Attributed Graph Alignment

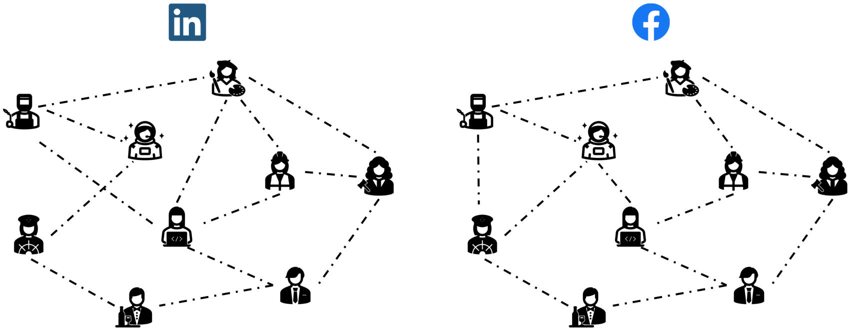
Lele Wang

University of British Columbia

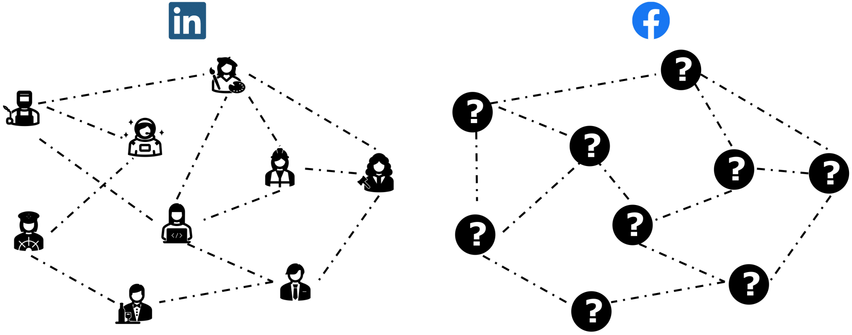
BIRS Workshop on Algorithmic Structures for Uncoordinated Communications
and Statistical Inference in Exceedingly Large Spaces
Banff, Canada, March 2024

Joint work with Ziao Wang (UBC), Ning Zhang (Oxford), and Weina Wang (CMU)

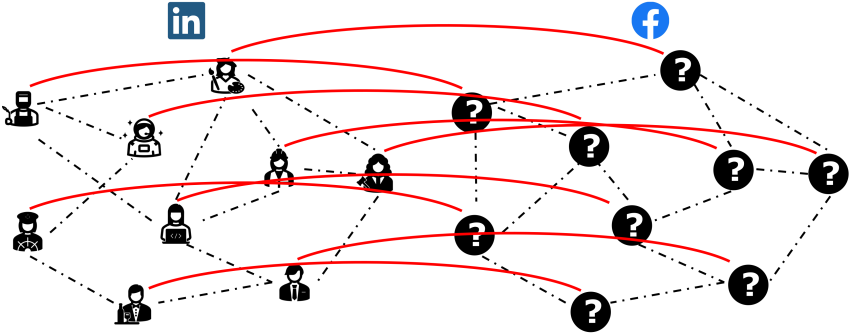
Motivation: Correlated social networks



Motivation: Correlated social networks



Motivation: Correlated social networks



Social network de-anonymization (Narayanan and Shmatikov 2008)

Combinatorial optimization formulation

- Let A and B be the adjacency matrices of the two (simple) graphs
- Quadratic assignment problem (Koopmans and Beckmann 1957)

$$\hat{\pi} = \arg\max_{\sigma} \sum_{i<j} A_{i,j} B_{\sigma(i),\sigma(j)}$$

Combinatorial optimization formulation

- Let A and B be the adjacency matrices of the two (simple) graphs
- Quadratic assignment problem (Koopmans and Beckmann 1957)

$$\hat{\pi} = \arg\max_{\sigma} \sum_{i<j} A_{i,j} B_{\sigma(i),\sigma(j)}$$

- Zero-error alignment: guarantee $\hat{\pi} = \pi$ in the worst case

Combinatorial optimization formulation

- Let A and B be the adjacency matrices of the two (simple) graphs
- Quadratic assignment problem (Koopmans and Beckmann 1957)

$$\hat{\pi} = \arg\max_{\sigma} \sum_{i<j} A_{i,j} B_{\sigma(i),\sigma(j)}$$

- Zero-error alignment: guarantee $\hat{\pi} = \pi$ in the worst case
- Complexity: NP-hard

Combinatorial optimization formulation

- Let A and B be the adjacency matrices of the two (simple) graphs
- Quadratic assignment problem (Koopmans and Beckmann 1957)

$$\hat{\pi} = \arg\max_{\sigma} \sum_{i<j} A_{i,j} B_{\sigma(i),\sigma(j)}$$

- Zero-error alignment: guarantee $\hat{\pi} = \pi$ in the worst case
- Complexity: NP-hard
- What about the average case?
  - Can we solve the problem for most typical practical networks?
  - What if we are fine with a small but vanishing amount of error?

Combinatorial optimization formulation

- Let A and B be the adjacency matrices of the two (simple) graphs
- Quadratic assignment problem (Koopmans and Beckmann 1957)

$$\hat{\pi} = \arg\max_{\sigma} \sum_{i<j} A_{i,j} B_{\sigma(i),\sigma(j)}$$

- Zero-error alignment: guarantee $\hat{\pi} = \pi$ in the worst case
- Complexity: NP-hard
- What about the average case?
  - Can we solve the problem for most typical practical networks? $\rightarrow$ a new random graph model
  - What if we are fine with a small but vanishing amount of error?

Combinatorial optimization formulation

- Let A and B be the adjacency matrices of the two (simple) graphs
- Quadratic assignment problem (Koopmans and Beckmann 1957)

$$\hat{\pi} = \arg\max_{\sigma} \sum_{i<j} A_{i,j} B_{\sigma(i),\sigma(j)}$$

- Zero-error alignment: guarantee $\hat{\pi} = \pi$ in the worst case
- Complexity: NP-hard
- What about the average case?
  - Can we solve the problem for most typical practical networks? $\rightarrow$ a new random graph model
  - What if we are fine with a small but vanishing amount of error? $\rightarrow$ a new performance metric

Combinatorial optimization formulation

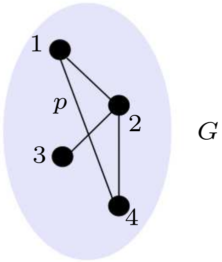
- Let A and B be the adjacency matrices of the two (simple) graphs
- Quadratic assignment problem (Koopmans and Beckmann 1957)

$$\hat{\pi} = \arg\max_{\sigma} \sum_{i<j} A_{i,j} B_{\sigma(i),\sigma(j)}$$

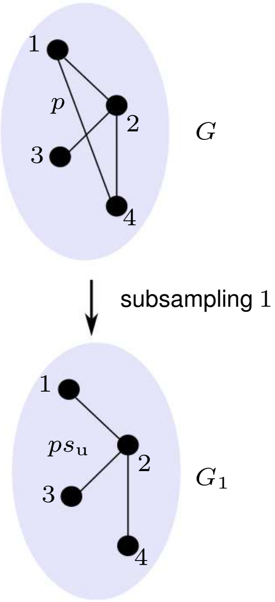
- Zero-error alignment: guarantee $\hat{\pi} = \pi$ in the worst case
- Complexity: NP-hard
- What about the average case?
  - Can we solve the problem for most typical practical networks? $\rightarrow$ a new random graph model
  - What if we are fine with a small but vanishing amount of error? $\rightarrow$ a new performance metric

Graph alignment problem

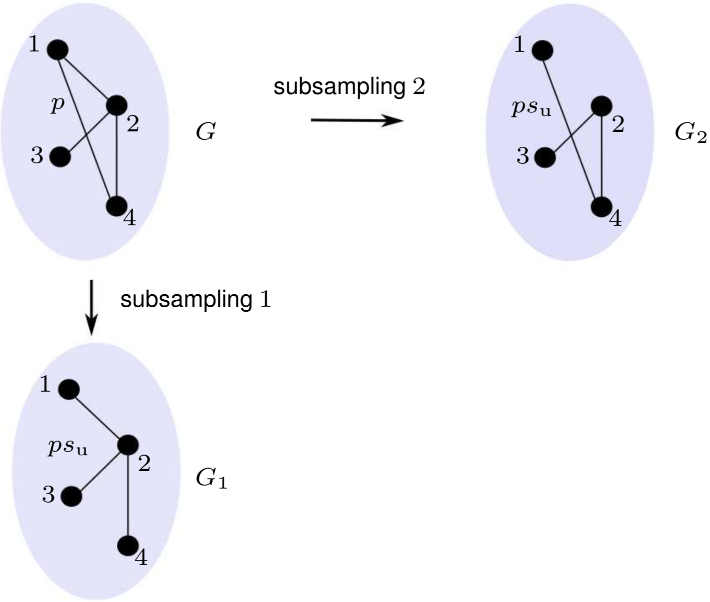
Correlated Erdős–Rényi graph pair $(G_1, G_2') \sim \mathcal{G}(n, p, s_\mathrm{u})$



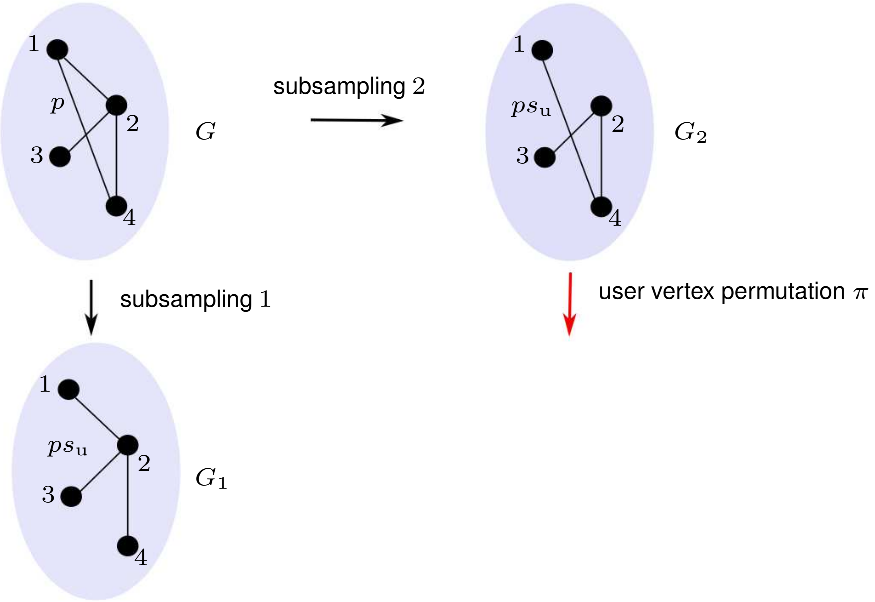
Correlated Erdős–Rényi graph pair $(G_1, G_2') \sim \mathcal{G}(n, p, s_u)$



