iNEAT: Incomplete Network Alignment

Presented by Si Zhang (ASU)

Si Zhang Hanghang Tong Jie Tang Jiejun Xu Wei Fan

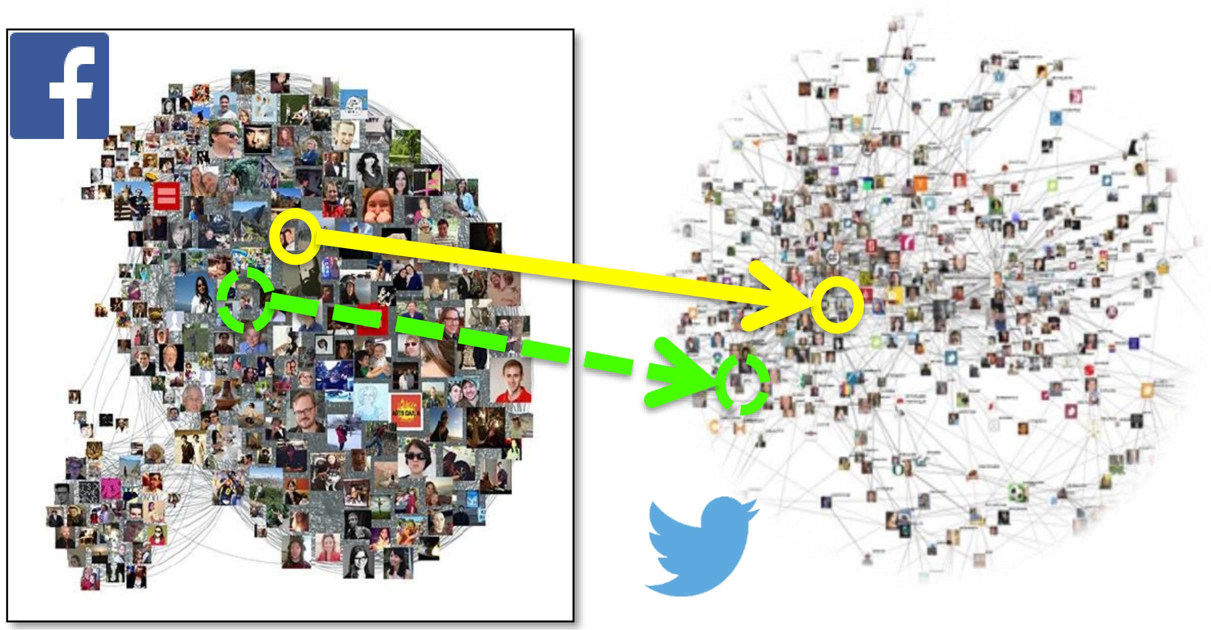
Why Incomplete Network Alignment?

- Networks are multi-sourced (**Variety**)

