AI4OPT Tutorial Lectures: Recent Results in Planted Assignment Problems

Lecture 4: Low-degree polynomials and limit of local algorithms

Lecturer: Jiaming Xu *Scribe: Jiaming Xu, April 21, 2023*

Initiated by [HS17], one versatile approach to understanding the computational limit of statistical problems is via low-degree polynomial approximation of the likelihood ratio. This approach not only gives predictions of the computational limit but also leads to efficient algorithms. As a case study, let us consider the random graph matching problem. We will show that the low-degree polynomial method gives efficient algorithms for both detection and recovery when the correlation is above an explicit constant known as Otter's constant. In the negative direction, we will show that all local algorithms fail below Otter's constant.

4.1 Low-degree polynomial approximation of likelihood ratio

Recall the hypothesis testing problem for graph matching

$$\begin{cases} H_0: & A \text{ and } B \text{ are two independent } \mathcal{G}(n,q) \\ H_1: & A \text{ and } B \text{ are two } \rho\text{-correlated } \mathcal{G}(n,q) \end{cases}.$$

Specifically, under H_1, there exists a random uniform permutation $\pi \in \mathcal{S}_n$ such that conditional on π, $\{A_{ij}, B_{\pi(i)\pi(j)}\}$ are i.i.d. pairs of Bern(q) with correlation ρ, that is

$$\bar{A}_{ij} \triangleq \frac{A_{ij}-q}{\sqrt{q(1-q)}}, \quad \bar{B}_{ij} \triangleq \frac{B_{ij}-q}{\sqrt{q(1-q)}}, \quad \mathbb{E}\left[\bar{A}_{ij}\bar{B}_{ij}\right] = \rho.$$

Let Q and P denote the joint distribution of (A,B) under H_0 and H_1, respectively. By the Neyman-Pearson lemma, the optimal test is given by the likelihood ratio test, that is,

$$L(A,B) \triangleq \frac{P(A,B)}{Q(A,B)} = \mathbb{E}_\pi\left[\frac{P(A,B \mid \pi)}{Q(A,B)}\right]_\pi.$$

Note that the above expectation invovles $n!$ permutations and hence is computationally intractable to evaluate. To obtain a computationally efficient test, an approach initiatied by and further developed by is known as the low-degree polynomial approximation, which projects the likelihood ratio function L onto the space spanned by the low-degree polynomials. To this end, consider a space consisting of all functions $f : \{0,1\}^{2\times\binom{n}{2}} \to \mathbb{R}$ endowed with inner product $\langle f, g\rangle \triangleq \mathbb{E}_Q[f(A,B)g(A,B)]$. Next, we introduce an orthonormal polynomial basis, indexed by subsets of $\binom{[n]}{2}$ or equivalently the edge-induced subgraph of the complete graph K_n, that is,

$$\phi_S(A,B) = \prod_{(i,j)\in S_1} \bar{A}_{ij} \prod_{(i,j)\in S_2} \bar{B}_{ij},$$

where $S \triangleq (S_1, S_2)$ and $S_1, S_2 \subset \binom{[n]}{2}$. Note that $\{\phi_S\}_{S\subset\binom{[n]}{2}\times\binom{[n]}{2}}$ is a Fourier basis for functions on the hypercube $\{0,1\}^{2\binom{n}{2}}$. In particular, it is easy to check that

- $\mathbb{E}_Q[\phi_S] = 0$ for all $S \neq \emptyset$
- $\langle \phi_S, \phi_T \rangle = \mathbb{1}\{S = T\}$
- ϕ_S is a degree-$|S|$ polynomial of the entries of A and B, where $|S| = |S_1| + |S_2|$.

Now, we are ready to introduce the "optimal" degree-$2K$ polynomial, that is,

$$f^* \in \underset{f:\deg(f)\leq 2K}{\arg\max} \frac{\mathbb{E}_P[f]}{\sqrt{\mathbb{E}_Q[f^2]}}.$$

Note that the quantity to be maximized can be viewed as a signal-to-noise ratio SNR for the test statistic f. By the change of measure $\mathbb{E}_P[f] = \langle L, f \rangle$ and moreover, $\mathbb{E}_Q[f^2] = \langle f, f \rangle = \|f\|_2^2$. Thus, by Cauchy-Schwartz inequality,

$$\langle L, f \rangle = \left\langle L^{\leq 2K}, f \right\rangle \leq \left\| L^{\leq 2K} \right\|_2 \|f\|_2,$$

where $L^{\leq 2K}$ is the projection of L on the subspace spanned by the basis $\{\phi_S : |S| \leq 2K\}$ and the inequality is met with equality when $f \propto L^{\leq 2K}$. Hence,

$$f^* = L^{\leq 2K} = \sum_{S:|S|\leq 2K} \langle L, \phi_S \rangle \phi_S$$

It is easy to check that

$$\mathbb{E}_P[f] = \mathbb{E}_Q[f^2] = \sum_{S:|S|\leq 2K} \langle L, \phi_S \rangle^2$$

and hence

$$\frac{\mathbb{E}_P[f^*]}{\sqrt{\mathbb{E}_Q[(f^*)^2]}} = \sqrt{\sum_{S:|S|\leq 2K} \langle L, \phi_S \rangle^2}.$$

Before proceeding, we remark that the above low-degree polynomial approximation approach works for general hypothesis testing problems, as long as the polynomial basis under the null distribution can be properly defined.

