qquad (1)$$

- Lower bound the posterior mass of matchings far away from M^*:

$$\frac{\mu_W(\mathcal{M}_{\rm far})}{\mu_W(M^*)} \ge e^{14\epsilon\delta n} \qquad (2)$$

Analysis of posterior distribution



- Upper bound the posterior mass of matchings near M^*:

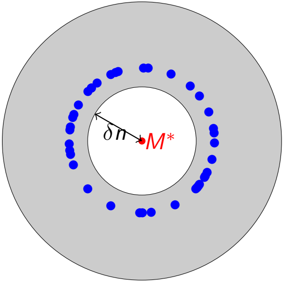
$$\frac{\mu_W(\mathcal{M}_{\text{near}})}{\mu_W(M^*)} \leq e^{7\epsilon\delta n} \qquad (1)$$

- Lower bound the posterior mass of matchings far away from M^*:

$$\frac{\mu_W(\mathcal{M}_{\text{far}})}{\mu_W(M^*)} \geq e^{14\epsilon\delta n} \qquad (2)$$

- Proof of (1) is straightforward: truncated first moment

Analysis of posterior distribution



- Upper bound the posterior mass of matchings near M^*:

$$\frac{\mu_W(\mathcal{M}_{\rm near})}{\mu_W(M^*)} \le e^{7\epsilon\delta n} \qquad (1)$$

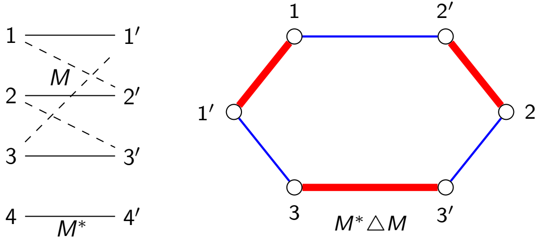
- Lower bound the posterior mass of matchings far away from M^*:

$$\frac{\mu_W(\mathcal{M}_{\rm far})}{\mu_W(M^*)} \ge e^{14\epsilon\delta n} \qquad (2)$$

- Proof of (1) is straightforward: truncated first moment
- Proof of (2) is constructive: find exponentially many matchings $M \in \mathcal{M}_{\rm far}$ whose likelihood exceeds that of M^*

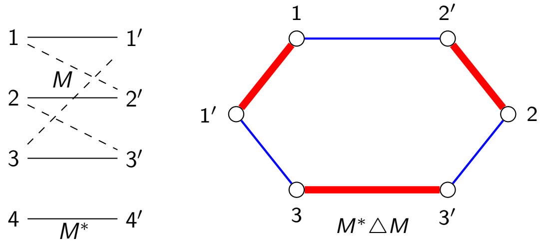
Lower bound: Augmenting alternating cycles

For perfect matching M, $M \triangle M^* =$ disjoint union of alternating cycles



Lower bound: Augmenting alternating cycles

For perfect matching M, $M \triangle M^* =$ disjoint union of alternating cycles

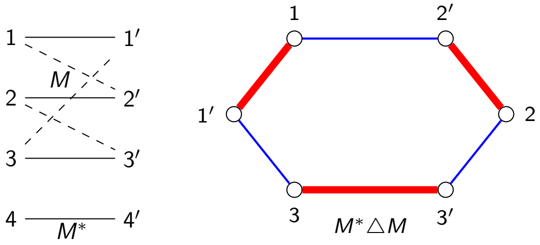


Goal: Find exponentially many long alternating cycles C that are augmenting:

$$\sum_{e \in E_{\text{blue}}(C)} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e) \geq \sum_{e \in E_{\text{red}}(C)} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e).$$

Lower bound: Augmenting alternating cycles

For perfect matching M, $M \triangle M^* =$ disjoint union of alternating cycles



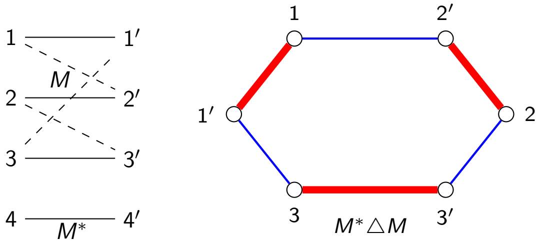
Goal: Find exponentially many long alternating cycles C that are augmenting:

$$\sum_{e \in E_{\mathrm{blue}}(C)} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e) \geq \sum_{e \in E_{\mathrm{red}}(C)} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e).$$

- Augmenting alternating cycles are rare;

Lower bound: Augmenting alternating cycles

For perfect matching M, $M \triangle M^* =$ disjoint union of alternating cycles



Goal: Find exponentially many long alternating cycles C that are augmenting:

$$\sum_{e \in E_{\text{blue}}(C)} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e) \geq \sum_{e \in E_{\text{red}}(C)} \log \frac{\mathcal{P}}{\mathcal{Q}}(W_e).$$

- Augmenting alternating cycles are rare; but there are many alternating cycles

Failure of second-moment in counting alternating cycles

Natural attempt: first and second moment method.

- Let S be the set of augmenting alternating cycles in G of length at least cn. Then $\mathbb{E}|S| = e^{\Omega(n)}$.

Failure of second-moment in counting alternating cycles

Natural attempt: first and second moment method.

- Let S be the set of augmenting alternating cycles in G of length at least cn. Then $\mathbb{E}|S| = e^{\Omega(n)}$.
- If $\mathbb{E}(|S|^2) \lesssim (\mathbb{E}|S|)^2$, then $|S| = e^{\Omega(n)}$ with constant probability.

Failure of second-moment in counting alternating cycles

Natural attempt: first and second moment method.

- Let S be the set of augmenting alternating cycles in G of length at least cn. Then $\mathbb{E}|S| = e^{\Omega(n)}$.
- If $\mathbb{E}(|S|^2) \lesssim (\mathbb{E}|S|)^2$, then $|S| = e^{\Omega(n)}$ with constant probability.
- However, $\mathbb{E}(|S|^2) \gg (\mathbb{E}|S|)^2$ due to the excessive correlation between long cycles.

Failure of second-moment in counting alternating cycles

Natural attempt: first and second moment method.

- Let S be the set of augmenting alternating cycles in G of length at least cn. Then $\mathbb{E}|S| = e^{\Omega(n)}$.
- If $\mathbb{E}(|S|^2) \lesssim (\mathbb{E}|S|)^2$, then $|S| = e^{\Omega(n)}$ with constant probability.
- However, $\mathbb{E}(|S|^2) \gg (\mathbb{E}|S|)^2$ due to the excessive correlation between long cycles.

Key idea

First find many disjoint short paths, then connect the paths into long cycles [Aldous '98, Ding '13, ...]

Existence of many long augmenting alternating cycles

Two-stage cycle-finding scheme

Reserve a set V of γn vertices for some small $\gamma > 0$.