Network Alignment

⬇

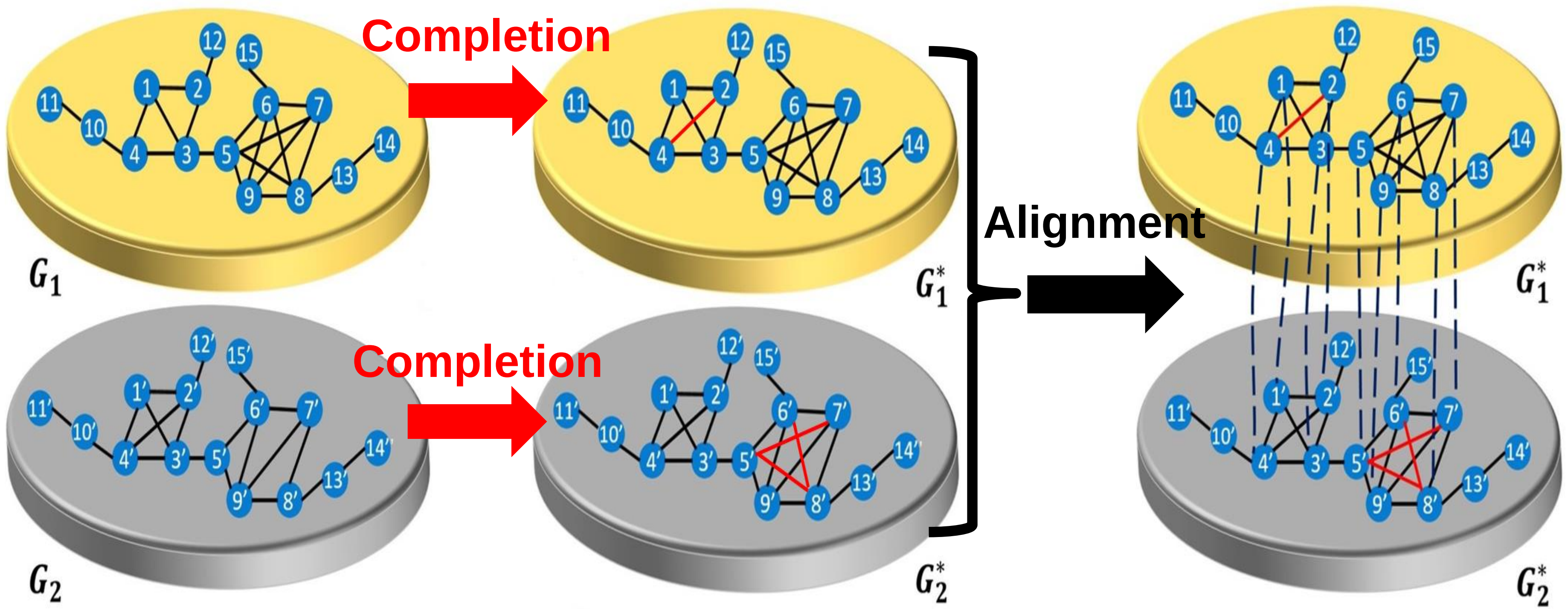
Find Someone Like You



- Networks are often incomplete (**Veracity**)
  - Missing edges in multiple networks
- **Q:** how to align networks with missing edges?

How about Completion-Then-Alignment?

- A natural choice



- Completion ideally provides higher-quality networks

Completion-Then-Alignment: Limitations

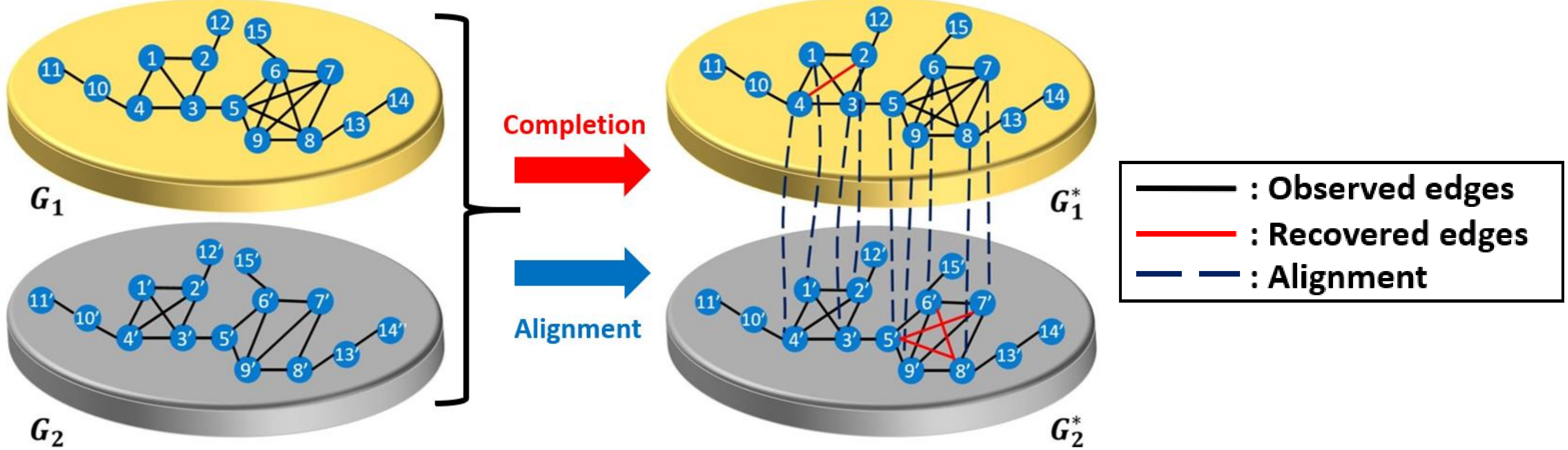
- Limitation 1: alignment efficiency
  - Many alignment methods have $O(mn)$ complexity

    # of edges
  - Densified networks after completion intensify computation
  - **Q1:** how to make alignment and completion more efficient?