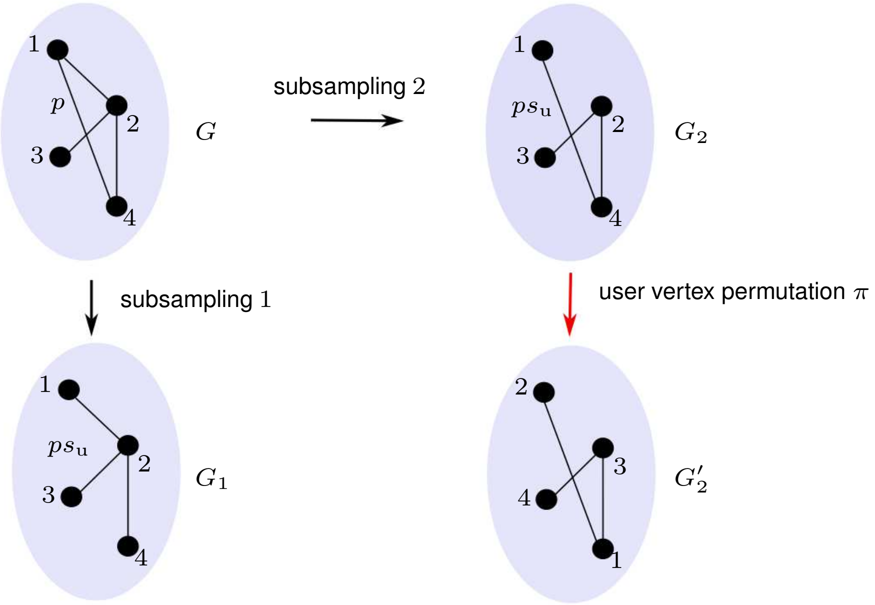
Correlated Erdős–Rényi graph pair $(G_1, G_2') \sim \mathcal{G}(n, p, s_\mathrm{u})$



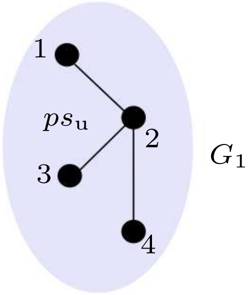
Correlated Erdős–Rényi graph pair $(G_1, G_2') \sim \mathcal{G}(n, p, s_\mathrm{u})$

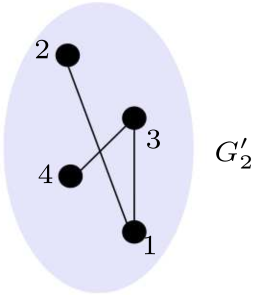


Correlated Erdős–Rényi graph pair $(G_1, G_2') \sim \mathcal{G}(n, p, s_\mathrm{u})$



Correlated Erdős–Rényi graph pair $(G_1, G'_2) \sim \mathcal{G}(n, p, s_{\mathrm{u}})$





- Exact alignment: Estimation $\hat{\pi}(G_1, G_2')$ such that for uniform Π,

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

- Exact alignment: Estimation $\hat{\pi}(G_1, G_2')$ such that for uniform Π,

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

- Achievability: Set of (n, p, s_u) s.t. exact alignment is achievable

- Exact alignment: Estimation $\hat{\pi}(G_1, G_2')$ such that for uniform Π,

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

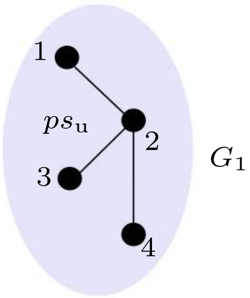
- Achievability: Set of (n, p, s_u) s.t. exact alignment is achievable
  - Information-theoretic limits
  - Efficient algorithms

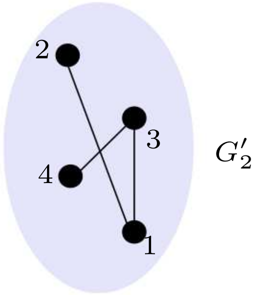
Correlated Erdős–Rényi graph pair $(G_1, G_2') \sim \mathcal{G}(n, p, s_u)$

- Exact alignment: Estimation $\hat{\pi}(G_1, G_2')$ such that for uniform Π,

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

- Achievability: Set of (n, p, s_u) s.t. exact alignment is achievable
  - Information-theoretic limits
  - Efficient algorithms
- Converse: Set of (n, p, s_u) s.t. no algorithm achieves exact alignment





Special case: Random graph isomorphism problem ($s_\mathrm{u} = 1$)

- $G \sim \mathrm{ER}(n, p)$
- $G_1 = G_2 = G$
- G_2' isomorphic to G_1

Special case: Random graph isomorphism problem ($s_{\mathrm{u}} = 1$)

- $G \sim \mathrm{ER}(n, p)$
- $G_1 = G_2 = G$
- G_2' isomorphic to G_1

Theorem (Babai, Erdős, and Selkow 1980, Czajka and Pandurangan 2008)

Assume $p \leq 1/2$

- If $np \geq \log n + \omega(1)$, $\exists$ a polynomial-time algorithm
- If $np \leq \log n - \omega(1)$, no algorithms exist

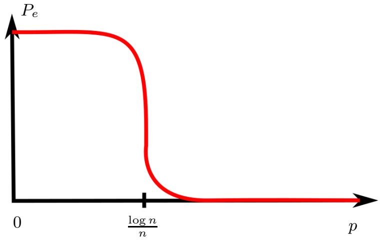
Special case: Random graph isomorphism problem ($s_\mathrm{u} = 1$)

- $G \sim \mathrm{ER}(n, p)$
- $G_1 = G_2 = G$
- G_2' isomorphic to G_1

Theorem (Babai, Erdős, and Selkow 1980, Czajka and Pandurangan 2008)

Assume $p \leq 1/2$

- If $np \geq \log n + \omega(1)$, $\exists$ a polynomial-time algorithm
- If $np \leq \log n - \omega(1)$, no algorithms exist



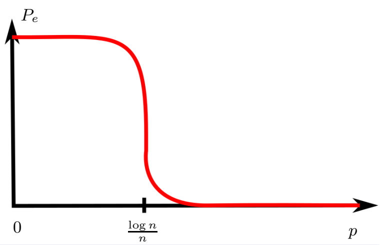
Special case: Random graph isomorphism problem ($s_u = 1$)

- $G \sim \mathrm{ER}(n, p)$
- $G_1 = G_2 = G$
- G_2' isomorphic to G_1

Theorem (Babai, Erdős, and Selkow 1980, Czajka and Pandurangan 2008)

Assume $p \leq 1/2$

- If $np \geq \log n + \omega(1)$, $\exists$ a polynomial-time algorithm
- If $np \leq \log n - \omega(1)$, no algorithms exist



Conjecture for correlated Erdős–Rényi alignment

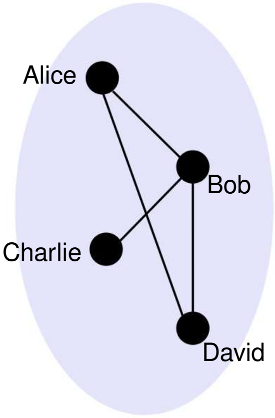
- If $nps_u^2 \geq \log n + \omega(1)$, $\exists$ an algorithm
- If $nps_u^2 \leq \log n - \omega(1)$, no algorithms exist
- No polynomial-time algorithms achieve the IT limit

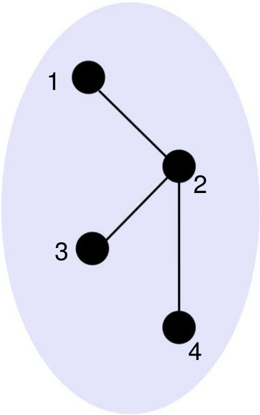
- **Correlated Erdős–Rényi graph alignment**
  - Information-theoretic limit: Pedarsani and Grossglauser (2011), Cullina and Kiyavash (2016), Cullina and Kiyavash (2017), Wu, Xu, and Yu (2021)
  - Polynomial-time algorithm: Lyzinski, Fishkind, Fiori, Vogelstein, Priebe, and Sapiro (2015), Nassar, Veldt, Mohammadi, Grama, and Gleich (2018), Feizi, Quon, Recamonde-Mendoza, Medard, Kellis, and Jadbabaie (2019), Fan, Mao, Wu, and Xu (2020), Onaran and Villar (2020), Barak, Chou, Lei, Schramm, and Sheng (2019a), Mao, Rudelson, and Tikhomirov (2021), Mao, Wu, Xu, and Yu (2022, 2023), Ding and Li (2023)

Related works

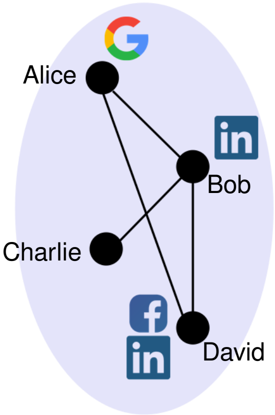
- **Correlated Erdős–Rényi graph alignment**
  - Information-theoretic limit: Pedarsani and Grossglauser (2011), Cullina and Kiyavash (2016), Cullina and Kiyavash (2017), Wu, Xu, and Yu (2021)
  - Polynomial-time algorithm: Lyzinski, Fishkind, Fiori, Vogelstein, Priebe, and Sapiro (2015), Nassar, Veldt, Mohammadi, Grama, and Gleich (2018), Feizi, Quon, Recamonde-Mendoza, Medard, Kellis, and Jadbabaie (2019), Fan, Mao, Wu, and Xu (2020), Onaran and Villar (2020), Barak, Chou, Lei, Schramm, and Sheng (2019a), Mao, Rudelson, and Tikhomirov (2021), Mao, Wu, Xu, and Yu (2022, 2023), Ding and Li (2023)
- **Seeded graph alignment**
  - Information-theoretic limit: converse: Mossel and Xu (2020)
  - Polynomial-time algorithm: Yartseva and Grossglauser (2013), Korula and Lattanzi (2014), Lyzinski, Fishkind, and Priebe (2014), Fishkind, Adali, Patsolic, Meng, Singh, Lyzinski, and Priebe (2019), Shirani, Garg, and Erkip (2017), Mossel and Xu (2020)
- **Bipartite graph alignment**
  - Information-theoretic limit: Cullina, Mittal, and Kiyavash (2018)
  - Polynomial-time algorithm: Hungarian algorithm
- **Many others...**

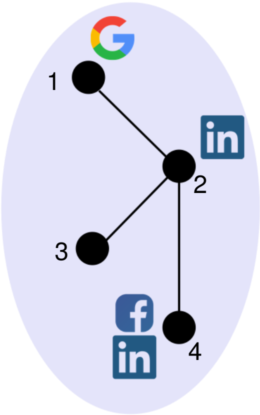
What if graph structure is not enough?



