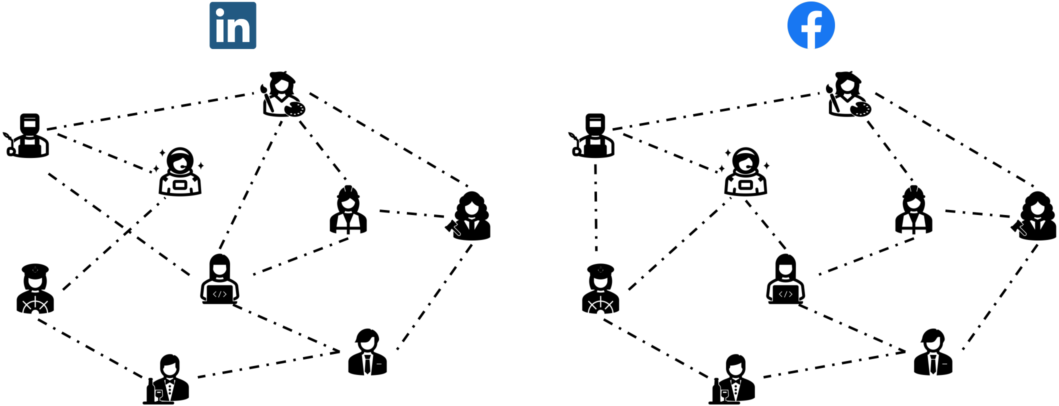
Attributed Graph Alignment

Ning Zhang * Weina Wang † Lele Wang *

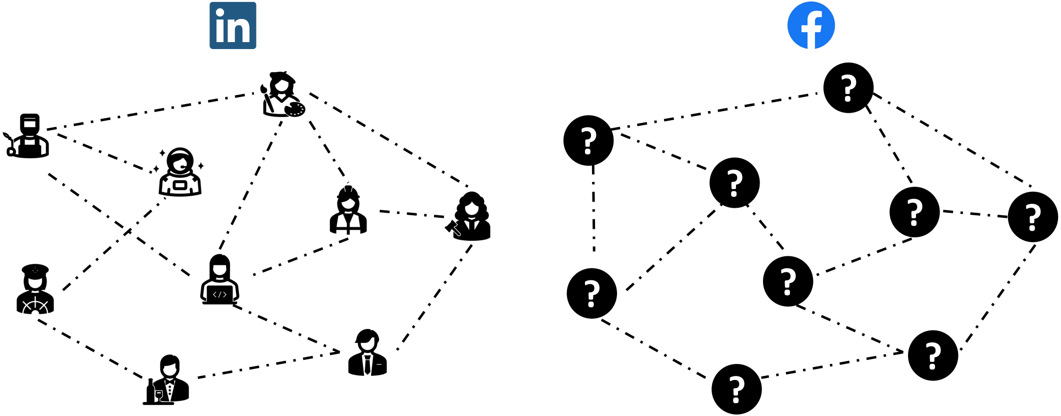
*University of British Columbia †Carnegie Mellon University

