The Planted Matching Problem: Sharp Threshold and Infinite-order Phase Transition

Dana Yang

Duke $\Longrightarrow$ Simons $\Longrightarrow$ Cornell

Nov 5, 2021

Joint IFML/CCSI Symposium

The Planted Matching Problem

Joint work with **Jian Ding** [1], **Yihong Wu** [2] and **Jiaming Xu** [3].

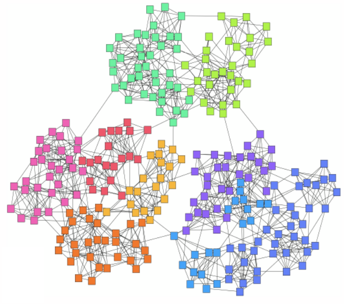
[1] Department of Statistics, The Wharton School, University of Pennsylvania

[2] Department of Statistics and Data Science, Yale University

[3] The Fuqua School of Business, Duke University

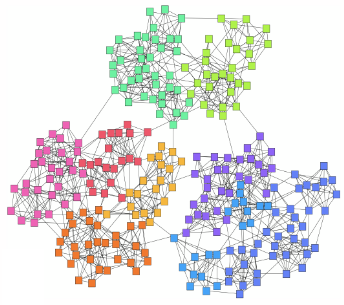
Planted models and recovery

Model : planted structure + noise.
Question : When/how can one recover the planted structure from its noisy observation ?



Planted models and recovery

Model : planted structure + noise.
Question : When/how can one recover the planted structure from its noisy observation ?

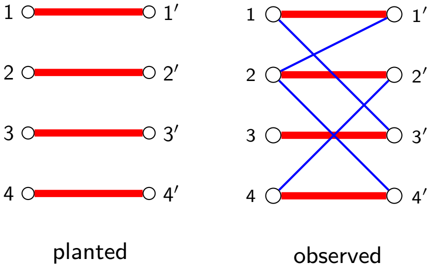


Examples

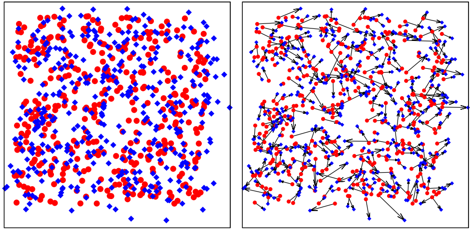
- Recovery of planted clique in Erdős-Rényi graphs.
- Community detection under the Stochastic Block Model.

The planted matching problem

Planted structure : perfect matching in bipartite graph.

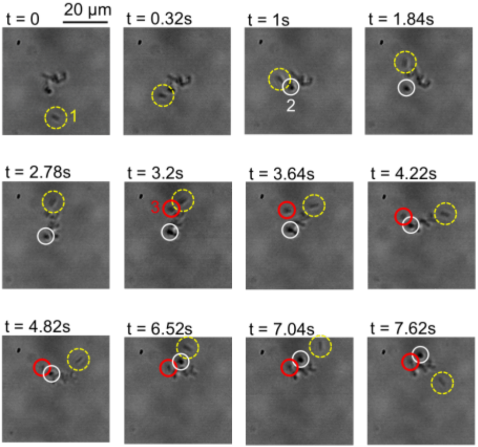


Motivating application : particle tracking



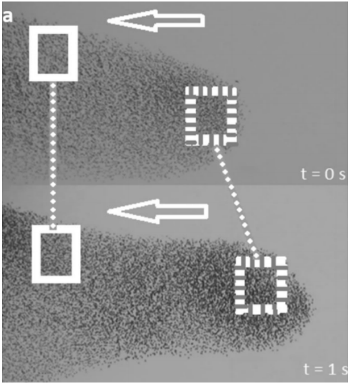
[Chertkov-Kroc-Krzakala-Vergassola-Zdeborová PNAS'10]

Motivating application : particle tracking



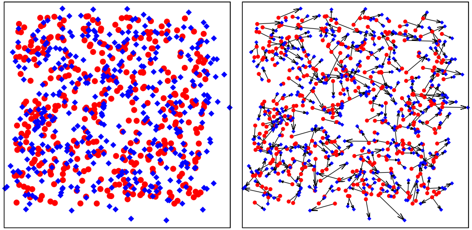
[Sinibaldi-Iebba-Chinappi MicrobiologyOpen'18]