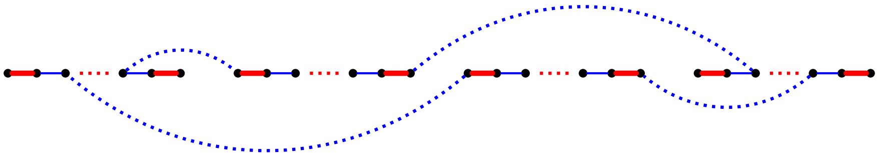
1. Stage 1 (path construction): Find $\Theta(n)$ disjoint short (constant length) augmenting alternating paths, using vertices in V^c.

Existence of many long augmenting alternating cycles

Two-stage cycle-finding scheme

Reserve a set V of γn vertices for some small $\gamma > 0$.

1. Stage 1 (path construction): Find $\Theta(n)$ disjoint short (constant length) augmenting alternating paths, using vertices in V^c.
2. Stage 2 (sprinkling): Connect the paths into long cycles, using vertices in V.

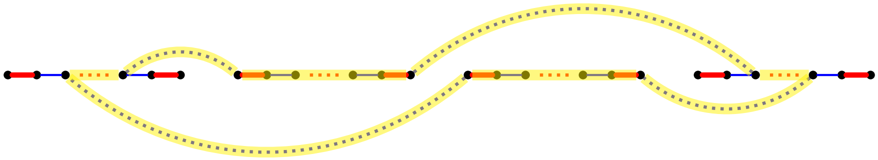


Existence of many long augmenting alternating cycles

Two-stage cycle-finding scheme

Reserve a set V of γn vertices for some small $\gamma > 0$.

1. Stage 1 (path construction): Find $\Theta(n)$ disjoint short (constant length) augmenting alternating paths, using vertices in V^c.
2. Stage 2 (sprinkling): Connect the paths into long cycles, using vertices in V.

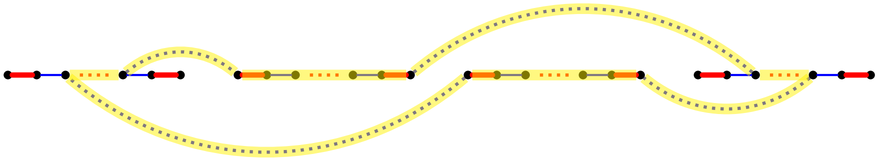


Existence of many long augmenting alternating cycles

Two-stage cycle-finding scheme

Reserve a set V of γn vertices for some small $\gamma > 0$.

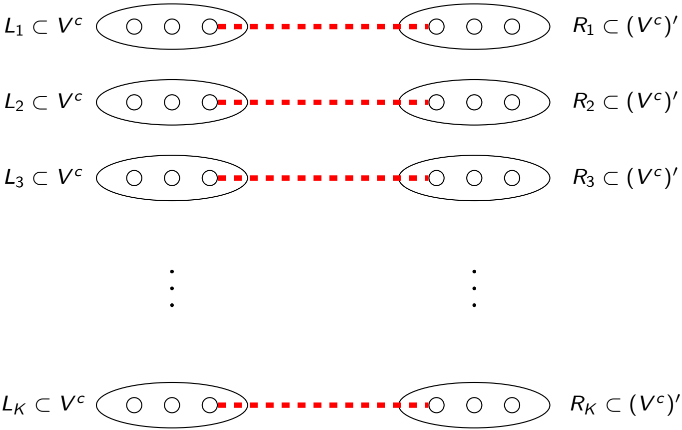
1. Stage 1 (path construction): Find $\Theta(n)$ disjoint short (constant length) augmenting alternating paths, using vertices in V^c.
2. Stage 2 (sprinkling): Connect the paths into long cycles, using vertices in V.



Caution: need to ensure alternating colors in sprinkling

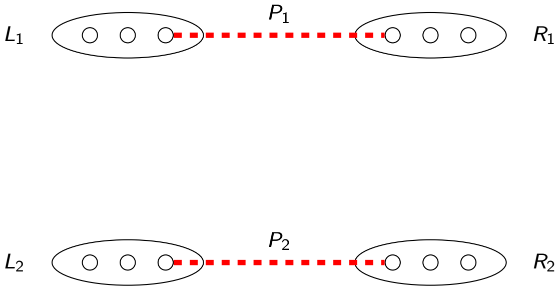
Two-stage cycle-finding scheme

Stage 1 (path construction): Find $\{L_k, R_k\}_{k=1}^{K}$ for $K = \Omega(n)$ such that every vertex in L_k is connected to every vertex in R_k via an augmenting alternating path (length = large constant)



Two-stage cycle-finding scheme

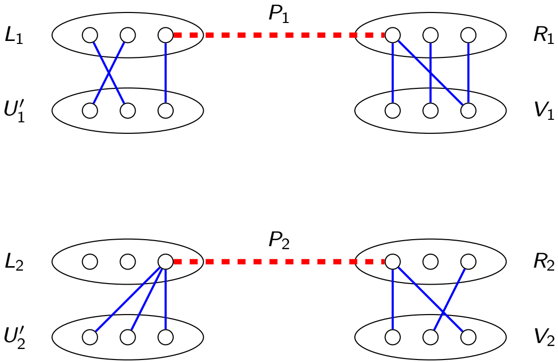
Stage 2 (sprinkling):



Two-stage cycle-finding scheme

Stage 2 (sprinkling):

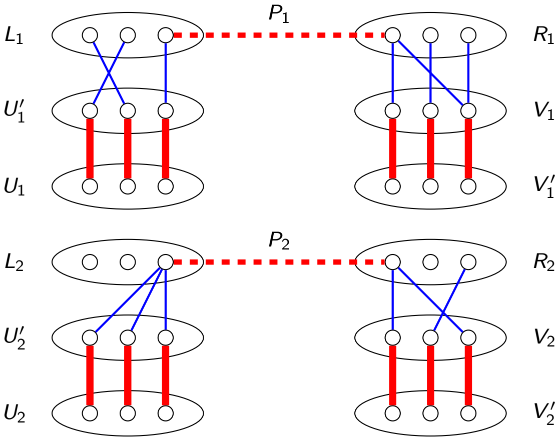
1. Let U'_k be set of reserved vertices connecting to L_k
   Let V_k be set of reserved vertices connecting to R_k



Two-stage cycle-finding scheme

Stage 2 (sprinkling):