Next, we focus on the graph matching problem and evaluate the coefficient $\langle L, \phi_S \rangle$. In particular,

$$\langle L, \phi_S \rangle = \mathbb{E}_P[\phi_S(A, B)] = \mathbb{E}_\pi \mathbb{E}_{P|\pi} \left[\prod_{(i,j)\in S_1} \bar{A}_{ij} \prod_{(k,\ell)\in S_2} \bar{B}_{k\ell} \right]$$

Due to the centering, crucially

$$\mathbb{E}_{P|\pi} \left[\prod_{(i,j)\in S_1} \bar{A}_{ij} \prod_{(k,\ell)\in S_2} \bar{B}_{k\ell} \right] = \begin{cases} 0 & \text{if } \pi(S_1) \neq S_2 \\ \rho^{|S_1|} & \text{o.w.} \end{cases}.$$

Therefore,

$$\langle L, \phi_S \rangle = \begin{cases} 0 & \text{if } S_1 \ncong S_2 \\ \rho^{|H|} \mathbb{P}\left[\pi(S_1) = S_2\right] & \text{if } S_1 \cong S_2 \cong H, \end{cases}$$

where H denotes an unlabeled graph or more formally an isomorphic class. Note that when $S_1 \cong S_2 \cong H$, $\mathbb{P}\left[\pi(S_1) = S_2\right] = \frac{1}{\mathrm{sub}_n(H)}$, where $\mathrm{sub}_n(H)$ denotes the number of copies of H in K_n. Define

$$a_H = \rho^{|H|} \times \frac{1}{\mathrm{sub}_n(H)}.$$

Then we have

$$f^* = \sum_{H:|H|\le K} a_H \sum_{S_1 \cong H} \sum_{S_2 \cong H} \phi_S = \sum_{H:|H|\le K} a_H \underbrace{\sum_{S_1\cong H} \prod_{(i,j)\in S_1} \bar{A}_{ij}}_{W_H(\bar{A})} \underbrace{\sum_{S_2\cong H} \prod_{(i,j)\in S_2} \bar{B}_{ij}}_{W_H((\bar{B}))}$$

Note that $W_H(A)$ is simply the subgaph count of H in A and hence $W_H(\bar{A})$ is known as the *signed* subgraph count due to the centering. Let $W((\bar{A}) = \left(\sqrt{a_H} W_H(\bar{A})\right)_H$ denote a graph feature vector where each coordinate corresponds to a weighted signed subgraph count. Then, the optimal low-degree polynomial f^* has an appealing interpretation in terms of graph kernel, that is

$$f^* = \left\langle W(\bar{A}), W(\bar{B}) \right\rangle.$$

Intuitively, f^* captures the inherent edge correlation between A and B by counting the co-occurrences of signed graphs for a family of non-isormophis graphs.

It is postulated in [Hop18, KWB19] that, if the signal-to-noise ratio $\frac{\mathbb{E}_{\mathcal{P}}[f^*]}{\sqrt{\mathbb{E}_{\mathcal{Q}}[(f^*)^2]}}$ stays bounded for $K = \mathrm{polylog}(n)$ as $n \to \infty$, then no polynomial-time algorithm can distinguish between $\mathcal{P}$ and $\mathcal{Q}$ with vanishing error. This is known as *low-degree hardness conjecture*.

Our above result shows that

$$\frac{\mathbb{E}_P[f^*]}{\sqrt{\mathbb{E}_Q[(f^*)^2]}} = \sum_{H:|H|\le K} \sum_{S_1\cong H} \prod_{(i,j)\in S_1} a_H^2 = \sum_{H:|H|\le K} \rho^{2H},$$

where the last inequlaity holds because $|\{S_1 : S_1 \cong H\}| = \mathrm{sub}_n(H)$. Since an unlabeled graph $[H] \in \mathcal{H}^*$ with k edges has at most $2k$ vertices, the number of such graphs is at most $\binom{(2k)^2}{k} \le (4ek)^k$ and hence

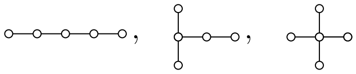
$$\sum_{[H]\in\mathcal{H}^*} \rho^{2|E(H)|} \le \sum_{k=1}^{K} (4ek)^k \rho^{2k} = O(1)$$

for $K = \mathrm{polylog}(n)$, provided that $\rho^2 \le \frac{1}{\mathrm{polylog}(n)}$. Therefore, if the squared correlation ρ^2 is smaller than $\frac{1}{\mathrm{polylog}(n)}$, then the signal-to-noise ratio for any degree-polylog(n) polynomial test is bounded, in which case the testing problem is conjectured to be computationally hard. In view of the close connection between hypothesis testing and estimation, we further conjecture the graph matching problem (namely, recovering the latent permutation π under the correlated Erdős-Rényi model $\mathcal{G}(n, q, \rho)$) is computationally hard when $\rho^2 \le \frac{1}{\mathrm{polylog}(n)}$. Note that these conjectures are consistent with the state-of-the-art results for which no polynomial-time test or matching algorithm is known when $\rho^2 \le \frac{1}{\mathrm{polylog}(n)}$.