- Limitation 2: alignment accuracy
  - Potential introduced noisy links could affect alignment
  - **Q2:** how to complete networks to benefit alignment?

Prob. Def: Incomplete Network Alignment

- **Given:**
  - (1) Adjacency matrices $\boldsymbol{A}_1, \boldsymbol{A}_2$ of two incomplete networks
  - (2) a prior alignment preference matrix $\boldsymbol{H}$
- **Output:**
  - (1) Alignment matrix $\boldsymbol{S}$
  - (2) Adjacency matrices $\boldsymbol{A}_1^*, \boldsymbol{A}_2^*$ of the complete networks



Preliminaries

- Network Alignment (FINAL, IsoRank)

$$\min_{\boldsymbol{s}} \ \alpha \boldsymbol{s}^T(\boldsymbol{D}_1 \otimes \boldsymbol{D}_2 - \boldsymbol{A}_1 \otimes \boldsymbol{A}_2)\boldsymbol{s} + (1-\alpha)\|\boldsymbol{s} - \boldsymbol{h}\|_2^2$$

  – Intuition: Alignment consistency

  – Complexity: $O(mn)$ or $O(n^2)$ with approximation

- Network Completion (Factorization-Based)

$$\min_{\boldsymbol{U}_1,\boldsymbol{V}_1,\boldsymbol{U}_2,\boldsymbol{V}_2} J_1 = \frac{1}{2}\left\|P_{\Omega_1}(\boldsymbol{A}_1 - \boldsymbol{U}_1\boldsymbol{V}_1^T)\right\|_F^2 + \frac{\lambda}{2}(\|\boldsymbol{U}_1\|_F^2 + \|\boldsymbol{V}_1\|_F^2) + \frac{1}{2}\left\|P_{\Omega_2}(\boldsymbol{A}_2 - \boldsymbol{U}_2\boldsymbol{V}_2^T)\right\|_F^2 + \frac{\lambda}{2}(\|\boldsymbol{U}_2\|_F^2 + \|\boldsymbol{V}_2\|_F^2)$$

within-network completion

  – Intuition: low-rank characteristics of real networks

Zhang, Si, and Hanghang Tong. "FINAL: Fast Attributed Network Alignment." *KDD*. 2016.
Singh, Rohit, Jinbo Xu, and Bonnie Berger. "Global alignment of multiple protein interaction networks with application to functional orthology detection." *National Academy of Sciences*, (2008).
Rennie, Jasson DM, and Nathan Srebro. "Fast maximum margin matrix factorization for collaborative prediction." *ICML*. ACM, 2005.

Roadmap

- Motivation and Background ✓
- Q1: Network Completion Helps Alignment
- Q2: Network Alignment Helps Completion
- iNEAT: Optimization Algorithm
- Experiments
- Conclusions

Network Completion Helps Alignment

- Alignment across complete networks

$$\min_{\boldsymbol{s}} \; J_2 = \alpha \boldsymbol{s}^T(\boldsymbol{D} - \boldsymbol{A}_1^* \otimes \boldsymbol{A}_2^*)\boldsymbol{s} + (1-\alpha)\|\boldsymbol{Ds} - \boldsymbol{h}\|_2^2$$

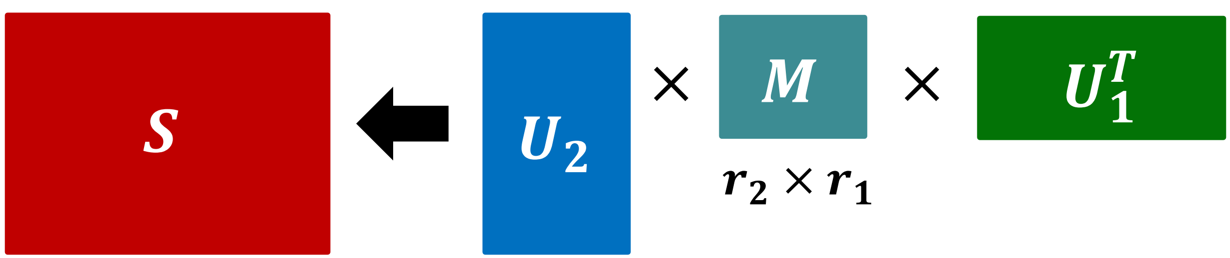
  – $\boldsymbol{A}_1^* = \boldsymbol{U}_1\boldsymbol{V}_1^T$, $\boldsymbol{A}_2^* = \boldsymbol{U}_2\boldsymbol{V}_2^T$ are complete adjacency matrices

