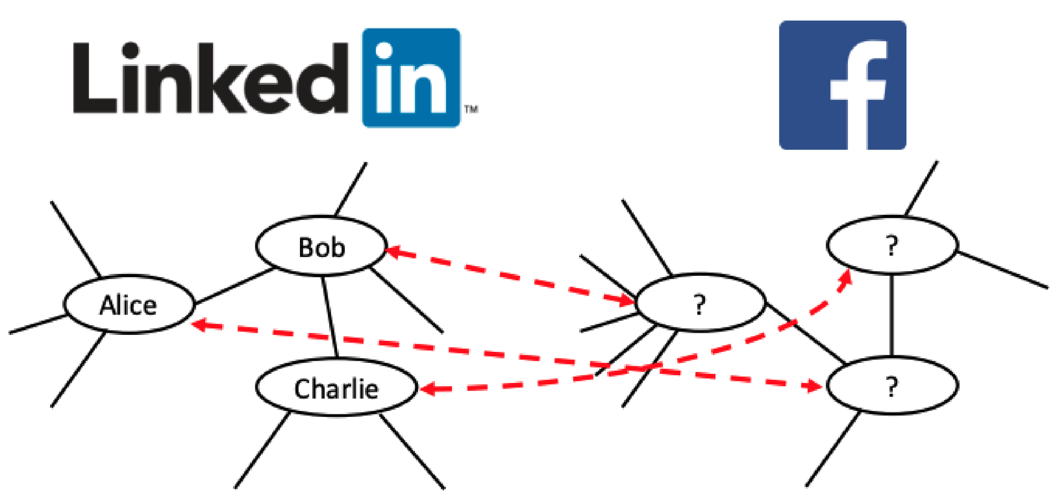
Information-Theoretic Thresholds for the Alignments of Partially Correlated Graphs

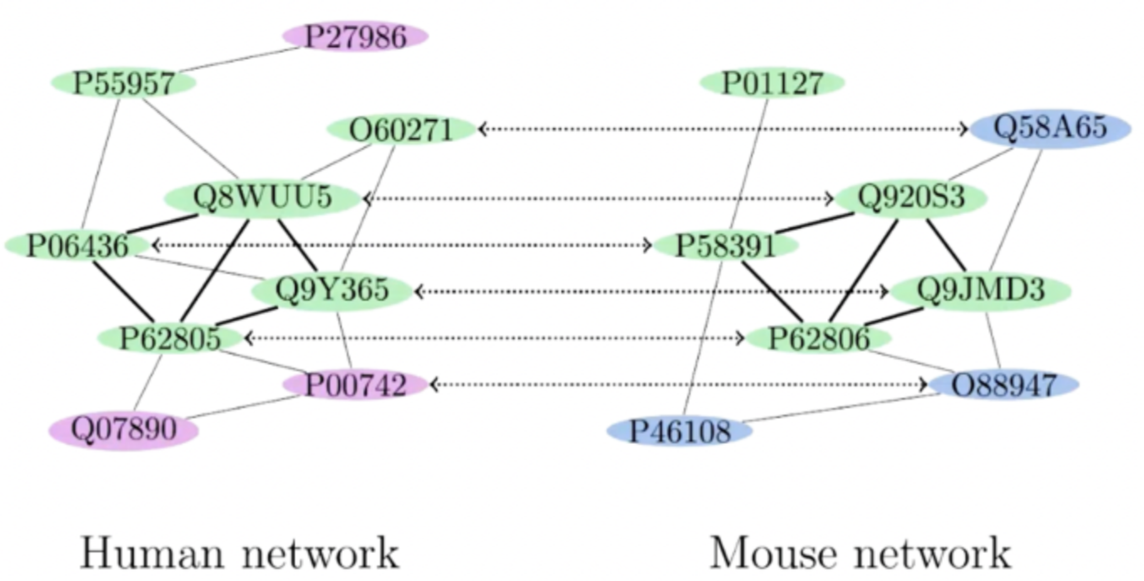
Dong Huang[1], Xianwen Song[2], and Pengkun Yang[1]
Email: hd23@mails.tsinghua.edu.cn
Department of Statistics and Data Science, Tsinghua University[1], Department of Statistics, University of Chicago[2]

Backgrounds and motivations

- Consider a pair of graphs with **latent structure**, we are interesting in *recovering the latent structure*.
- Questions arise from many fields:
  - Social network de-anonymization [1]:



  - Protein interaction network [2]:



  - Computer vision, natural language processing...
- Consider partially correlated Erdős-Rényi graphs due to **the lack of complete correlation** in real-world networks.