4.2 Efficient testing via counting trees

In the positive direction, we can restrict subgraph count to tree count and obtain an efficient testing procedure. To this end, let $\mathcal{T}$ denote the set of all unlabeled trees with K edges. For example, for

$K = 4$, $\mathcal{T}$ consists of three trees shown in pictograms below (see [? , App. I] for bigger examples)

$\mathcal{T} = \Big\{$  $\Big\}$

A celebrated result of Otter [Ott48] is that the number of unlabeled trees grows exponentially with

$$|\mathcal{T}| \asymp \frac{1}{K^{3/2}} (1/\alpha)^K , \tag{4.1}$$

where $\alpha \approx 0.33833$ is Otter's constant. Define

$$f_{\mathcal{T}} = \sum_{H \in \mathcal{T}} a_H W_H(\bar{A}) W_H(\bar{B}).$$

Theorem 4.1 ([MWXY21])**.** *Suppose $q \leq 1/2$,*

$$nq \geq n^{-o(1)}, \quad \rho^2 > \alpha, \quad \omega(1) \leq K \leq \frac{\log n}{16 \log\log n \vee 2 \log\left(\frac{1}{nq}\right)} . \tag{4.2}$$

Then the testing error satisfies

$$Q(f_{\mathcal{T}}(A, B) \geq \tau) + P(f_{\mathcal{T}}(A, B) \leq \tau) = o(1), \tag{4.3}$$

where the threshold is chosen as $\tau = C\mathbb{E}_P[f_{\mathcal{T}}(A, B)]$ for any fixed constant $0 < C < 1$.

Note that assuming $q \leq 1/2$ is without loss of generality, as we can equivalently test the correlation between the complement graphs of the observed graphs and replace q by $1 - q$. The condition $nq \geq n^{-o(1)}$ in fact applies to the very sparse regime of vanishing average degrees, as long as they are slower than any polynomial in n. This condition is necessary for the existence of trees with $K = \omega(1)$ edges.

Proof sketch of Theorem 4.1. We have shown that $\mathbb{E}_Q[f_{\mathcal{T}}] = 0$ and

$$\mathbb{E}_P[f_{\mathcal{T}}] = \mathbb{E}_Q[f_{\mathcal{T}}^2] = \sum_{H \in \mathcal{T}} \rho^{2|H|} = \rho^{2k} |\mathcal{T}| \to \infty,$$

where the last assertion holds by (4.1) and the assumptions that $\rho^2 > \alpha$ and $K \to \infty$. It remains to further show that $\mathbb{E}_P[f_{\mathcal{T}}]/\sqrt{\mathrm{Var}_P[f_{\mathcal{T}}]} \to \infty$ and the proof is complete by invoking Chebyshev's inequality. To bound the variance, it suffices to control

$$\begin{aligned} \mathbb{E}_P[f_{\mathcal{T}}^2] &= \sum_{H, I \in \mathcal{T}} a_H a_I \mathbb{E}_P[W_H(\bar{A}) W_H(\bar{B}) W_I(\bar{A}) W_I(\bar{B})] \\ &= \sum_{H, I \in \mathcal{T}} a_H a_I \sum_{S_1, S_2 \cong H} \sum_{T_1, T_2 \cong I} \underbrace{\mathbb{E}_P[\bar{A}_{S_1} \bar{B}_{S_2} \bar{A}_{T_1} \bar{B}_{T_2}]}_{(I)} . \end{aligned}$$

We can bound term (I) in terms of the overlaping pattern of the 4-tuple (S_1, S_2, T_1, T_2), and then enumerate the 4-tuples. We omit the details here and refer the interested reader to [MWXY21]. Note that this is the place where we crucially use the tree property and the upper bound to K. □

Having established the statistical guarantee of $f_{\mathcal{T}}$, next we design an efficient algorithm to approximately compute $f_{\mathcal{T}}$. In particular, the key subroutine is to compute a weighted tree count $W_H(M) = \sum_{S \cong H} \prod_{(i,j)\in S} M_{ij}$ for $H \in \mathcal{T}$. Note that computing $W_H(M)$ via a naive exhaustive search takes $n^{\Omega(K)}$ time, which is superpolynomial when $K \to \infty$. To resolve this computational issue, we design an $n^2 e^{O(K)}$-time algorithm to compute an approximation of $W_H(M)$ using the strategy of *color coding* [AYZ95]. The main steps are as follows.

1. Assign random coloring μ to each vertex from $[K+1]$ uniformly at random.
2. Define the colorful subgraph count

$$Y_H(M,\mu) = \sum_{S \cong H} \mathbb{1}\left\{S \text{ has } K+1 \text{ colors}\right\} \prod_{(i,j)\in S} M_{ij}$$

   Then $\mathbb{E}_\mu[Y_H(M,\mu] = rW_H(M)$, where $r = \frac{(K+1)!}{(K+1)^{K+1}}$.

3. Generate $O(1/r)$ independent random colorings μ_t and compute

$$\widehat{W}_H(M) = \sum_{t=1}^{1/r} Y_H(M,\mu_t)$$

4. Compute $Y_H(M,\mu)$ via dynamic programming in $n^2 e^{O(K)}$ time.

Step 2 shows that $Y_H(M,\mu)/r$ is an unbiased estimator of $W_H(M)$. To further reduce the variance, in Step 3 we average over many independent colorings and obtain a more accurate unbiased estimator $\widehat{W}_H(M)$. Now crucially each colorful subgraph count $Y_H(M,\mu)$ can be efficiently computed via dynamic programming using the recursive tree property. The high-level idea is as follows. Pick any edge (u,v) in the tree H, which divides the tree into two subtrees, one rooted at u, say H_u, and the other rooted at v, say H_v . Suppose the colorful subtree counts have already been recursively computed. Then we can multify the colorful subtree count for H_u, the colorful subtree count for H_v, and M_{uv}, subject to the constraints that the color sets for H_u and H_v are disjoint. This should give the colorful tree count for H. Note that here the colorful property ensures that the vertex sets of H_u and H_v must be disjoint, so that we obtain a valid tree H when patching H_u and H_v together.