  – Benefit: higher-quality of input networks

- Low-rank structure of alignment matrix

  – Low rank of networks ⟶ low rank of alignment matrix

  – $\boldsymbol{S} = \alpha \boldsymbol{U}_2 \boldsymbol{M} \boldsymbol{U}_1^T + (1-\alpha)\boldsymbol{H}$ (proof in paper)



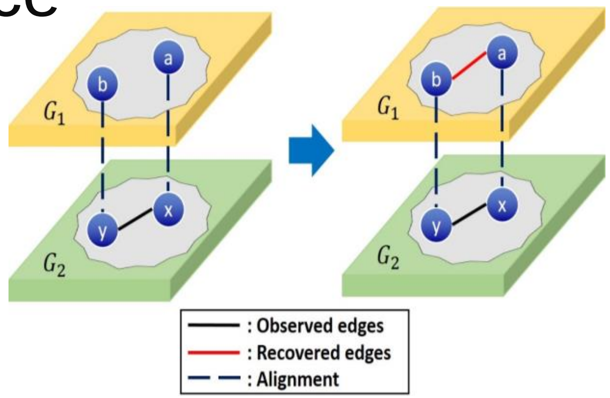
  – Benefit: an alignment algorithm with a ***linear*** complexity

Roadmap

- Motivation and Background ✓
- Q1: Network Completion Helps Alignment ✓
- Q2: Network Alignment Helps Completion
- iNEAT: Optimization Algorithm
- Experiments
- Conclusions

Network Alignment Helps Completion

- Auxiliary confidence of edge existence
  - (1) node-a and node-x are aligned
  - (2) node-b and node-y are aligned
  - (3) node-x and node-y are connected
  - Infer: a potential edge (a, b)



$$\boldsymbol{A}_1^*(a,b) \approx \sum_{x,y}^{n_2} \boldsymbol{S}(a,b)\boldsymbol{S}(b,y)\boldsymbol{A}_2(x,y) = (\boldsymbol{S}^T\boldsymbol{A}_2\boldsymbol{S})(a,b)$$

- Mathematically, we have

$$\min_{\boldsymbol{U}_1,\boldsymbol{V}_1,\boldsymbol{U}_2,\boldsymbol{V}_2,\boldsymbol{M}} \boxed{\begin{aligned} J_3 = &\frac{\beta}{2}\left\|P_{\bar{\Omega}_1}(\boldsymbol{U}_1\boldsymbol{V}_1^T - \boldsymbol{U}_1\boldsymbol{M}^T\boldsymbol{U}_2^T\boldsymbol{A}_2\boldsymbol{U}_2\boldsymbol{M}\boldsymbol{U}_1^T)\right\|_F^2 \\ &+\frac{\beta}{2}\left\|P_{\bar{\Omega}_2}(\boldsymbol{U}_2\boldsymbol{V}_2^T - \boldsymbol{U}_2\boldsymbol{M}\boldsymbol{U}_1^T\boldsymbol{A}_1\boldsymbol{U}_1\boldsymbol{M}^T\boldsymbol{U}_2^T)\right\|_F^2 \end{aligned}}$$

cross-network completion

Roadmap

- Motivation and Background ✓
- Q1: Network Completion Helps Alignment ✓
- Q2: Network Alignment Helps Completion ✓
- iNEAT: Optimization Algorithm
- Experiments
- Conclusions

iNEAT: Optimization Algorithm

$$J_1 = \frac{1}{2}\left\|P_{\Omega_1}(\boldsymbol{A}_1 - \boldsymbol{U}_1\boldsymbol{V}_1^T)\right\|_F^2 + \frac{\lambda}{2}(\|\boldsymbol{U}_1\|_F^2 + \|\boldsymbol{V}_1\|_F^2) + \frac{1}{2}\left\|P_{\Omega_2}(\boldsymbol{A}_2 - \boldsymbol{U}_2\boldsymbol{V}_2^T)\right\|_F^2 + \frac{\lambda}{2}(\|\boldsymbol{U}_2\|_F^2 + \|\boldsymbol{V}_2\|_F^2)$$

within-network completion

$$J_2 = \alpha \boldsymbol{s}^T(\boldsymbol{D} - \boldsymbol{A}_1^* \otimes \boldsymbol{A}_2^*)\boldsymbol{s} + (1-\alpha)\|\boldsymbol{D}\boldsymbol{s} - \boldsymbol{h}\|_2^2$$