Motivating application : particle tracking



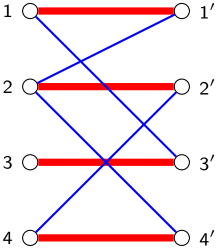
[Hemelrijk-Costanzo-Hildenbrandt-Carere Behavioral Ecology and Sociobiology'19]

Motivating application : particle tracking



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

Planted matching model



- Planted matching M^* uniformly distributed in all perfect matchings.
- Edges not in M^* appear independently w.p. $\frac{d}{n}$.
- Edge weight

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

Main results

Theorem (Ding-Wu-Xu-Y '21)

Sharp threshold for almost perfect recovery :

- If $\sqrt{d}B(\mathcal{P},\mathcal{Q}) \leq 1$, then some $\widehat{M}$ achieves

$$\frac{1}{n}\mathbb{E}\left|\widehat{M}\triangle M^*\right| \to 0 \text{ as } n \to \infty.$$

- If $\sqrt{d}B(\mathcal{P},\mathcal{Q}) \geq 1+\epsilon$ for $\epsilon > 0$, then for all $\widehat{M}$ and some constant c,

$$\frac{1}{n}\mathbb{E}\left|\widehat{M}\triangle M^*\right| \geq c.$$

Bhattacharyya coefficient (Hellinger affinity) $B(\mathcal{P},\mathcal{Q}) \triangleq \int \sqrt{d\mathcal{P}d\mathcal{Q}}$.

Main results

Theorem (Ding-Wu-Xu-Y '21)

Sharp threshold for almost perfect recovery :

- If $\sqrt{d}B(\mathcal{P},\mathcal{Q}) \leq 1$, then some $\widehat{M}$ achieves

$$\frac{1}{n}\mathbb{E}\left|\widehat{M}\triangle M^*\right| \to 0 \text{ as } n \to \infty.$$

- If $\sqrt{d}B(\mathcal{P},\mathcal{Q}) \geq 1+\epsilon$ for $\epsilon > 0$, then for all $\widehat{M}$ and some constant c,

$$\frac{1}{n}\mathbb{E}\left|\widehat{M}\triangle M^*\right| \geq c.$$

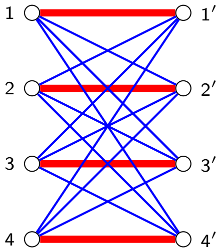
Bhattacharyya coefficient (Hellinger affinity) $B(\mathcal{P},\mathcal{Q}) \triangleq \int \sqrt{d\mathcal{P}d\mathcal{Q}}$.

- Result works for both sparse (d bounded) and dense ($d \to \infty$) graphs.
- In the dense $d \to \infty$ regime, need certain scaling assumptions on Q.

Examples

Exponential model (dense regime)[Maharrami-Moore-Xu '19, Semerjian-Sicuro-Zdeborová '20] :

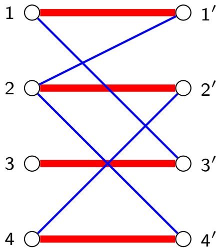
- $d = n$, $P = \text{Exp}(\lambda)$, $Q = \text{Exp}(1/n)$.
- Sharp threshold : $\lambda = 4$.



Examples

Unweighted model (sparse regime) :

- d =const, $P = Q$.
- The edge weights do not offer any information for recovery.
- Sharp threshold : $d = 1$.



Analysis (unweighted model)

1. Establish positive direction ($d \leq 1$) by analyzing the maximum likelihood estimator;
2. Establish negative direction ($d \geq 1 + \epsilon$) by analyzing the posterior distribution.

Analysis for $d \leq 1$

1. Analyze the MLE :

$$\widehat{M}_{\mathrm{ML}} \in \{\text{perfect matchings in } G\}\,.$$

Analysis for $d \leq 1$

1. Analyze the MLE :

$$\widehat{M}_{\mathrm{ML}} \in \{\text{perfect matchings in } G\}.$$

2. First moment analysis :

$$\mathbb{P}\left\{\left|\widehat{M}_{\mathrm{ML}} \triangle M^*\right| \geq \beta n\right\}$$
$$\leq \mathbb{P}\left\{\exists \text{ perfect matching } M \text{ in } G,\ s.t.\ |M \triangle M^*| \geq \beta n\right\}$$
$$\leq \sum_{t \geq \beta n/2} \binom{n}{t} t! \left(\frac{d}{n}\right)^t$$
$$\to 0 \text{ for some } \beta = o(1) \text{ when } d \leq 1.$$

Analysis (unweighted model)

1. Establish positive direction ($d \leq 1$) by analyzing the maximum likelihood estimator ;
2. Establish negative direction ($d \geq 1 + \epsilon$) by analyzing the posterior distribution.

