Growing a Graph Matching from a Handful of Seeds

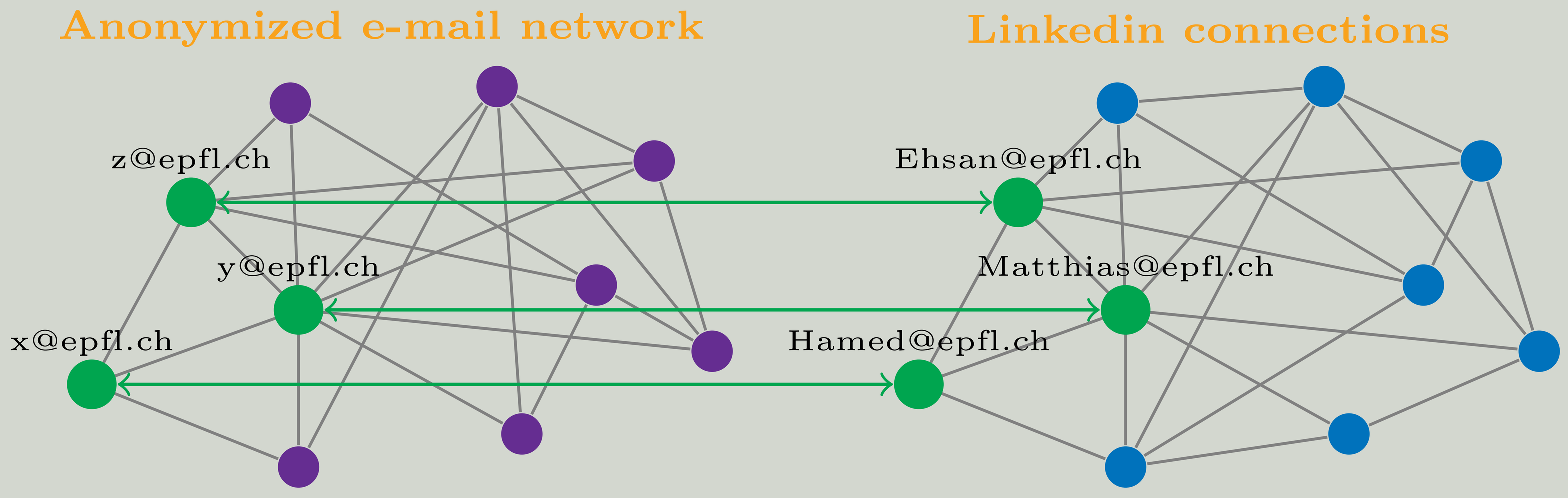
Ehsan Kazemi*, Seyed Hamed Hassani, and Matthias Grossglauser
*ehsan.kazemi@epfl.ch
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

1. Motivation

- Network reconciliation: matching two networks in similar domains
- Bioinformatics: protein–protein–interaction networks alignment
- Image databases: matching graph segments of two scenes

Example: de-anonymization of social networks:

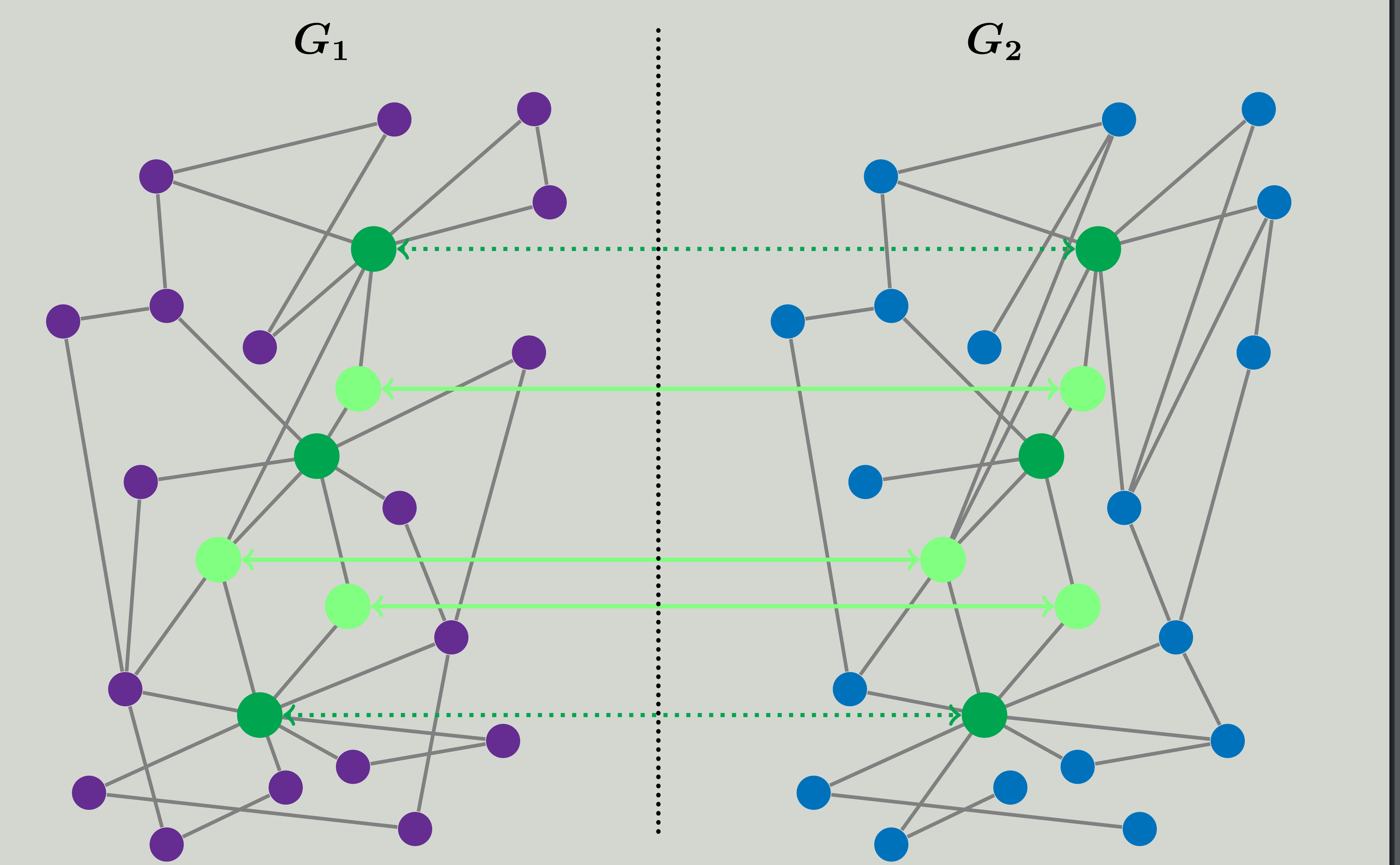
- ▷ Anonymized network = unlabeled graph
- ▷ Side information = noisy labeled version of the same graph