International Symposium on Information Theory (ISIT)
June 2021

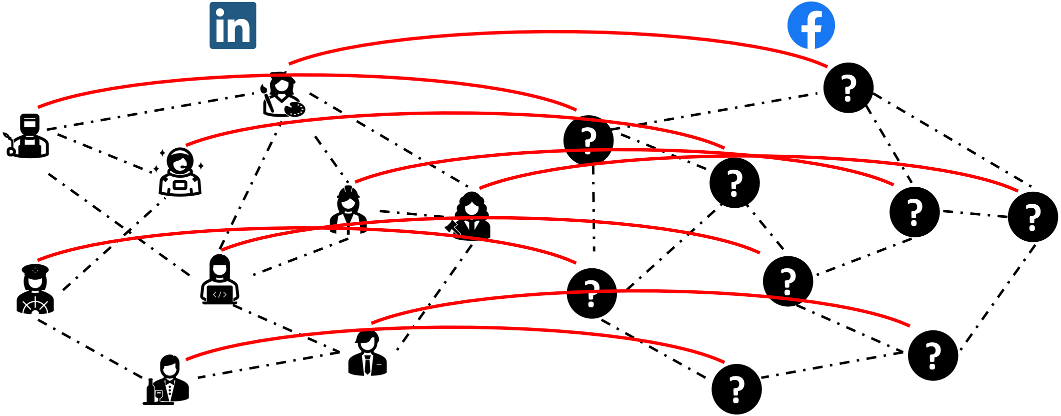
Social network de-anonymization



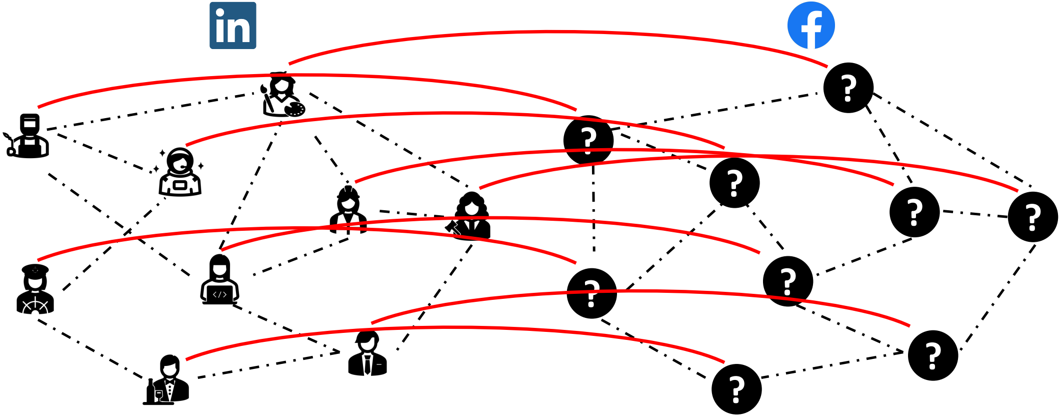
Social network de-anonymization



Social network de-anonymization



Social network de-anonymization



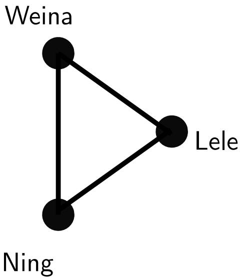
Other applications: computer vision, biological network analysis, natural language processing...

Related works

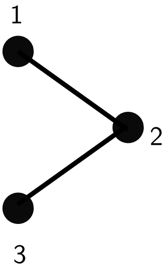
- **Information-theoretic limit:**
  Pedarsani and Grossglauser (2011), Cullina and Kiyavash (2016), Cullina and Kiyavash (2017), Wu et al. (2021)

- **Efficient algorithms:**
  - **Convex optimization:** Lyzinski et al. (2015);
  - **Spectral method:** Nassar et al. (2018), Feizi et al. (2019), Fan et al. (2020) etc. ;
  - **Belief propagation:** Onaran and Villar (2020)
  - **Others:** Barak et al. (2019)

What if graph structure is not enough?



G_1



G'_2

We do know more ...

Ning Zhang

Intro

Studies ECE at **University of British Columbia**

Studied Physics Major at Nankai University, China

Lives in **Vancouver, British Columbia**

From **Shijiazhuang Shi, Hebei, China**

Joined February 2017

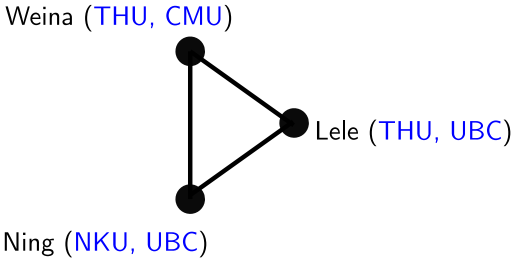
Edit Details

Hiking **Badminton**

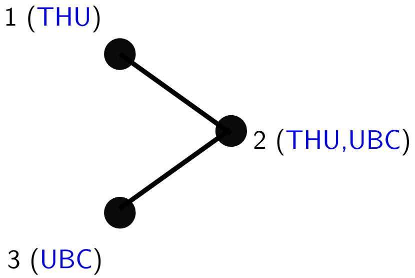
Ning Zhang

Information theory; MASc@UBC; Vancouver

We do know more ...

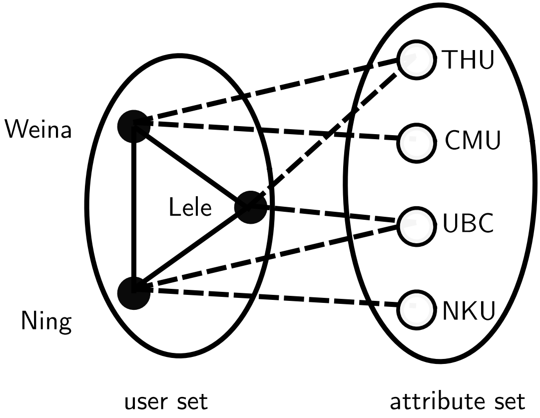


G_1



G_2'

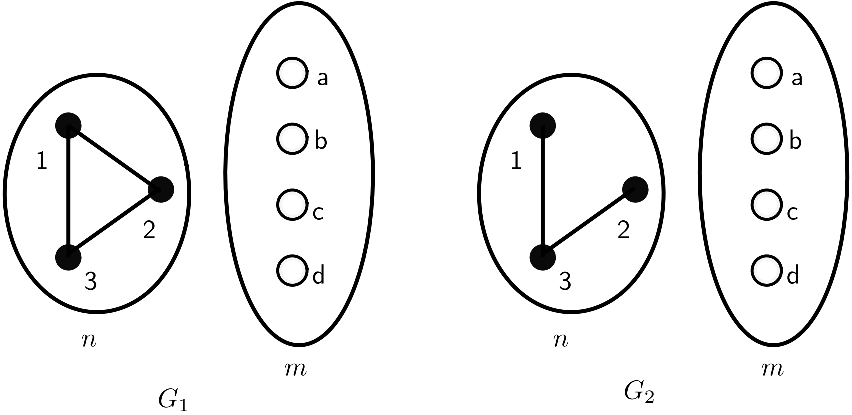
Attributes as vertices



How much benefit can vertex attributes bring?

