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)

October 17, 2022
INFORMS Annual Meeting
Motivating application: particle tracking

- Tracking particles advected by turbulent fluid flow
- **Goal**: recover the latent correspondence between particles

[Chertkov-Kroc-Krzakala-Vergassola-Zdeborová PNAS'10]
The planted matching model

- A complete bipartite graph
- A hidden perfect matching $M^*$
- Edge weight

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

Goal: recover $M^*$ from $W$.

Our work: $P = \text{Exp}(\lambda)$, $Q = \text{Exp}(\frac{1}{n})$ (mean $1/\lambda$ vs. $n$).

Minimum-weight matching is Maximum Likelihood Estimator $\hat{M}_{\text{ML}}$.

How much does $\hat{M}_{\text{ML}}$ have in common with $M^*$?
The planted matching model

- A complete bipartite graph
- A hidden perfect matching $M^*$
- Edge weight

\[ W_e \sim_{\text{ind.}} \begin{cases} P & e \in M^* \\ Q & e \notin M^* \end{cases} \]

- Goal: recover $M^*$ from $W$
The planted matching model

- A complete bipartite graph
- A hidden perfect matching $M^*$
- Edge weight $W_e \overset{\text{ind.}}{\sim} \begin{cases} P & e \in M^* \\ Q & e \notin M^* \end{cases}$
- Goal: recover $M^*$ from $W$

- Our work: $P = \text{Exp}(\lambda), Q = \text{Exp}(1/n)$ (mean $1/\lambda$ vs. $n$)
The planted matching model

- A complete bipartite graph
- A hidden perfect matching $M^*$
- Edge weight

\[ W_e \sim \begin{cases} P & e \in M^* \\ Q & e \notin M^* \end{cases} \]

- Goal: recover $M^*$ from $W$

- Our work: $P = \text{Exp} (\lambda), Q = \text{Exp} (1/n)$ (mean $1/\lambda$ vs. $n$)
- Minimum-weight matching is Maximum Likelihood Estimator $\hat{M}_{\text{ML}}$
The planted matching model

- A complete bipartite graph
- A hidden perfect matching $M^*$
- Edge weight $w_e \sim \begin{cases} P & e \in M^* \\ Q & e \notin M^* \end{cases}$
- Goal: recover $M^*$ from $W$

- Our work: $P = \text{Exp}(\lambda), Q = \text{Exp}(1/n)$ (mean $1/\lambda$ vs. $n$)
- Minimum-weight matching is Maximum Likelihood Estimator $\hat{M}_{ML}$
- How much does $\hat{M}_{ML}$ have in common with $M^*$?
Main result: phase transition at $\lambda = 4$

Theorem (Moharrami-Moore-X. AAP’21)

overlap: $\lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ \left| \hat{M}_{ML} \cap M^* \right| \right] = \begin{cases} 1 & \text{if } \lambda \geq 4 \\ \alpha(\lambda) & \text{if } 0 < \lambda < 4 \end{cases}$

where $\alpha(\lambda) = 1 - 2 \int_0^\infty (1 - F(x)) (1 - G(x)) V(x)W(x) \, dx$, 
Main result: phase transition at $\lambda = 4$

Theorem (Moharrami-Moore-X. AAP’21)

$$\text{overlap: } \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ |\hat{M}_{\text{ML}} \cap M^*| \right] = \begin{cases} 1 & \text{if } \lambda \geq 4 \\ \alpha(\lambda) & \text{if } 0 < \lambda < 4 \end{cases}$$

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

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

Boundary conditions: $F(x), V(x), G(-x), W(-x) \to \begin{cases} 1 & x \to +\infty \\ 0 & x \to -\infty \end{cases}$
Main result: phase transition at $\lambda = 4$

**Theorem (Moharrami-Moore-X. AAP’21)**

\[
\text{overlap: } \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ |\hat{M}_{ML} \cap M^*| \right] = \begin{cases} 
1 & \text{if } \lambda \geq 4 \\
\alpha(\lambda) & \text{if } 0 < \lambda < 4 
\end{cases}
\]

where $\alpha(\lambda) = 1 - 2 \int_0^\infty (1 - F(x)) (1 - G(x)) V(x) W(x) \, dx,$
Main result: phase transition at $\lambda = 4$

