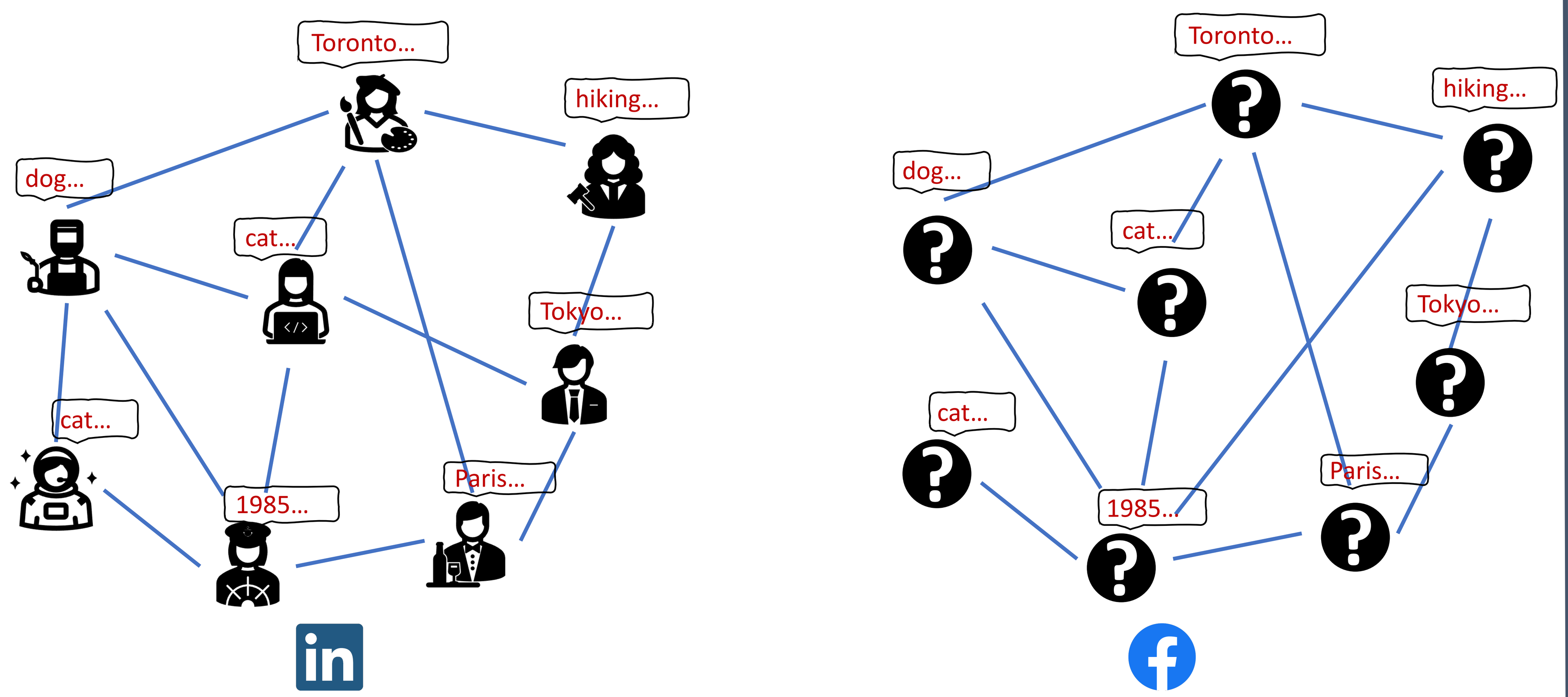
Attributed Graph Alignment

Ning Zhang*, Weina Wang†, Lele Wang*
* The University of British Columbia, † Carnegie Mellon University