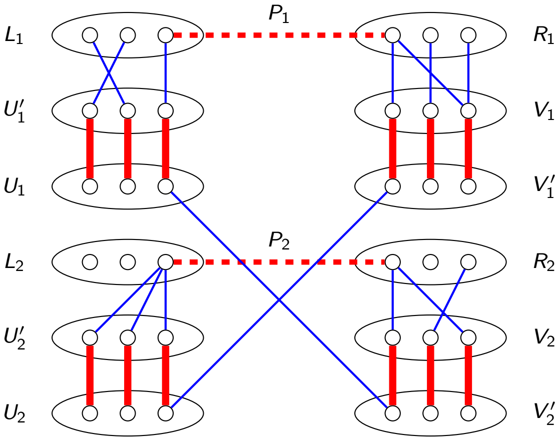
1. Let U'_k be set of reserved vertices connecting to L_k
   Let V_k be set of reserved vertices connecting to R_k



Two-stage cycle-finding scheme

Stage 2 (sprinkling):

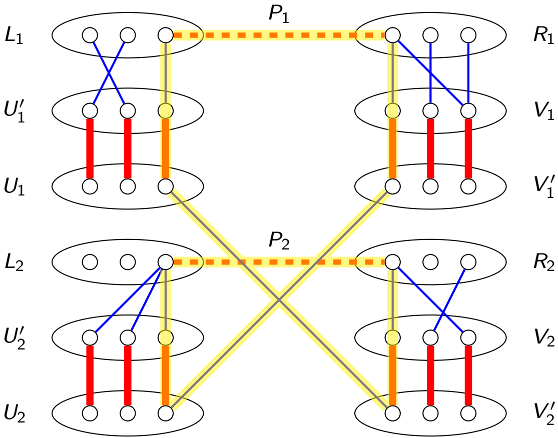
1. Let U'_k be set of reserved vertices connecting to L_k
   Let V_k be set of reserved vertices connecting to R_k
2. Find blue edges connecting $\{U_k\}, \{V'_k\}$.



Two-stage cycle-finding scheme

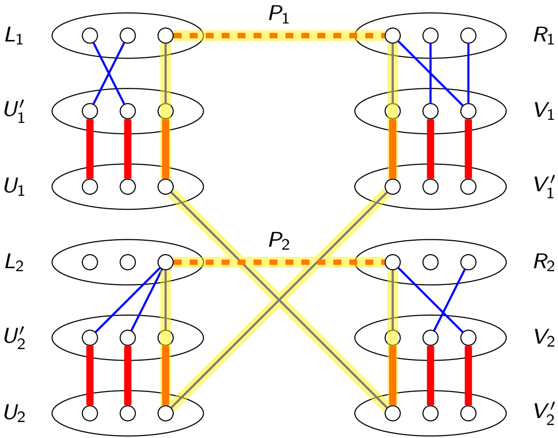
Stage 2 (sprinkling):

1. Let U'_k be set of reserved vertices connecting to L_k
   Let V_k be set of reserved vertices connecting to R_k
2. Find blue edges connecting $\{U_k\}, \{V'_k\}$.



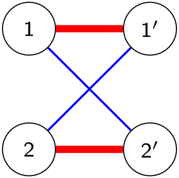
Two-stage cycle-finding scheme

Super graph: Define G_{super} on $[K] \times [K]'$, such that (k, k') is a red edge for all k, and (i, j') is a blue edge iff U_i and V_j' is connected by at least one blue edge.



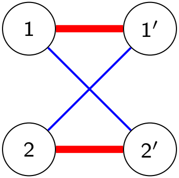
Two-stage cycle-finding scheme

Super graph: Define G_{super} on $[K] \times [K]'$, such that (k, k') is a red edge for all k, and (i, j') is a blue edge iff U_i and V_j' is connected by at least one blue edge.



Two-stage cycle-finding scheme

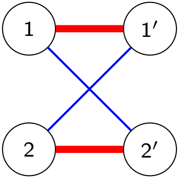
Super graph: Define G_{super} on $[K] \times [K]'$, such that (k, k') is a red edge for all k, and (i, j') is a blue edge iff U_i and V_j' is connected by at least one blue edge.



1. Each alternating cycle on G_{super} expands into an augmenting alternating cycle in G

Two-stage cycle-finding scheme

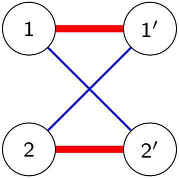
Super graph: Define G_{super} on $[K] \times [K]'$, such that (k, k') is a red edge for all k, and (i, j') is a blue edge iff U_i and V'_j is connected by at least one blue edge.



1. Each alternating cycle on G_{super} expands into an augmenting alternating cycle in G
2. G_{super} is a very supercritical Erdős-Rényi bipartite graph with a planted perfect matching

Two-stage cycle-finding scheme

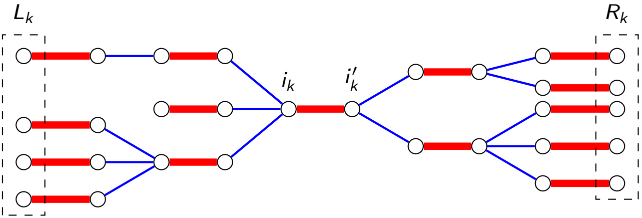
Super graph: Define G_{super} on $[K] \times [K]'$, such that (k, k') is a red edge for all k, and (i, j') is a blue edge iff U_i and V'_j is connected by at least one blue edge.



1. Each alternating cycle on G_{super} expands into an augmenting alternating cycle in G
2. G_{super} is a very supercritical Erdős-Rényi bipartite graph with a planted perfect matching
3. G_{super} contains $e^{\Omega(K)} = e^{\Omega(n)}$ alternating cycles of length $\Omega(K) = \Omega(n)$ (standard DFS argument [Krivelevich-Lee-Sudakov '13])

Path construction via neighborhood exploration process

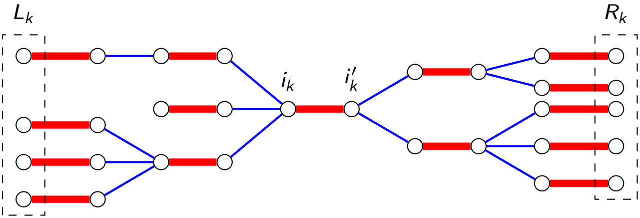
Two-sided tree:



Starting from i_k, grow a tree, remove the inspected vertices, and then grow another tree from i'_k

Path construction via neighborhood exploration process

Two-sided tree:

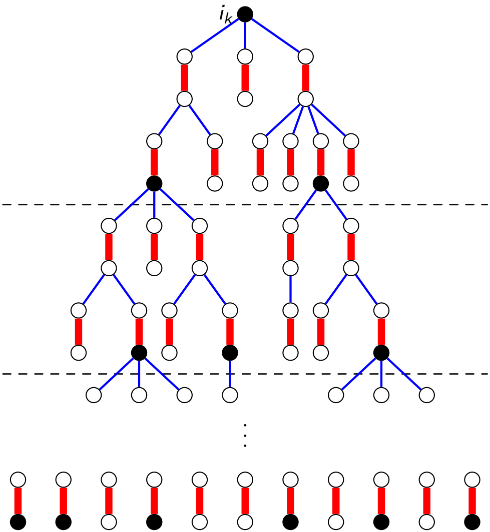


Starting from i_k, grow a tree, remove the inspected vertices, and then grow another tree from i'_k

Key challenge

How to ensure large L_k, R_k connected via augmenting alternating paths. while not using up too many vertices?