Theorem (Moharrami-Moore-X. AAP’21)

\[
\text{overlap: } \lim_{n \to \infty} \frac{1}{n} \mathbb{E}\left[|\hat{M}_{ML} \cap M^*|\right] = \begin{cases} 
1 & \text{if } \lambda \geq 4 \\
\alpha(\lambda) & \text{if } 0 < \lambda < 4
\end{cases}
\]

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

\[
\text{overlap } \alpha(\lambda)
\]
When $\lambda \geq 4$: count augmenting cycles

- Probability that an alternating cycle of length $\ell$ is augmenting:
  $$P[Erlang(\ell, \lambda) \geq Erlang(\ell, 1/n)] \leq (\lambda n 4^{-\ell} - \ell)$$

- There are $\binom{n \ell}{\ell} \leq n^{\ell} e^{-\ell^2/2n}$ alternating cycles of length $\ell$

- Expected number of such augmenting cycles is at most $(\lambda/4) - \ell e^{-\ell^2/2n}$

- Sum over $\ell$:
  $$\mathbb{E}[\left|\hat{M}_{ML} - \Delta M^*\right|] = o(n)$$
When $\lambda \geq 4$: count augmenting cycles

- Probability that an alternating cycle of length $2\ell$ is augmenting:

$$\mathbb{P} \left[ \text{Erlang}(\ell, \lambda) \geq \text{Erlang}(\ell, 1/n) \right] \leq \left( \frac{\lambda n}{4} \right)^{-\ell}$$
When $\lambda \geq 4$: count augmenting cycles

- Probability that an alternating cycle of length $2\ell$ is augmenting:

$$\mathbb{P}[\text{Erlang}(\ell, \lambda) \geq \text{Erlang}(\ell, 1/n)] \leq \left(\frac{\lambda n}{4}\right)^{-\ell}$$

- There are $\binom{n}{\ell} \ell! \leq n^\ell e^{-\ell^2/2n}$ alternating cycles of length $2\ell$
When $\lambda \geq 4$: count augmenting cycles

- Probability that an alternating cycle of length $2\ell$ is augmenting:
  \[
  \mathbb{P} \left[ \text{Erlang}(\ell, \lambda) \geq \text{Erlang}(\ell, 1/n) \right] \leq \left( \frac{\lambda n}{4} \right)^{-\ell}
  \]

- There are \( \binom{n}{\ell} \ell! \leq n^\ell e^{-\ell^2/2n} \) alternating cycles of length $2\ell$

  \[\Rightarrow\] Expected $#$ of such augmenting cycles is at most \( (\lambda/4)^{-\ell} e^{-\ell^2/2n} \)
When $\lambda \geq 4$: count augmenting cycles

- Probability that an alternating cycle of length $2\ell$ is augmenting:
  \[
  \mathbb{P}[\text{Erlang}(\ell, \lambda) \geq \text{Erlang}(\ell, 1/n)] \leq \left(\frac{\lambda n}{4}\right)^{-\ell}
  \]

- There are $\binom{n}{\ell} \ell! \leq n^\ell e^{-\ell^2/2n}$ alternating cycles of length $2\ell$

$\Rightarrow$ Expected # of such augmenting cycles is at most $(\lambda/4)^{-\ell} e^{-\ell^2/2n}$

$\Rightarrow$ Sum over $\ell$ $\Rightarrow$ $\mathbb{E}\left[|\hat{M}_{\text{ML}} \Delta \hat{M}^*|\right] = o(n)$
Warmup: the (un-planted) random assignment problem

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

\[ \min_{M \in \mathcal{M}} \frac{1}{n} \sum_{e \in M} W_e = 1 + \frac{1}{4} + \frac{1}{9} + \cdots + \frac{1}{n^2} \to \frac{\pi^2}{6} \]
Warmup: the (un-planted) random assignment problem