4.3 Efficient recovery via counting chandeliers

In this subsection, we turn to recovery. Analogous to testing, an immediate question is: can we define a feature vector based on subgraph count for each node? The answer is yes using the *rooted* subgraph count. As such, for node i in graph A, define

$$W_{i,H}(\bar{A}) = \sum_{S(i)\cong H} \prod_{e \in S_i} \bar{A}_e,$$

where H is a rooted subgraph and $S(i)$ denote a subgraph rooted at i. The isomorphism between two rooted graphs ensures that the isomorphism mapping must map between the two roots. Now, given a family $\mathcal{H}$ of non-isomrophic graphs, define the feature (signature) vector,

$$s_i \triangleq \left(\sqrt{a_H} W_{i,H}(\bar{A})\right)_{H\in\mathcal{H}}.$$

Similarly define t_j for node j in graph B. Then we can define a similary measure between node i in A and node j in B as

$$\Phi_{ij} = \langle s_i, t_j \rangle = \sum_{H \in \mathcal{H}} a_H W_{i,H}(\bar{A}) W_{j,H}(\bar{B}).$$

Finally, for each $i \in [n]$, if there exists a unique $j \in [n]$ such that $\Phi_{ij}^{\mathcal{H}} \geq \tau$, let $\widehat{\pi}(i) = j$ and include i in set I.

It remains to choose this collection of subgraphs $\mathcal{H}$. Ideally, we would like Φ_{ij} to be maximized at $j = \pi(i)$, at least on average.

Mean separation. Let's first compute the mean. In the following calculation, let us assume π is the identity permutation for simplicity. Then

$$\begin{aligned} \mathbb{E}\left[\Phi_{ij}\right] &= \sum_{H \in \mathcal{H}} a_H \mathbb{E}\left[W_{i,H}(\bar{A}) W_{j,H}(\bar{B})\right] \\ &= \sum_{H \in \mathcal{H}} a_H \sum_{S(i) \cong H} \sum_{T(j) \cong H} \mathbb{E}\left[\bar{A}_{S(i)} \bar{B}_{T(j)}\right]. \end{aligned}$$

Again, note that

$$\mathbb{E}\left[\bar{A}_{S(i)} \bar{B}_{T(j)}\right] = \rho^{|H|} \mathbb{1}\left\{S(i) = T(j)\right\}.$$

Therefore,

$$\mathbb{E}\left[\Phi_{ij}\right] = \sum_{H \in \mathcal{H}} a_H \rho^{|H|} \sum_{S(i) \cong H} \sum_{T(j) \cong H} \mathbb{1}\left\{S(i) = T(j)\right\}$$

To ensure that $\mathbb{E}\left[\Phi_{ij}\right]$ has zero mean for fake pairs $j \neq i$, we now restrict the graph H to be uniquely rooted, i.e., H is non-ismorphic to itsef if we change its root to any other vertex. For example, ●–○ is not uniquely rooted, while ○–●–○ is uniquely rooted.

Now, thanks to the uniquely rooted property of H, if $S(i) \cong H$, $T(j) \cong H$, then $S(i) = T(j)$ implies $i = j$. Hence,

$$\mathbb{E}\left[\Phi_{ij}\right] = \sum_{H \in \mathcal{H}} a_H \rho^{|H|} \mathrm{sub}_n(H) \mathbb{1}\left\{i = j\right\} = \sum_{H \in \mathcal{H}} \rho^{2|H|} \mathbb{1}\left\{i = j\right\},$$

where the last equality holds by defining $a_H = \rho^{|H|}/\mathrm{sub}_n(H)$ and $\mathrm{sub}_n(H)$ is the number of copies of the rooted graph H in the complete graph K_n.

Variance bound. It remains to bound $\mathrm{Var}(\Phi_{ij})$. Note that

$$\mathrm{Var}(\Phi_{ij}) = \sum_{H, I \in \mathcal{H}} a_H a_I \mathrm{Cov}\left(W_{i,H}(\bar{A}) W_{j,H}(\bar{B}), W_{i,I}(\bar{A}) W_{j,I}(\bar{B})\right).$$

To get a sense, let us first ignore the cross correlation for distinct H and I. Then for fake pairs $j \neq i$, it turns out that we can approximate the variance as:

$$\mathrm{Var}(\Phi_{ij}) \approx \sum_{H \in \mathcal{H}} a_H^2 \underbrace{\mathrm{Var}\left(W_{i,H}(\bar{A}) W_{j,H}(\bar{B})\right)}_{\approx \mathrm{sub}_n^2(H)} \approx \sum_{H \in \mathcal{H}} \rho^{2|H|}.$$

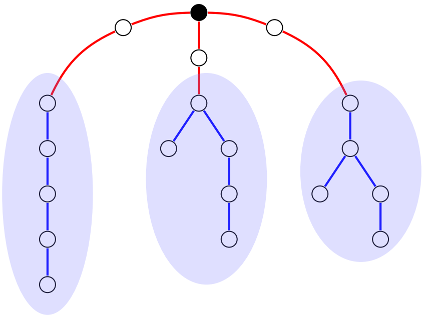


Figure 4.1: A chandelier with $L = 3$, $M = 2$, $N = 4$, rooted at the solid vertex. The wires are shown in red, and the bulbs in blue. In this case $R = 1$ since each bulb has no non-trivial automorphism (as rooted graphs).