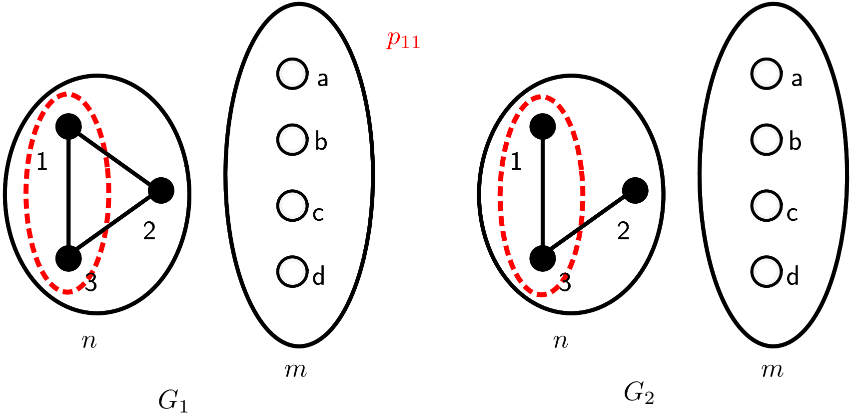
Model: attributed Erdős–Rényi graph pair

- $(G_1, G_2) \sim \mathcal{G}(n, \boldsymbol{p}; m, \boldsymbol{q})$
  - For $e \in \mathcal{E}_u$, $(\mathbb{1}\{e \in E_1\}, \mathbb{1}\{e \in E_2\}) \overset{\text{i.i.d}}{\sim} \boldsymbol{p}$, where $\boldsymbol{p} = (p_{11}, p_{10}, p_{01}, p_{00})$.



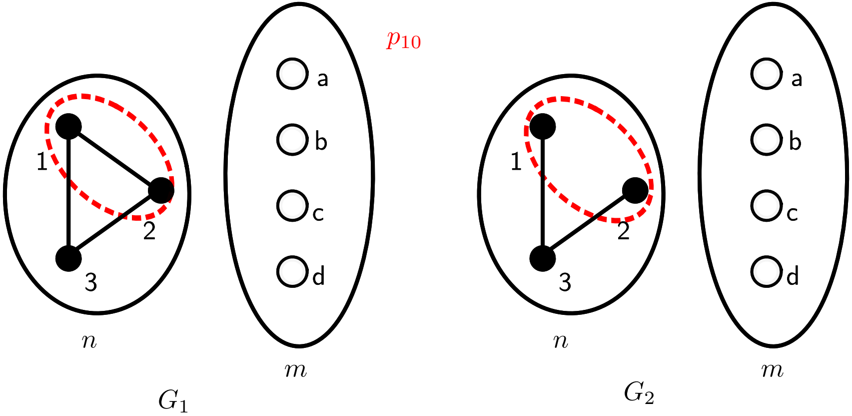
Model: attributed Erdős–Rényi graph pair

- $(G_1, G_2) \sim \mathcal{G}(n, \boldsymbol{p}; m, \boldsymbol{q})$
  - For $e \in \mathcal{E}_u$, $(\mathbb{1}\{e \in E_1\}, \mathbb{1}\{e \in E_2\}) \overset{\text{i.i.d}}{\sim} \boldsymbol{p}$, where $\boldsymbol{p} = (p_{11}, p_{10}, p_{01}, p_{00})$.



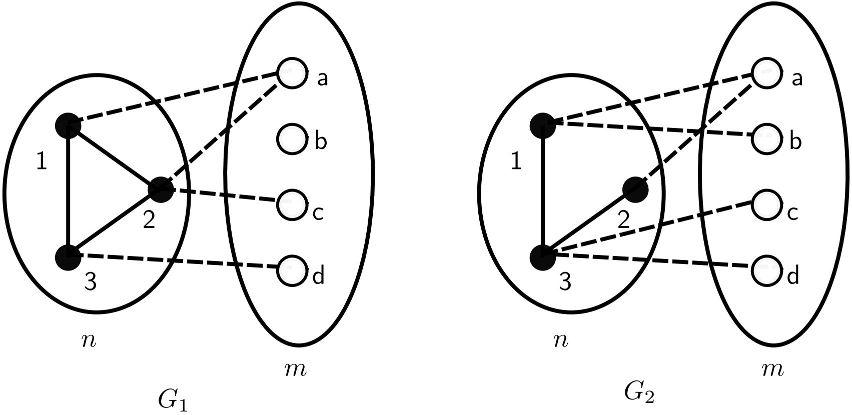
Model: attributed Erdős–Rényi graph pair

- $(G_1, G_2) \sim \mathcal{G}(n, \boldsymbol{p}; m, \boldsymbol{q})$
  - For $e \in \mathcal{E}_u$, $(\mathbb{1}\{e \in E_1\}, \mathbb{1}\{e \in E_2\}) \overset{\text{i.i.d}}{\sim} \boldsymbol{p}$, where $\boldsymbol{p} = (p_{11}, p_{10}, p_{01}, p_{00})$.