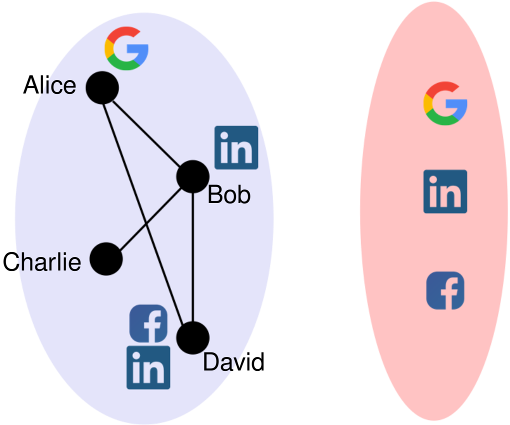


We do know more ...

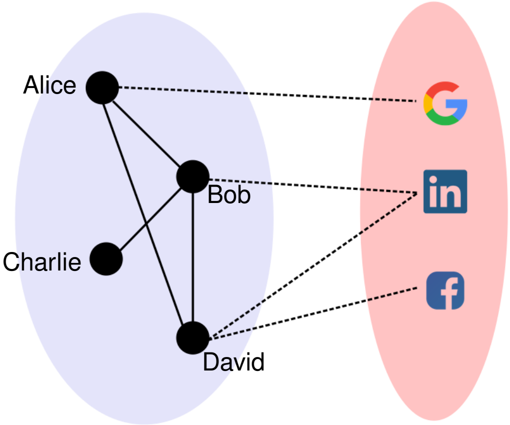




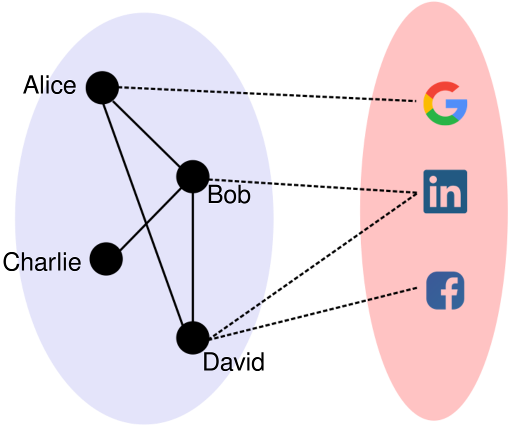
Attributes as vertices



Attributes as vertices



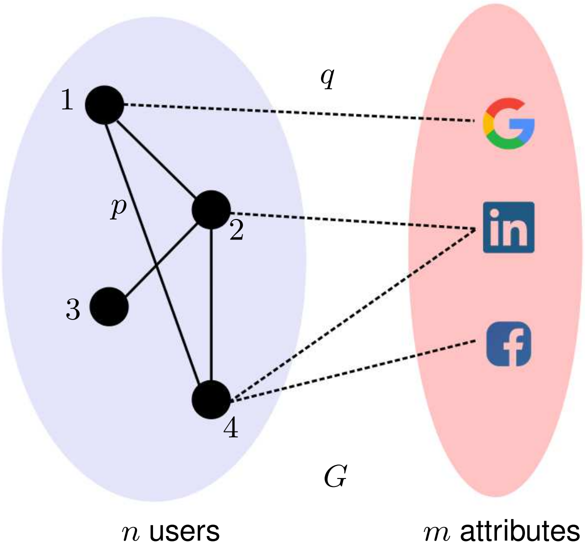
Attributes as vertices



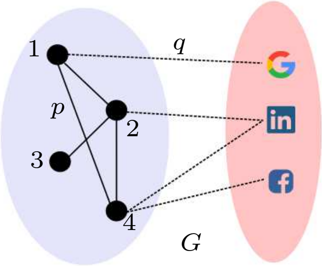
How much benefit can vertex attributes bring?

Model: Attributed Erdős–Rényi graph pair $(G_1, G'_2) \sim \mathcal{G}(n, p, s_\mathrm{u}; m, q, s_\mathrm{a})$

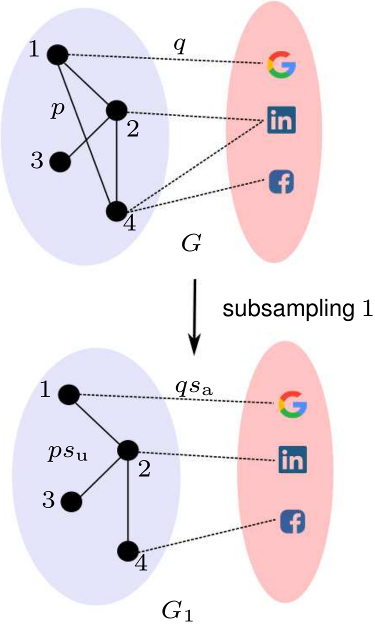
- Base graph G generation



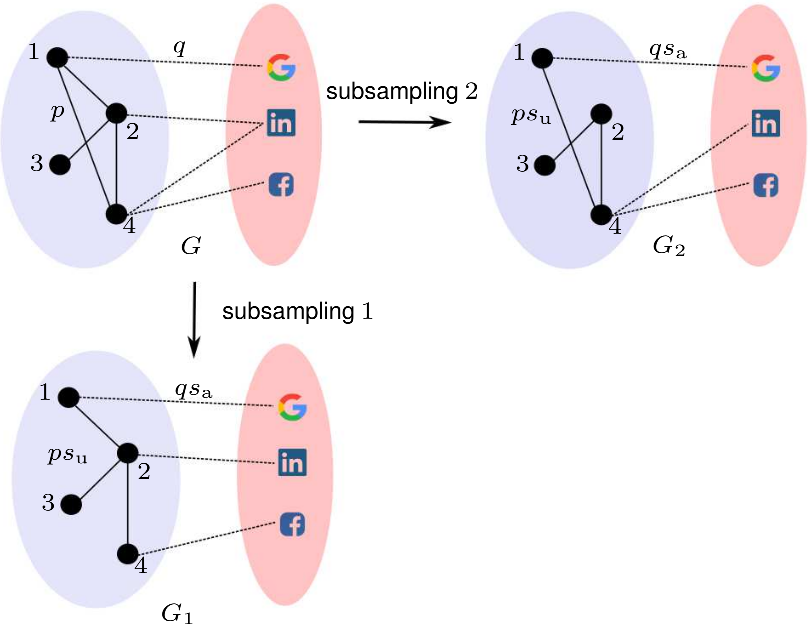
Model: Attributed Erdős–Rényi graph pair $(G_1, G'_2) \sim \mathcal{G}(n, p, s_\mathrm{u}; m, q, s_\mathrm{a})$



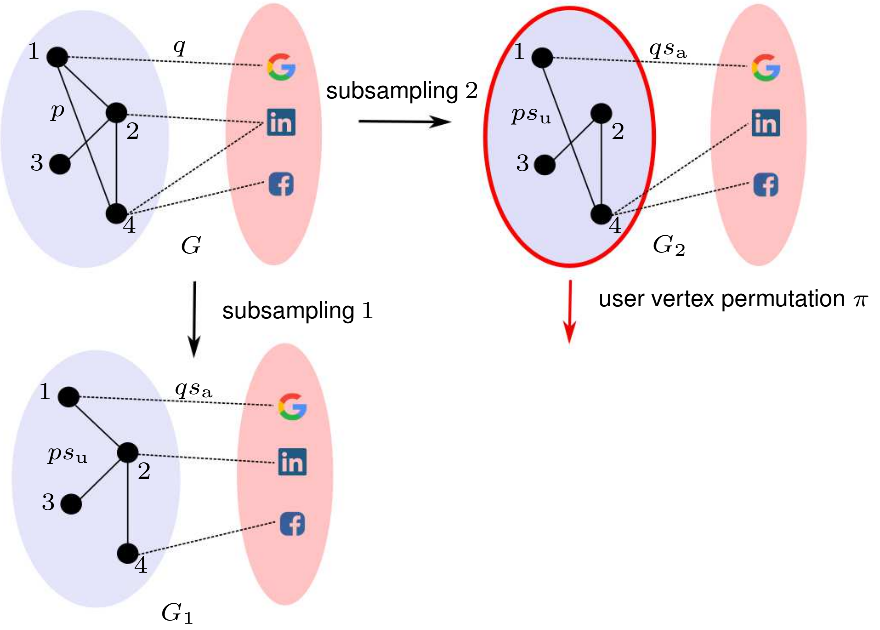
Model: Attributed Erdős–Rényi graph pair $(G_1, G_2') \sim \mathcal{G}(n, p, s_\mathrm{u}; m, q, s_\mathrm{a})$



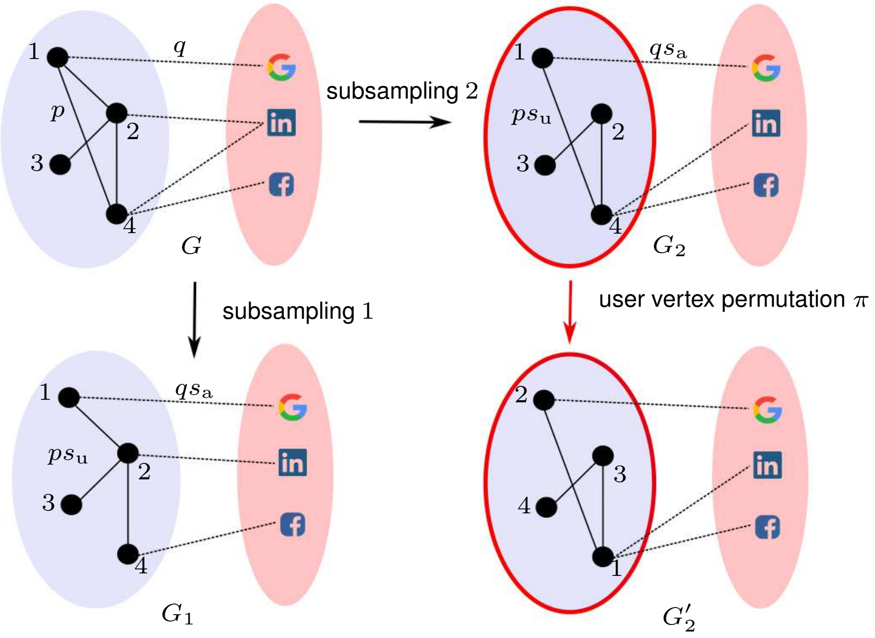
Model: Attributed Erdős–Rényi graph pair $(G_1, G_2') \sim \mathcal{G}(n, p, s_{\mathrm{u}}; m, q, s_{\mathrm{a}})$