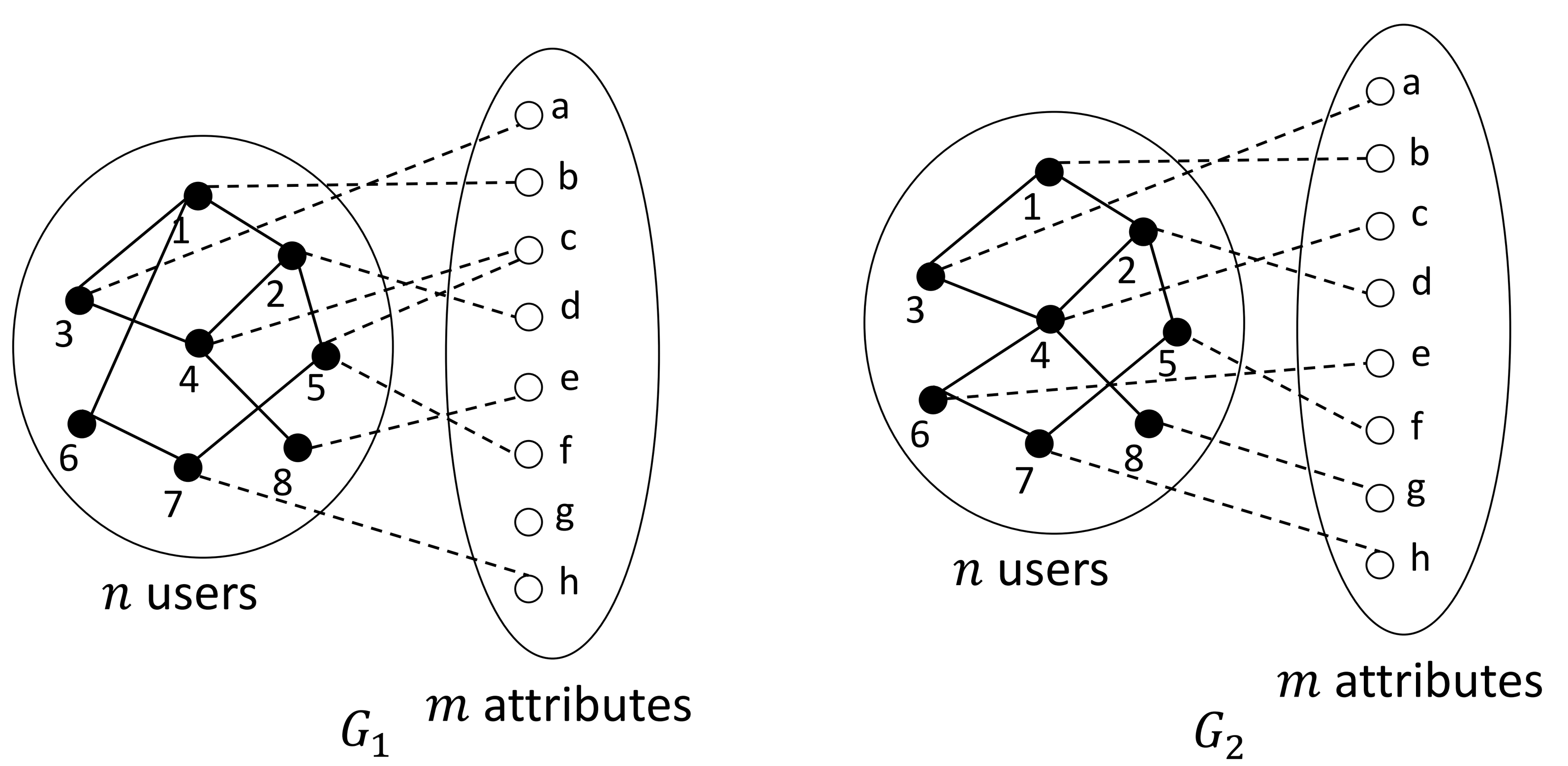
Social network deanonymization



- Find the user correspondence using both **graph topology information** and **attribute information**.