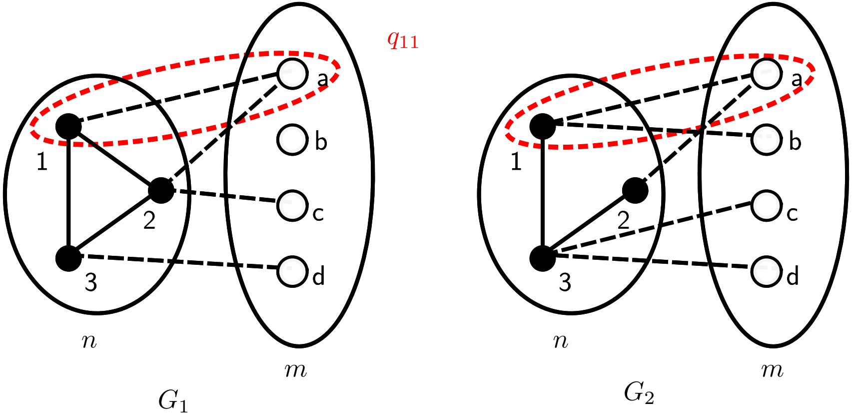
Model: attributed Erdős–Rényi graph pair

- $(G_1, G_2) \sim \mathcal{G}(n, \boldsymbol{p}; m, \boldsymbol{q})$
  - For $e \in \mathcal{E}_u$, $(\mathbb{1}\{e \in E_1\}, \mathbb{1}\{e \in E_2\}) \overset{\text{i.i.d}}{\sim} \boldsymbol{p}$, where $\boldsymbol{p} = (p_{11}, p_{10}, p_{01}, p_{00})$.
  - For $e \in \mathcal{E}_a$, $(\mathbb{1}\{e \in E_1\}, \mathbb{1}\{e \in E_2\}) \overset{\text{i.i.d}}{\sim} \boldsymbol{q}$, where $\boldsymbol{q} = (q_{11}, q_{10}, q_{01}, q_{00})$.



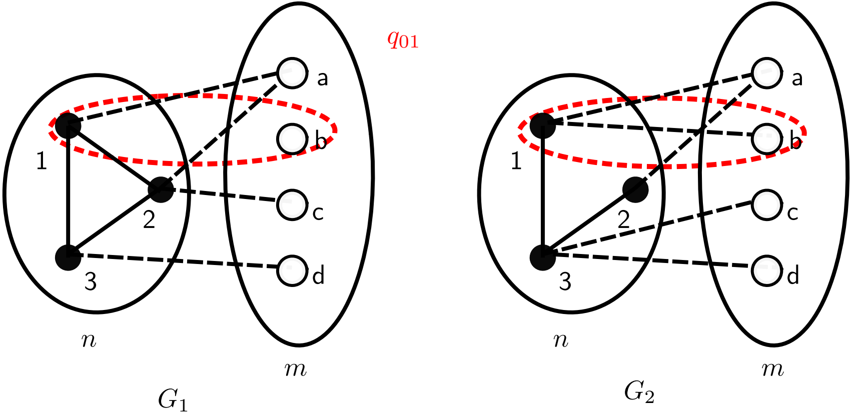
Model: attributed Erdős–Rényi graph pair

- $(G_1, G_2) \sim \mathcal{G}(n, \boldsymbol{p}; m, \boldsymbol{q})$
  - For $e \in \mathcal{E}_u$, $(\mathbb{1}\{e \in E_1\}, \mathbb{1}\{e \in E_2\}) \overset{\text{i.i.d}}{\sim} \boldsymbol{p}$, where $\boldsymbol{p} = (p_{11}, p_{10}, p_{01}, p_{00})$.
  - For $e \in \mathcal{E}_a$, $(\mathbb{1}\{e \in E_1\}, \mathbb{1}\{e \in E_2\}) \overset{\text{i.i.d}}{\sim} \boldsymbol{q}$, where $\boldsymbol{q} = (q_{11}, q_{10}, q_{01}, q_{00})$.



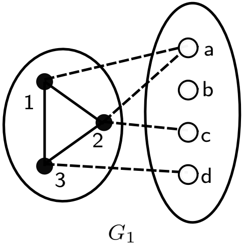
Model: attributed Erdős–Rényi graph pair

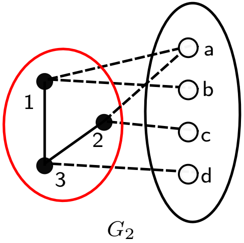
- $(G_1, G_2) \sim \mathcal{G}(n, \boldsymbol{p}; m, \boldsymbol{q})$
  - For $e \in \mathcal{E}_u$, $(\mathbb{1}\{e \in E_1\}, \mathbb{1}\{e \in E_2\}) \overset{\text{i.i.d}}{\sim} \boldsymbol{p}$, where $\boldsymbol{p} = (p_{11}, p_{10}, p_{01}, p_{00})$.
  - For $e \in \mathcal{E}_a$, $(\mathbb{1}\{e \in E_1\}, \mathbb{1}\{e \in E_2\}) \overset{\text{i.i.d}}{\sim} \boldsymbol{q}$, where $\boldsymbol{q} = (q_{11}, q_{10}, q_{01}, q_{00})$.



