A Polynomial-time Iterative Algorithm for Random Graph Matching with Non-vanishing Correlation

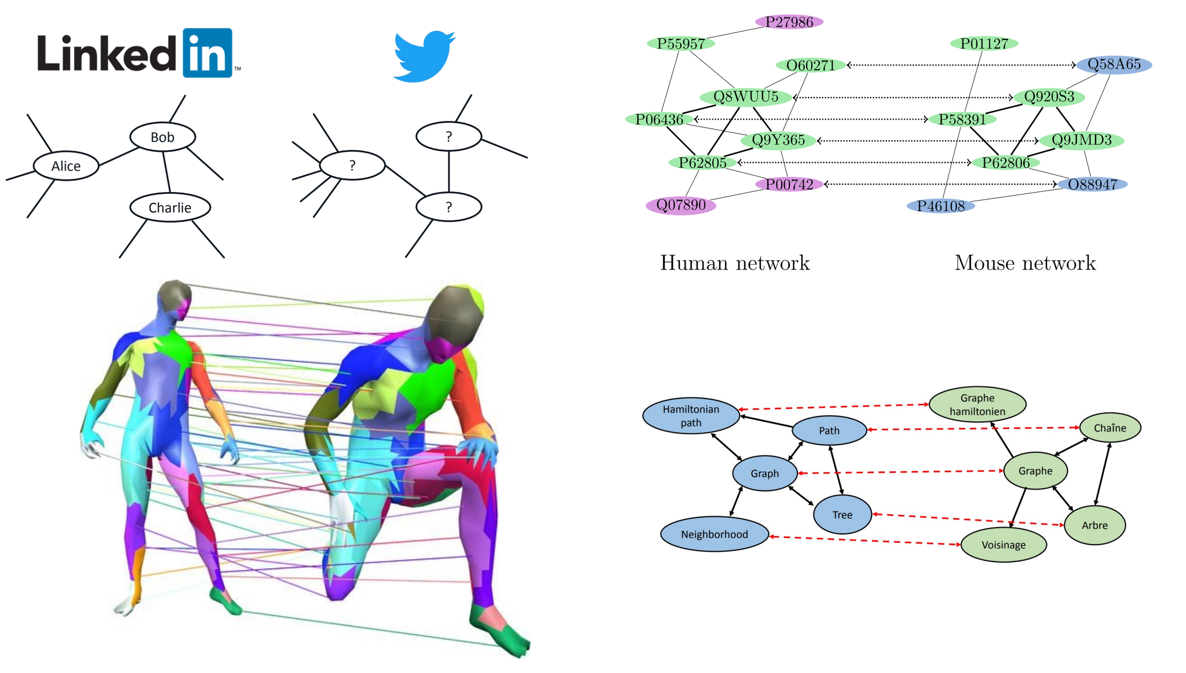
Jian Ding, Zhangsong Li

Peking University

1.Motivation

- Consider a pair of graphs with **(positive) correlated structure** through certain vertex correspondence.
- In most scenarios, the labels of vertices are lost and the only observable is the **topology structure** of graphs.
- People are concerned about recovering the hidden correspondence based on such sheer information, as well as the closely related correlation detection problem.
- Questions of this type are important combinatorial optimization problems which have applications in various fields:

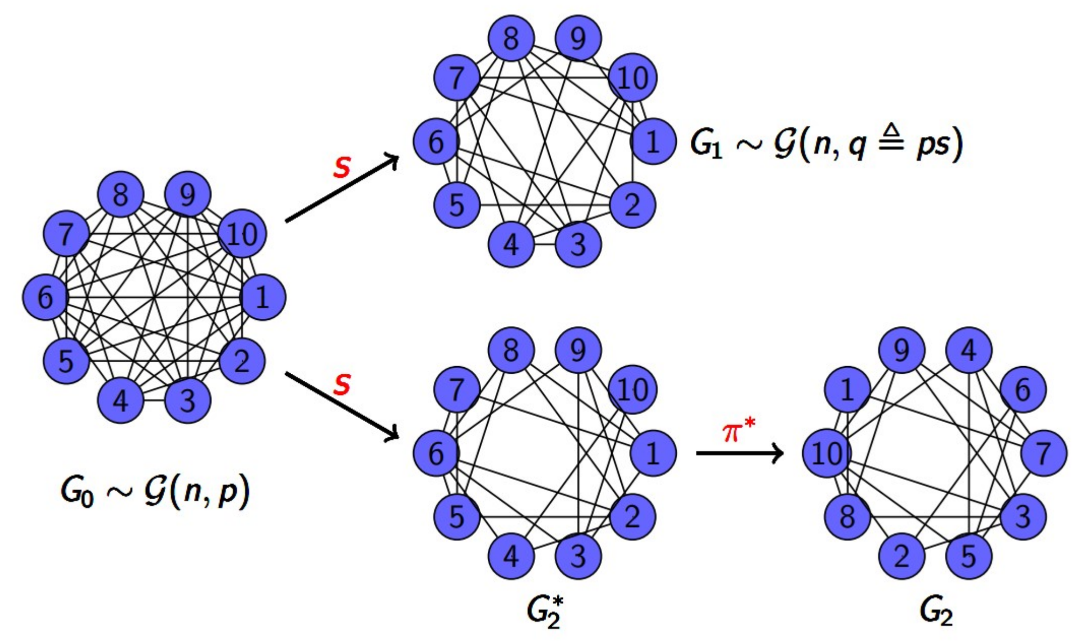
  social network de-anonymization;
  protein-protein interaction;
  computer vision;
  machine translation.



- In our works [DL22+] and [DL23+], we consider the algorithmic perspective for this problem in *correlated Erdős-Rényi graphs model*.

2.Mathematical settings

- Denote $\mathcal{G}(n,p)$ for the law of Erdős-Rényi graph on n vertices with edge density p (i.e. each edge is kept independently with probability p).
- Consider the following idealistic model:
  - fix $n \in \mathbb{N}$, $p, s \in (0,1)$;
  - Sample a "mother graph" $G_0 \sim \mathcal{G}(n,p)$;
  - Independently subsample $G_1, G_2^* \subset G_0$ with subsampling probability s;
  - Relabel G_2^* by a uniform permutation π^* to obtain G_2.



- $\mathbb{P}$: the law of (G_1, G_2) generated as above;
  $\mathbb{Q}$: the law of two independent $\mathcal{G}(n, ps)$ graphs.
- The fundamental problems for this model:
  - **Detection**: given (G_1, G_2), determine whether it is sampled from $\mathbb{P}$ or $\mathbb{Q}$. (e.g. applications in distinguishing objects by computer.)
  - **Recovery**: given $(G_1, G_2) \sim \mathbb{P}$, recover π^* as good as possible. (e.g., applications in de-anonymizing social networks.)
- Believed to exhibit **information-computation gap**, a major challenge in many random combinatorial optimization problems.
- By the collective effort of the community (for example, see [WXY21], [DD22a] and [DD22b]), now we have a fairly complete understanding of the information thresholds for the problem of correlation detection and matching recovery. In contrast, the understanding of the computational aspect is far from being complete.
- Intuitively, the matching π^* that maximizes the common edge between two graphs (i.e. the MLE) should be the most effective estimator for recovering the latent matching π.
- This intuition helps us to reduce the problem into the **network alignment** problem of correlated random graphs.
- Unfortunately, the classical graph alignment problem is a **NP-hard** optimization problem, so we must seek help from randomness.
- Previously, arguably the best result is the recent work [MWXY23] that obtained a polynomial time algorithm that succeeds as long as the correlation is above the square root of the Otter's constant.

3. Previous results

- The existing algorithms are essentially of two types:
  - the optimization-based method that relies on "convex relaxation and rounding": original optimization problem is hard to solve, but feasible if we enlarge the space of potential solutions ([FMWX22a, FMWX22b]);
  - the signature-based method that relies on "computing and comparing signatures": for each vertex, compute a "signature" and match pairs of vertices with similar signatures. ([BCL+19, DMWX21]).
- In a later breakthrough [MRT23], the authors found the first polynomial time algorithm (based on some sophisticated partition tree) that succeeds for exact matching with correlation a constant close to 1;
- In a recent breakthrough [MWXY23], the authors substantially improved [MRT23] and obtained a polynomial time algorithm which succeeds as long as the correlation is above $\sqrt{\alpha}$, where $\alpha \approx 0.338$ is the Otter's constant.