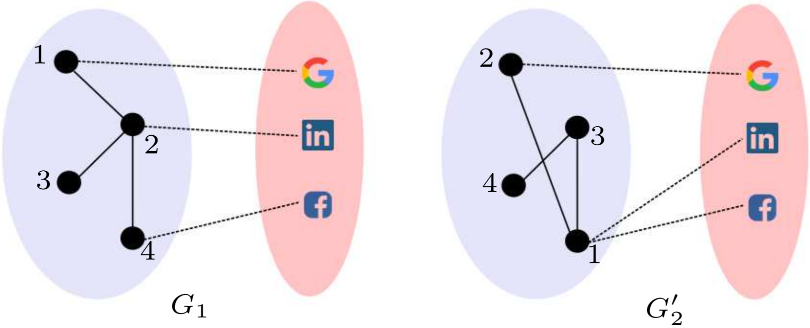
Model: Attributed Erdős–Rényi graph pair $(G_1, G_2') \sim \mathcal{G}(n, p, s_\mathrm{u}; m, q, s_\mathrm{a})$



Model: Attributed Erdős–Rényi graph pair $(G_1, G_2') \sim \mathcal{G}(n, p, s_\mathrm{u}; m, q, s_\mathrm{a})$

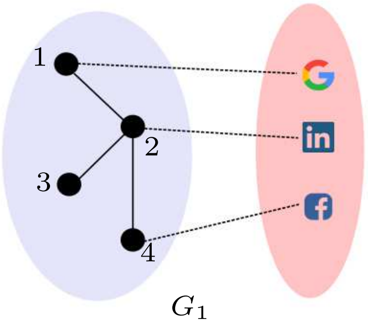


Model: Attributed Erdős–Rényi graph pair $(G_1, G_2') \sim \mathcal{G}(n, p, s_{\rm u}; m, q, s_{\rm a})$



- Exact alignment: Estimation $\hat{\pi}(G_1, G_2')$ such that for uniform Π,

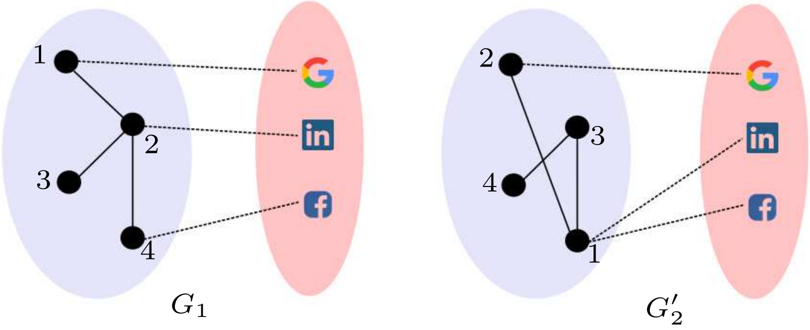
$$\lim_{n\to\infty} \mathsf{P}(\hat{\pi}(G_1, G_2') = \Pi) = 1$$



- Exact alignment: Estimation $\hat{\pi}(G_1, G_2')$ such that for uniform Π,

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

- Achievability: Set of $(n, p, s_\mathrm{u}; m, q, s_\mathrm{a})$ s.t. exact alignment is achievable
  - Information-theoretic limits
  - Efficient algorithms
- Converse: Set of $(n, p, s_\mathrm{u}; m, q, s_\mathrm{a})$ s.t. no algorithm achieves exact alignment



Result 1: IT limits (simplified)

Under mild conditions
$1-p=\Theta(1), 1-q=\Theta(1)$,
$s_{\mathrm{u}}=\Omega(\frac{(\log n)^2}{\sqrt{n}})$,
$s_{\mathrm{a}}=\Omega(\frac{(\log n)^{1.5}}{\sqrt{m}})$

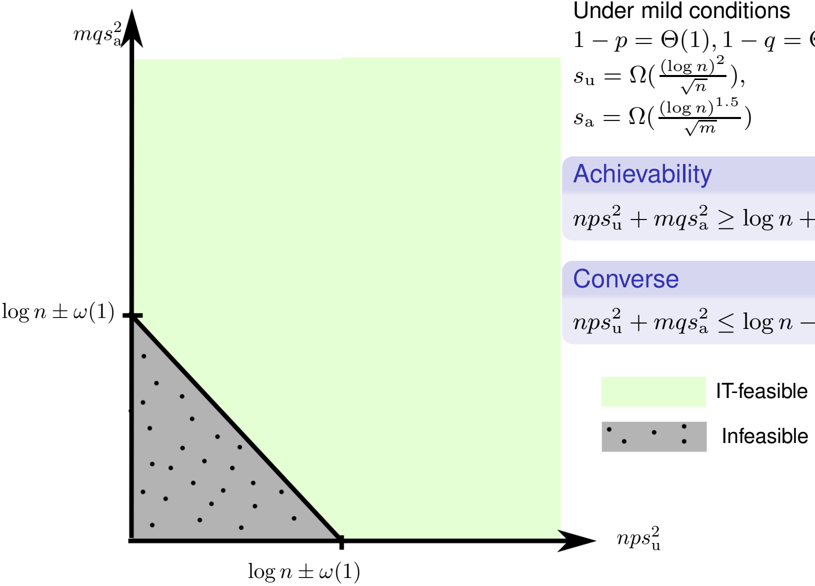
Achievability

$$nps_{\mathrm{u}}^2+mqs_{\mathrm{a}}^2 \geq \log n+\omega(1)$$

Converse

$$nps_{\mathrm{u}}^2+mqs_{\mathrm{a}}^2 \leq \log n-\omega(1)$$

https://arxiv.org/abs/2102.00665



Under mild conditions
$1 - p = \Theta(1), 1 - q = \Theta(1)$,
$s_{\mathrm{u}} = \Omega(\frac{(\log n)^2}{\sqrt{n}})$,
$s_{\mathrm{a}} = \Omega(\frac{(\log n)^{1.5}}{\sqrt{m}})$

Achievability

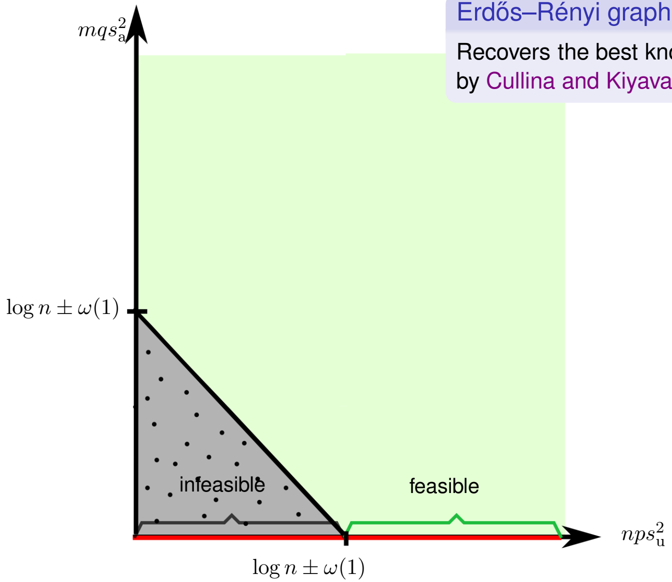
$$nps_{\mathrm{u}}^2 + mqs_{\mathrm{a}}^2 \geq \log n + \omega(1)$$

Converse

$$nps_{\mathrm{u}}^2 + mqs_{\mathrm{a}}^2 \leq \log n - \omega(1)$$

https://arxiv.org/abs/2102.00665

Result 1: IT limits (simplified)

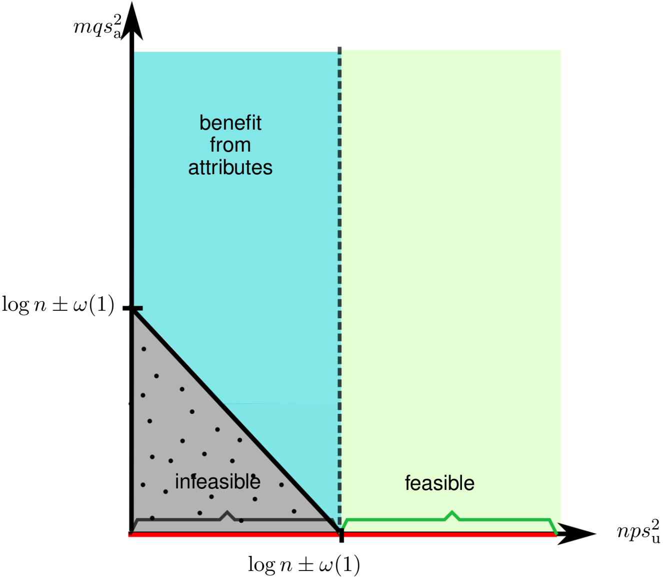


Erdős–Rényi graph alignment

Recovers the best known IT limits by Cullina and Kiyavash (2017)

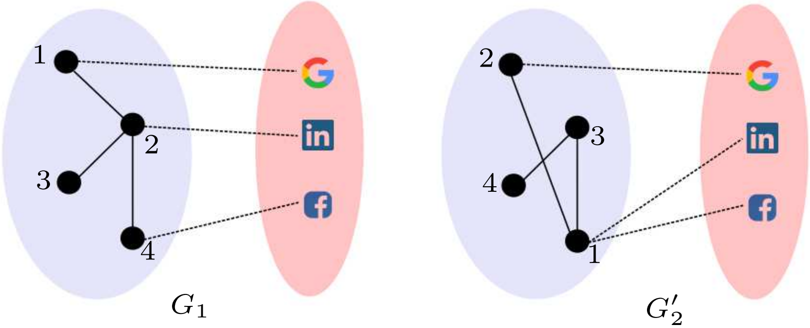
https://arxiv.org/abs/2102.00665

Benefit from attribute information



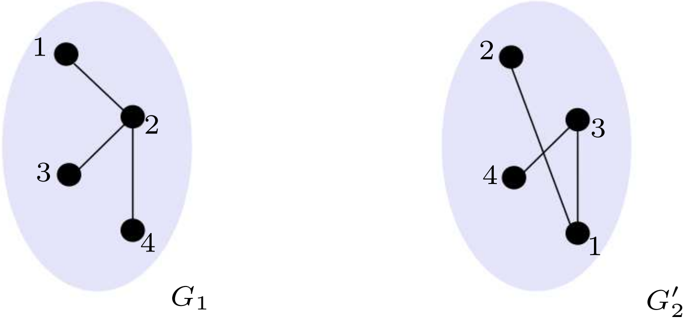
https://arxiv.org/abs/2102.00665

Relation to other models



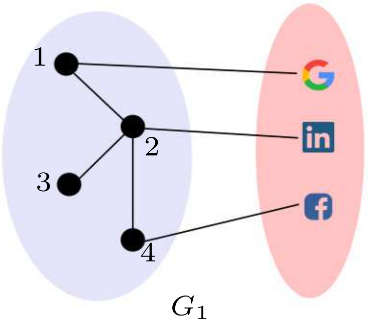
Relation to other models

- When $m = 0$ or $qs_{\mathrm{a}} = 0$, reduces to correlated Erdős–Rényi graph alignment