Model: attributed Erdős–Rényi graph pair

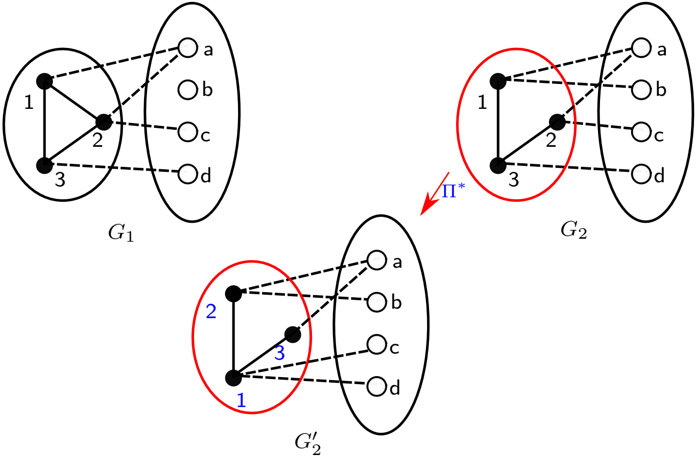
- $(G_1, G_2) \sim \mathcal{G}(n, \boldsymbol{p}; m, \boldsymbol{q})$
- **Anonymization:** Obtain G_2' by applying random Π^* on user vertex set





Model: attributed Erdős–Rényi graph pair

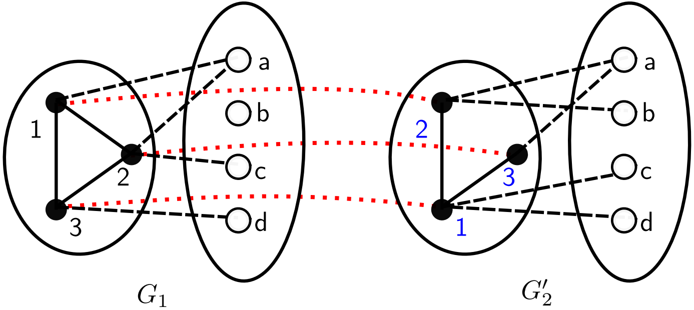
- $(G_1, G_2) \sim \mathcal{G}(n, \boldsymbol{p}; m, \boldsymbol{q})$
- **Anonymization:** Obtain G_2' by applying random Π^* on user vertex set



Model: attributed Erdős–Rényi graph pair

- $(G_1, G_2) \sim \mathcal{G}(n, \boldsymbol{p}; m, \boldsymbol{q})$
- **Anonymization:** Obtain G_2' by applying random Π^* on user vertex set
- **Exact alignment:** Recover $\hat{\pi}(G_1, G_2')$ s.t.

$$\lim_{n \to \infty} \mathsf{P}(\hat{\pi}(G_1, G_2') = \Pi^*) = 1$$



Model: attributed Erdős–Rényi graph pair

- $(G_1, G_2) \sim \mathcal{G}(n, \boldsymbol{p}; m, \boldsymbol{q})$
- **Anonymization:** Obtain G_2' by applying random Π^* on user vertex set
- **Exact alignment:** Recover $\hat{\pi}(G_1, G_2')$ s.t.

$$\lim_{n \to \infty} \mathsf{P}(\hat{\pi}(G_1, G_2') = \Pi^*) = 1$$

- **Our goal:** characterize the information-theoretic limit

Achievability

Conditions on $(n, \boldsymbol{p}, m, \boldsymbol{q})$ to achieve exact alignment;

Converse

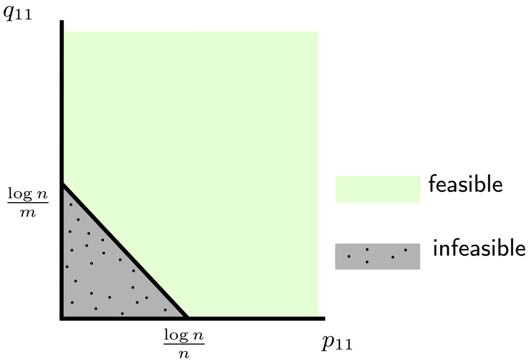
Conditions on $(n, \boldsymbol{p}, m, \boldsymbol{q})$ s.t. exact alignment is not possible.