Network alignment

$$J_3 = \frac{\beta}{2}\left\|P_{\bar{\Omega}_1}(\boldsymbol{U}_1\boldsymbol{V}_1^T - \boldsymbol{U}_1\boldsymbol{M}^T\boldsymbol{U}_2^T\boldsymbol{A}_2\boldsymbol{U}_2\boldsymbol{M}\boldsymbol{U}_1^T)\right\|_F^2 + \frac{\beta}{2}\left\|P_{\bar{\Omega}_2}(\boldsymbol{U}_2\boldsymbol{V}_2^T - \boldsymbol{U}_2\boldsymbol{M}\boldsymbol{U}_1^T\boldsymbol{A}_1\boldsymbol{U}_1\boldsymbol{M}^T\boldsymbol{U}_2^T)\right\|_F^2$$

cross-network completion

- Overall joint optimization problem

$$\min_{\boldsymbol{U}_1,\boldsymbol{V}_1,\boldsymbol{U}_2,\boldsymbol{V}_2,\boldsymbol{M}} \quad J = J_1 + J_2 + J_3$$
$$s.t. \quad \boldsymbol{U}_1, \boldsymbol{V}_1, \boldsymbol{U}_2, \boldsymbol{V}_2, \boldsymbol{M} \geq \boldsymbol{0}$$

- Optimization Algorithm
  - Block coordinate descent + Multiplicative update
  - Mathematical details in paper
  - Complexity: ***linear*** for both time and space complexity
- Output:
  - (1) Alignment matrix $\boldsymbol{S} = \alpha \boldsymbol{U}_2\boldsymbol{M}\boldsymbol{U}_1^T + (1-\alpha)\boldsymbol{H}$;
  - (2) Complete adjacency matrix $\boldsymbol{A}_1^* = \boldsymbol{U}_1\boldsymbol{V}_1^T, \boldsymbol{A}_2^* = \boldsymbol{U}_2\boldsymbol{V}_2^T$

Roadmap

- Motivation and Background ✓
- Q1: Network Completion Helps Alignment ✓
- Q2: Network Alignment Helps Completion ✓
- iNEAT: Optimization Algorithm ✓
- Experiments
- Conclusions

