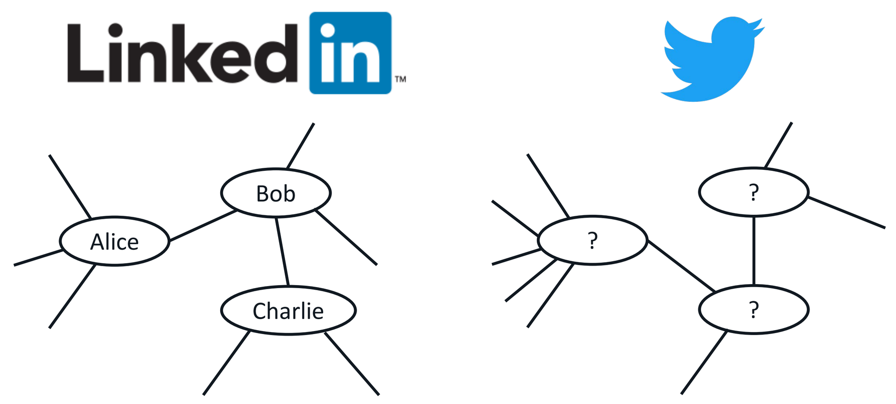
Detection and recovery thresholds for correlated random graphs

Jian Ding, Hang Du

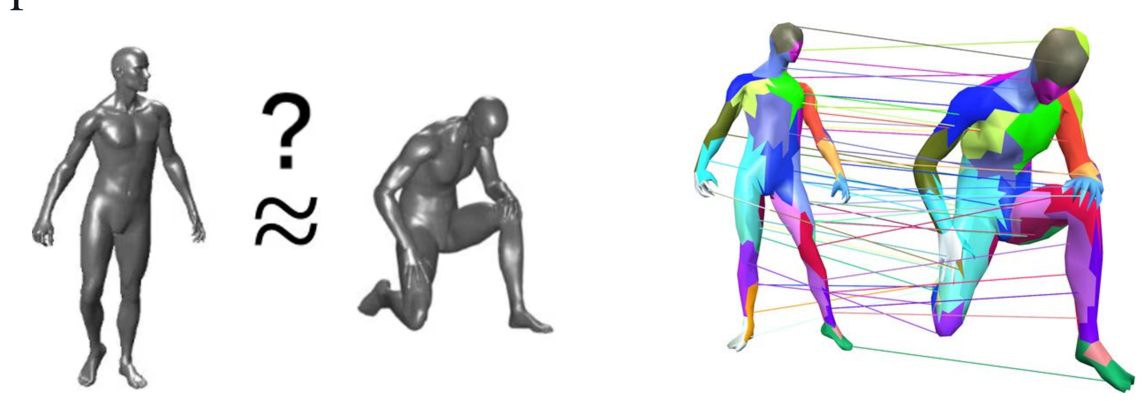
Peking University

1. Motivations

- Consider a pair of graphs with **latent correlated structure**, it is desirable to detect/recover the hidden correlation solely based on their **topological structures**.
- Questions from a wide range of applied fields:
  - Social network de-anonymization



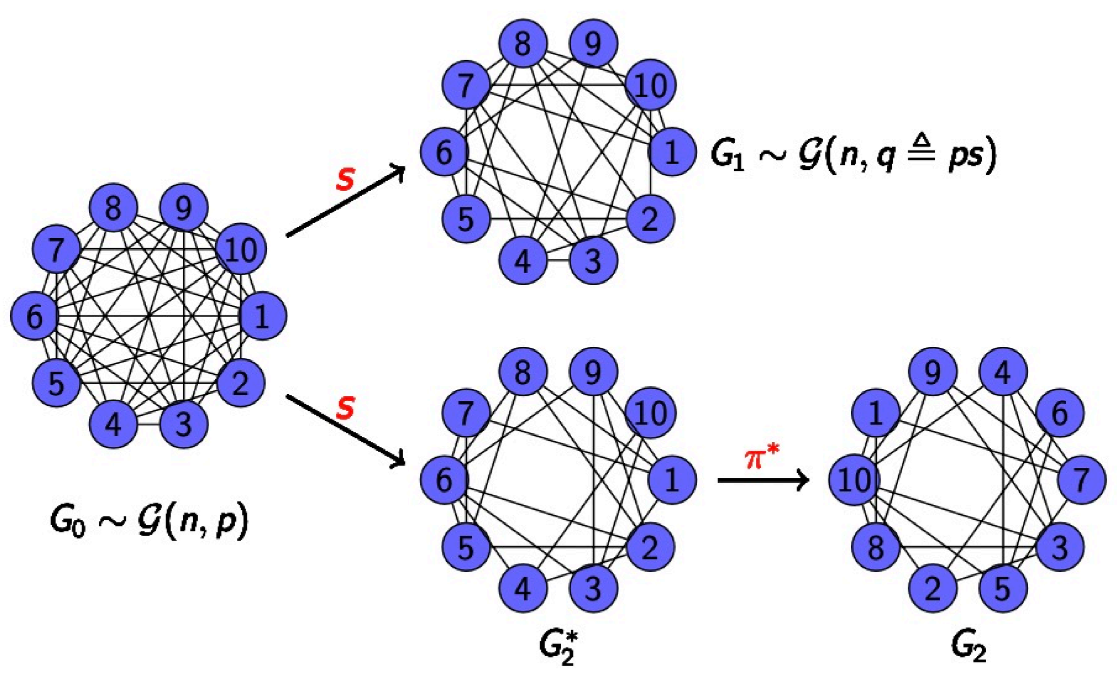
  - Computer vision



  - Biology computing, natural language processing, $\cdots$

2. Mathematical Setting

- $\mathcal{G}(n,p)$: Erdős-Rényi graph on n vertices with edge density p (i.e., each edge is kept independently with probability p).
- Fix $n \in \mathbb{N}, p, s \in (0,1)$, sample $(G_1, G_2) \sim \mathbb{P}$ as follow:
  - Sample a "parent" graph $G_0 \sim \mathcal{G}(n,p)$;
  - Independently sample $G_1, G_2^* \subset G_0$ with density s;
  - Relabel G_2^* by a unifrom permutaion π^* to get G_2.