Exploration + selection



- Explore via BFS in epochs, each epoch has H steps
- At the end of each epoch, select leaves whose paths to root are augmenting and continue growing
- Behaves as a branching process with average number of offsprings $\left(dB^2(\mathcal{P}, \mathcal{Q})\right)^H > 1$

Tight error lower bound under exponential model

- Recall exponential model: $d = n$, $\mathcal{P} = \exp(\lambda)$, $\mathcal{Q} = \exp(1/n)$
- Follow the two-stage cycle finding scheme
- However, the tree-based path construction is too wasteful (construct a fat tree, but ultimately uses one path)

Tight error lower bound under exponential model

- Recall exponential model: $d = n$, $\mathcal{P} = \exp(\lambda)$, $\mathcal{Q} = \exp(1/n)$
- Follow the two-stage cycle finding scheme
- However, the tree-based path construction is too wasteful (construct a fat tree, but ultimately uses one path)

Improved Path construction (exponential model)

1. Directly show the existence of many short augmenting alternating paths using first and second moment method
2. Extract a large collection of vertex-disjoint paths via Turán's Theorem

Follow the program in [Ding-Goswami '15] in a different context

First and second moment under exponential model

- Log-likelihood weight $\log \frac{\mathcal{P}}{\mathcal{Q}}(W_e)$ is scale-and-shift of $-W_e$:

$$\text{Alternating path } P \text{ is augmenting } \Leftrightarrow \mathsf{wt_r}(P) \geq \mathsf{wt_b}(P)$$

First and second moment under exponential model

- Log-likelihood weight $\log \frac{\mathcal{P}}{\mathcal{Q}}(W_e)$ is scale-and-shift of $-W_e$:

$$\text{Alternating path } P \text{ is augmenting } \Leftrightarrow \mathsf{wt}_{\mathsf{r}}(P) \geq \mathsf{wt}_{\mathsf{b}}(P)$$

- Separately control the total red and blue edge weights of P:

$$\mathsf{wt}_{\mathsf{r}}(P) \approx \frac{2}{\lambda} \cdot |\mathsf{r}(P)|, \quad \mathsf{wt}_{\mathsf{b}}(P) \approx \frac{2-\epsilon}{\lambda} \cdot |\mathsf{b}(P)|$$

First and second moment under exponential model

- Log-likelihood weight $\log \frac{\mathcal{P}}{\mathcal{Q}}(W_e)$ is scale-and-shift of $-W_e$:

$$\text{Alternating path } P \text{ is augmenting } \Leftrightarrow \mathsf{wt_r}(P) \geq \mathsf{wt_b}(P)$$

- Separately control the total red and blue edge weights of P:

$$\mathsf{wt_r}(P) \approx \frac{2}{\lambda} \cdot |\mathsf{r}(P)|, \quad \mathsf{wt_b}(P) \approx \frac{2-\epsilon}{\lambda} \cdot |\mathsf{b}(P)|$$

- Further need *uniformity* [Ding '13, Ding-Sun-Wilson'15] to reduce correlations among different P:

$$\text{Deviation of } \mathsf{wt_r}(Q) \text{ and } \mathsf{wt_b}(Q) \text{ in every subpath } Q \text{ is } O\left(\frac{1}{\sqrt{\epsilon}}\right)$$

First and second moment under exponential model

- Log-likelihood weight $\log \frac{\mathcal{P}}{\mathcal{Q}}(W_e)$ is scale-and-shift of $-W_e$:

$$\text{Alternating path } P \text{ is augmenting } \Leftrightarrow \mathsf{wt}_{\mathsf{r}}(P) \geq \mathsf{wt}_{\mathsf{b}}(P)$$

- Separately control the total red and blue edge weights of P:

$$\mathsf{wt}_{\mathsf{r}}(P) \approx \frac{2}{\lambda} \cdot |\mathsf{r}(P)|, \quad \mathsf{wt}_{\mathsf{b}}(P) \approx \frac{2-\epsilon}{\lambda} \cdot |\mathsf{b}(P)|$$

- Further need *uniformity* [Ding '13, Ding-Sun-Wilson'15] to reduce correlations among different P:

$$\text{Deviation of } \mathsf{wt}_{\mathsf{r}}(Q) \text{ and } \mathsf{wt}_{\mathsf{b}}(Q) \text{ in every subpath } Q \text{ is } O\left(\frac{1}{\sqrt{\epsilon}}\right)$$

- Let S_ℓ denote the set of such alternating paths of length ℓ:

$$\mathsf{Var}(|S_\ell|) \leq (\mathbb{E}[|S_\ell|])^2 \, \frac{\ell^2 e^{\Theta(1/\sqrt{\epsilon})}}{n}$$

Extract vertex-disjoint alternating paths via Turán

- Define graph H
  - Vertex: alternating path in S_ℓ
  - Edge: if two alternating paths share common vertices
- Independent set $\Leftrightarrow$ collection of vertex-disjoint alternating paths

Extract vertex-disjoint alternating paths via Turán

- Define graph H
  - Vertex: alternating path in S_ℓ
  - Edge: if two alternating paths share common vertices
- Independent set $\Leftrightarrow$ collection of vertex-disjoint alternating paths

Turán's Theorem

Let $H = (V, E)$ be any simple graph. Then H contains an independent subset of size at least $|V|^2/(2|E| + |V|)$.

Extract vertex-disjoint alternating paths via Turán

- Define graph H
  - Vertex: alternating path in S_ℓ
  - Edge: if two alternating paths share common vertices
- Independent set $\Leftrightarrow$ collection of vertex-disjoint alternating paths

Turán's Theorem

Let $H = (V, E)$ be any simple graph. Then H contains an independent subset of size at least $|V|^2/(2|E| + |V|)$.

- Apply Turán's Theorem with $|V| \approx \mathbb{E}[|S_\ell|]$ and $|E| \approx \mathrm{Var}(|S_\ell|)$

$\Rightarrow$ There exist $\frac{n}{\ell^2 e^{\Theta(1/\sqrt{\epsilon})}}$ vertex-disjoint alternating paths of length ℓ

Extract vertex-disjoint alternating paths via Turán

- Define graph H
  - Vertex: alternating path in S_ℓ
  - Edge: if two alternating paths share common vertices
- Independent set $\Leftrightarrow$ collection of vertex-disjoint alternating paths

Turán's Theorem

Let $H = (V, E)$ be any simple graph. Then H contains an independent subset of size at least $|V|^2/(2|E| + |V|)$.