Then we deduce that

$$\frac{\mathbb{E}\left[\Phi_{ii}\right]^2}{\mathrm{Var}(\Phi_{ij})} = \sum_{H \in \mathcal{H}} \rho^{2|H|}$$

If we still restrict $\mathcal{H}$ to be the set $\mathcal{T}$ of unlabeled, uniquely rooted trees with K edges, then the above quantity diverges when $\rho^2 > \alpha$ and $K \to \infty$. In order to apply a union bound over all the $n(n-1)$ fake pairs, we further need $K \asymp \log n$, so that

$$\frac{\mathbb{E}\left[\Phi_{ii}\right]^2}{\mathrm{Var}(\Phi_{ij})} = \sum_{H \in \mathcal{H}} \rho^{2|H|} = \omega(n^2).$$

Note that in contrast to testing which only needs $K = O(\log n / \log\log n)$, here we need much larger trees with $\Theta(\log n)$ edges. This will incur significant analysis challenges when we bound $\mathrm{Var}(\Phi_{ij})$.

Unfortunately, the cross-correlation plays a significant role and cannot be ignored. In fact, we are unable to bound $\mathrm{Var}(\Phi_{ij})$ when we restrict $\mathcal{H}$ to the set of all unlabeled, uniquely rooted trees. To resolve this challenge, we propose to count a special family $\mathcal{T}^*$ of unlabeled rooted trees, which we call chandeliers.

Definition 4.1 (Chandelier)**.** An (L, M, N, R)-*chandelier* is a rooted tree with L branches, each of which consists of a path with M edges (which we call an M-*wire*) followed by a rooted tree with N edges (which we call a N-*bulb*); the N-bulbs are non-isomorphic to each other and each of them has at most R automorphisms.

Each (L, M, N, R)-chandelier has $K = L(N + M)$ edges in total. The chandelier structure plays a crucial role in curbing the undesired correlation between different tree counts. Moreover, even though chandeliers only occupy a vanishing fraction of all trees, by choosing the parameters appropriately, we can ensure that $|\mathcal{T}^*| = (1/\alpha + o(1))^K$, which grows almost at the same rate as the entire family of trees. We refer the interested reader to [MWXY22] for more details. Below, we summarize the final variance bounds.

$$\frac{\mathbb{E}\left[\Phi_{ii}\right]^2}{\mathrm{Var}(\Phi_{ij})} \lesssim \begin{cases} |\mathcal{T}^*|\rho^{2K} & \overset{\rho^2>\alpha, K \asymp \log n}{=} \omega\left(n^2\right) & \text{if } j = i \\ \frac{nq}{L^2} & \underset{L^2 = o(nq)}{=} \omega(1) & \text{if } j \neq i. \end{cases}$$

With the above variane bound, we are able to establish the following main result.

Theorem 4.2 ([MWXY22]). *Assume that $0 < q \leq 1/2$ and $\rho^2 > \alpha$. Choose $\tau = C\mathbb{E}[\Phi_{ii}]$ for any constant $0 < C < 1$. The following holds:*

- (Partial recovery) *For any constant $\delta \in (0,1)$, there is a constant $C(\rho, \delta) > 0$ depending only on ρ and δ such that if $nq \geq C(\rho, \delta)$, there is a subset $I \subset [n]$ and a map $\widehat{\pi} : I \to [n]$ satisfy that $\widehat{\pi} = \pi|_I$ with high probability and $\mathbb{E}[|I|] \geq (1-\delta)n$.*
- (Almost exact recovery) *If $nq = \omega(1)$, there is a subset $I \subset [n]$ and a map $\widehat{\pi} : I \to [n]$ such that $\widehat{\pi} = \pi|_I$ and $|I| = (1 - o(1))n$ with high probability.*
- (Exact recovery) *If $\rho > 0$ and $nq(q + \rho(1-q)) \geq (1+\epsilon)\log n$ for any constant $\epsilon > 0$,*[1] *then the almost exact recovery can be further made exact with high probability using a seeded graph matching based on the seed set I.*

4.4 Limit of local algorithms

In the previous sections, we have estalished that efficient testing and recovery is possible if $\rho^2 > \alpha$. You may wonder if there is any fundamental computational barrier at the Otter's threshold. Well, it turns out that $\rho^2 > \alpha$ corresponds to the limit of local algorithms.

Definition 4.2. We say an estimator is a d-local algorithm if it outputs a set S of node pairs such that

$$\mathbb{1}\{(i,j) \in S\} = f\left(N_d^A(i), N_d^B(j)\right),$$

where f is a boolen function which maps two rooted subgraphs to $\{0,1\}$, and $N_d^A(i)$ (resp. $N_d^B(j)$) is the rooted subgraphs induced by vertices whose distance from i (resp. j) is at most d in A (resp. B).

We say a local algorithm succeeds in partial recovery, if

- $\mathbb{P}[(i, \pi(i)) \in S] = \Omega(1)$;
- $\mathbb{P}[(i,j) \in S] = o(1)$ for any $j \neq \pi(i)$.

