GRAPH ALIGNMENT PROBLEM FOR TWO INDEPENDENT ERDŐS-RÉNYI GRAPHS: INFORMATIONAL AND COMPUTATIONAL THRESHOLDS

Hang Du (MIT), Jian Ding (PKU), Shuyang Gong (PKU), Rundong Huang (PKU)

1.Motivation

- **Graph alignment problem** is an important combinatorial optimization problem which has applications in various fields.
- The problem is **NP-hard** in the worst case, and even finding near optimal solution is computationally intractable in general.
- There are extensive studies for the problem (properties of optimal solutions, efficient algorithms...) on graphs of specific types, e.g., sparse graphs, correlated random graphs...
- In our works [DDG22] and [DGH23], we consider the maximal overlap of two **independent Erdős-Rényi graphs**.

3.Informational results

- It is clear that $\mathbb{E}\,\mathrm{O}(\pi) = \binom{n}{2}p^2$ for any π.
- There is a **phase transition at $\mathbf{p_c} := \sqrt{\log \mathbf{n}/\mathbf{n}}$**: when $p \ll p_c$, $\Gamma_{\mathrm{OPT}} \gg \binom{n}{2}p^2$, and when $p \gg p_c$, $\Gamma_{\mathrm{OPT}} \sim \binom{n}{2}p^2$, w.h.p.
- Out first theorem characterize the asymptotics of Γ_{OPT} in the sparse regime, and the second order asymptotics in the dense regime.

Theorem (Informational results)**.** *let* $p_c = \sqrt{\log n/n}$.
(Sparse regime) For $\log n/n \ll p \ll p_c$, $S_{n,p} := \frac{n\log n}{\log\left(\frac{\log n}{np^2}\right)}$,

$$\frac{\Gamma_{\mathrm{OPT}}}{S_{n,p}} \xrightarrow{\text{in probability}} 1.$$

(Dense regime) For $p_c \ll p \ll 1$, $D_{n,p} := \sqrt{n^3p^2\log n}$,

$$\frac{\Gamma_{\mathrm{OPT}} - \binom{n}{2}p^2}{D_{n,p}} \xrightarrow{\text{in probability}} 1.$$

Basic proof strategy:

- Upper bound: the first moment method.
- Lower bound for dense regime: the second moment method + concentration inequality.
- Lower bound for sparse regime: a constructive proof via analyzing a greedy type algorithm.

5.Hardness result for online algorithms

- We justify that $\widetilde{\Gamma}_{\mathrm{ALG}} = \sqrt{8/9}\cdot\widetilde{\Gamma}_{\mathrm{OPT}}$ by proving a hardness result for online algorithms.
- **Online algorithms**: assume that G_1 is coonstructed vertex by vertex online, while G_2 is off-line saved. An online matching algorithm requires to match a vertex of G_1 immediately at its construction.
- The iterative greedy matching algorithms is an online algorithm.

Theorem (Hardness for online algorithms)**.** *For any* $\varepsilon > 0$*, there exists* $c > 0$ *such that for any online matching algorithm, its output* π^* *satisfies that*

$$\mathbb{P}\left[\mathrm{O}(\pi^*) \ge \binom{n}{2}p^2 + (\sqrt{8/9}+\varepsilon)D_{n,p}\right] \le \exp(-cn\log n)\,.$$

- The proof employs the **branching-OGP** framework (where OGP stands for the overlap gap property) introduced in [HS21].

2.Mathematical settings

- Erdős-Rényi graph: a random graph with each edge in K_n preserved independently with probability p.
- Let $\mathbb{P} = \mathbb{P}_{n,p}$ be the law of a pair of independent Erdős-Rényi graphs $(\mathsf{G}_1, \mathsf{G}_2)$ with n vertices and edge density p.
- **Question**: Find a bijection between the vertex sets such that the size of overlap is as large as possible.
- Formally, for a bijection $\pi : \mathsf{V}_1 \to \mathsf{V}_2$,

$$\mathrm{O}(\pi) = \sum_{u\neq v} \mathbf{1}_{(u,v)\in E_1}\mathbf{1}_{(\pi(u),\pi(v))\in E_2}.$$

Our focus is twofold:

- The asymptotics of $\mathbf{\Gamma}_{\mathrm{OPT}} = \max_\pi \mathrm{O}(\pi)$ under $\mathbb{P}$.
- The best performance of **efficient algorithms $\mathbf{\Gamma}_{\mathrm{ALG}}$** under $\mathbb{P}$.

4. Algorithmic results

- The **greedy iterative matching algorithm**: successively for each $i \in V(\mathsf{G}_1)$, set $\pi(i)$ to be an unmatched $j \in V(\mathsf{G}_2)$ that maximizes
$$\sum_{k\prec i} \mathbf{1}_{(k,i)\in E(G_1)}\mathbf{1}_{(\pi(k),j)\in E(G_2)}\,.$$
- This simple algorithm turns out to reach the heart of the computational aspect to this random optimization problem.
- In the sparse regime, variants of the iterative greedy matching algorithm suggests $\mathbf{\Gamma}_{\mathrm{ALG}} = \mathbf{\Gamma}_{\mathrm{OPT}}$.

Theorem (PTAS in sparse regime)**.** *For* $\log n/n \ll p \ll p_c$*, for any fixed* $\varepsilon > 0$*, there exists a polynomial-time algorithm which takes* G_1 *and* G_2 *as input and outputs a bijection* π^* *such that,*

$$\mathbb{P}\left[\mathrm{O}(\pi^*) > (1-\varepsilon)S_{n,p}\right] = 1 - o(1).$$

- In the dense regime, let $\widetilde{\Gamma}_{\mathrm{OPT}} = \Gamma_{\mathrm{OPT}} - \binom{n}{2}p^2$, $\widetilde{\Gamma}_{\mathrm{ALG}} = \Gamma_{\mathrm{ALG}} - \binom{n}{2}p^2$.
- The above algorithm gives $\widetilde{\mathbf{\Gamma}}_{\mathrm{ALG}} \ge \sqrt{\mathbf{8/9}}\cdot\widetilde{\mathbf{\Gamma}}_{\mathrm{OPT}}$.

Theorem. *For* $p_c \ll p \ll 1/(\log n)^3$*, the output of the iterative greedy matching algorithm* π^* *satisfies that*

$$\mathbb{P}\left[\mathrm{O}(\pi^*) \ge \binom{n}{2}p^2 + (\sqrt{8/9} - o(1))D_{n,p}\right] = 1 - o(1).$$

References

[DDG22] J. Ding, H. Du and S. Gong, "A polynomial-time approximation scheme for the maximal overlap of two independent Erdős-Rényi graphs", *Random Structures & Algorithms* (2024), pp. 1-38.

[DGH23] H. Du, S. Gong and R. Huang, "The algorithmic phase transition of random graph alignment problem", preprint, *arXiv:2307.06590*.

[HS21] B. Huang and M. Sellke, "Tight Lipschitz Hardness for optimizing Mean Field Spin Glasses", *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, Denver, CO, USA, 2022, pp. 312-322.

LaTeX TikZposter