3. Basic Questions

- $\mathbb{P}$: correlated law sampled as before;
  $\mathbb{Q}$: indepndent law of two $\mathbf{G}(n, ps)$ graphs.
- The fundamental problem for this model is twofold:
  - **Detection**: given (G_1, G_2), do hypothesis testing:
  $$(G_1,G_2) \sim \mathbb{P} \text{ v.s. } (G_1,G_2) \sim \mathbb{Q}.$$
  - **Recovery:** recover the hidden π^* for $(G_1, G_2) \sim \mathbb{P}$.

4. Previous Results

- A natural estimator (MLE):
$$\hat{\pi} = \operatorname{argmax} \mathrm{C}(\pi),$$
where $\mathrm{C}(\pi)$ is the total number of edges in the intersection graph of G_1 and G_2 thorugh π.
- In the **dense regime** ($p = n^{-o(1)}$), Wu, Xu and Yu showed that $\hat{\pi}$ is indeed optimal, and the transition threshold for s is given by
$$s_0 = \sqrt{\frac{2\log n}{np\left(\log\frac{1}{p} - 1 + p\right)}}.$$
- More precisely, in this regime for any $\varepsilon > 0$,
  - if $s > (1+\varepsilon)s_0$, then for some explicit τ it holds
  $$\mathbb{P}[\mathrm{C}(\hat{\pi}) < \tau] + \mathbb{Q}[\mathrm{C}(\hat{\pi}) > \tau] = o(1)\,,$$
  and for any $\delta < 1$ it holds
  $$\mathbb{P}[\mathrm{O}(\hat{\pi}, \pi^*) > \delta n] = 1 - o(1)\,.$$
  - If $s < (1-\varepsilon)s_0$, then $\mathrm{TV}(\mathbb{P},\mathbb{Q}) = o(1)$ (so detection is impossible), and for any estimator $\tilde{\pi}$ and $\delta > 0$,
  $$\mathbb{P}[\mathrm{O}(\tilde{\pi}, \pi^*) > \delta n] = o(1)\,.$$
- In the **sparse regime**, Wu, Xu and Yu also used $\hat{\pi}$ to determine the transition thereshold **up to a multiplicative constant**.
- Furthermore, they determined the threshold for the possibility of **exact recovery** (when this can be done, it is also achieved by $\hat{\pi}$).

5. Our Results

- We establish sharp informational thresholds for **detection** and **partial recovery** in the **sparse regime**.
- Our novel estimator:
$$\check{\pi} = \operatorname{argmax} D(\pi),$$
where $\mathrm{D}(\pi)$ stands for the **maximal edge-vertex ratio** over all subgraphs of the intersection graph of G_1 and G_2 through π.
- The main technical imput for us is the following:

Theorem (Anatharam-Salez'16)**.**
There is a function $\varrho : [1,\infty) \to [1,\infty)$ such that $\forall c \geq 1$, for $G \sim \mathcal{G}(n.c/n)$, it holds that
$$\max_{\emptyset \neq H \subset G} \frac{|E(H)|}{|V(H)|} \to \varrho(c) \text{ in probability as } n \to \infty.$$

- We obtain the sharp thereshold in terms of ϱ:

Theorem (D.-D.'22)**.**
Assume $p = n^{-\alpha+o(1)}$ for some constant $\alpha \in (0,1)$, and let $\lambda_* = \varrho^{-1}(\alpha^{-1})$. Then for any $\varepsilon > 0$,
  - *If $nps^2 \geq \lambda_* + \varepsilon$, then for some τ it holds*
  $$\mathbb{P}[\mathrm{D}(\check{\pi}) < \tau] + \mathbb{Q}[\check{\pi} > \tau] = o(1)\,,$$
  and these exists some $\delta > 0$ such that
  $$\mathbb{P}[\mathrm{O}(\check{\pi}, \pi^*) > \delta n] = 1 - o(1)\,.$$
  - *If $nps^2 \leq \lambda_* - \varepsilon$, then $\mathrm{TV}(\mathbb{P},\mathbb{Q}) = o(1)$, and for any estimator $\tilde{\pi}$ and $\delta > 0$,*
  $$\mathbb{P}[\mathrm{O}(\tilde{\pi}, \pi^*) > \delta n] = o(1)\,.$$

References

J. Ding and H. Du, *Detection threshold for correlated Erdős-Rényi graphs via densest subgraphs*, IEEE Trans. Inf. Theory.

J. Ding and H. Du, *Matching Recovery Threshold for Correlated Random Graphs*, Ann. Stat.

Y. Wu, J. Xu and S. H. Yu, *Testing Correlation of Unlabeled Random Graphs*, Ann. Appl. Probab.

Y. Wu, J. Xu, S. H. Yu, *Setting the Sharp Reconstruction Thresholds for Random Graph Matching*, IEEE Trans. Inf. Theory.