Problem formulation

- **Model: The attributed Erdős-Rényi pair $\mathcal{G}(n, \mathbf{p}; m, \mathbf{q})$**



- For $i, j \in [n]$, $\left(\mathbf{1}_{\{(i,j)\in E_1\}}, \mathbf{1}_{\{(i,j)\in E_2\}}\right) \sim \mathbf{p} = (p_{11}, p_{10}, p_{01}, p_{00})$
- For $i \in [n]$, $j' \in [m]$, $\left(\mathbf{1}_{\{(i,j')\in E_1\}}, \mathbf{1}_{\{(i,j')\in E_2\}}\right) \sim \mathbf{q} = (q_{11}, q_{10}, q_{01}, q_{00})$

- **Anonymization:** Apply unknown permutation Π^* on user set of G_2 and the resulting graph is $G_2{}'$.
- **Attributed graph alignment:** Given G_1 and $G_2{}'$, find a permutation $\hat{\pi}(G_1, G_2'): [n] \to [n]$, s.t., $\mathrm{P}(\hat{\pi}(G_1, G_2') = \Pi^*) = 1 - o(1)$.

Main results

- **Maximum a Posterior estimator** (optimal)
  $\hat{\pi}_{\mathrm{MAP}} = \mathrm{argmin}_{\pi}\{\, w_1 \Delta^u(G_1, \pi \circ G_2) + w_2 \Delta^a(G_1, \pi \circ G_2)\}$

$\Delta_u(G_1, \pi \circ G_2)$: Hamming distance of user-user edges between G_1 and $\pi \circ G_2$
$\Delta_a(G_1, \pi \circ G_2)$: Hamming distance of user-attributed edges between G_1, $\pi \circ G_2$

- **(Simplified) achievability condition** (feasible region in Fig1&2)

If $\mathcal{G}(n, \mathbf{p}; m, \mathbf{q})$ satisfies
$$np_{11} + m(\sqrt{q_{11}q_{00}} - \sqrt{q_{10}q_{01}}) - \log n \to \infty,$$
then $\mathrm{P}(\hat{\pi}_{\mathrm{MAP}} = \Pi^*) = 1 - o(1)$.

- **Converse condition** (infeasible region in Fig1&2)

If $\mathcal{G}(n, \mathbf{p}; m, \mathbf{q})$ satisfies
$$np_{11} + mq_{11} - \log n \to -\infty,$$
then no algorithm guarantees recovering Π^* with high probability.