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

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

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

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

\[
\begin{align*}
X_\emptyset & = \min_{i \geq 1} \{ W_\emptyset, i - X_i \} \\
X_d & = \min_{i \geq 1} \{ \zeta_i - X_i \}
\end{align*}
\]

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

$X_v \triangleq \text{cost of min matching on } T_v - \text{cost of min matching on } T_v \setminus \{v\}$
Poisson-weighted infinite tree approximation

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

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

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

$X_\emptyset = \min_{i \geq 1} \{W_\emptyset,i - X_i\}$
Poisson-weighted infinite tree approximation

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

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

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

$X_{\emptyset} = \min_{i \geq 1} \{W_{\emptyset,i} - X_i\} \implies X \overset{d}{=} \min_{i \geq 1} \{\zeta_i - X_i\}$
From distributional to differential equations

\[ X \overset{d}{=} \min \{ \zeta_i - X_i \} \text{ where } \zeta_i \text{ are Poisson arrivals} \]
From distributional to differential equations

\[ X \overset{d}{=} \min \{ \zeta_i - X_i \} \text{ where } \zeta_i \text{ are Poisson arrivals} \]

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

\[ X \overset{d}{=} \min \{ \zeta_i - X_i \} \text{ where } \zeta_i \text{ are Poisson arrivals} \]

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

Generate pairs \((\zeta_i, X_i)\): two-dimensional Poisson process with density \( F' \)
From distributional to differential equations

\[ X \overset{d}{=} \min \{ \zeta_i - X_i \} \text{ where } \zeta_i \text{ are Poisson arrivals} \]

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

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

\[
\bar{F}(x) = \exp\left( - \int_{-x}^{\infty} \bar{F}(t) \, dt \right) \Rightarrow \frac{dF(x)}{dx} = F(x)F(-x)
\]
\[ \frac{dF(x)}{dx} = F(x)F(-x) \quad \Rightarrow \quad F(x) = \frac{e^x}{1 + e^x} \]
From distributional to differential equations, cont’d

\[
\frac{dF(x)}{dx} = F(x)F(-x) \implies F(x) = \frac{e^x}{1 + e^x}
\]

Contribution of a single edge:

\[
\int_0^\infty w \mathbb{P}[Z + Z' \geq w] dw = \frac{1}{4} \text{Var}[Z + Z'] = \frac{1}{2} \text{Var}[Z] = \frac{\pi^2}{6}
\]
Planted Poisson-weighted infinite tree

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

\[ X_v \triangleq \text{cost of min matching in } T_v - \text{cost of min matching on } T_v \setminus \{v\} \]
Partner in planted matching is either parent or child 0, other children sorted 1, 2, 3, …

\[ x_v \triangleq \text{cost of min matching in } T_v - \text{cost of min matching on } T_v \setminus \{v\} \]

**Recursion:**

\[ x_\emptyset = \min \left\{ W_\emptyset,0 - x_0, \min_{i \geq 1} \{ W_\emptyset,i - x_i \} \right\} \]

\[ x_0 = \min_{i \geq 1} \{ W_{0,0i} - x_{0i} \} \]
Planted Poisson-weighted infinite tree

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

\[ X_v \triangleq \text{cost of min matching in } T_v - \text{cost of min matching on } T_v \setminus \{v\} \]

Recursion:

\[ X_{\emptyset} = \min \left\{ W_{\emptyset,0} - X_0, \min_{i \geq 1} \{ W_{\emptyset,i} - X_i \} \right\} \]

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

\[ Y \overset{d}{=} \min \{ \eta - Z, Z' \} \]

\[ Z \overset{d}{=} \min_i \{ \zeta_i - Y_i \} \]
From distributional to differential equations, redux

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

where \( \eta \sim \text{Exp}(\lambda) \) and \( \zeta_i \) are Poisson arrivals
From distributional to differential equations, redux

\[ Y \overset{d}{=} \min \{ \eta - Z, Z' \} \]

\[ Z \overset{d}{=} \min \{ \zeta_i - Y_i \}_{i=1}^{\infty} \]

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

\[ F(x) = \mathbb{P}[Z < x], \quad G(x) = F(-x), \quad V(x) = \mathbb{E}[F(x + \eta)], \quad W(x) = V(-x) \]
From distributional to differential equations, redux

\[
Y_d \triangleq \min \{ \eta - Z, Z' \} \\
Z_d \triangleq \min \{ \zeta_i - Y_i \} \quad \forall i = 1
\]

where \( \eta \sim \text{Exp}(\lambda) \) and \( \zeta_i \) are Poisson arrivals.