3.Our results

Theorem 1 (D.-L.'2022)**.** *There is a polynomial time iterative algorithm for matching Gaussian matrices as long as the correlation is non-vanishing.*

- Strongly suggest the barrier at Otter's constant in [MWXY23] is not real in **dense region**;
- New feature: signal is stored in a vector where each coordinate is a pair of sets, and signal per coordinate decreases with iteration but compensated by increase in dimension;
- Might shed light on many other matching problems too.

Sketch of our iterative algorithm:

(1) Obtain a seed $u_i, \pi(u_i) : 1 \leq i \leq K$ via brutal-force searching, where K is a large constant;

(2) At each time t, construct pairs of sets $\left(\Gamma_i^{(t)}, \Pi_i^{(t)}\right), 1 \leq i \leq K_t$ according to the edge weights of each vertex v to the seeds or $\left(\Gamma_i^{(t-1)}, \Pi_i^{(t-1)}\right), 1 \leq i \leq K_{t-1}$.

(3) Main observation: although the signal in each pair $(\Gamma_k^{(t)}, \Pi_k^{(t)})$ will decrease, we may take many linear combinations of those edge weights to increase the number of paired sets, so the total signal is increasing in t.

Theorem 2 (D.-L.'2023)**.** *There is a polynomial time iterative algorithm for matching dense Erdős-Rényi graphs(i.e., when the edge-density q of Erdős-Rényi graphs satisfies $q \geq n^{-1+\alpha+o(1)}$ for some constant $\alpha > 0$) as long as the correlation is non-vanishing.*

- Demonstrates the robustness of the iterative matching algorithm we proposed in [DL22+], this type of **"algorithmic universality"** is closely related to the universality phenomenon in random matrix theory.
- Expected to be sharp in the following sense: as the correlation tends to 0, no polynomial time algorithm with a fixed power would be able to match two random graphs.

References

[BCL+19] *B. Barak, C.-N. Chou, Z. Lei, T. Schramm, and Y. Sheng.* (nearly) efficient algorithms for the graph matching problem on correlated random graphs. In *Nueral IPS*, 2019.

[DD22a] *J. Ding and H. Du.* Detection threshold for correlated erdos-renyi graphs via densest subgraph. In *IEEE Transactions on Information Theory*, online 2023.

[DD22b] *J. Ding and H. Du.* Matching recovery threshold for correlated random graphs. *to appear in Annals of Statistics.*

[DL22+] *J. Ding and Z. Li.* A polynomial time iterative algorithm for matching Gaussian matrices with non-vanishing correlation. *Preprint, arXiv:2212.13677.*

[DL23+] *J. Ding and Z. Li.* A polynomial-time iterative algorithm for random graph matching with non-vanishing correlation. *Preprint, arXiv:2306.00266.*

[DMWX21] *J. Ding, Z. Ma, Y. Wu, and J. Xu.* Efficient random graph matching via degree profiles. In *Probability Theory and Related Fields*, 2021.

[GML22+] *L. Ganassali, L. Massoulié and M. Lelarge.* Statistical limits of correlation detection in trees. *Preprint, arXiv:2209.13723.*

[MRT23] *C. Mao, M. Rudelson, and K. Tikhomirov.* Exact matching of random graphs with constant correlation. In *Probability Theory and Related Fields*, 2023.

[MWXY23] *C. Mao, Y. Wu, J. Xu and S. H. Yu.* Random graph matching at Otter's threshold via counting chandeliers. *to appear in STOC 2023.*

[WXY21] *Y. Wu, J. Xu, and S. H. Yu.* Settling the sharp reconstruction thresholds of random graph matching. In *IEEE Transactions on Information Theory*, 2022

LaTeX TikZposter