- Apply Turán's Theorem with $|V| \approx \mathbb{E}[|S_\ell|]$ and $|E| \approx \mathrm{Var}(|S_\ell|)$

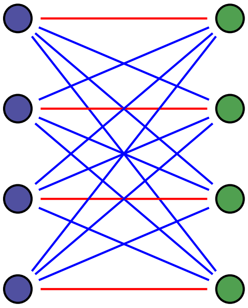
$\Rightarrow$ There exist $\frac{n}{\ell^2 e^{\Theta(1/\sqrt{\epsilon})}}$ vertex-disjoint alternating paths of length ℓ

- Choose $\ell = e^{\Theta(1/\sqrt{\epsilon})}$ and get desired augmenting alternating cycles of length $ne^{-\Theta(1/\sqrt{\epsilon})}$ via sprinkling

Analysis

- Proof of positive result via maximum likelihood
- Proof of negative result via analyzing posterior distribution
- Proof of tight error lower bound under exponential model
- Proof of overlap of MLE under exponential model

Exponential model

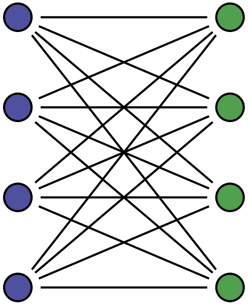


- A complete bipartite graph
- A hidden perfect matching M
- Edge weight

$$W_{ij} \overset{\text{ind.}}{\sim} \begin{cases} \text{Exp}(\lambda) & e \in M \\ \text{Exp}(1/n) & e \notin M \end{cases}$$

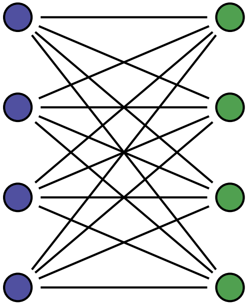
Minimum-weight matching $M_{\min}$ is the Maximum Likelihood Estimator

Warmup: the (un-planted) random assignment problem



- A complete bipartite graph
- Weights are i.i.d. $\mathrm{Exp}(1/n)$
- Cost of minimum matching?

Warmup: the (un-planted) random assignment problem



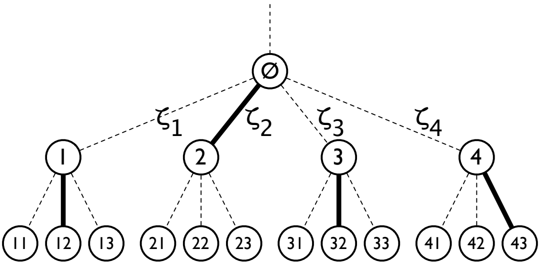
- A complete bipartite graph
- Weights are i.i.d. Exp$(1/n)$
- Cost of minimum matching?

[Walkup'79, Mézard-Parisi'87, Steele'97, Aldous'01, Nair-Prabhakar-Sharma'05, Wästlund'09]

$$\mathbb{E}\left[\min_{M \in \mathcal{M}} \frac{1}{n} \sum_{e \in M} W_e\right] = 1 + \frac{1}{4} + \frac{1}{9} + \cdots + \frac{1}{n^2} \to \frac{\pi^2}{6}$$

Poisson-weighted infinite tree approximation

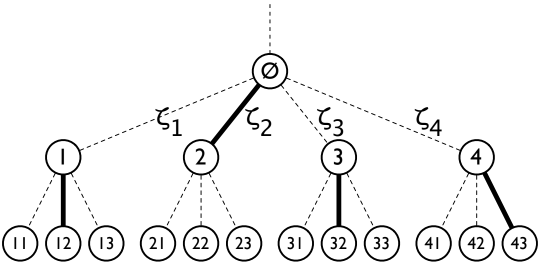
Cavity method: model as a tree [Mézard-Parisi '87, Aldous'00]



sort edge weights $W_{\varnothing,1}, W_{\varnothing,2}, \ldots$ from smallest to largest: arrivals $\zeta_1, \zeta_2, \ldots$ of a Poisson process with rate 1

Poisson-weighted infinite tree approximation

Cavity method: model as a tree [Mézard-Parisi '87, Aldous'00]

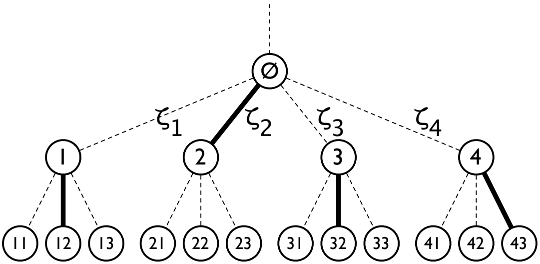


sort edge weights $W_{\varnothing,1}, W_{\varnothing,2}, \ldots$ from smallest to largest: arrivals $\zeta_1, \zeta_2, \ldots$ of a Poisson process with rate 1

$X_v \triangleq$ cost of min matching on T_v $-$ cost of min matching on $T_v \setminus \{v\}$

Poisson-weighted infinite tree approximation

Cavity method: model as a tree [Mézard-Parisi '87, Aldous'00]



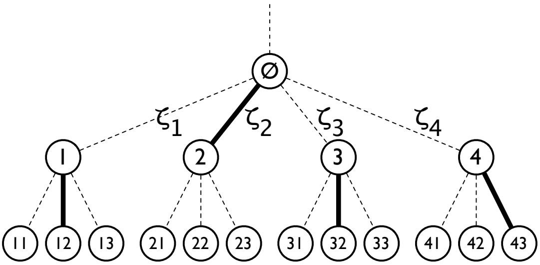
sort edge weights $W_{\varnothing,1}, W_{\varnothing,2}, \ldots$ from smallest to largest: arrivals $\zeta_1, \zeta_2, \ldots$ of a Poisson process with rate 1

$X_v \triangleq$ cost of min matching on T_v $-$ cost of min matching on $T_v \setminus \{v\}$

$$X_\varnothing = \min_{i \geq 1} \{W_{\varnothing,i} - X_i\}$$

Poisson-weighted infinite tree approximation

Cavity method: model as a tree [Mézard-Parisi '87, Aldous'00]



sort edge weights $W_{\varnothing,1}, W_{\varnothing,2}, \ldots$ from smallest to largest: arrivals $\zeta_1, \zeta_2, \ldots$ of a Poisson process with rate 1

$X_v \triangleq$ cost of min matching on T_v $-$ cost of min matching on $T_v \setminus \{v\}$

$$X_\varnothing = \min_{i \geq 1} \{W_{\varnothing,i} - X_i\} \implies X \stackrel{d}{=} \min_{i \geq 1} \{\zeta_i - X_i\}$$