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

\[
\dot{F} = (1 - F)(1 - G)V \\
\dot{G} = -(1 - F)(1 - G)W \\
\dot{V} = \lambda(V - F) \\
\dot{W} = -\lambda(W - G)
\]
From distributional to differential equations, redux

\[ Y \overset{d}{=} \min \{ \eta - Z, Z' \} \]

\[ Z \overset{d}{=} \min \{ \zeta_i - Y_i \}_{i=1}^{\infty} \]

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

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

\[ \dot{F} = (1 - F)(1 - G)V \]

\[ \dot{G} = -(1 - F)(1 - G)W \]

\[ \dot{V} = \lambda(V - F) \]

\[ \dot{W} = -\lambda(W - G) \]

\( \dot{V} \) and \( \dot{W} \) from \( \eta \sim \text{Exp}(\lambda) \), integration by parts
From distributional to differential equations, redux

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

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

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

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

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

Boundary conditions: \( F(x), V(x), G(-x), W(-x) \rightarrow \begin{cases} 1 & x \to +\infty \\ 0 & x \to -\infty \end{cases} \)
Phase transition of ODE at $\lambda = 4$

\[ \begin{align*}
\dot{F} &= (1 - F)(1 - G)V \\
\dot{G} &= -(1 - F)(1 - G)W \\
\dot{V} &= \lambda(V - F) \\
\dot{W} &= -\lambda(W - G)
\end{align*} \]

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

Lemma

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

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

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

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

$$= 1 - 2 \int_{0}^{+\infty} (1 - F(x))(1 - G(x)) V(x) W(x) \, dx$$
• Construct a \textit{spatially invariant} $M_{\text{opt}}$ on $T_\infty$ using message passing

• Show $(K_{n,n}, M_{\text{min}})$ converges locally to $(T_\infty, M_{\text{opt}})$
  ▶ Local treelikeness of light edges
  ▶ Almost-doubly-stochastic matrix
Conclusion

- Sharp threshold for almost perfect recovery: $\lambda = 4$
- Exact characterization of overlap of MLE by system of ODEs
Conclusion

• Sharp threshold for almost perfect recovery: $\lambda = 4$
• Exact characterization of overlap of MLE by system of ODEs

Extensions

• $\lambda = 4$ is the sharp information-theoretic threshold [Ding-Wu-X.-Yang '21]:
  
  $\lambda = 4 - \epsilon$: Optimal reconstruction error is $\exp(-\Theta(1/\sqrt{\epsilon}))$

• Sharp threshold in general $(d, P, Q)$ model [Ding-Wu-X.-Yang '21]:
  
  $\sqrt{d} \int \sqrt{dPdQ} = 1$

• Non-IID spatial model [Kunisky-Niles-Weed '22, Wang-Wu-X.-Yolou '22]
Conclusion

- Sharp threshold for almost perfect recovery: $\lambda = 4$
- Exact characterization of overlap of MLE by system of ODEs

Extensions

- $\lambda = 4$ is the sharp information-theoretic threshold [Ding-Wu-X.-Yang '21]:
  
  $$\lambda = 4 - \epsilon: \text{Optimal reconstruction error is } \exp \left( -\Theta \left( 1/\sqrt{\epsilon} \right) \right)$$

- Sharp threshold in general $(d, P, Q)$ model [Ding-Wu-X.-Yang '21]:
  $$\sqrt{d} \int \sqrt{dPdQ} = 1$$

- Non-IID spatial model [Kunisky-Niles-Weed '22, Wang-Wu-X.-Yolou '22]

Reference