Problem setting

Models

- Erdős-Rényi graph $\mathcal{G}(n,p)$: Graph on n vertices where each edge connects with probability $0<p<1$ independently.
- **Correlated Erdős-Rényi graphs** $\mathcal{G}(n,p,\rho)$:
  - $G_1, G_2 \sim \mathcal{G}(n,p)$;
  - Latent bijective mapping π^* on vertices set: $V(G_1) \mapsto V(G_2)$;
  - $\mathrm{Corr}(uv, \pi^*(u)\pi^*(v)) = \rho \in [0,1]$ for any $u,v \in V(G_1)$.
- **Partially correlated Erdős-Rényi graphs** $\mathcal{G}(n,p,\rho,m)$:
  - $G_1, G_2 \sim \mathcal{G}(n,p)$;
  - Latent injective mapping $\pi^* : S^* \subseteq V(G_1) \mapsto V(G_2)$ with $|S^*| = m$;
  - $\mathrm{Corr}(uv, \pi^*(u)\pi^*(v)) = \rho \in [0,1]$ for any $u,v \in S^*$.

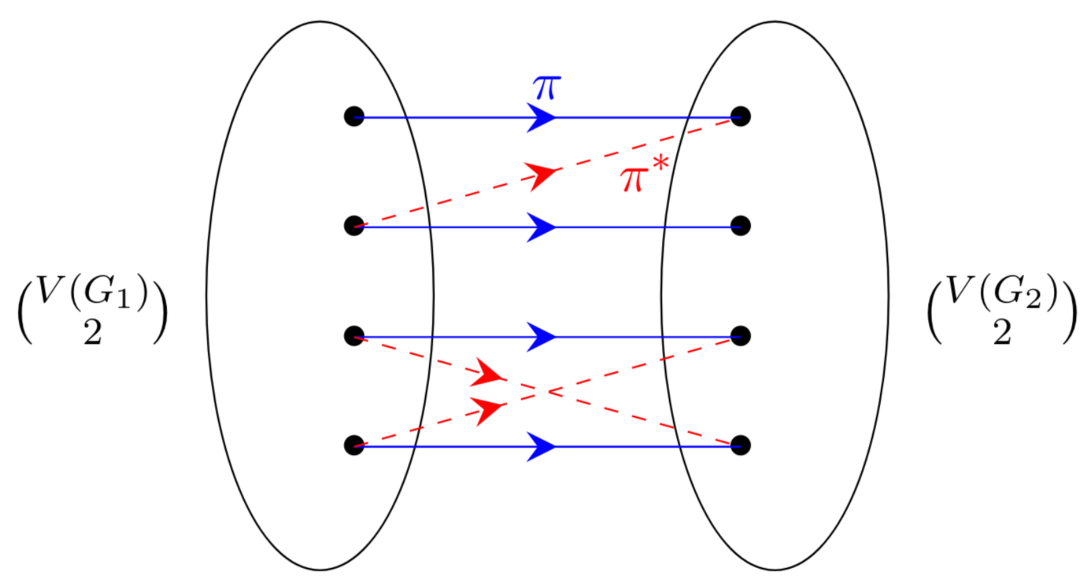
Goals

- Two recovery regimes: Let $\hat{\pi} : \hat{S} \subseteq V(G_1) \to V(G_2)$ be the estimator and $\mathrm{overlap}(\pi^*, \hat{\pi}) \triangleq \frac{|v \in S^* \cap \hat{S} : \pi^*(v) = \hat{\pi}(v)|}{|S^*|}$:
  - *Partial recovery*: $\mathrm{overlap}(\pi^*, \hat{\pi}) \geq \delta$ for $\delta \in (0,1)$;
  - *Exact recovery*: $\mathrm{overlap}(\pi^*, \hat{\pi}) = 1$.
- **Goals**: Information-theoretic thresholds $m^*_\delta(n,p,\rho)$ for $\delta \in (0,1]$:

$$m^*_\delta(n,p,\rho) \triangleq \min\{m : \exists \hat{\pi}, \forall (G_1, G_2) \in \mathcal{G}(n,p,\rho,m), \mathbb{P}\left[\mathrm{overlap}(\pi^*, \hat{\pi}) \geq \delta\right] \geq 0.9\}$$

Challenges

- Unknown location: correlated vertices set S^*;
- Unknown permutation: latent mapping $\pi^* : S^* \subseteq V(G_1) \mapsto V(G_2)$;
- Asymmetric structures due to partial correlation:



Main results

Denote $\gamma \triangleq \frac{\rho(1-p)}{p}$ and $\phi(\gamma) \triangleq (1+\gamma)\log(1+\gamma) - \gamma$. **Sharp information-theoretic thresholds**:

- *Partial recovery*: Fix a constant $0 < \delta < 1$.