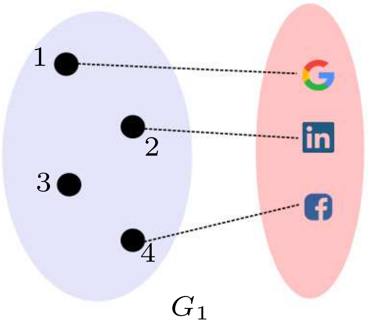
Relation to other models

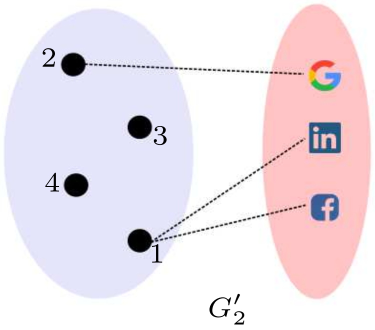
- When $m = 0$ or $qs_{\mathrm{a}} = 0$, reduces to correlated Erdős–Rényi graph alignment
- When $p = q$ and $s_{\mathrm{u}} = s_{\mathrm{a}}$, reduces to seeded graph alignment



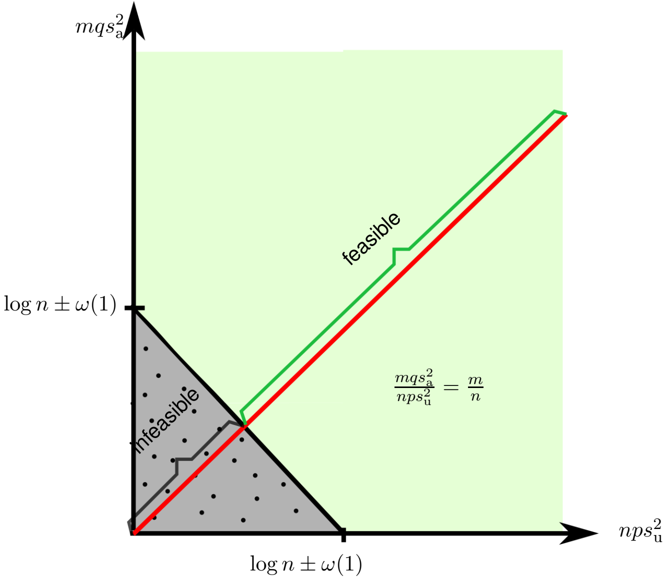
Relation to other models

- When $m = 0$ or $qs_{\mathrm{a}} = 0$, reduces to correlated Erdős–Rényi graph alignment
- When $p = q$ and $s_{\mathrm{u}} = s_{\mathrm{a}}$, reduces to seeded graph alignment
- When $ps_{\mathrm{a}} = 0$, reduces to bipartite graph alignment





Specialization to seeded graph alignment: $p = q, s_u = s_a$



Specialization to seeded graph alignment: $p = q, s_{\mathrm{u}} = s_{\mathrm{a}}$

- For fair comparison, assume n unmatched vertices and m seeds

Specialization to seeded graph alignment: $p = q, s_{\mathrm{u}} = s_{\mathrm{a}}$

- For fair comparison, assume n unmatched vertices and m seeds

Best-known results

- Achievability: unseeded achievability by Cullina and Kiyavash (2017)

$$(m+n)ps_{\mathrm{u}}^2 \geq \log(m+n) + \omega(1)$$

- Converse: Mossel and Xu (2020) for $m = O(n)$

$$(m+n)ps_{\mathrm{u}}^2 \leq \log(m+n) + O(1)$$

Specialization to seeded graph alignment: $p = q, s_{\mathrm{u}} = s_{\mathrm{a}}$

- For fair comparison, assume n unmatched vertices and m seeds

Best-known results

- Achievability: unseeded achievability by Cullina and Kiyavash (2017)

$$(m+n)ps_{\mathrm{u}}^2 \geq \log(m+n) + \omega(1)$$

- Converse: Mossel and Xu (2020) for $m = O(n)$

$$(m+n)ps_{\mathrm{u}}^2 \leq \log(m+n) + O(1)$$

Our result: Tight threshold

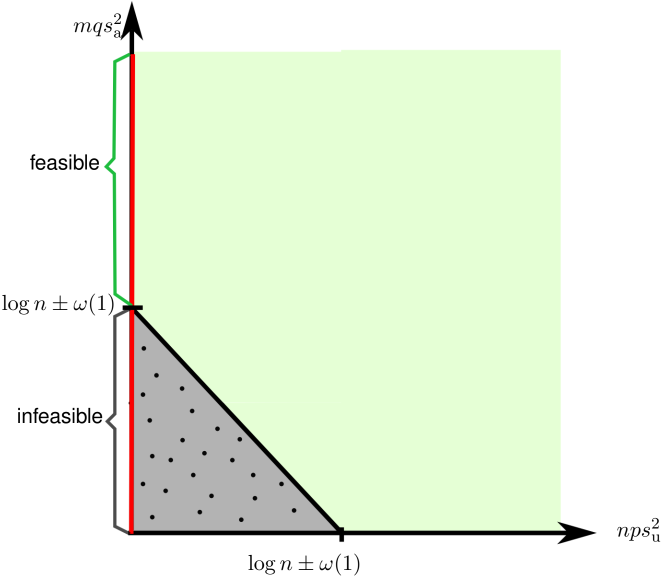
- Achievability: strict improvement

$$(m+n)ps_{\mathrm{u}}^2 \geq \log n + \omega(1)$$

- Converse: extension to $m = \omega(n)$

$$(m+n)ps_{\mathrm{u}}^2 \leq \log n - \omega(1)$$

Specialization to bipartite graph alignment: $ps_{\mathrm{u}} = 0$



Specialization to bipartite graph alignment: $ps_{\mathrm{u}} = 0$

- Studied in the more general setting of database alignment (Cullina et al. 2018)

Best-known results

- Achievability:

$$\frac{1}{2} I_2^{\circ}(Q^{\otimes m}) \geq \log n + \omega(1)$$

- Converse: for constant $\epsilon \in (0, 1)$

$$\frac{1}{2} I_2^{\circ}(Q^{\otimes m}) \leq (1 - \epsilon) \log n$$

where $Q = \begin{pmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{pmatrix}$, $I_2^{\circ}(A) = -\log \operatorname{tr}((ZZ^T)^2)$, $Z_{ij} = \sqrt{A_{ij}}$

Specialization to bipartite graph alignment: $ps_{\mathrm{u}} = 0$

- Studied in the more general setting of database alignment (Cullina et al. 2018)

Refined best-known results

- Achievability:

$$-\frac{m}{2}\log(1-2\psi_{\mathrm{a}}) \geq \log n + \omega(1)$$

- Converse: for constant $\epsilon \in (0,1)$

$$-\frac{m}{2}\log(1-2\psi_{\mathrm{a}}) \leq (1-\epsilon)\log n$$

where $\psi_{\mathrm{a}} = (\sqrt{q_{11}q_{00}} - \sqrt{q_{01}q_{10}})^2$

Specialization to bipartite graph alignment: $ps_{\mathrm{u}} = 0$

- Studied in the more general setting of database alignment (Cullina et al. 2018)

Refined best-known results

- Achievability:

$$-\frac{m}{2}\log(1-2\psi_{\mathrm{a}}) \geq \log n + \omega(1)$$

- Converse: for constant $\epsilon \in (0,1)$

$$-\frac{m}{2}\log(1-2\psi_{\mathrm{a}}) \leq (1-\epsilon)\log n$$

where $\psi_{\mathrm{a}} = (\sqrt{q_{11}q_{00}} - \sqrt{q_{01}q_{10}})^2$

Our result

- Achievability: recovers the best

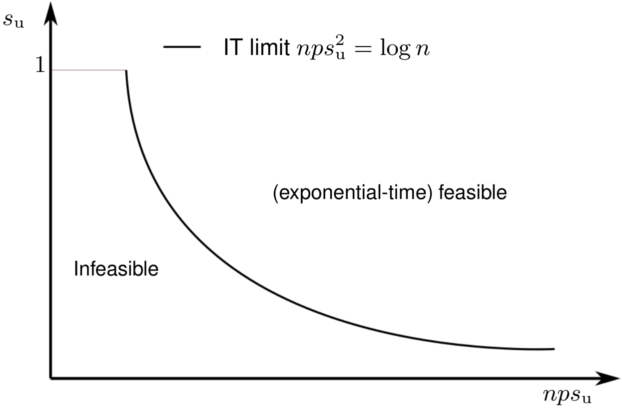
$$-\frac{m}{2}\log(1-2\psi_{\mathrm{a}}) \geq \log n + \omega(1)$$

