The Planted Matching Problem

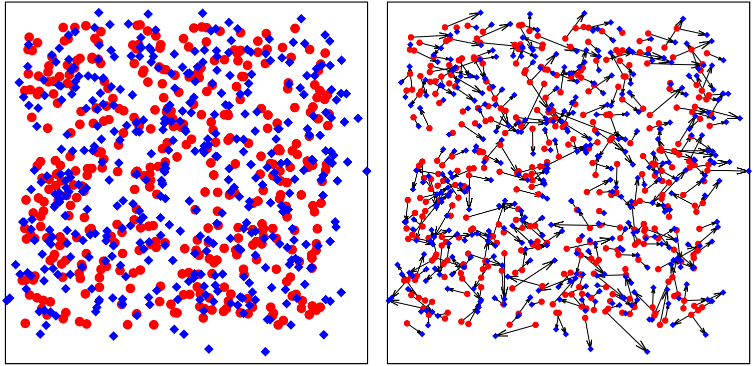
Jiaming Xu

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

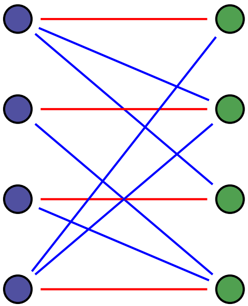
Motivating application: particle tracking



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

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

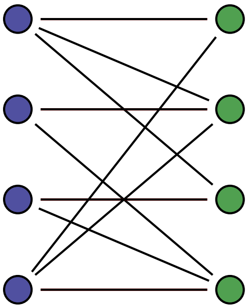
The planted matching model



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

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

The planted matching model

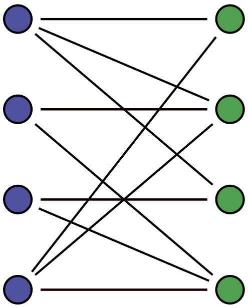


- A complete bipartite graph
- A hidden perfect matching M^*
- Edge weight

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

- Goal: recover M^* from W

The planted matching model



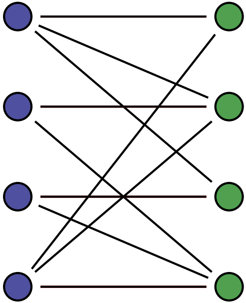
- A complete bipartite graph
- A hidden perfect matching M^*
- Edge weight

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

- 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



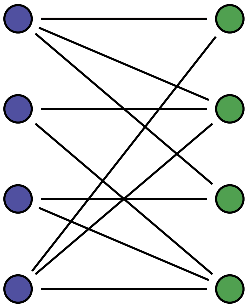
- A complete bipartite graph
- A hidden perfect matching M^*
- Edge weight

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

- Goal: recover M^* from W

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

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)
- Minimum-weight matching is Maximum Likelihood Estimator $\widehat{M}_{\rm ML}$
- How much does $\widehat{M}_{\rm ML}$ have in common with M^*?

Main result: phase transition at $\lambda = 4$

Theorem (Moharrami-Moore-X. AAP'21)

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

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)

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

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

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

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

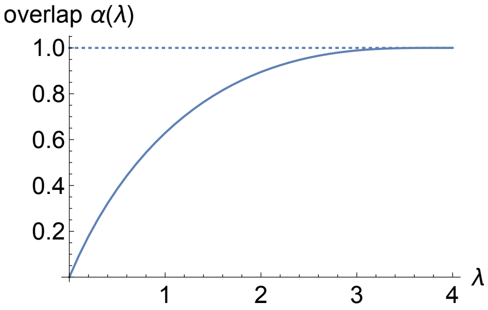
Boundary conditions: $F(x), V(x), G(-x), W(-x) \to \begin{cases} 1 & x \to +\infty \\ 0 & x \to -\infty \end{cases}$

Main result: phase transition at $\lambda = 4$

Theorem (Moharrami-Moore-X. AAP'21)