Note that the second bullet only requires that the probability of misclassifying a fake pair is $o(1)$. Since there are $n(n-1)$ fake pairs in total, S may still contain $o(n^2)$ fake pairs on average. Hence, this is a quite weak requirement. Nevertheless, we will show that in the sparse regime $q = \lambda/n$ for a constant λ, as long as $d = o(\log n)$, $\rho^2 > \alpha$ is necessary for any d-local algorithm to achieve the above notion of partial recovery.

To establish such a result, we first reduce the partial recovery problem to a hypothesis testing problem on trees. It is well-known that in the sparse regime, as long as $d = o(\log n)$, $N_d^A(i)$ and $N_d^B(j)$ can be coupled as two Galton-Watson trees t, t' with Poi(λ) offspring distribution. In particular, we have

[1] The condition $nq(q + \rho(1-q)) \geq (1+\epsilon)\log n$ is information-theoretically necessary, for otherwise the intersection graph between A and B (under the vertex correspondence π) contains isolated vertices with high probability and exact recovery is impossible.

- $j = \pi(i)$: t and t' are correlated;
- $j \neq \pi(i)$: t and t' are independent.

Thus, the recovery problem reduces to a problem of detecting correlation in two Galton-Watson trees. To formally define the hypothesis testing problem, we first rigorously define the Galton-Watson measure on unlabeled, rooted trees. The following material is from [GMS22].

Definition 4.3 (Unlabeled rooted tree)**.** Let $\mathcal{X}_d$ denote the set of all unlabeled rooted trees of depth at most d defined recursively as follows.

- $\mathcal{X}_0 = \{\bullet\}$.
- For $d \geq 1$, any $t \in \mathcal{X}_d$ can be represented as $t = (N_\tau)_{\tau \in \mathcal{X}_{d-1}}$, where N_τ is the number of children of the root whose subtrees are equal to τ.

Example:

- If $d = 1$ and $t = \circ\!\!-\!\!\bullet\!\!-\!\!\circ$, then $t = (N_\bullet = 2)$.
- If $d = 2$ and $t = \substack{\circ\\ \circ}\!\!>\!\!\circ\!\!-\!\!\bullet\!\!-\!\!\circ$, then $t = (N_\bullet = 1, N_{\circ\!-\!\bullet\!-\!\circ} = 1)$.

Definition 4.4 (Galton-Watson distribution)**.** Let GW_d^λ denote the Galton-Watson distribution on $\mathcal{X}_d$ defined recursively as follows:

- $\mathrm{GW}_d^\lambda = \delta_\bullet$, where δ denotes a delta measure;
- Let $t = (N_\tau)_{\tau \in \mathcal{X}_{d-1}} \sim \mathrm{GW}_d^\lambda$, if $N_\tau \overset{\text{i.i.d.}}{\sim} \mathrm{Poi}\left(\lambda \mathrm{GW}_{d-1}^\lambda(\tau)\right)$.

Example:

- If $d = 1$ and $t = (N_\bullet) \sim \mathrm{GW}_1^\lambda$, then $N_\bullet \sim \mathrm{Poi}(\lambda)$.
- If $d = 2$ and $t = (N_\tau)_{\tau \in \mathcal{X}_1}$, then $N_{\circ\!-\!\bullet\!-\!\circ} \overset{\text{i.i.d.}}{\sim} \mathrm{Poi}\left(\lambda \mathrm{GW}_1^\lambda(\circ\!\!-\!\!\bullet\!\!-\!\!\circ)\right)$.

Definition 4.5 (Correlated Galton-Watson trees)**.** Let $P_d^{\lambda,s}$ denote the distribution of the two correlated Galton-Watson trees defined recursively as follows:

- $P_0^{\lambda,s} = \mathrm{GW}_0^\lambda \otimes \mathrm{GW}_0^\lambda$;
- Let $(t, t') \sim P_d^{\lambda,s}$ where $t = (N_\tau)_{\tau \in \mathcal{X}_{d-1}}$ and $t' = (N_{\tau'})_{\tau' \in \mathcal{X}_{d-1}}$, if

  $$N_\tau = \Delta_\tau + \sum_{\tau' \in \mathcal{X}_{d-1}} M_{\tau,\tau'}$$
  $$N_{\tau'} = \Delta'_{\tau'} + \sum_{\tau \in \mathcal{X}_{d-1}} M_{\tau,\tau'},$$

  where $\{\Delta_\tau\}, \{\Delta'_{\tau'}\}, \{M_{\tau,\tau'}\}$ are mutually independent, and

  $\Delta_\tau \overset{\text{i.i.d.}}{\sim} \mathrm{Poi}\left(\lambda(1-s)\mathrm{GW}_{d-1}^\lambda(\tau)\right)$, $\Delta'_{\tau'} \overset{\text{i.i.d.}}{\sim} \mathrm{Poi}\left(\lambda(1-s)(\mathrm{GW}_{d-1}^\lambda(\tau'))\right)$, $M_{\tau,\tau'} \overset{\text{i.i.d.}}{\sim} \mathrm{Poi}\left(\lambda s P_{d-1}^{\lambda,s}(\tau,\tau')\right)$.

Note that $P_d^{\lambda,0} = \mathrm{GW}_d^\lambda \otimes \mathrm{GW}_d^\lambda$, which is the joint distribution of two independent Galton-Watson trees.