From distributional to differential equations

$X \stackrel{d}{=} \min\{\zeta_i - X_i\}$ where ζ_i are Poisson arrivals

From distributional to differential equations

$X \stackrel{d}{=} \min\{\zeta_i - X_i\}$ where ζ_i are Poisson arrivals

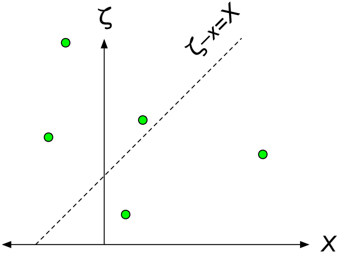
Define the ccdf $\bar{F}(x) = 1 - F(x) = \mathbb{P}[X > x] = \mathbb{P}[\forall i : \zeta_i - x > X_i]$

From distributional to differential equations

$X \stackrel{d}{=} \min\{\zeta_i - X_i\}$ where ζ_i are Poisson arrivals

Define the ccdf $\bar{F}(x) = 1 - F(x) = \mathbb{P}[X > x] = \mathbb{P}[\forall i : \zeta_i - x > X_i]$

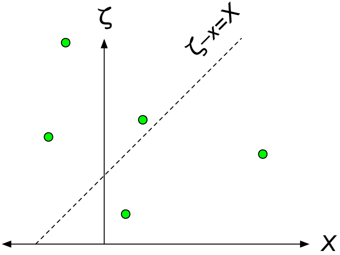
Generate pairs (ζ_i, X_i): two-dimensional Poisson process with density F'



From distributional to differential equations

$X \stackrel{d}{=} \min\{\zeta_i - X_i\}$ where ζ_i are Poisson arrivals

Define the ccdf $\bar{F}(x) = 1 - F(x) = \mathbb{P}[X > x] = \mathbb{P}[\forall i : \zeta_i - x > X_i]$
Generate pairs (ζ_i, X_i): two-dimensional Poisson process with density F'



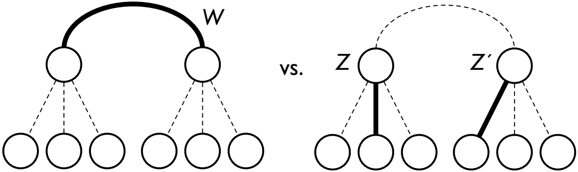
$$\bar{F}(x) = \exp\left(-\int_{-x}^{\infty} \bar{F}(t)\, dt\right) \quad \Rightarrow \quad \frac{dF(x)}{dx} = F(x)F(-x)$$

From distributional to differential equations, cont'd

$$\frac{dF(x)}{dx} = F(x)F(-x) \quad \Longrightarrow F(x) = \frac{e^x}{1+e^x}$$

From distributional to differential equations, cont'd

$$\frac{dF(x)}{dx} = F(x)F(-x) \quad \Longrightarrow F(x) = \frac{e^x}{1+e^x}$$

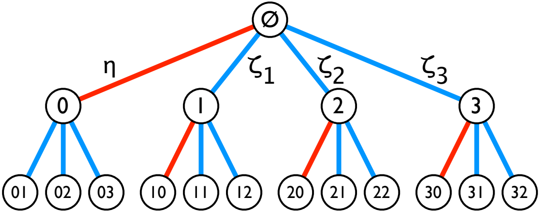


Contribution of a single edge:

$$\int_0^\infty w\mathbb{P}\left[Z + Z' \geq w\right]\mathrm{d}w = \frac{1}{4}\mathrm{Var}[Z+Z'] = \frac{1}{2}\mathrm{Var}[Z] = \frac{\pi^2}{6}$$

Planted poission-weighted infinite tree

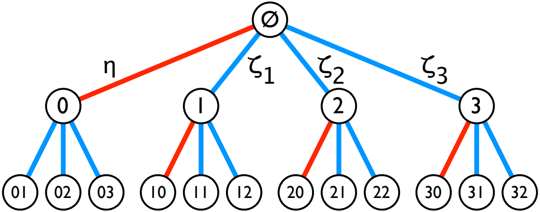
Partner in planted matching is either parent or child 0,
other children sorted $1, 2, 3, \ldots$



$X_v \triangleq$ cost of min matching in T_v $-$ cost of min matching on $T_v \setminus \{v\}$

Planted poission-weighted infinite tree

Partner in planted matching is either parent or child 0,
other children sorted $1, 2, 3, \ldots$



$X_v \triangleq$ cost of min matching in T_v $-$ cost of min matching on $T_v \setminus \{v\}$

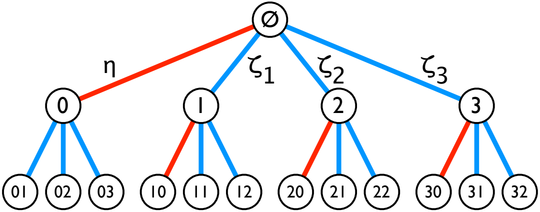
Recursion:

$$X_\varnothing = \min \left\{ W_{\varnothing,0} - X_0, \ \min_{i \geq 1} \{ W_{\varnothing,i} - X_i \} \right\}$$

$$X_0 = \min_{i \geq 1} \{ W_{0,0i} - X_{0i} \}$$

Planted poission-weighted infinite tree

Partner in planted matching is either parent or child 0, other children sorted $1, 2, 3, \ldots$



$X_v \triangleq$ cost of min matching in T_v $-$ cost of min matching on $T_v \setminus \{v\}$

Recursion:

$$X_\varnothing = \min\left\{ W_{\varnothing,0} - X_0,\ \min_{i\geq 1}\{W_{\varnothing,i} - X_i\}\right\}$$

$$X_0 = \min_{i\geq 1}\{W_{0,0i} - X_{0i}\}$$

$$Y \stackrel{d}{=} \min\{\eta - Z, Z'\}$$

$$Z \stackrel{d}{=} \min_i \{\zeta_i - Y_i\}$$

From distributional to differential equations, redux

$$Y \overset{d}{=} \min\left\{\eta - Z, Z'\right\}$$
$$Z \overset{d}{=} \min\left\{\zeta_i - Y_i\right\}_{i=1}^{\infty}$$

where $\eta \sim \text{Exp}(\lambda)$ and ζ_i are Poisson arrivals

From distributional to differential equations, redux

$$Y \stackrel{d}{=} \min\left\{\eta - Z, Z'\right\}$$
$$Z \stackrel{d}{=} \min\left\{\zeta_i - Y_i\right\}_{i=1}^{\infty}$$

where $\eta \sim \text{Exp}(\lambda)$ and ζ_i are Poisson arrivals

$F(x) = \mathbb{P}\left[Z < x\right], G(x) = F(-x), V(x) = \mathbb{E}[F(x+\eta)], W(x) = V(-x)$