Experiments

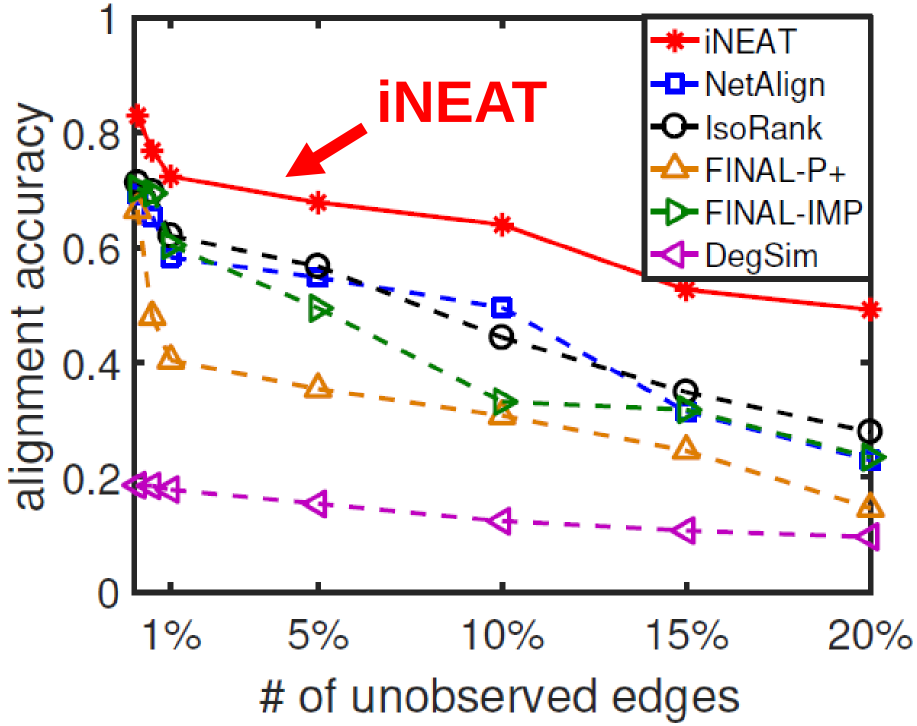
- Datasets

Category	Network	# of Nodes	# of Edges
Collaboration	GrQc	5,241	14,484
Infrastructure	Oregon	7,352	15,665
Social	Google+	23,628	39,194
Social	Youtube	1,134,890	2,987,624

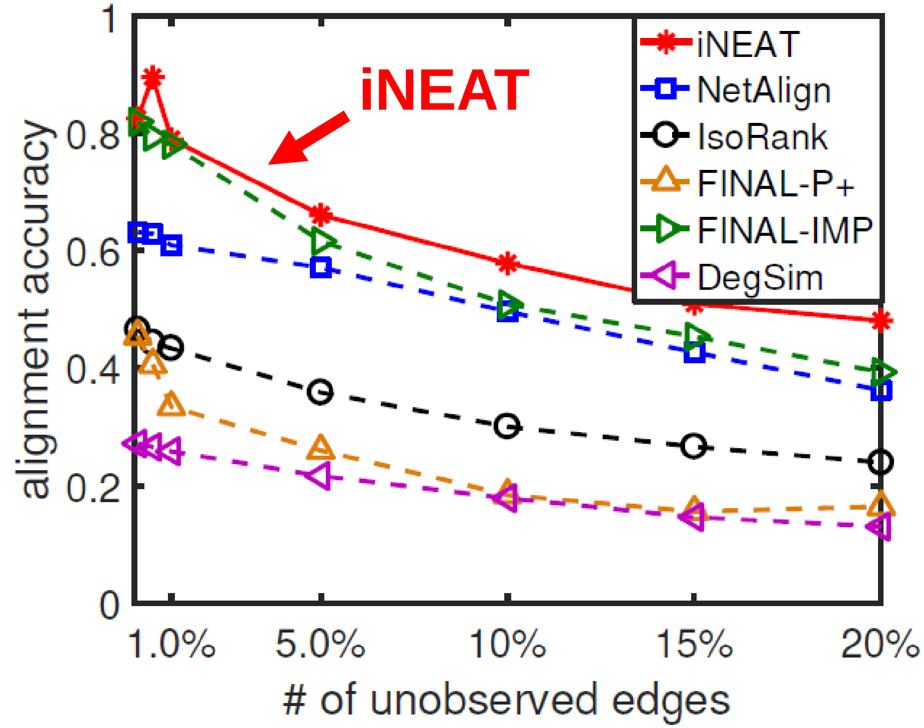
- Evaluation objectives
  - Effectiveness of aligning the incomplete networks
  - Effectiveness of multiple network completion
  - Efficiency and scalability