Now, we are ready to formally define the tree correlation detection problem:

$$\begin{cases} H_0 : & (t,t') \sim P_d^{\lambda,0} \triangleq P_d^\lambda \\ H_1 : & (t,t') \sim P_d^{\lambda,s} \end{cases}.$$

Remark 4.1. Note that $P_d^{\lambda,s}(\bullet,\bullet) = e^{-2\lambda+\lambda s}$. It follows that

$$\frac{1}{2}\left(e^{-2\lambda+\lambda s} - e^{-2\lambda}\right) \le \mathrm{TV}(P_d^{\lambda,s}, P_d^\lambda) \le 1 - e^{-2\lambda}.$$

Therefore, the weak detection always holds and the strong detection never holds. Inspired by the partial recovery requirement, the right notion to consider is the so-called one-sided detection, that is,

$$\begin{aligned} &\mathbb{P}_{H_1}(T(t,t') = 1) = \omega(1) \quad \text{positive power} \\ &\mathbb{P}_{H_0}(T(t,t') = 1) = o(1) \quad \text{vanishing type-I error.} \end{aligned}$$

The statistical limit for the one-sided detection is established in [GMS22].

Theorem 4.3 ([GMS22])**.**

- *If $s \le \sqrt{\alpha}$, then the one-sided detection is impossible for all λ;*
- *If $s > \sqrt{\alpha}$, then there exists $\lambda(s) > 0$ such that for all $\lambda \ge \lambda(s)$, the one-sided detection is feasible.*

In this lecture, we focus on proving the negative result.

Proof of the negative direction. Let

$$L_d(t,t') = \frac{P_d^{\lambda,s}(t,t')}{P_d^\lambda(t,t')}$$

denote the likelihood ratio. Let $\mathbb{E}_d^\lambda$ denote the expectation taken under the measure P_d^λ. Then it suffices to show that

$$\mathbb{E}_d^\lambda\left[L_d^2\right] < +\infty,$$

which implies that $P_d^{\lambda,s}$ is asymptotically contiguous to P_d^λ and hence the one-sided detection is impossible. To bound the second moment, the key is to exploit an orthogonal decompsotion of L_d in certain polynomial basis.

Lemma 4.1 ([GMS22])**.** *For $d \ge 0$, there exists a collection of $\{f_{d,\beta}^\lambda\}_{\beta \in \mathcal{X}_d}$ with $f_{d,\beta}^\lambda : \mathcal{X}_d \to \mathbb{R}$ such that*

$$L_d(t,t') = \sum_{\beta \in \mathcal{X}_d} s^{|\beta|} f_{d,\beta}^\lambda(t) f_{d,\beta}^\lambda(t'),$$

where

- *$f_{d,\bullet} \equiv 1$;*

- $\mathbb{E}_{t\sim \mathrm{GW}_d^\lambda}[f_{d,\beta}^\lambda(t) f_{d,\beta'}^\lambda(t)] = \mathbb{1}\{\beta = \beta'\}$.

It turns out that given $\beta = (\beta_\tau)_{\tau\in\mathcal{X}_{d-1}}$, $f_{d,\beta}^\lambda(t)$ is a polynomial of entries $\{t_\tau\}_{\tau\in\mathcal{X}_{d-1}}$ with degree at most $\sum_{\tau\in\mathcal{X}_{d-1}} \beta_\tau$. With this lemma, the second moment can be readily bounded as follows:

$$\mathbb{E}_d^\lambda\left[L_d^2\right] = \sum_{\beta\in\mathcal{X}_d} s^{2|\beta|} \le \sum_{n\ge 0} s^{2n} A_n < +\infty,$$

where $A_n \lesssim n^{-3/2}\alpha^{-n}$ denote the number of unlabeled, rooted trees with n edge due to (4.1), and the last inequality holds by the assumption that $s^2 \le \alpha$. □

It remains to prove Lemma 4.1. The proof follows from the induction on depth d. Here for simplicity we verify the lemma for $d = 1$. The induction from d to $d+1$ follows from similar derivations but with more tedious calculations. We refer the interested reader to [GMS22].

Proof of Lemma 4.1 for $d = 1$. Recall that any $t = (N_\bullet) \in \mathcal{X}_1$ can be identifies as an integer $N_\bullet$. Thus we can map two trees (t, t') to two integers (ℓ, ℓ'). Define the characteristic function of $P_1^{\lambda,s} : \mathbb{R}^2 \to \mathbb{R}$ as

$$\widehat{P}_1^{\lambda,s}(\theta,\theta') \triangleq \mathbb{E}_{(\ell,\ell')\sim P_1^{\lambda,s}}\left[e^{\mathbf{i}\theta\ell + \mathbf{i}\theta'\ell'}\right].$$