From distributional to differential equations, redux

$$\boxed{\begin{aligned} Y &\overset{d}{=} \min\left\{\eta - Z, Z'\right\} \\ Z &\overset{d}{=} \min\left\{\zeta_i - Y_i\right\}_{i=1}^{\infty} \end{aligned}}$$

where $\eta \sim \text{Exp}(\lambda)$ and ζ_i are Poisson arrivals

$F(x) = \mathbb{P}\left[Z < x\right]$, $G(x) = F(-x)$, $V(x) = \mathbb{E}[F(x+\eta)]$, $W(x) = V(-x)$

$$\boxed{\begin{aligned} \dot{F} &= (1-F)(1-G)V \\ \dot{G} &= -(1-F)(1-G)W \\ \dot{V} &= \lambda(V-F) \\ \dot{W} &= -\lambda(W-G) \end{aligned}}$$

From distributional to differential equations, redux

$$
\boxed{\begin{aligned} Y &\stackrel{d}{=} \min\{\eta - Z, Z'\} \\ Z &\stackrel{d}{=} \min\{\zeta_i - Y_i\}_{i=1}^{\infty} \end{aligned}}
$$

where $\eta \sim \mathrm{Exp}(\lambda)$ and ζ_i are Poisson arrivals

$F(x) = \mathbb{P}[Z < x]$, $G(x) = F(-x)$, $V(x) = \mathbb{E}[F(x + \eta)]$, $W(x) = V(-x)$

$$
\boxed{\begin{aligned} \dot{F} &= (1 - F)(1 - G)V \\ \dot{G} &= -(1 - F)(1 - G)W \\ \dot{V} &= \lambda(V - F) \\ \dot{W} &= -\lambda(W - G) \end{aligned}}
$$

$\dot{V}$ and $\dot{W}$ from $\eta \sim \mathrm{Exp}(\lambda)$, integration by parts

From distributional to differential equations, redux

$$
\boxed{\begin{aligned} Y &\overset{d}{=} \min\{\eta - Z, Z'\} \\ Z &\overset{d}{=} \min\{\zeta_i - Y_i\}_{i=1}^{\infty} \end{aligned}}
$$

where $\eta \sim \text{Exp}(\lambda)$ and ζ_i are Poisson arrivals

$F(x) = \mathbb{P}[Z < x]$, $G(x) = F(-x)$, $V(x) = \mathbb{E}[F(x+\eta)]$, $W(x) = V(-x)$

$$
\boxed{\begin{aligned} \dot{F} &= (1-F)(1-G)V \\ \dot{G} &= -(1-F)(1-G)W \\ \dot{V} &= \lambda(V-F) \\ \dot{W} &= -\lambda(W-G) \end{aligned}}
$$

$\dot{V}$ and $\dot{W}$ from $\eta \sim \text{Exp}(\lambda)$, integration by parts

Boundary conditions: $F(x), V(x), G(-x), W(-x) \to \begin{cases} 1 & x \to +\infty \\ 0 & x \to -\infty \end{cases}$

Phase transition of ODE at $\lambda = 4$

$$\boxed{\begin{aligned} \dot{F} &= (1-F)(1-G)V \\ \dot{G} &= -(1-F)(1-G)W \\ \dot{V} &= \lambda(V-F) \\ \dot{W} &= -\lambda(W-G) \end{aligned}}$$

Boundary conditions: $F(x), V(x), G(-x), W(-x) \to \begin{cases} 1 & x \to +\infty \\ 0 & x \to -\infty \end{cases}$

Lemma

There is a unique solution if and only if $\lambda < 4$.

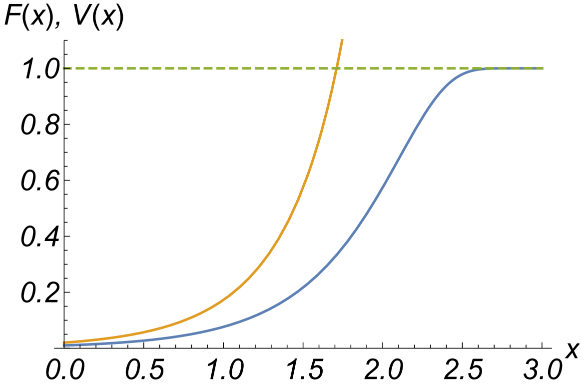
No solution for $\lambda \geq 4$

At least no sensible one. . .

No solution for $\lambda \geq 4$

At least no sensible one...

Want $F(+\infty) = V(+\infty) = 1$. But...



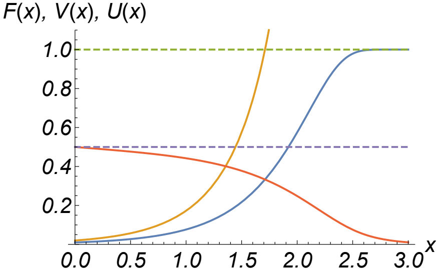
No solution for $\lambda \geq 4$

Conservation law: $FW + GV - VW = 0 \;\Rightarrow\; V(0) = 2F(0)$

No solution for $\lambda \geq 4$

Conservation law: $FW + GV - VW = 0 \Rightarrow V(0) = 2F(0)$

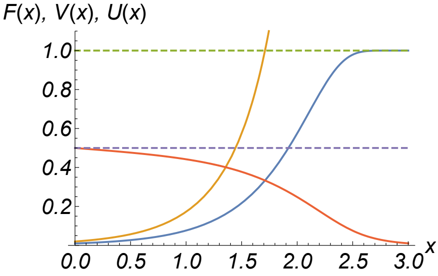
Let $U(x) = F(x)/V(x)$. Then $U(0) = 1/2$, want $U(+\infty) = 1\ldots$



No solution for $\lambda \geq 4$

Conservation law: $FW + GV - VW = 0 \Rightarrow V(0) = 2F(0)$

Let $U(x) = F(x)/V(x)$. Then $U(0) = 1/2$, want $U(+\infty) = 1\ldots$

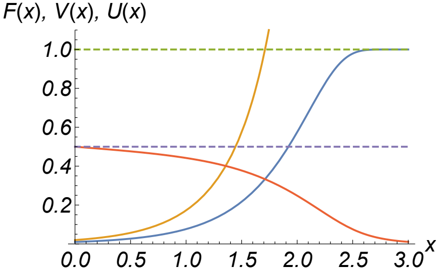


$$\dot{U} = -\lambda U(1-U) + (1-F)(1-G) \leq -\lambda U(1-U) + 1$$

No solution for $\lambda \geq 4$

Conservation law: $FW + GV - VW = 0 \Rightarrow V(0) = 2F(0)$

Let $U(x) = F(x)/V(x)$. Then $U(0) = 1/2$, want $U(+\infty) = 1 \ldots$



$$\dot{U} = -\lambda U(1-U) + (1-F)(1-G) \leq -\lambda U(1-U) + 1$$

If $\lambda \geq 4$, $\dot{U}(1/2) \leq 0 \implies U(x)$ can never exceed $1/2$

A unique solution when $\lambda < 4$

$(F, G, V, W) \Longleftrightarrow (U, V, W)$: three-dimensional dynamical system

$$\boxed{\begin{aligned} &\dot{U} = -\lambda U(1-U) + (1-UV)\left(1-(1-U)W\right) \\ &\dot{V} = \lambda V(1-U) \\ &\dot{W} = \lambda WU \\ &\text{Initial conditions: } U(0) = \frac{1}{2}, V(0) = W(0) = \delta \end{aligned}}$$