Analysis for $d \geq 1 + \epsilon$

Crucial observations

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

Analysis for $d \geq 1 + \epsilon$

Crucial observations

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

Proof : Let $\widetilde{M}$ be sampled from the posterior distribution, then for any estimator $\widehat{M}$,

$$\mathbb{E}\left|\widetilde{M} \triangle M^*\right| \leq \mathbb{E}\left|\widetilde{M} \triangle \widehat{M}\right| + \mathbb{E}\left|\widetilde{M} \triangle M^*\right| = 2\mathbb{E}\left|\widehat{M} \triangle M^*\right|,$$

where the equality is because $\mathcal{L}(G, \widetilde{M}) = \mathcal{L}(G, M^*)$.

Analysis for $d \geq 1 + \epsilon$

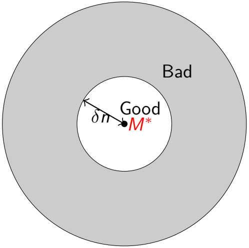
Crucial observations

1. Sampling from the posterior distribution is close to optimal.
2. The posterior distribution is uniform over the set of all perfect matchings in G.

Analysis for $d \geq 1 + \epsilon$

Crucial observations

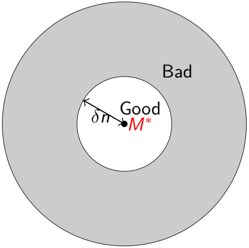
1. Sampling from the posterior distribution is close to optimal.
2. The posterior distribution is uniform over the set of all perfect matchings in G.



Analysis for $d \geq 1 + \epsilon$

Crucial observations

1. Sampling from the posterior distribution is close to optimal.
2. The posterior distribution is uniform over the set of all perfect matchings in G.



Claim : number of "bad" solutions is $\exp(\Omega(n))$.

For $d \geq 1 + \epsilon$

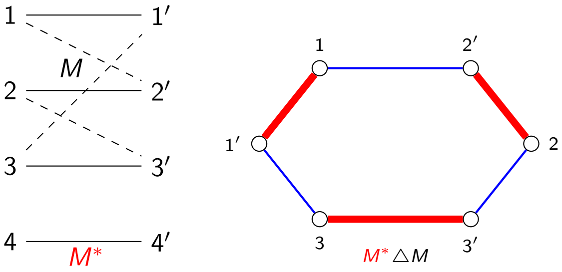
Crucial observations

1. Sampling from the posterior distribution is close to optimal.
2. The posterior distribution is uniform over the set of all perfect matchings in G.
3. For all perfect matching M, $M \triangle M^*$ consists of a disjoint union of alternating cycles.

For $d \geq 1 + \epsilon$

Crucial observations

1. Sampling from the posterior distribution is close to optimal.
2. The posterior distribution is uniform over the set of all perfect matchings in G.
3. For all perfect matching M, $M \triangle M^*$ consists of a disjoint union of alternating cycles.



For $d \geq 1 + \epsilon$

Crucial observations

1. Sampling from the posterior distribution is close to optimal.
2. The posterior distribution is uniform over the set of all perfect matchings in G.
3. For all perfect matching M, $M \triangle M^*$ consists of a disjoint union of alternating cycles.

Conclusion : It suffices to show the existence of $e^{\Omega(n)}$ distinct alternating cycles in G of length $\Omega(n)$.

Existence of many long alternating cycles

Natural attempt : first and second moment method.

Existence of many long alternating cycles

Natural attempt : first and second moment method.

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

Existence of many long alternating cycles

Natural attempt : first and second moment method.

- Let S be the set of 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 some constant probability.

Existence of many long alternating cycles

Natural attempt : first and second moment method.

- Let S be the set of 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 some constant probability.
- However, $\mathbb{E}(|S|^2) \gg (\mathbb{E}|S|)^2$ due to the excessive correlation between long cycles.

Existence of many long alternating cycles

Natural attempt : first and second moment method.

- Let S be the set of 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 some 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-Goswami '15, ...].

Existence of many long alternating cycles

Two-stage cycle-finding scheme

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

Existence of many long 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 $\Omega(n)$ disjoint short (constant length) alternating paths, using vertices in V^c.