- Converse: strict improvement

$$mqs_{\mathrm{a}}^2 \leq \log n - \omega(1)$$

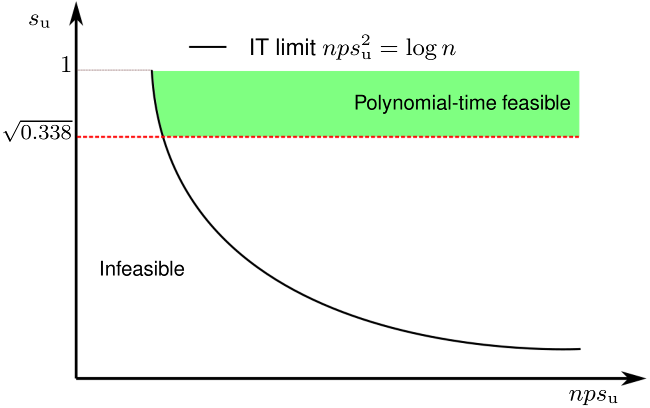
Efficient Algorithms

Conjectured information-computation gap in correlated Erdős–Rényi model



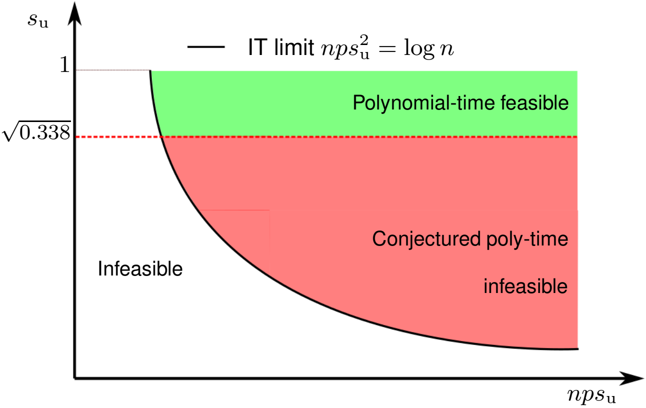
- IT limits: Cullina and Kiyavash (2017)

Conjectured information-computation gap in correlated Erdős–Rényi model



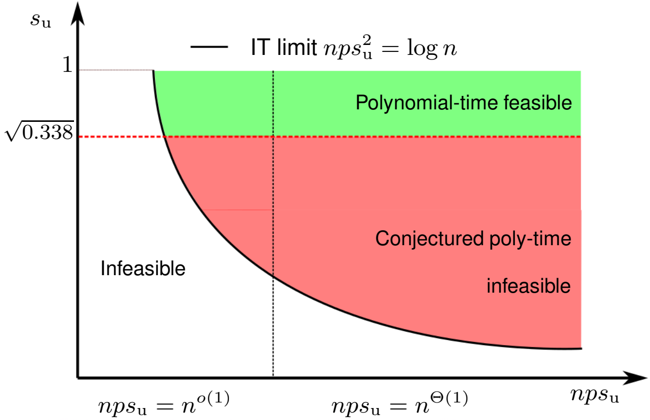
- IT limits: Cullina and Kiyavash (2017)
- Poly-time algorithm with correlation at Otter's constant: Mao, Wu, Xu, and Yu (2023)

Conjectured information-computation gap in correlated Erdős–Rényi model



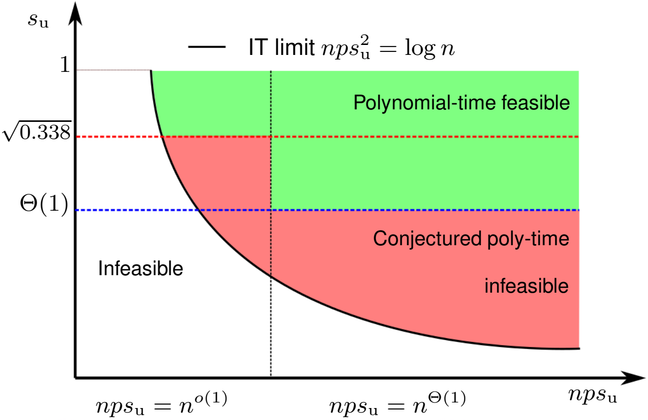
- IT limits: Cullina and Kiyavash (2017)
- Poly-time algorithm with correlation at Otter's constant: Mao, Wu, Xu, and Yu (2023)
- Information-computation gap conjecture: Yu (2023), Mao, Wu, Xu, and Yu (2022)

Conjectured information-computation gap in correlated Erdős–Rényi model



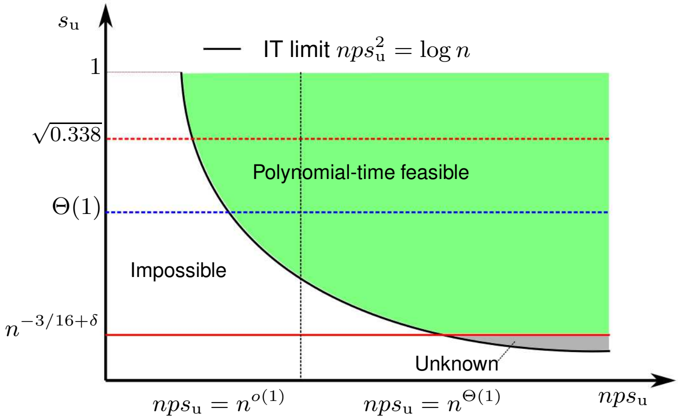
- IT limits: Cullina and Kiyavash (2017)
- Poly-time algorithm with correlation at Otter's constant: Mao, Wu, Xu, and Yu (2023)
- Information-computation gap conjecture: Yu (2023), Mao, Wu, Xu, and Yu (2022)
- Poly-time algorithm in the dense regime: Ding and Li (2023)

Conjectured information-computation gap in correlated Erdős–Rényi model



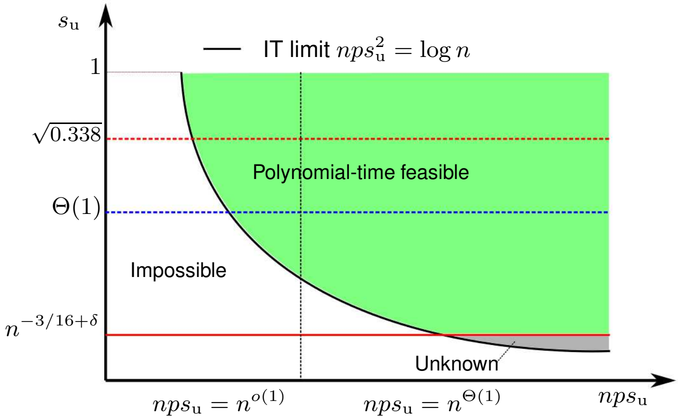
- IT limits: Cullina and Kiyavash (2017)
- Poly-time algorithm with correlation at Otter's constant: Mao, Wu, Xu, and Yu (2023)
- Information-computation gap conjecture: Yu (2023), Mao, Wu, Xu, and Yu (2022)
- Poly-time algorithm in the dense regime: Ding and Li (2023)

Result 2: Efficient algorithms for attributed graph alignment



https://arxiv.org/abs/2201.10106 https://arxiv.org/abs/2308.09210

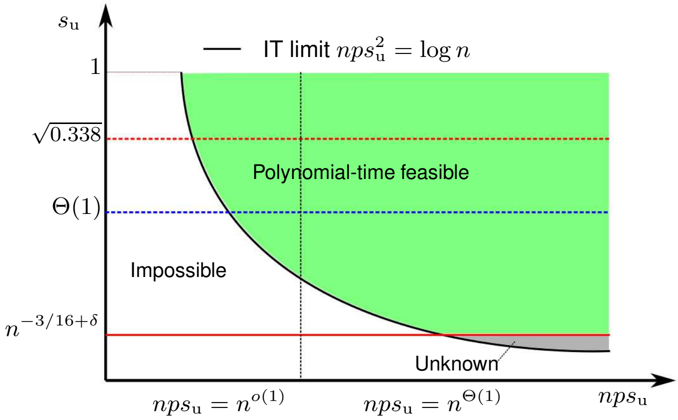
Result 2: Efficient algorithms for attributed graph alignment



With a tiny bit of attribute information (e.g. $mqs_a^2 = 1/\sqrt{\log n}$),
poly-time algorithms can achieve exact alignment with vanishing correlation!

https://arxiv.org/abs/2201.10106 https://arxiv.org/abs/2308.09210

Result 2: Efficient algorithms for attributed graph alignment

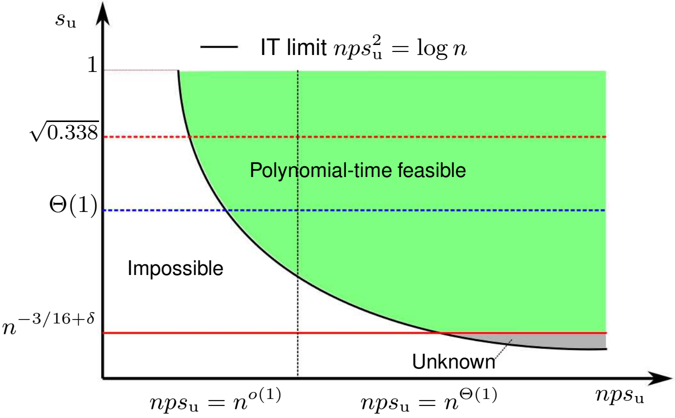


Specialization to seeded graph alignment ($p = q, s_u = s_a$)

Strictly improve the best known achievable region for poly-time algorithms by Shirani, Garg, and Erkip (2017), Mossel and Xu (2020)

https://arxiv.org/abs/2201.10106 https://arxiv.org/abs/2308.09210

Result 2: Efficient algorithms for attributed graph alignment



Specialization to bipartite graph alignment ($ps_{\mathrm{u}} = 0$)

Alternative poly-time algorithm for the Hungarian algorithm with a smaller time complexity when $m = o(n)$

https://arxiv.org/abs/2201.10106 https://arxiv.org/abs/2308.09210

Efficient algorithms by subgraph counting

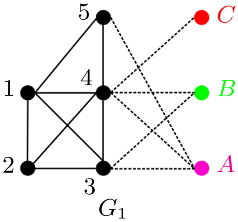
- Idea: use the occurrences of a chosen graph structure as vertex feature
  - Identifying clusters in graphs: Mossel et al. (2014)
  - Testing correlation between two graphs: Mao et al. (2022)
  - Graph alignment: Barak et al. (2019b), Mao et al. (2023)

Efficient algorithms by subgraph counting

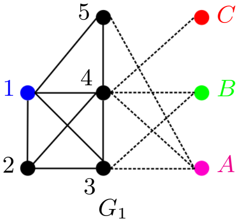
- Idea: use the occurrences of a chosen graph structure as vertex feature
  - Identifying clusters in graphs: Mossel et al. (2014)
  - Testing correlation between two graphs: Mao et al. (2022)
  - Graph alignment: Barak et al. (2019b), Mao et al. (2023)

- For attributed graphs: We identify a *rooted* subgraph involving both attributes and users