Effectiveness of network alignment

Oregon infrastructure network



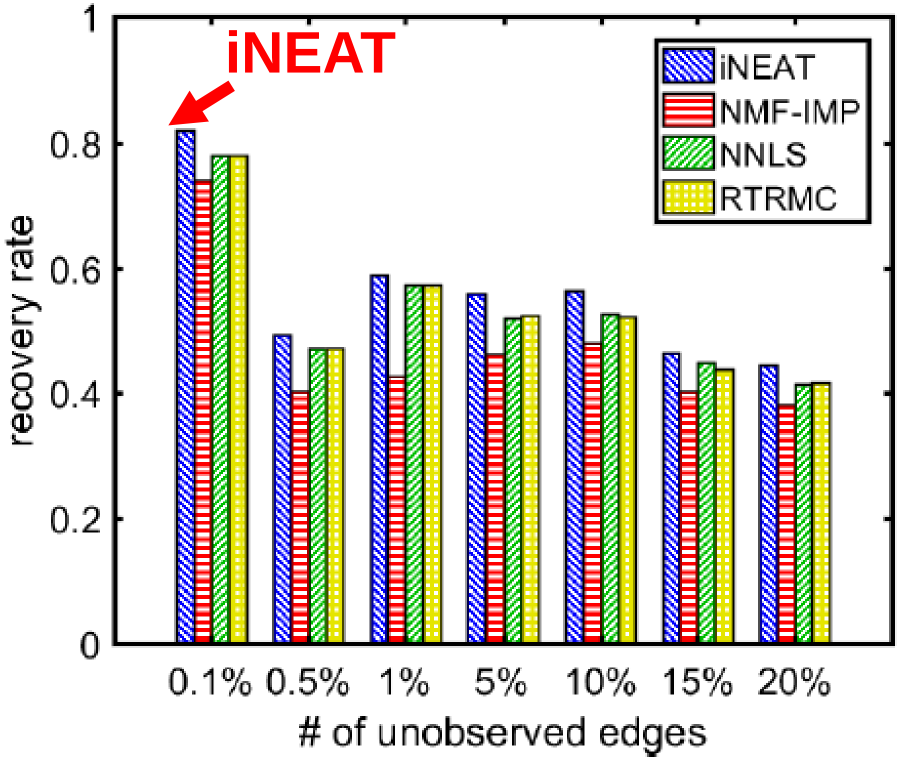
Google+ social network



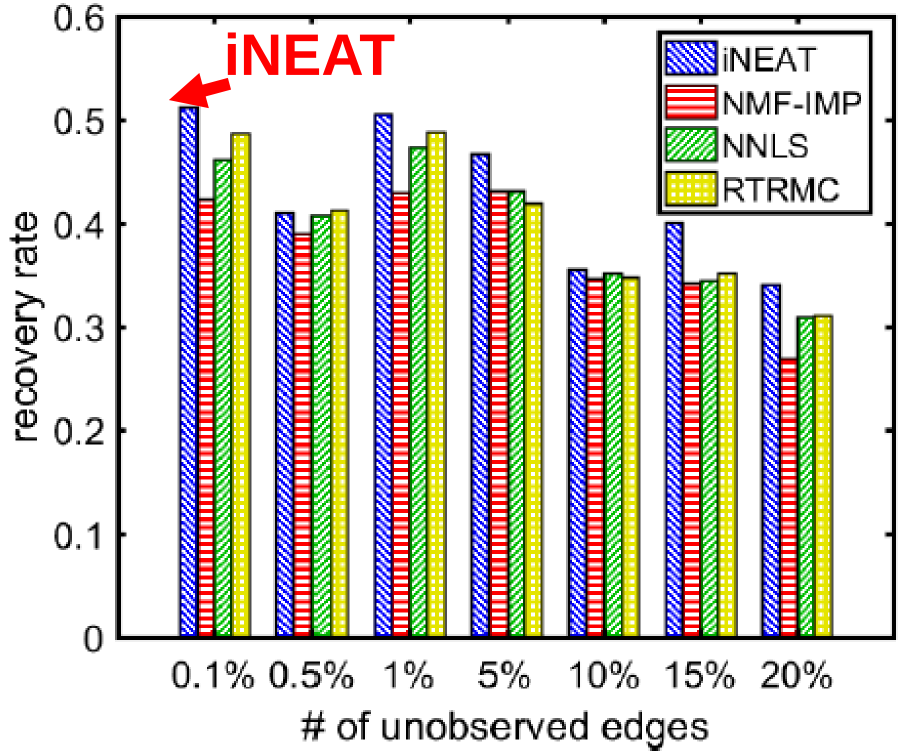
- Observations:
  - (1) iNEAT achieves a better alignment accuracy
  - (2) completion-then-alignment (FINAL-IMP) might be worse