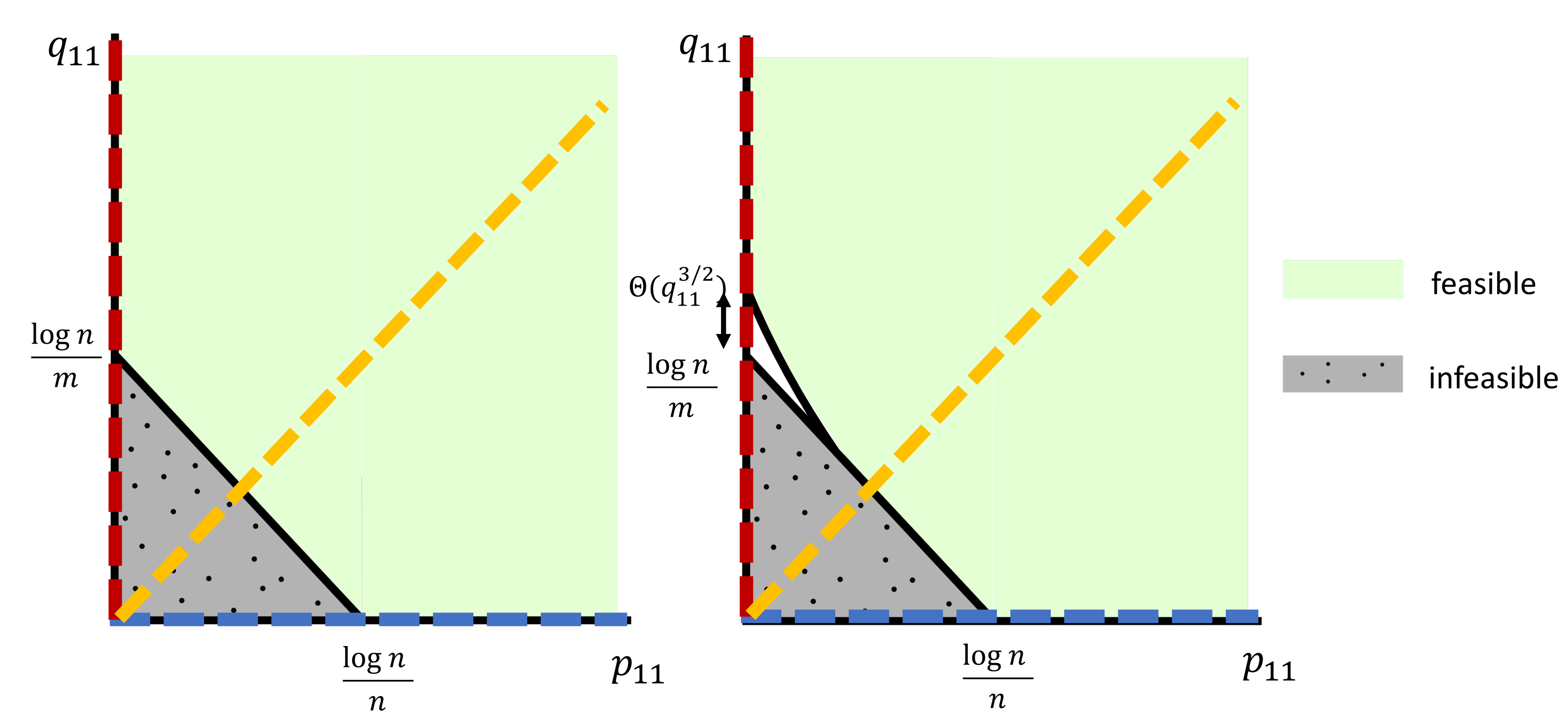


Fig1. $m = \Omega((\log n)^3)$ Fig2. $m = o((\log n)^3)$

*We added a few mild assumptions to demonstrate the result in 2D plane. For general achievability results, please check our paper.
** The dot lines correspond to three specialized settings in the next section.

Connection with prior work

Erdős-Rényi graph alignment

Setting: Only user-user connection information is available. (Pedarsani& Grossglauser2011)

Our model: By removing user-attribute edges, our model specializes to correlated Erdős-Rényi graph pair and the corresponding information-theoretic limit is shown as the blue lines in Fig 1&2.

✓ Our results recover information-theoretic limits given by Cullina&Kiyavash in 2017

Bipartite graph alignment

Setting: Only user-attribute connection information is available. (Narayanan&Shmatikov 2008)

Our model: By removing user-user edges, our model specializes to correlated random bipartite graph pair and the corresponding information-theoretic limit is shown as the red lines in Fig 1&2.

✓ Our results improve the previous information-theoretic limits from Dai et al. 2019

Seeded graph alignment

Setting: Only user-user connection is available, and part of the vertices are correctly pre-aligned. (Yartseva&Grossglauser 2013, Shirani et al.2017)

Our model: We treat attributes as pre-aligned users and set the edge probability between user-attribute to be the same as user-user. The corresponding information-theoretic limit is shown as the yellow lines in Fig 1&2.

✓ Our results provide the tight achievability and converse conditions.

Future directions

- ❑ Tight achievability and converse conditions?
- ❑ Aligning graphs with community structure?

Conference paper Full version

Paper links: https://arxiv.org/abs/2102.00665v1
https://arxiv.org/abs/2102.00665v2