Achievability (simplified)

- $m = \Omega((\log n)^3)$: $np_{11} + mq_{11} - \log n \to \infty$
- $m = o((\log n)^3)$:
  $np_{11} + mq_{11} - O(mq_{11}^{3/2}) - \log n \to \infty$

Converse

$np_{11} + mq_{11} - \log n \to -\infty$



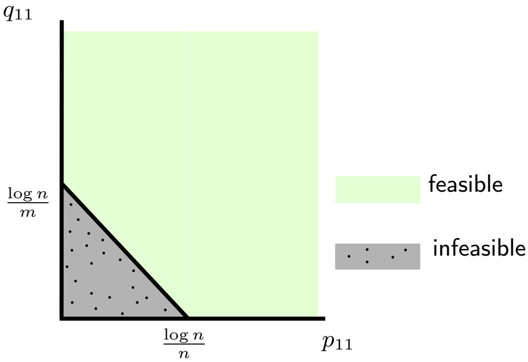
Main results

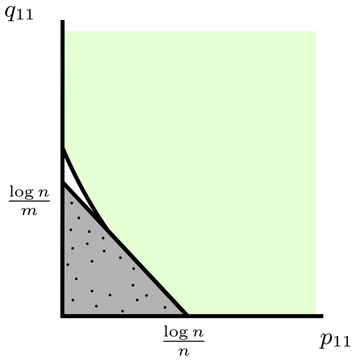
Achievability (simplified)

- $m = \Omega((\log n)^3)$: $np_{11} + mq_{11} - \log n \to \infty$
- $m = o((\log n)^3)$:
  $np_{11} + mq_{11} - O(mq_{11}^{3/2}) - \log n \to \infty$

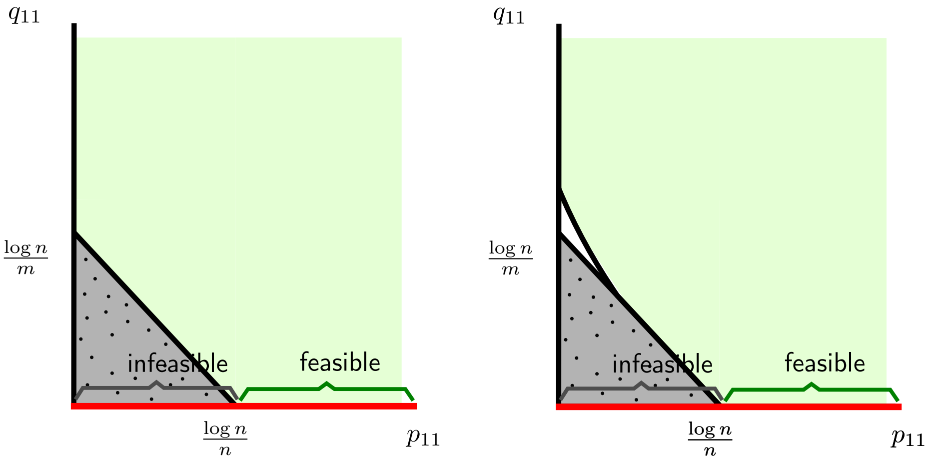
Converse

$np_{11} + mq_{11} - \log n \to -\infty$



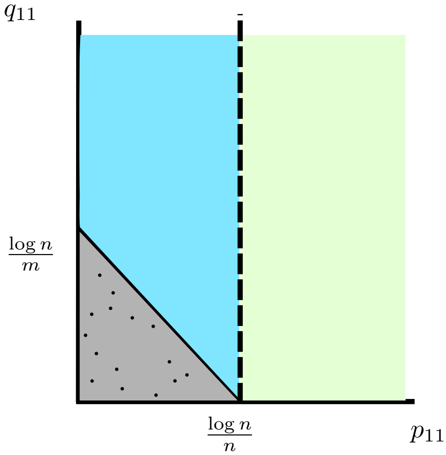


1. **Erdős–Rényi graph alignment** ($q_{00} = 1$) Cullina and Kiyavash (2017)

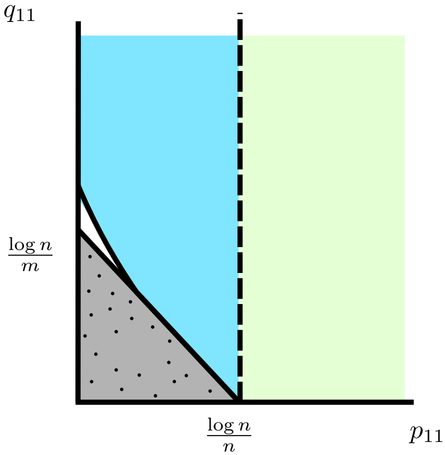


(a) $m = \Omega((\log n)^3)$ (b) $m = o((\log n)^3)$

1. **Erdős–Rényi graph alignment** ($q_{00} = 1$) Cullina and Kiyavash (2017)

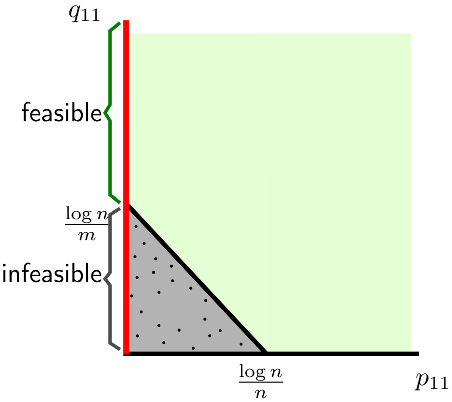


(a) $m = \Omega((\log n)^3)$

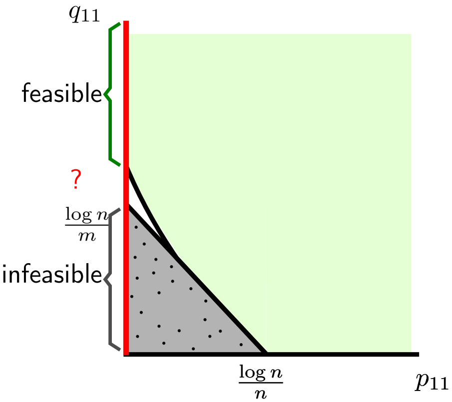


(b) $m = o((\log n)^3)$

2. **Bipartite graph alignment** $(p_{00} = 1)$ Dai et al. (2019)