Existence of many long 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 $\Omega(n)$ disjoint short (constant length) 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 alternating cycles

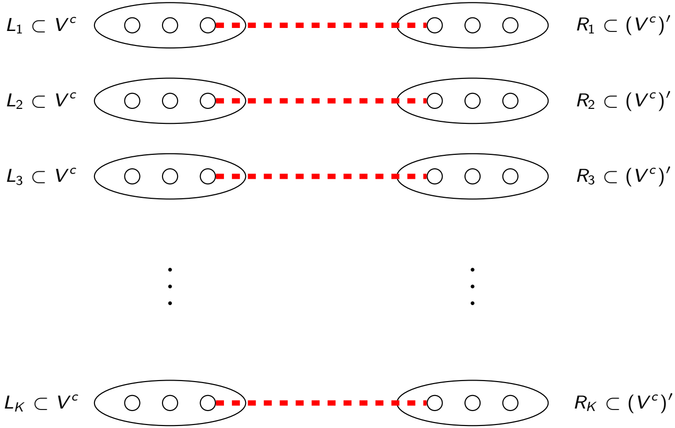
Two-stage cycle-finding scheme

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

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

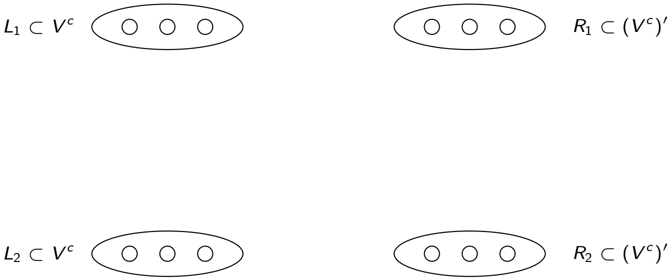
Two-stage cycle-finding scheme

Stage 1 (path construction) :



Two-stage cycle-finding scheme

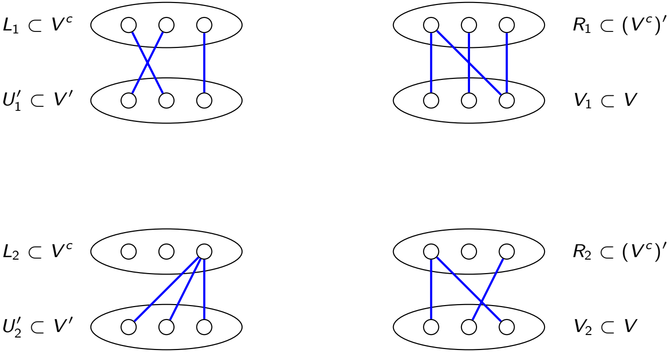
Stage 2 (sprinkling) :



Two-stage cycle-finding scheme

Stage 2 (sprinkling) :

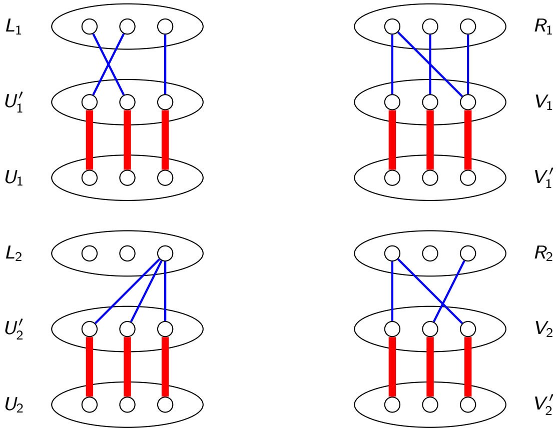
1. Let $U'_k = \{v' \in V' : (v', u) \in E(G) \text{ for some } u \in L_k\}$,
   $V_k = \{v \in V : (v, u') \in E(G) \text{ for some } u' \in R_k\}$.



Two-stage cycle-finding scheme

Stage 2 (sprinkling) :

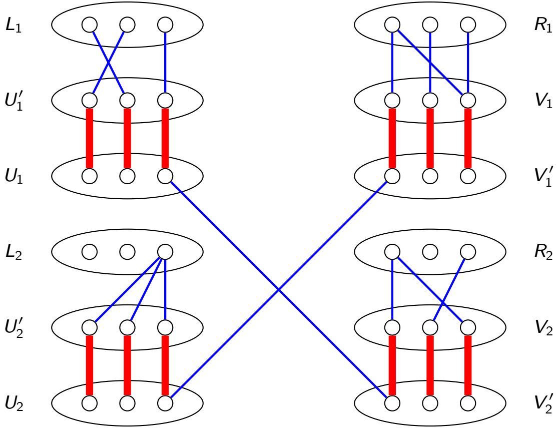
1. Let $U'_k = \{v' \in V' : (v', u) \in E(G) \text{ for some } u \in L_k\}$, $V_k = \{v \in V : (v, u') \in E(G) \text{ for some } u' \in R_k\}$.