A unique solution when $\lambda < 4$

$(F, G, V, W) \Longleftrightarrow (U, V, W)$: three-dimensional dynamical system

$$
\boxed{
\begin{aligned}
&\dot{U} = -\lambda U(1-U) + (1-UV)\,(1-(1-U)W) \\
&\dot{V} = \lambda V(1-U) \\
&\dot{W} = \lambda WU \\
&\text{Initial conditions: } U(0) = \frac{1}{2}, V(0) = W(0) = \delta
\end{aligned}
}
$$

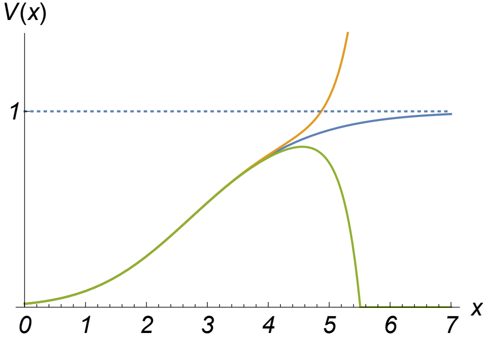
Lemma (Moharrami-Moore-X. '19)

If $\lambda < 4$, there is a unique $\delta_0 \in (0, 1)$ such that

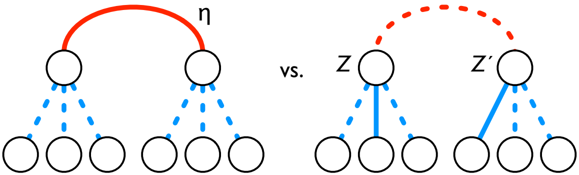
- *If $\delta \in [0, \delta_0)$, $U(x) \to +\infty$*
- *If $\delta = \delta_0$, $U(x) \to 1$ and $V(x) \to 1$*
- *If $\delta \in (\delta_0, 1]$, $V(x) \to +\infty$*

Geometric interpretation of uniqueness

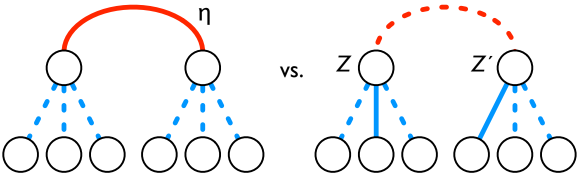
When $\lambda < 4$, $(F = 1, G = 0, V = 1, W = 0)$ is a saddle point:
There exists a unique initial condition from which we approach the saddle along its unstable manifold



Finally, computing the overlap for $\lambda < 4$



Finally, computing the overlap for $\lambda < 4$



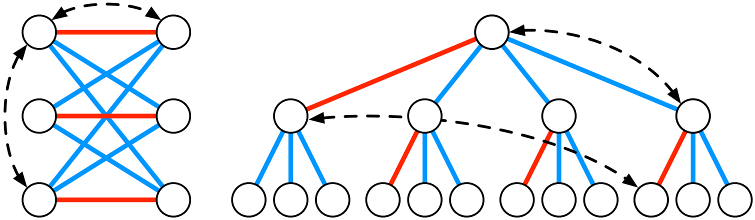
$$\alpha(\lambda) = \mathbb{P}\left[\eta < Z + Z'\right] = 1 - \mathbb{E}_\eta \int_{-\infty}^{+\infty} f(x) F(\eta - x)\, dx$$

$$= 1 - \int_{-\infty}^{+\infty} f(x)\, \mathbb{E}_\eta F(\eta - x)\, dx$$

$$= 1 - \int_{-\infty}^{+\infty} (1 - F(x))(1 - G(x)) V(x) W(x)\, dx$$

$$= 1 - 2\int_{0}^{+\infty} (1 - F(x))(1 - G(x)) V(x) W(x)\, dx$$

Proving it: Local weak convergence (Aldous 1992, 2001)



- Construct a *spatially invariant* M_{opt} on T_∞ using message passing
- Show $(K_{n,n}, M_{\mathsf{min}})$ converges locally to $(T_\infty, M_{\mathrm{opt}})$
  - Local treelikeness of light edges
  - Almost-doubly-stochastic matrix

Conclusion

- Sharp threshold for almost perfect recovery: $\sqrt{d}\, B(\mathcal{P}, \mathcal{Q}) = 1$
- Infinite-order phase transition under the exponential model: Optimal reconstruction error is $\exp\left(-\Theta(1/\sqrt{\epsilon})\right)$ when $\lambda = 4 - \epsilon$
- Key idea: two-stage cycle finding (path construction + sprinkling)
- Characterization of overlap of MLE by system of ODEs

Conclusion

- Sharp threshold for almost perfect recovery: $\sqrt{d}\, B(\mathcal{P}, \mathcal{Q}) = 1$
- Infinite-order phase transition under the exponential model: Optimal reconstruction error is $\exp\left(-\Theta(1/\sqrt{\epsilon})\right)$ when $\lambda = 4 - \epsilon$
- Key idea: two-stage cycle finding (path construction + sprinkling)
- Characterization of overlap of MLE by system of ODEs

Reference

- M. Moharrami, C. Moore, & J. Xu, *The planted matching problem: Phase transitions and exact results*. *Annals of Applied Probability*, 2021.
- J. Ding, Y. Wu, J. Xu, & D. Yang, *The planted matching problem: Sharp threshold and infinite-order phase transition*. arXiv:2103.09383.

Open problems

1. Optimal error for general distributions? in entire parameter range? Interpolation method [Coja-Oghlan-Krzakala-Perkins-Zdeborová '18]?
2. Extension to planted k-factor model
   Conjecture: $\sqrt{kd}\, B(\mathcal{P}, \mathcal{Q}) = 1$ [Sicuro-Zdeborová '20]
3. Extension to k-hypergraphs [Adomaityte-Toshniwal-Sicuro-Zdeborová '22]
   Observe first-order phase transition when $k > 2$
4. Finite-dimensional Euclidean space? [Kunisky-Niles-Weed '22]
5. Planted feature matching [Dai-Cullina-Kiyavash '19, Wang-Wu-X.-Yolou '22]
6. Other planted structures: spanning trees, traveling salespeople [Bagaria-Ding-Tse-Wu-X. '18]?