By the definition of $P_1^{\lambda,s}$, we can write $ell = \Delta + M$ and $\ell' = \Delta' + M$, where $\Delta \sim \mathrm{Poi}(\lambda(1-s))$, $\Delta' \sim \mathrm{Poi}(\lambda(1-s))$, and $M \sim \mathrm{Poi}(\lambda s)$ are mutually independent. Therefore,

$$\begin{aligned}
\widehat{P}_1^{\lambda,s}(\theta,\theta') &= \mathbb{E}\left[e^{\mathbf{i}\theta\Delta}\right]\mathbb{E}\left[e^{\mathbf{i}\theta'\Delta'}\right]\mathbb{E}\left[e^{\mathbf{i}(\theta+\theta')M}\right] \\
&= \exp\left(\lambda(1-s)(e^{\mathbf{i}\theta}-1) + \lambda(1-s)(e^{\mathbf{i}\theta'}-1) + \lambda s(e^{\mathbf{i}(\theta+\theta')}-1)\right) \\
&= \exp\left(\lambda(e^{\mathbf{i}\theta}-1) + \lambda(e^{\mathbf{i}\theta'}-1) + \lambda s(e^{\mathbf{i}\theta}-1)(e^{\mathbf{i}\theta'}-1)\right) \\
&= \exp\left(\lambda(e^{\mathbf{i}\theta}-1) + \lambda(e^{\mathbf{i}\theta'}-1)\right)\sum_{m\ge 0}\frac{(\lambda s)^m}{m!}(e^{\mathbf{i}\theta}-1)^m(e^{\mathbf{i}\theta'}-1)^m \\
&= \sum_{m\ge 0} s^m \widehat{g}_{1,m}^\lambda(\theta)\widehat{g}_{1,m}^\lambda(\theta'),
\end{aligned}$$

where

$$\widehat{g}_{1,m}^\lambda(\theta) = \sqrt{\frac{(\lambda s)^m}{m!}}(e^{\mathbf{i}\theta}-1)^m \exp\left(\lambda(e^{\mathbf{i}\theta}-1)\right).$$

Invertning the fourier transform, we get that

$$P_1^{\lambda,s}(t,t') = \sum_{m\ge 0} s^m g_{1,m}^\lambda(\ell) g_{1,m}^\lambda(\ell'),$$

where

$$g_{1,m}^\lambda(\ell) = \int_0^{2\pi}\frac{\mathrm{d}\theta}{2\pi}e^{-\mathbf{i}\theta\ell}\widehat{g}_{1,m}^\lambda(\theta) = \sqrt{m!}[x^m]e^{-\lambda-\sqrt{\lambda}x}\frac{(\lambda+x\sqrt{\lambda})^\ell}{\ell!},$$

where $[x^m]$ denotes the coefficient of the monomial x^m in the following power series expansion. Note that

$$P_1^\lambda(t,t') = P_1^{\lambda,0}(t,t') = g_{1,0}^\lambda(\ell) g_{1,0}^\lambda(\ell').$$

Therefore,

$$L_1(t, t') = \sum_{m \geq 0} s^m f^\lambda_{1,m}(\ell) f^\lambda_{1,m}(\ell'),$$

where

$$f^\lambda_{1,m}(\ell) = \frac{g^\lambda_{1,m}(\theta)}{g^\lambda_{1,0}(\theta)} = \sqrt{m!}[x^m]e^{-\sqrt{\lambda}x}\left(1 + \frac{x}{\sqrt{\lambda}}\right)^\ell.$$

Note that $f^\lambda_{1,m}(\ell)$ is known as Charlier polynomial, which is the orthogonal polynomial under the $\mathrm{Poi}(\lambda)$ distribution. In particular, it is easy to verify that

- $f^\lambda_{1,0}(\ell) \equiv 1$;
- $\mathbb{E}_{\ell \sim \mathrm{Poi}(\lambda)}\left[f^\lambda_{1,m}(\ell) f^\lambda_{1,m'}(\ell)\right] = \mathbb{1}\{m = m'\}$.

□

4.5 Open questions

There are a number of interesting questions left open.

- (Dis)prove the computational hardness conjecture below Otter's threshold in the sparse regime $nq = \mathsf{polylog}(n)$;
- Imrpove over the Otter's threshold in the dense regime with $nq = n^{\Omega(1)}$. Note that partial progress has been made in [DL22] under the Gaussian model in which efficient recovery is achieved for $\rho \geq \epsilon$ for any small constant ϵ.
- Low-degree polynomial method beyond the Erdős-Rényi model.

References

[AYZ95] Noga Alon, Raphael Yuster, and Uri Zwick. Color-coding. *Journal of the ACM (JACM)*, 42(4):844–856, 1995.

[DL22] Jian Ding and Zhangsong Li. A polynomial time iterative algorithm for matching gaussian matrices with non-vanishing correlation. *arXiv preprint arXiv:2212.13677*, 2022.

[GMS22] Luca Ganassali, Laurent Massoulié, and Guilhem Semerjian. Statistical limits of correlation detection in trees. *arXiv preprint arXiv:2209.13723*, 2022.

[Hop18] Samuel Hopkins. *Statistical inference and the sum of squares method.* PhD thesis, Cornell University, 2018.

[HS17] Samuel B Hopkins and David Steurer. Efficient bayesian estimation from few samples: Community detection and related problems. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 379–390, Los Alamitos, CA, USA, Oct 2017. IEEE Computer Society.

[KWB19] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. *arXiv preprint arXiv:1907.11636*, 2019.

[MWXY21] Cheng Mao, Yihong Wu, Jiaming Xu, and Sophie H Yu. Testing network correlation efficiently via counting trees. *arXiv preprint arXiv:2110.11816. Annals of Statistics*, 2021.

[MWXY22] Cheng Mao, Yihong Wu, Jiaming Xu, and Sophie H Yu. Random graph matching at otter's threshold via counting chandeliers. *arXiv preprint arXiv:2209.12313. STOC 2023*, 2022.

[Ott48] Richard Otter. The number of trees. *Annals of Mathematics*, pages 583–599, 1948.