Proposed subgraph counting algorithm

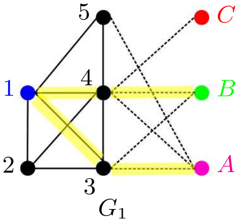


Proposed subgraph counting algorithm



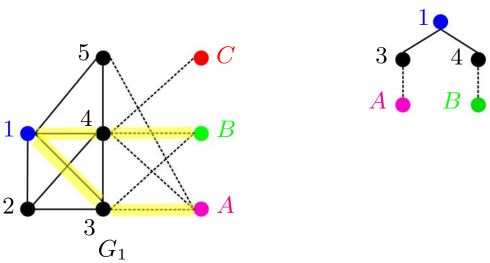
$$W_{1,\{A,B\}}(G_1)$$

Proposed subgraph counting algorithm



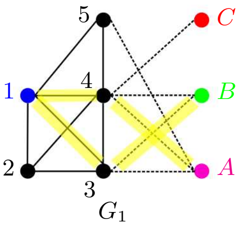
$$W_{1,\{A,B\}}(G_1)$$

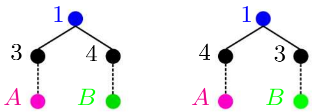
Proposed subgraph counting algorithm



$$W_{1,\{A,B\}}(G_1)$$

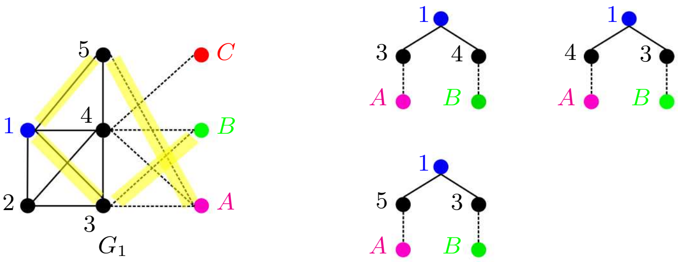
Proposed subgraph counting algorithm





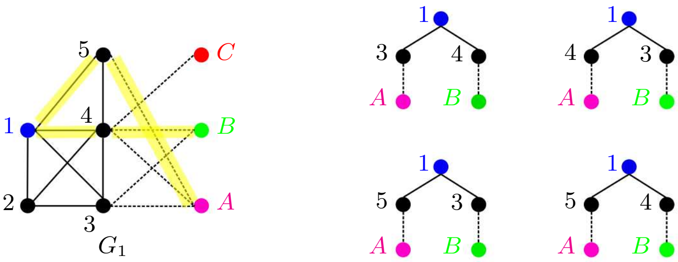
$$W_{1,\{A,B\}}(G_1)$$

Proposed subgraph counting algorithm



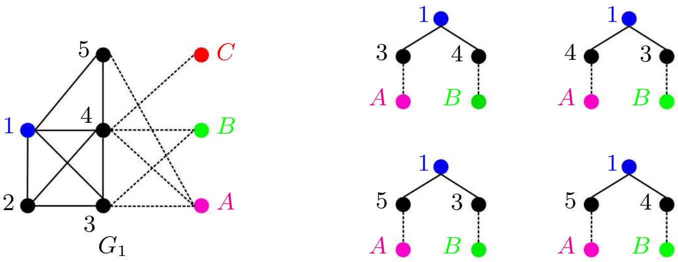
$$W_{1,\{A,B\}}(G_1)$$

Proposed subgraph counting algorithm



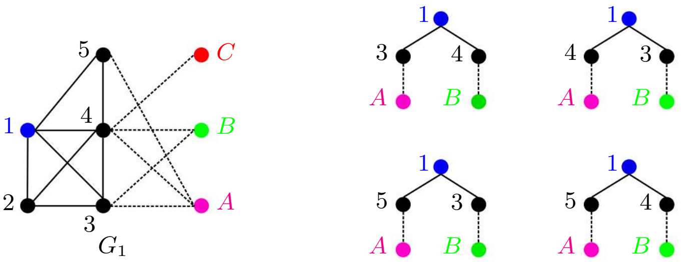
$W_{1,\{A,B\}}(G_1)$

Proposed subgraph counting algorithm



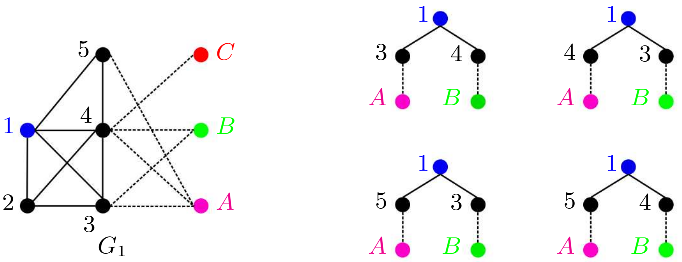
$$W_{1,\{A,B\}}(G_1) = 4$$

Proposed subgraph counting algorithm



$$W_{1,\{A,B\}}(G_1) = 4$$

- Construct feature vector for each user vertex



$$W_{1,\{A,B\}}(G_1) = 4$$

- Construct feature vector for each user vertex
  e.g.: $X_1 = (W_{1,\{A,B\}}(G_1), W_{1,\{A,C\}}(G_1), W_{1,\{B,C\}}(G_1))$

Proposed subgraph counting algorithm

- Similarity score between user i from G_1 and j from G_2'

$$\Gamma_{ij} \triangleq X_i \cdot X_j = \sum_{\mathcal{T}:|\mathcal{T}|=k} W_{i,\mathcal{T}}(G_1) W_{j,\mathcal{T}}(G_2').$$

- Key observation: For any wrong pair $j \neq \Pi(i)$,

$$\mathsf{E}[\underbrace{W_{i,\mathcal{T}}(G_1) W_{j,\mathcal{T}}(G_2')}_{\text{almost independent}}] < \mathsf{E}[\underbrace{W_{i,\mathcal{T}}(G_1) W_{\Pi(i),\mathcal{T}}(G_2')}_{\text{positively correlated}}],$$

Proposed subgraph counting algorithm

- Similarity score between user i from G_1 and j from G_2'

$$\Gamma_{ij} \triangleq X_i \cdot X_j = \sum_{\mathcal{T}:|\mathcal{T}|=k} W_{i,\mathcal{T}}(G_1) W_{j,\mathcal{T}}(G_2').$$

- Key observation: For any wrong pair $j \neq \Pi(i)$,

$$\mathsf{E}[\underbrace{W_{i,\mathcal{T}}(G_1) W_{j,\mathcal{T}}(G_2')}_{\text{almost independent}}] < \mathsf{E}[\underbrace{W_{i,\mathcal{T}}(G_1) W_{\Pi(i),\mathcal{T}}(G_2')}_{\text{positively correlated}}],$$

which further implies

$$\mathsf{E}[\Gamma_{ij}] < \mathsf{E}[\Gamma_{i,\Pi(i)}].$$

Proposed subgraph counting algorithm

- Similarity score between user i from G_1 and j from G_2'

$$\Gamma_{ij} \triangleq X_i \cdot X_j = \sum_{\mathcal{T}:|\mathcal{T}|=k} W_{i,\mathcal{T}}(G_1) W_{j,\mathcal{T}}(G_2').$$

- Key observation: For any wrong pair $j \neq \Pi(i)$,

$$\mathsf{E}[\underbrace{W_{i,\mathcal{T}}(G_1) W_{j,\mathcal{T}}(G_2')}_{\text{almost independent}}] < \mathsf{E}[\underbrace{W_{i,\mathcal{T}}(G_1) W_{\Pi(i),\mathcal{T}}(G_2')}_{\text{positively correlated}}],$$

which further implies

$$\mathsf{E}[\Gamma_{ij}] < \mathsf{E}[\Gamma_{i,\Pi(i)}].$$

Proposed subgraph counting algorithm

- Similarity score between user i from G_1 and j from G'_2

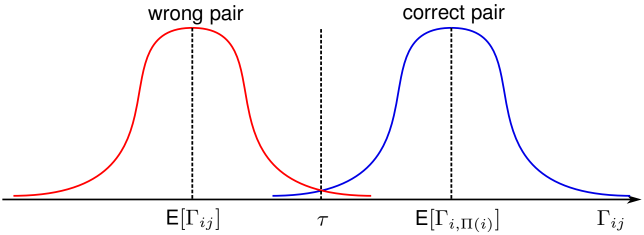
$$\Gamma_{ij} \triangleq X_i \cdot X_j = \sum_{\mathcal{T}:|\mathcal{T}|=k} W_{i,\mathcal{T}}(G_1) W_{j,\mathcal{T}}(G'_2).$$

- Key observation: For any wrong pair $j \neq \Pi(i)$,

$$\mathsf{E}[\underbrace{W_{i,\mathcal{T}}(G_1) W_{j,\mathcal{T}}(G'_2)}_{\text{almost independent}}] < \mathsf{E}[\underbrace{W_{i,\mathcal{T}}(G_1) W_{\Pi(i),\mathcal{T}}(G'_2)}_{\text{positively correlated}}],$$

which further implies

$$\mathsf{E}[\Gamma_{ij}] < \mathsf{E}[\Gamma_{i,\Pi(i)}].$$



Summary

- Propose attributed Erdős–Rényi graph pair model
  - Understand the benefit of attributes
  - Unify existing models
- Characterize the information-theoretic limits
  - Improve IT limits for existing models
- Propose polynomial-time algorithms
  - Improve efficient algorithms for existing models
  - Shed new light on information-computation gap

More details: arXiv:2102.006655 arXiv:2201.10106 arXiv:2308.09210

Summary

- Propose attributed Erdős–Rényi graph pair model
  - Understand the benefit of attributes
  - Unify existing models
- Characterize the information-theoretic limits
  - Improve IT limits for existing models
- Propose polynomial-time algorithms
  - Improve efficient algorithms for existing models
  - Shed new light on information-computation gap

Thank you!