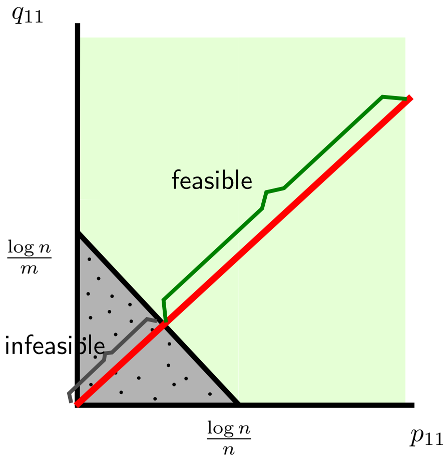
(a) $m = \Omega((\log n)^3)$



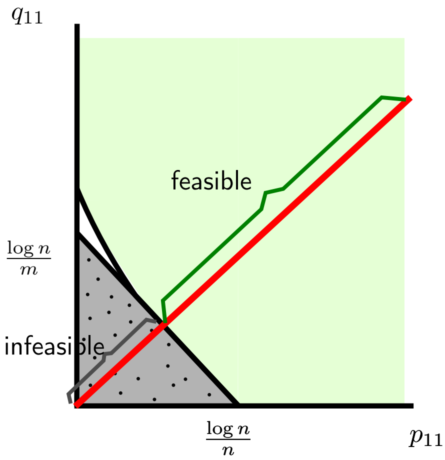
(b) $m = o((\log n)^3)$

Improve the seeded graph alignment

3. **Seeded graph alignment** $(p = q)$ Yartseva and Grossglauser (2013)



(a) $m = \Omega((\log n)^3)$



(b) $m = o((\log n)^3)$

Proof of achievability (sketch)

Achievability

- $m = \Omega((\log n)^3)$: $np_{11} + mq_{11} - \log n \to \infty$
- $m = o((\log n)^3)$: $np_{11} + mq_{11} - O(mq_{11}^{3/2}) - \log n \to \infty$

Key ideas:

- MAP estimator (see Lemma1)

Proof of achievability (sketch)

Achievability

- $m = \Omega((\log n)^3)$: $np_{11} + mq_{11} - \log n \to \infty$
- $m = o((\log n)^3)$: $np_{11} + mq_{11} - O(mq_{11}^{3/2}) - \log n \to \infty$

Key ideas:

- MAP estimator (see Lemma1)
- Upper bound error probability of MAP (Cullina and Kiyavash (2017)):
  - Orbit decomposition
  - Generating functions

Proof of converse (sketch)

Converse

$np_{11} + mq_{11} - \log n \to -\infty$

Key ideas:

- $\mathsf{P}(\hat{\pi}_{\text{MAP}}(G_1, G_2') = \Pi^*) \leq \frac{1}{|\text{Aut}(G_1 \wedge G_2)|}$

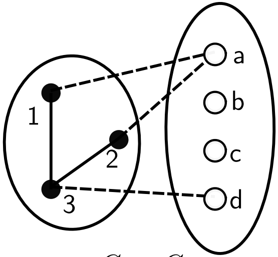
Proof of converse (sketch)

Converse

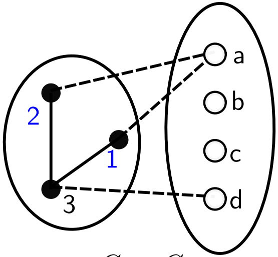
$np_{11} + mq_{11} - \log n \to -\infty$

Key ideas:

- $\mathsf{P}(\hat{\pi}_{\mathrm{MAP}}(G_1, G_2') = \Pi^*) \leq \frac{1}{|\mathrm{Aut}(G_1 \wedge G_2)|}$



$G_1 \wedge G_2$



$G_1 \wedge G_2$

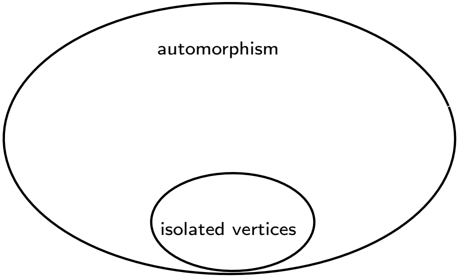
Proof of converse (sketch)

Converse

$np_{11} + mq_{11} - \log n \to -\infty$

Key ideas:

- $\mathsf{P}(\hat{\pi}_{\mathrm{MAP}}(G_1, G_2') = \Pi^*) \leq \frac{1}{|\mathrm{Aut}(G_1 \wedge G_2)|}$
- $|\mathrm{Aut}| \geq |\mathrm{Aut}_{\mathrm{iso}}|$