Two-stage cycle-finding scheme

Stage 2 (sprinkling) :

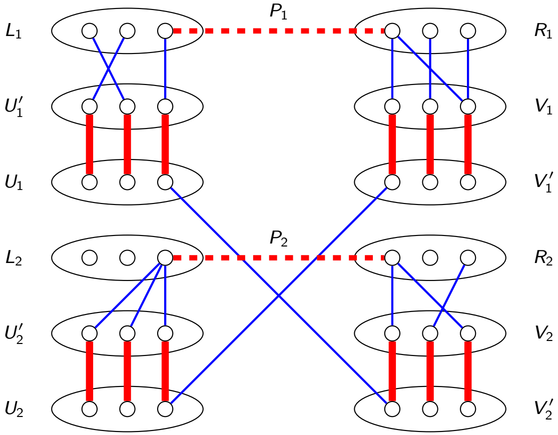
1. Let $U'_k = \{v' \in V' : (v', u) \in E(G) \text{ for some } u \in L_k\}$, $V_k = \{v \in V : (v, u') \in E(G) \text{ for some } u' \in R_k\}$.
2. Look for blue edges connecting $\{U_k\}, \{V'_k\}$.



Two-stage cycle-finding scheme

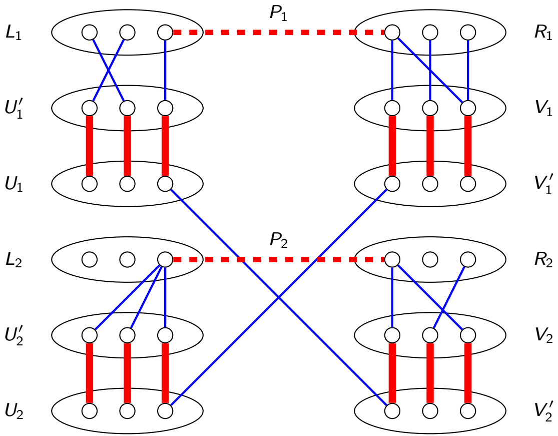
Stage 2 (sprinkling) :

1. Let $U'_k = \{v' \in V' : (v', u) \in E(G) \text{ for some } u \in L_k\}$, $V_k = \{v \in V : (v, u') \in E(G) \text{ for some } u' \in R_k\}$.
2. Look for blue edges connecting $\{U_k\}_{k \le K}, \{V'_k\}_{k \le K}$.



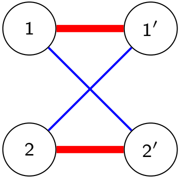
Two-stage cycle-finding scheme

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.



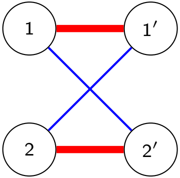
Two-stage cycle-finding scheme

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.

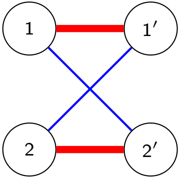


Existence of exponentially many long alternating cycles in G

1. Each alternating cycle on G_{super} expands into a 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.

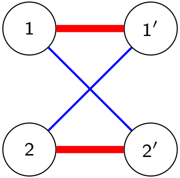


Existence of exponentially many long alternating cycles in G

1. Each alternating cycle on G_{super} expands into a alternating cycle in G.
2. (Informal) G_{super} is a very supercritical E-R 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.

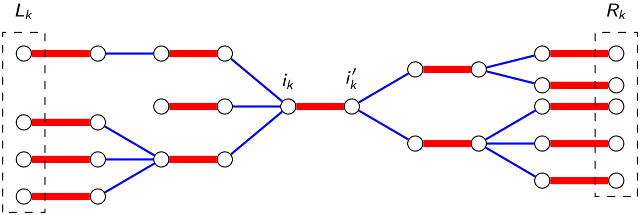


Existence of exponentially many long alternating cycles in G

1. Each alternating cycle on G_{super} expands into a alternating cycle in G.
2. (Informal) G_{super} is a very supercritical E-R bipartite graph with a planted perfect matching.
3. G_{super} contains $e^{\Omega(n)}$ alternating cycles of length $\Omega(n)$.