Effectiveness of network completion

GrQc collaboration network



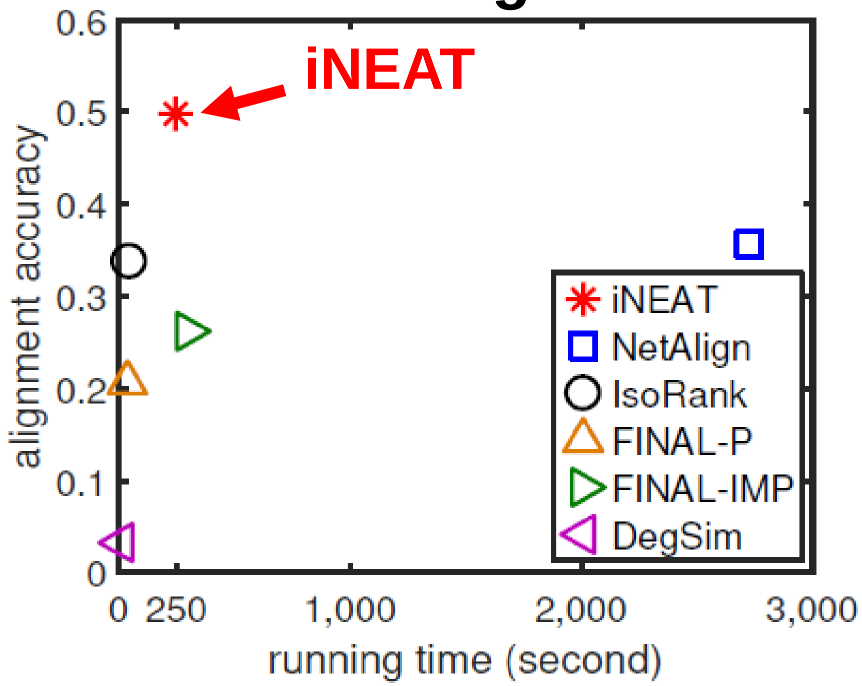
Google+ social network



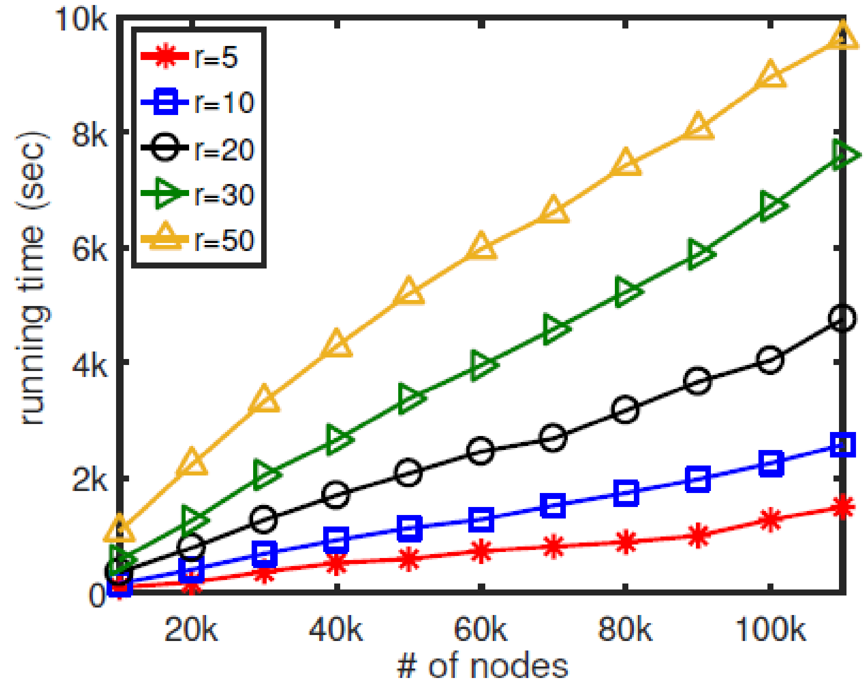
- Observation:
  - Multiple networks completion is benefitted from alignment

Efficiency and Scalability

Balance between accuracy & running time



Scalability



- Observations:
  - (1) iNEAT has a better accuracy-time trade-off;
  - (2) iNEAT has a *__linear__* complexity w.r.t # of nodes

Roadmap

- Motivation and Background ✓
- Q1: Network Completion Helps Alignment ✓
- Q2: Network Alignment Helps Completion ✓
- iNEAT: Optimization Algorithm ✓
- Experiments ✓
- Conclusions

Conclusions

- Incomplete Network Alignment
  - Q1: Alignment efficiency
  - A1: Network completion helps alignment
  - Q2: Alignment accuracy
  - A2: Network alignment helps completion
- Results
  - Better network alignment and completion performance
  - Fast, and ***linear*** complexity
- More in paper
  - Details of the optimization algorithm
  - Proof of low-rank structure of alignment matrix

Thank You!