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

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

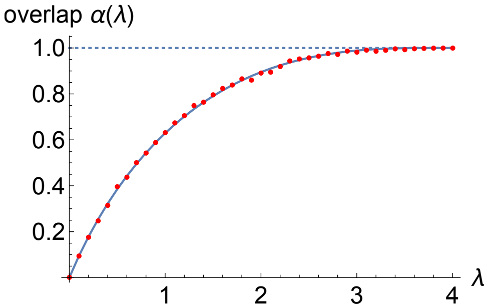


Main result: phase transition at $\lambda = 4$

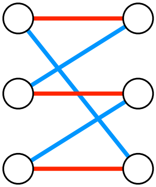
Theorem (Moharrami-Moore-X. AAP'21)

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

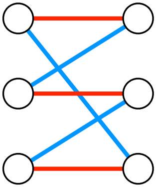
where $\alpha(\lambda) = 1 - 2\int_0^\infty (1 - F(x))\,(1 - G(x))\,V(x)W(x)\,dx,$



When $\lambda \geq 4$: count augmenting cycles



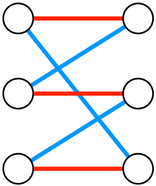
When $\lambda \geq 4$: count augmenting cycles



- Probability that an alternating cycle of length 2ℓ is augmenting:

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

When $\lambda \geq 4$: count augmenting cycles

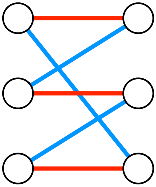


- Probability that an alternating cycle of length 2ℓ is augmenting:

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

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

When $\lambda \geq 4$: count augmenting cycles



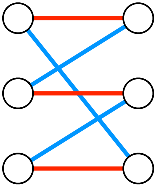
- Probability that an alternating cycle of length 2ℓ is augmenting:

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

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

$\Rightarrow$ 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 2ℓ is augmenting:

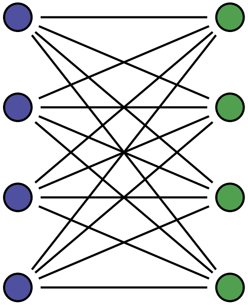
$$\mathbb{P}\left[\mathrm{Erlang}(\ell, \lambda) \geq \mathrm{Erlang}(\ell, 1/n)\right] \leq \left(\frac{\lambda n}{4}\right)^{-\ell}$$

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

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

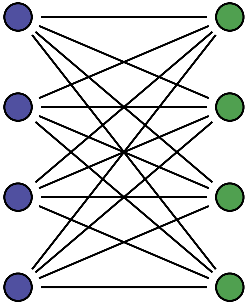
$\Rightarrow$ Sum over $\ell \Longrightarrow \mathbb{E}\left[\left|\widehat{M}_{\mathrm{ML}} \triangle M^*\right|\right] = o(n)$

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



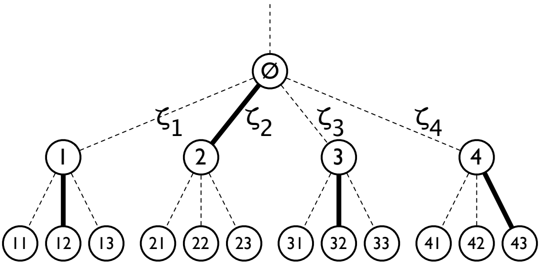
- A complete bipartite graph
- Weights are i.i.d. $\text{Exp}(1/n)$
- Cost of minimum matching?

[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

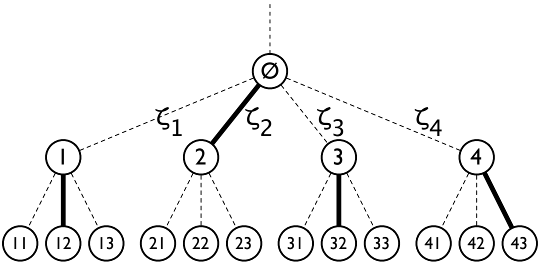
Cavity method: model as a tree [Mézard-Parisi '87, Aldous'00]



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]

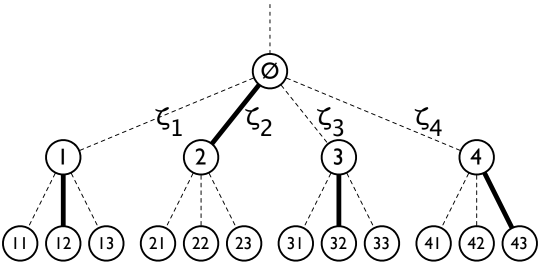


sort edge weights $W_{\varnothing,1}, W_{\varnothing,2}, \ldots$ from smallest to largest: arrivals $\zeta_1, \zeta_2, \ldots$ of a Poisson process with rate 1