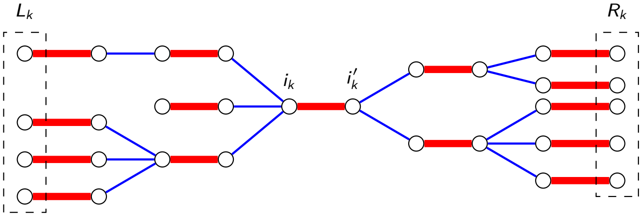
Path construction via breadth-first search

Two-sided tree (“bush") :



Path construction via breadth-first search

Two-sided tree (“bush") :



When $d \geq 1 + \epsilon$, the branching processes survive with constant probability.

Exponential model

Unplanted version (Random assignment model) [Mézard-Parisi 87', Aldous 01', ...] :

- Observe W_e on the complete bipartite graph, where $W_e \stackrel{i.i.d.}{\sim} \exp(1)$.
- Minimum weight matching has average weight $\pi^2/6$.

Exponential model

Unplanted version (Random assignment model) [Mézard-Parisi 87', Aldous 01', ...] :

- Observe W_e on the complete bipartite graph, where $W_e \overset{i.i.d.}{\sim} \exp(1)$.
- Minimum weight matching has average weight $\pi^2/6$.

Planted version :

- $d = n$, $\mathcal{P} = \exp(\lambda)$, $\mathcal{Q} = \exp(1/n)$.

Exponential model

Unplanted version (Random assignment model) [Mézard-Parisi 87', Aldous 01', ...] :

- Observe W_e on the complete bipartite graph, where $W_e \overset{i.i.d.}{\sim} \exp(1)$.
- Minimum weight matching has average weight $\pi^2/6$.

Planted version :

- $d = n$, $\mathcal{P} = \exp(\lambda)$, $\mathcal{Q} = \exp(1/n)$.
- Minimum weight matching achieves almost perfect recovery iff $\lambda \geq 4$[Maharrami-Moore-Xu '19].

Exponential model

Unplanted version (Random assignment model) [Mézard-Parisi 87', Aldous 01', ...] :

- Observe W_e on the complete bipartite graph, where $W_e \overset{i.i.d.}{\sim} \exp(1)$.
- Minimum weight matching has average weight $\pi^2/6$.

Planted version :

- $d = n$, $\mathcal{P} = \exp(\lambda)$, $\mathcal{Q} = \exp(1/n)$.
- Minimum weight matching achieves almost perfect recovery iff $\lambda \geq 4$[Maharrami-Moore-Xu '19].
- Sharp threshold : $\lambda = 4$.

Exponential model

Unplanted version (Random assignment model) [Mézard-Parisi 87', Aldous 01', ...] :

- Observe W_e on the complete bipartite graph, where $W_e \overset{i.i.d.}{\sim} \exp(1)$.
- Minimum weight matching has average weight $\pi^2/6$.

Planted version :

- $d = n$, $\mathcal{P} = \exp(\lambda)$, $\mathcal{Q} = \exp(1/n)$.
- Minimum weight matching achieves almost perfect recovery iff $\lambda \geq 4$[Maharrami-Moore-Xu '19].
- Sharp threshold : $\lambda = 4$.
- For $\lambda = 4 - \epsilon$,

$$e^{-c_1/\sqrt{\epsilon}} \leq \inf_{\widehat{M}} \frac{1}{n}\mathbb{E}\left|\widehat{M}\triangle M^*\right| \leq e^{-c_2/\sqrt{\epsilon}},$$

  revealing an infinite-order phase transition, conjectured in [Semerjian-Sicuro-Zdeborová '20].
- For the lower bound proof, we resort to ~~"bushes"~~ paths.

Conclusion

- Sharp threshold for almost perfect recovery under the planted matching model : $\sqrt{d}B(\mathcal{P}, \mathcal{Q}) = 1$.

Conclusion

- Sharp threshold for almost perfect recovery under the planted matching model : $\sqrt{d}B(\mathcal{P}, \mathcal{Q}) = 1$.
- Infinite-order phase transition under the exponential model : optimal reconstruction error is $\exp(-\Theta(1/\sqrt{\epsilon}))$ when $\lambda = 4 - \epsilon$.

Conclusion

- Sharp threshold for almost perfect recovery under the planted matching model : $\sqrt{d}B(\mathcal{P}, \mathcal{Q}) = 1$.
- Infinite-order phase transition under the exponential model : optimal reconstruction error is $\exp(-\Theta(1/\sqrt{\epsilon}))$ when $\lambda = 4 - \epsilon$.
- Key idea : two-stage cycle finding (path construction + sprinkling).