More details: arXiv:2102.006655 arXiv:2201.10106 arXiv:2308.09210

Proof Sketch for IT limits

Key ideas in achievability

- Correlated Erdős–Rényi model

$$\hat{\pi}_{\text{MAP}} = \operatorname{argmin}_{\pi} \text{edge misalignment between } G_1 \text{ and } \pi^{-1}(G_2')$$

Key ideas in achievability

- Correlated Erdős–Rényi model

$$\hat{\pi}_{\text{MAP}} = \text{argmin}_{\pi} \text{edge misalignment between } G_1 \text{ and } \pi^{-1}(G_2')$$

- Attributed Erdős–Rényi model

MAP estimator = weighted minimum misalignment

$$\hat{\pi}_{\text{MAP}} = \text{argmin}_{\pi}\{w_1 \Delta_\pi^{\text{u}} + w_2 \Delta_\pi^{\text{a}}\},$$

where
Δ_π^{u} : user-user edge misalignment between G_1 and $\pi^{-1}(G_2')$
Δ_π^{a} : user-attribute edge misalignment between G_1 and $\pi^{-1}(G_2')$
$w_1 = \log\left(\frac{p_{11}p_{00}}{p_{10}p_{01}}\right)$, $w_2 = \log\left(\frac{q_{11}q_{00}}{q_{10}q_{01}}\right)$

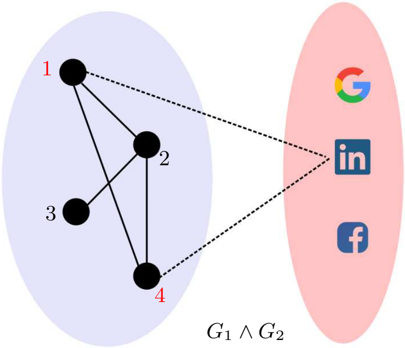
- Error bounding techniques (Cullina and Kiyavash (2017)):
  - Orbit decomposition
  - Generating functions

Key ideas in converse

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

Key ideas in converse

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

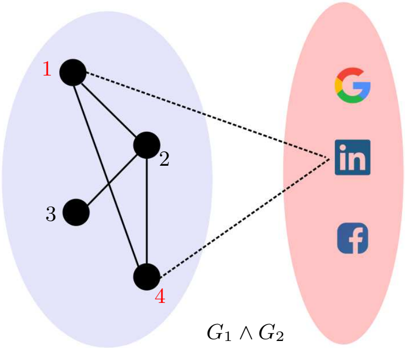


Key ideas in converse

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

- Correlated Erdős–Rényi model

$$|\mathrm{Aut}| \geq |\mathrm{Aut}_{\mathrm{iso}}|$$



Key ideas in converse

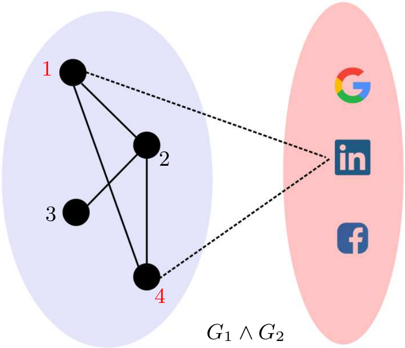
- $\mathsf{P}(\hat{\pi}_{\mathrm{MAP}} = \Pi) \leq \frac{1}{|\mathrm{Aut}(G_1 \wedge G_2)|}$

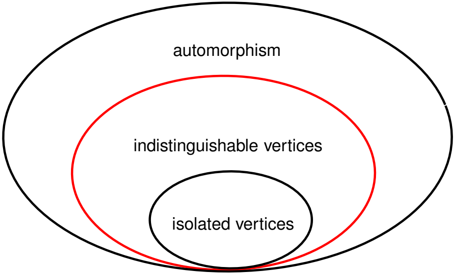
- Correlated Erdős–Rényi model

$$|\mathrm{Aut}| \geq |\mathrm{Aut}_{\mathrm{iso}}|$$

- Attributed Erdős–Rényi model

$$|\mathrm{Aut}| \geq |\mathrm{Aut}_{\mathrm{ind}}| \geq |\mathrm{Aut}_{\mathrm{iso}}|$$





i and j are indistinguishable if for all $k \in \mathcal{V} \setminus \{i, j\}$, $i \sim k$ iff $j \sim k$

Key ideas in converse

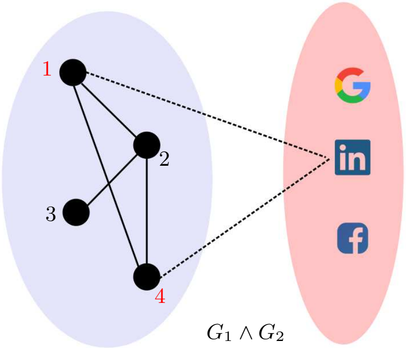
- $\mathsf{P}(\hat{\pi}_{\mathrm{MAP}} = \Pi) \le \frac{1}{|\mathrm{Aut}(G_1 \wedge G_2)|}$

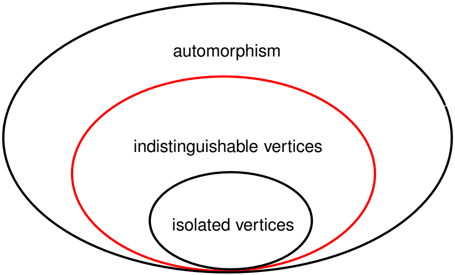
- Correlated Erdős–Rényi model

$$|\mathrm{Aut}| \ge |\mathrm{Aut}_{\mathrm{iso}}|$$

- Attributed Erdős–Rényi model

$$|\mathrm{Aut}| \ge |\mathrm{Aut}_{\mathrm{ind}}| \ge |\mathrm{Aut}_{\mathrm{iso}}|$$

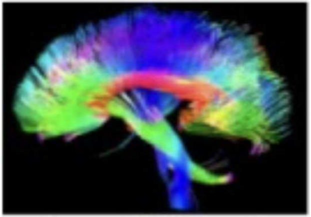


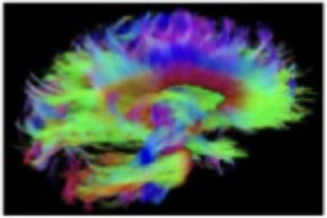


i and j are indistinguishable if for all $k \in \mathcal{V} \setminus \{i, j\}$, $i \sim k$ iff $j \sim k$

Side result: threshold of the existence of indistinguishable pairs in attributed graphs

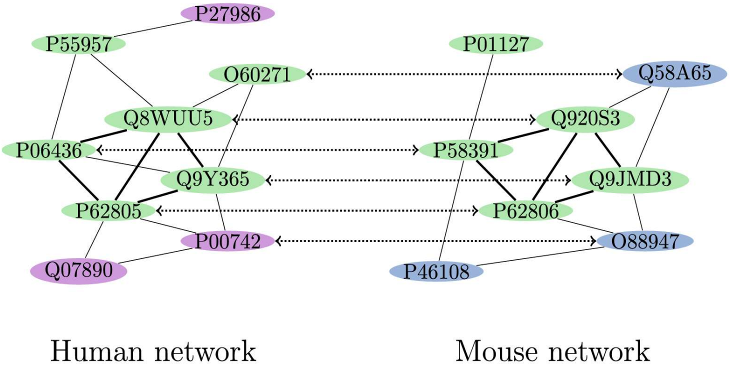
Motivation 2: Biomedical image analysis from multiple views





Brain connectome network analysis (Zhang, He, Chen, Luo, Zhou, and Wang 2018)

Motivation 3: Protein with similar functions across different species



Uncover relation and transfer biological knowledge between different species (Kazemi, Hassani, Grossglauser, and Modarres 2016)

References

L. Babai, P. Erdős, and S. Selkow. Random graph isomorphism. *SIAM J. Comput.*, 9: 628–635, 08 1980. doi: 10.1137/0209047.

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., 2019a.

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*, volume 32. Curran Associates, Inc., 2019b.

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.

D. Cullina, P. Mittal, and N. Kiyavash. Fundamental limits of database alignment. In *Proc. IEEE Int. Symp. Information Theory*, pages 651–655, 2018. doi: 10.1109/ISIT.2018.8437908.

References (cont.)

T. Czajka and G. Pandurangan. Improved random graph isomorphism. *Journal of Discrete Algorithms*, 6(1):85–92, 2008. ISSN 1570-8667. doi: https://doi.org/10.1016/j.jda.2007.01.002. URL https://www.sciencedirect.com/science/article/pii/S1570866707000147. Selected papers from AWOCA 2005.

J. Ding and Z. Li. A polynomial-time iterative algorithm for random graph matching with non-vanishing correlation. 2023. URL https://arxiv.org/abs/2306.00266.

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.

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.

D. E. Fishkind, S. Adali, H. G. Patsolic, L. Meng, D. Singh, V. Lyzinski, and C. E. Priebe. Seeded graph matching. *Pattern recognition*, 87:203–215, 2019.

E. Kazemi, H. Hassani, M. Grossglauser, and H. P. Modarres. Proper: global protein interaction network alignment through percolation matching. *BMC Bioinformatics*, 2016.

T. C. Koopmans and M. Beckmann. Assignment problems and the location of economic activities. *Econometrica*, 25(1):53–76, 1957. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1907742.

N. Korula and S. Lattanzi. An efficient reconciliation algorithm for social networks. *Proc. VLDB Endow.*, 7(5):377–388, Jan. 2014. ISSN 2150-8097. doi: 10.14778/2732269.2732274.

V. Lyzinski, D. E. Fishkind, and C. E. Priebe. Seeded graph matching for correlated erdös-rényi graphs. *J. Mach. Learn. Res.*, 15(1):3513–3540, 2014.

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.

C. Mao, M. Rudelson, and K. Tikhomirov. Exact matching of random graphs with constant correlation. 2021. URL https://arxiv.org/abs/2110.05000.

C. Mao, Y. Wu, J. Xu, and S. H. Yu. Testing network correlation efficiently via counting trees. 2022. URL https://arxiv.org/abs/2110.11816.

C. Mao, Y. Wu, J. Xu, and S. H. Yu. Random graph matching at otter's threshold via counting chandeliers. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, page 1345–1356, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450399135. doi: 10.1145/3564246.3585156.

References (cont.)

E. Mossel and J. Xu. Seeded graph matching via large neighborhood statistics. *Random Struct. & Algorithms*, 57(3):570–611, 2020. doi: 10.1002/rsa.20934.

E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162, 07 2014. doi: 10.1007/s00440-014-0576-6.

A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *Proc. IEEE Symp. Security and Privacy*, pages 111–125, 2008. doi: 10.1109/SP.2008.33.

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.

S. Oya and F. Kerschbaum. IHOP: Improved statistical query recovery against searchable symmetric encryption through quadratic optimization. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 2407–2424, Boston, MA, Aug. 2022. USENIX Association. ISBN 978-1-939133-31-1. URL https://www.usenix.org/conference/usenixsecurity22/presentation/oya.

References (cont.)

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.

F. Shirani, S. Garg, and E. Erkip. Seeded graph matching: Efficient algorithms and theoretical guarantees. In *Proc. Asilomar Conf. Signals, Systems, and Computers*, pages 253–257. IEEE, 2017.

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

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.

S. H. Yu. *Matching in Networks: Fundamental Limits and Efficient Algorithms*. PhD Thesis, Duke University, 2023.

X. Zhang, L. He, K. Chen, Y. Luo, J. Zhou, and F. Wang. Multi-view graph convolutional network and its applications on neuroimage analysis for parkinson's disease. In *AMIA Annu Symp Proc.*, 2018.