$X_v \triangleq$ 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]



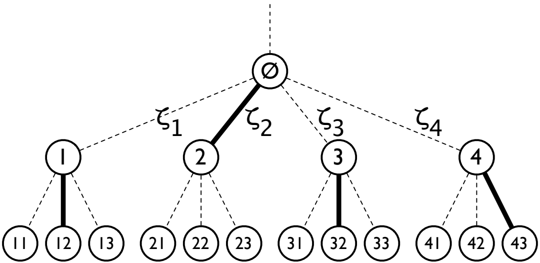
sort edge weights $W_{\varnothing,1}, W_{\varnothing,2}, \ldots$ from smallest to largest: arrivals $\zeta_1, \zeta_2, \ldots$ of a Poisson process with rate 1

$X_v \triangleq$ 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 $T_v \setminus \{v\}$

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

From distributional to differential equations

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

From distributional to differential equations

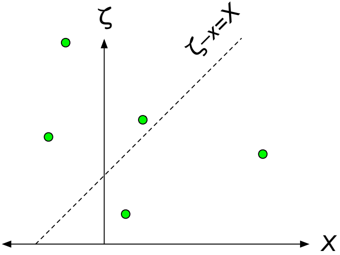
$X \overset{d}{=} \min\{\zeta_i - X_i\}$ where ζ_i are Poisson arrivals

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

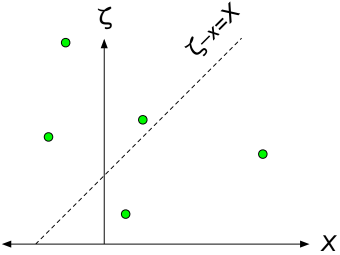
Define the ccdf $\bar{F}(x) = 1 - F(x) = \mathbb{P}[X > x] = \mathbb{P}[\forall i : \zeta_i - x > X_i]$
Generate pairs (ζ_i, X_i): two-dimensional Poisson process with density F'



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'



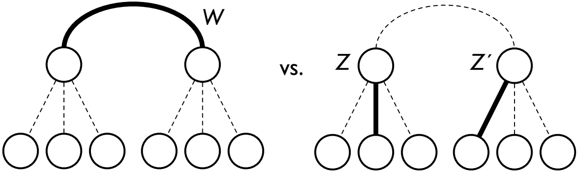
$$\bar{F}(x) = \exp\left(-\int_{-x}^{\infty} \bar{F}(t)\,dt\right) \quad \Rightarrow \quad \frac{dF(x)}{dx} = F(x)F(-x)$$

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

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

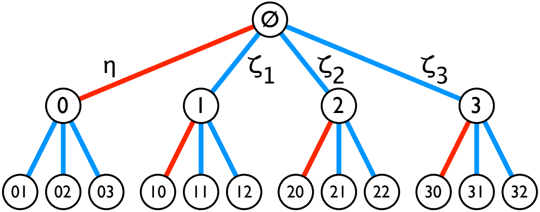


Contribution of a single edge:

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

Planted Poisson-weighted infinite tree

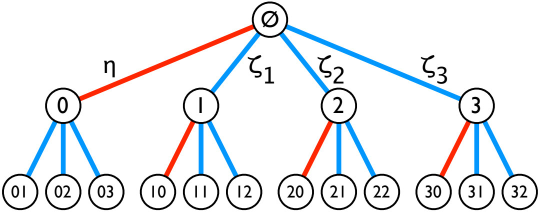
Partner in planted matching is either parent or child 0,
other children sorted $1, 2, 3, \ldots$



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

Planted Poisson-weighted infinite tree

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



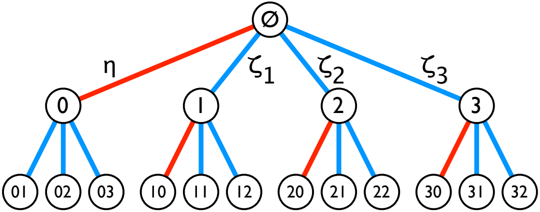
$X_v \triangleq$ cost of min matching in T_v $-$ cost of min matching on $T_v \setminus \{v\}$