Conclusion

- Sharp threshold for almost perfect recovery under the planted matching model : $\sqrt{d}B(\mathcal{P}, \mathcal{Q}) = 1$.
- Infinite-order phase transition under the exponential model : optimal reconstruction error is $\exp(-\Theta(1/\sqrt{\epsilon}))$ when $\lambda = 4 - \epsilon$.
- Key idea : two-stage cycle finding (path construction + sprinkling).

Open problems :

1. Phase transition for general distributions ?

Conclusion

- Sharp threshold for almost perfect recovery under the planted matching model : $\sqrt{d}B(\mathcal{P}, \mathcal{Q}) = 1$.
- Infinite-order phase transition under the exponential model : optimal reconstruction error is $\exp(-\Theta(1/\sqrt{\epsilon}))$ when $\lambda = 4 - \epsilon$.
- Key idea : two-stage cycle finding (path construction + sprinkling).

Open problems :

1. Phase transition for general distributions ?
2. Error characterization in entire parameter range ?

Conclusion

- Sharp threshold for almost perfect recovery under the planted matching model : $\sqrt{d}B(\mathcal{P}, \mathcal{Q}) = 1$.
- Infinite-order phase transition under the exponential model : optimal reconstruction error is $\exp(-\Theta(1/\sqrt{\epsilon}))$ when $\lambda = 4 - \epsilon$.
- Key idea : two-stage cycle finding (path construction + sprinkling).

Open problems :

1. Phase transition for general distributions ?
2. Error characterization in entire parameter range ?
3. Extension to planted k-factor model ?
   Conjecture[Sicuro-Zdeborová '20] : $\sqrt{kd}B(\mathcal{P}, \mathcal{Q}) = 1$.

Conclusion

- Sharp threshold for almost perfect recovery under the planted matching model : $\sqrt{d}B(\mathcal{P}, \mathcal{Q}) = 1$.
- Infinite-order phase transition under the exponential model : optimal reconstruction error is $\exp(-\Theta(1/\sqrt{\epsilon}))$ when $\lambda = 4 - \epsilon$.
- Key idea : two-stage cycle finding (path construction + sprinkling).

Open problems :

1. Phase transition for general distributions ?
2. Error characterization in entire parameter range ?
3. Extension to planted k-factor model ?
   Conjecture[Sicuro-Zdeborová '20] : $\sqrt{kd}B(\mathcal{P}, \mathcal{Q}) = 1$.

Thank you !

General model

Planted matching model (general)

True matching $M^* \sim \text{Unif}\{\text{all perfect matchings on } [n] \times [n]'\}$. Observed graph G contains all the edges in M^*, and for each $e \notin M^*$, $e \in G$ independently with probability d/n. Observe $(W_e)_{e \in G}$, where

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

General model

Planted matching model (general)

True matching $M^* \sim \text{Unif}\{\text{all perfect matchings on } [n] \times [n]'\}$. Observed graph G contains all the edges in M^*, and for each $e \notin M^*$, $e \in G$ independently with probability d/n. Observe $(W_e)_{e \in G}$, where

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

Sharp threshold : $\sqrt{d} B(\mathcal{P}, \mathcal{Q}) = 1$.

Proof of the positive result

1. Analyze the MLE :

$$\widehat{M}_{\mathrm{ML}} \in \{\text{perfect matchings in } G\}\,.$$

2. Union bound :

$$\begin{aligned}
&\mathbb{P}\left\{\left|\widehat{M}_{\mathrm{ML}} \triangle M^*\right| \geq \beta n\right\} \\
\leq &\mathbb{P}\left\{\exists \text{ perfect matching } M \text{ in } G,\ s.t.\ |M \triangle M^*| \geq \beta n\right\} \\
\leq &\sum_{t \geq \beta n/2} \binom{n}{t} t! \left(\frac{d}{n}\right)^t \\
\to &\ 0 \text{ for some } \beta = o(1) \text{ when } d \leq 1.
\end{aligned}$$

Proof of the positive result

1. Analyze the MLE :

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

Proof of the positive result

1. Analyze the MLE :

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

2. Union bound :

$$\mathbb{P}\left\{\left|\widehat{M}_{\mathrm{ML}} \triangle M^*\right| \geq \beta n\right\}$$

$$\leq \mathbb{P}\left\{\exists M \text{ in } G \text{ of higher likelihood than } M^*,\ s.t.\ |M \triangle M^*| \geq \beta n\right\}$$

$$\leq \sum_{t \geq \beta n/2} \binom{n}{t} t! \left(\frac{d}{n} B^2(\mathcal{P}, \mathcal{Q})\right)^t$$

$\to 0$ for some $\beta = o(1)$ when $\sqrt{d} B(\mathcal{P}, \mathcal{Q}) \leq 1$.