$$m^*_\delta(n,p,\rho) \asymp \frac{\log n}{p^2 \phi(\gamma)}.$$

- *Exact recovery*:

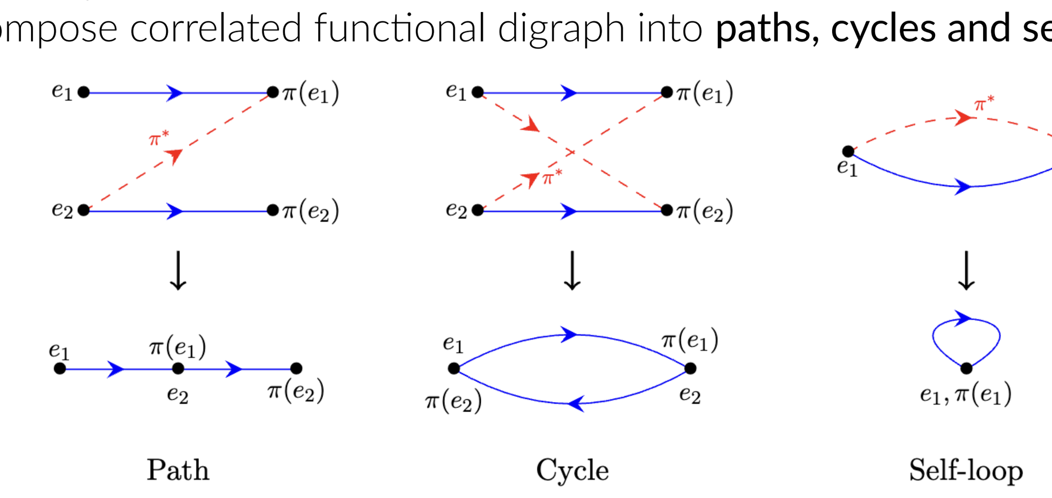
$$m^*_1(n,p,\rho) \asymp \frac{\log n}{p^2 \phi(\gamma)} \vee \frac{\log(1/p^2\gamma)}{p^2 \gamma}.$$

Key ideas

Upper bound

Pick an appropriate estimator $\hat{\pi}$ and divide $\{d(\pi^*, \hat{\pi}) = k\}$ into two events: $\{d(\pi^*, \hat{\pi}) = k\} \subseteq \mathcal{E}_{k,1} \cup \mathcal{E}_{k,2}$.

- $\mathcal{E}_{k,1}$: *bad event of signal*. Apply Chernoff bound for Binomial tail.
- $\mathcal{E}_{k,2}$: *bad event of noise*.
  - Extend mappings on vertices to **correlated functional digraph** on edges induced by π^* and $\hat{\pi}$;
  - Decompose correlated functional digraph into **paths, cycles and self-loops**;



  - Bound cumulant generating function on paths, cycles and self-loops.

Lower bound

Lower bound due to unknown location:

- Construct a packing set $\mathcal{M}_\delta$ of parameter space and bound $|\mathcal{M}_\delta|$;
- Choose a prior on π^* and bound mutual information $I(\pi^*; G_1, G_2)$;
- Apply generalized Fano's inequality [3, Lemma 20]:

$$\mathbb{P}\left[\mathrm{overlap}(\pi^*, \hat{\pi}) < \delta\right] \geq 1 - \frac{I(\pi^*; G_1, G_2)}{\log|\mathcal{M}_\delta|}.$$

Lower bound due to unknown permutation:

Reduction from lower bound in correlated Erdős-Rényi graphs [4, 5].

Future directions

- Refined results: sharp constants and optimal δ-dependency;
- Efficient algorithms and computation hardness;
- Detection problem: hypothesis testing between independent graphs and correlated graphs.

References

[1] A. Narayanan and V. Shmatikov, "De-anonymizing social networks," in *2009 30th IEEE symposium on security and privacy*, pp. 173–187, IEEE, 2009.

[2] E. Kazemi, H. Hassani, M. Grossglauser, and H. Pezeshgi Modarres, "Proper: global protein interaction network alignment through percolation matching," *BMC bioinformatics*, vol. 17, pp. 1–16, 2016.

[3] S. Banerjee, N. Hegde, and L. Massoulié, "The price of privacy in untrusted recommendation engines," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 920–927, IEEE, 2012.

[4] Y. Wu, J. Xu, and S. H. Yu, "Settling the sharp reconstruction thresholds of random graph matching," *IEEE Transactions on Information Theory*, vol. 68, no. 8, pp. 5391–5417, 2022.

[5] J. Ding and H. Du, "Matching recovery threshold for correlated random graphs," *The Annals of Statistics*, vol. 51, no. 4, pp. 1718–1743, 2023.