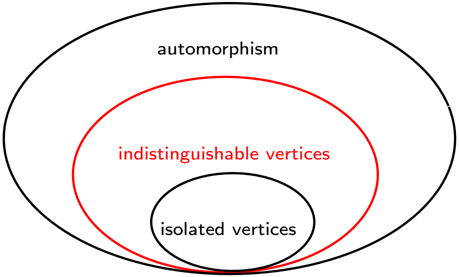
Proof of converse (sketch)

Converse

$np_{11} + mq_{11} - \log n \to -\infty$

Key ideas:

- $\mathsf{P}(\hat{\pi}_{\mathrm{MAP}}(G_1, G_2') = \Pi^*) \leq \frac{1}{|\mathrm{Aut}(G_1 \wedge G_2)|}$
- $|\mathrm{Aut}| \geq |\mathrm{Aut}_{\mathrm{ind}}| \geq |\mathrm{Aut}_{\mathrm{iso}}|$



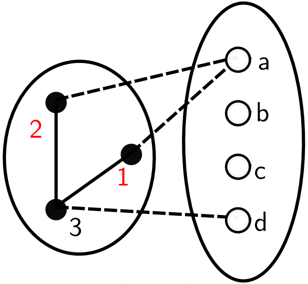
Proof of converse (sketch)

Converse

$np_{11} + mq_{11} - \log n \to -\infty$

Key ideas:

- $\mathsf{P}(\hat{\pi}_{\mathrm{MAP}}(G_1, G_2') = \Pi^*) \leq \frac{1}{|\mathrm{Aut}(G_1 \wedge G_2)|}$
- $|\mathrm{Aut}| \geq |\mathrm{Aut}_{\mathrm{ind}}| \geq |\mathrm{Aut}_{\mathrm{iso}}|$



$G_1 \wedge G_2$

$|\mathrm{Aut}_{\mathrm{iso}}| = 0$
$|\mathrm{Aut}_{\mathrm{ind}}| = 1$

Proof of converse (sketch)

Converse

$np_{11} + mq_{11} - \log n \to -\infty$

Key ideas:

- $\mathsf{P}(\hat{\pi}_{\text{MAP}}(G_1, G_2') = \Pi^*) \leq \frac{1}{|\text{Aut}(G_1 \wedge G_2)|}$
- $|\text{Aut}| \geq |\text{Aut}_{\text{ind}}| \geq |\text{Aut}_{\text{iso}}|$
- Use the second moment method to show that w.h.p. $|\text{Aut}(G_1 \wedge G_2)| \geq 2$
  In particular, we show # indistinguishable vertex pair >0.

Contributions

- Propose attributed Erdős–Rényi graph pair model
- Characterize the information-theoretic limits
- Unify existing models for graph alignment

Selected references

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 *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2019.

D. Cullina and N. Kiyavash. Improved achievability and converse bounds for Erdős-Rényi graph matching. *ACM SIGMETRICS Perform. Evaluation Rev.*, 44(1):63–72, 2016. doi: 10.1145/2964791.2901460.

D. Cullina and N. Kiyavash. Exact alignment recovery for correlated Erdős-Rényi graphs. *arXiv:1711.06783 [cs.IT]*, 2017.

O. E. Dai, D. Cullina, N. Kiyavash, and M. Grossglauser. Analysis of a canonical labeling algorithm for the alignment of correlated erdos-rényi graphs. *Proc. ACM Meas. Anal. Comput. Syst.*, 3(2), June 2019. doi: 10.1145/3341617.3326151.

Z. Fan, C. Mao, Y. Wu, and J. Xu. Spectral graph matching and regularized quadratic relaxations: Algorithm and theory. In *International Conference on Machine Learning*, pages 2985–2995. PMLR, 2020.

Selected references (cont.)

S. Feizi, G. Quon, M. Recamonde-Mendoza, M. Medard, M. Kellis, and A. Jadbabaie. Spectral alignment of graphs. *IEEE Transactions on Network Science and Engineering*, 7(3):1182–1197, 2019.

V. Lyzinski, D. E. Fishkind, M. Fiori, J. T. Vogelstein, C. E. Priebe, and G. Sapiro. Graph matching: Relax at your own risk. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):60–73, 2015.

H. Nassar, N. Veldt, S. Mohammadi, A. Grama, and D. F. Gleich. Low rank spectral network alignment. In *Proceedings of the 2018 World Wide Web Conference*, pages 619–628, 2018.

E. Onaran and S. Villar. Efficient belief propagation for graph matching. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 9060–9064, 2020. doi: 10.1109/ICASSP40776.2020.9053147.

P. Pedarsani and M. Grossglauser. On the privacy of anonymized networks. In *Proc. Ann. ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD)*, pages 1235–1243, 2011. doi: 10.1145/2020408.2020596.

Y. Wu, J. Xu, and S. H. Yu. Settling the sharp reconstruction thresholds of random graph matching, 2021.

Selected references (cont.)

L. Yartseva and M. Grossglauser. On the performance of percolation graph matching. In *Proceedings of the first ACM conference on Online social networks*, pages 119–130, 2013.