Recursion:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$\dot{V}$ 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 $\lambda = 4$

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

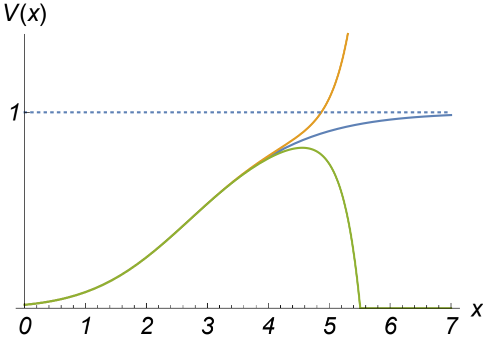
Boundary conditions: $F(x), V(x), G(-x), W(-x) \to \begin{cases} 1 & x \to +\infty \\ 0 & x \to -\infty \end{cases}$

Lemma

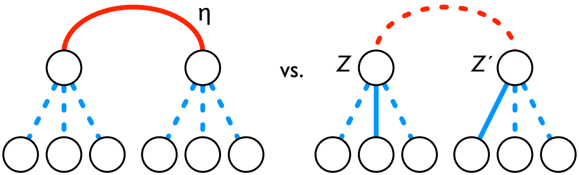
There is a unique solution if and only if $\lambda < 4$.

Geometric interpretation of uniqueness

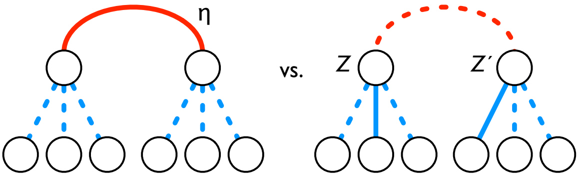
When $\lambda < 4$, $(F = 1, G = 0, V = 1, W = 0)$ is a saddle point:
There exists a unique initial condition from which we approach the saddle along its unstable manifold



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

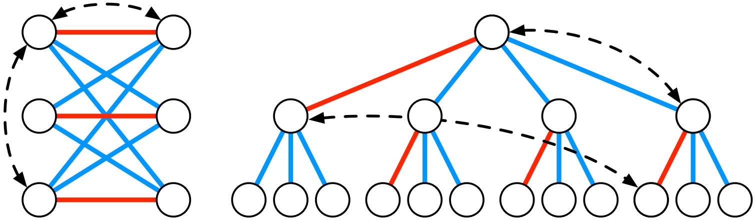


Finally, computing the overlap for $\lambda < 4$



$$
\begin{aligned}
\alpha(\lambda) = \mathbb{P}\left[\eta < Z + Z'\right] &= 1 - \mathbb{E}_\eta \int_{-\infty}^{+\infty} f(x) F(\eta - x)\, dx \\
&= 1 - \int_{-\infty}^{+\infty} f(x)\, \mathbb{E}_\eta F(\eta - x)\, dx \\
&= 1 - \int_{-\infty}^{+\infty} (1 - F(x))(1 - G(x)) V(x) W(x)\, dx \\
&= 1 - 2 \int_{0}^{+\infty} (1 - F(x))(1 - G(x)) V(x) W(x)\, dx
\end{aligned}
$$

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: $\lambda = 4$
- 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]:

$$\lambda = 4 - \epsilon: \text{ Optimal reconstruction error is } \exp\left(-\Theta(1/\sqrt{\epsilon})\right)$$

- Sharp threshold in general (d, P, Q) model [Ding-Wu-X.-Yang '21]:

$$\sqrt{d} \int \sqrt{\mathrm{d}P \mathrm{d}Q} = 1$$

- 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]:

$$\lambda = 4 - \epsilon\text{: Optimal reconstruction error is } \exp\left(-\Theta(1/\sqrt{\epsilon})\right)$$

- Sharp threshold in general (d, P, Q) model [Ding-Wu-X.-Yang '21]:

$$\sqrt{d} \int \sqrt{\mathrm{d}P\mathrm{d}Q} = 1$$

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

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.