Proof of the negative result

Crucial observations

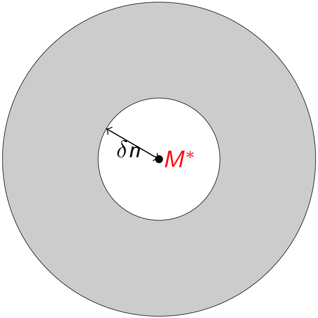
1. Sampling from the posterior distribution is close to optimal.
2. The posterior distribution is uniform over the set of all perfect matchings in G.
3. For all perfect matching M, $M \triangle M^*$ consists of a disjoint union of alternating cycles.

Proof of the negative result

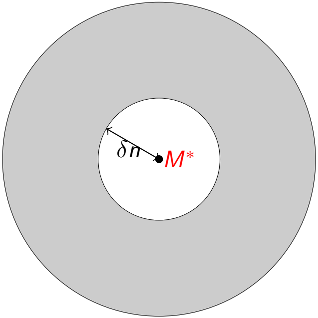
Crucial observations

1. Sampling from the posterior distribution is close to optimal.
2. ~~The posterior distribution is uniform over the set of all perfect matchings in G.~~
3. For all perfect matching M, $M \triangle M^*$ consists of a disjoint union of alternating cycles.

Analysis of the posterior distribution



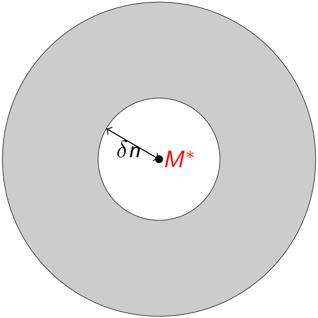
Analysis of the posterior distribution



Additional challenges for weighted graphs :

- Control the posterior mass of matchings close to M^*.

Analysis of the posterior distribution



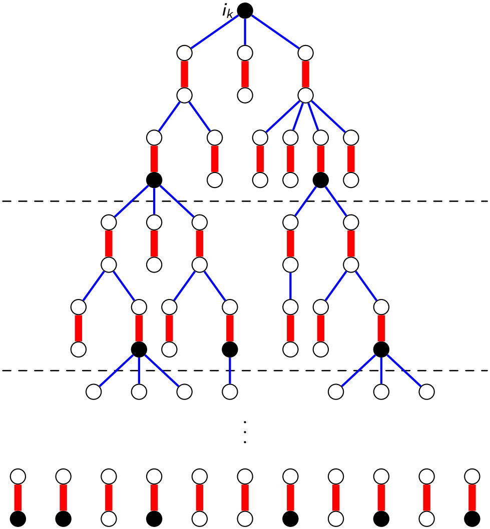
Additional challenges for weighted graphs :

- Control the posterior mass of matchings close to M^*.
- Find exponentially many long alternating cycles C that are augmenting :

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

Exploration + selection

Construction of augmenting paths :



Analysis (lower bound)

- Follow the two-stage cycle finding scheme.

Analysis (lower bound)

- Follow the two-stage cycle finding scheme.
- For the path construction stage, the exploration $+$ selection scheme is too wasteful.

Analysis (lower bound)

- Follow the two-stage cycle finding scheme.
- For the path construction stage, the exploration + selection scheme is too wasteful.

Path construction (exponential model)

1. Directly show the existence of many short augmenting alternating paths using first and second moment method.
   - Impose uniformity (bounded weight fluctuation on paths) to reduce second moment [Ding '13, Ding-Sun-Wilson '15].

Analysis (lower bound)

- Follow the two-stage cycle finding scheme.
- For the path construction stage, the exploration + selection scheme is too wasteful.

Path construction (exponential model)

1. Directly show the existence of many short augmenting alternating paths using first and second moment method.
   - Impose uniformity (bounded weight fluctuation on paths) to reduce second moment [Ding '13, Ding-Sun-Wilson '15].
2. Extract a large collection of vertex-disjoint paths via Turán's Theorem [Ding-Goswami '15].