Is it possible to use only the graph structures to establish the true matching between the nodes?

2. Percolation Graph Matching

Percolation graph matching (PGM) algorithms start from an initial **seed set** and iteratively match pairs with at least r **neighbouring** seed-pairs

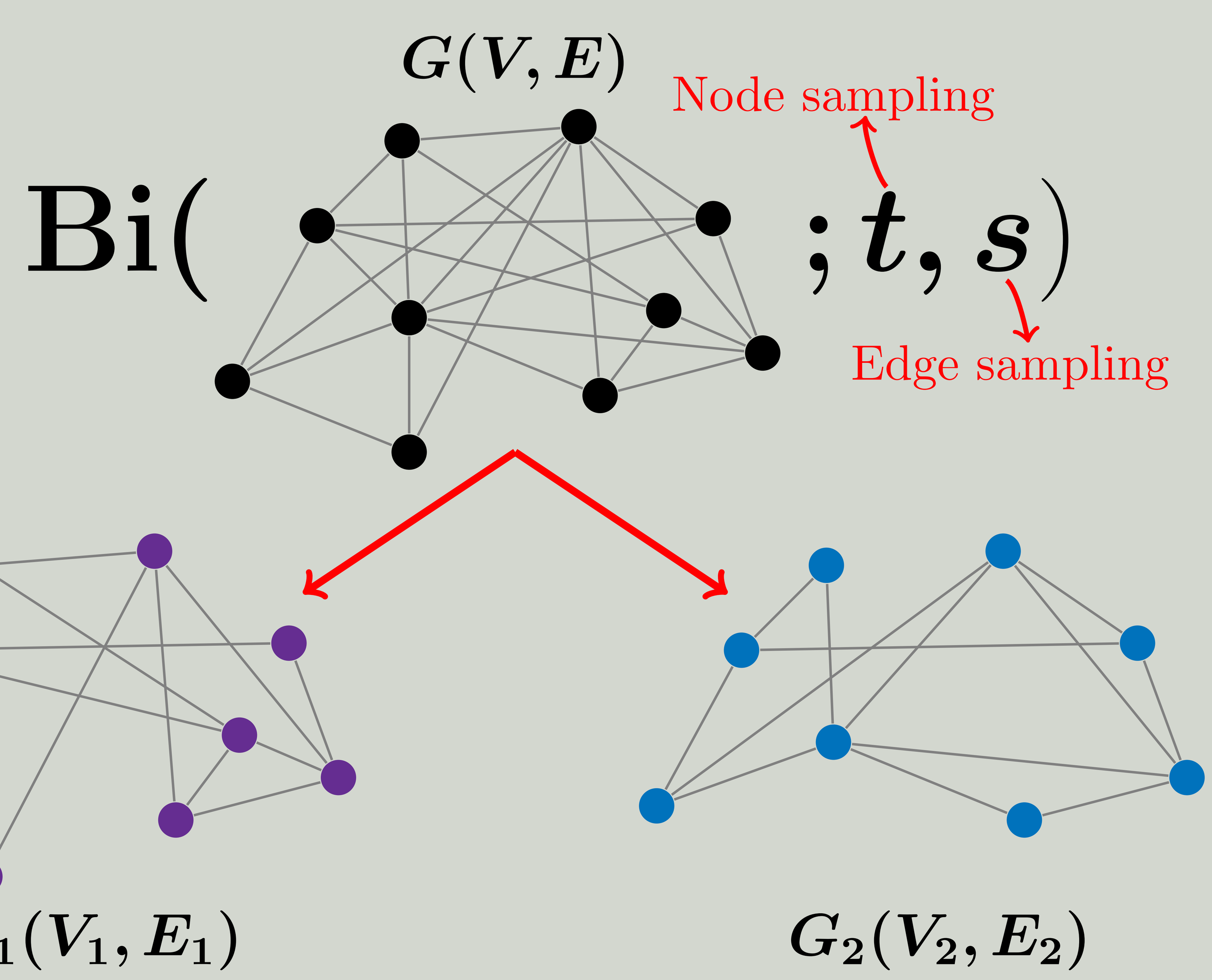


- Dark green nodes are initial seeds
- Light green nodes are the new matched pairs after the first three iterations

Question: Size of the final matching vs. size of the initial seed set?

3. Model and Performance Guarantee

- Using the theory of bootstrap percolation [Janson et al, 2010]
- $\text{Bi}(G;t,s)$: a random bigraph model to generate two correlated graphs with overlapping vertex sets [Pedarsani and Grossglauser, 2011; Kazemi et al., 2015]



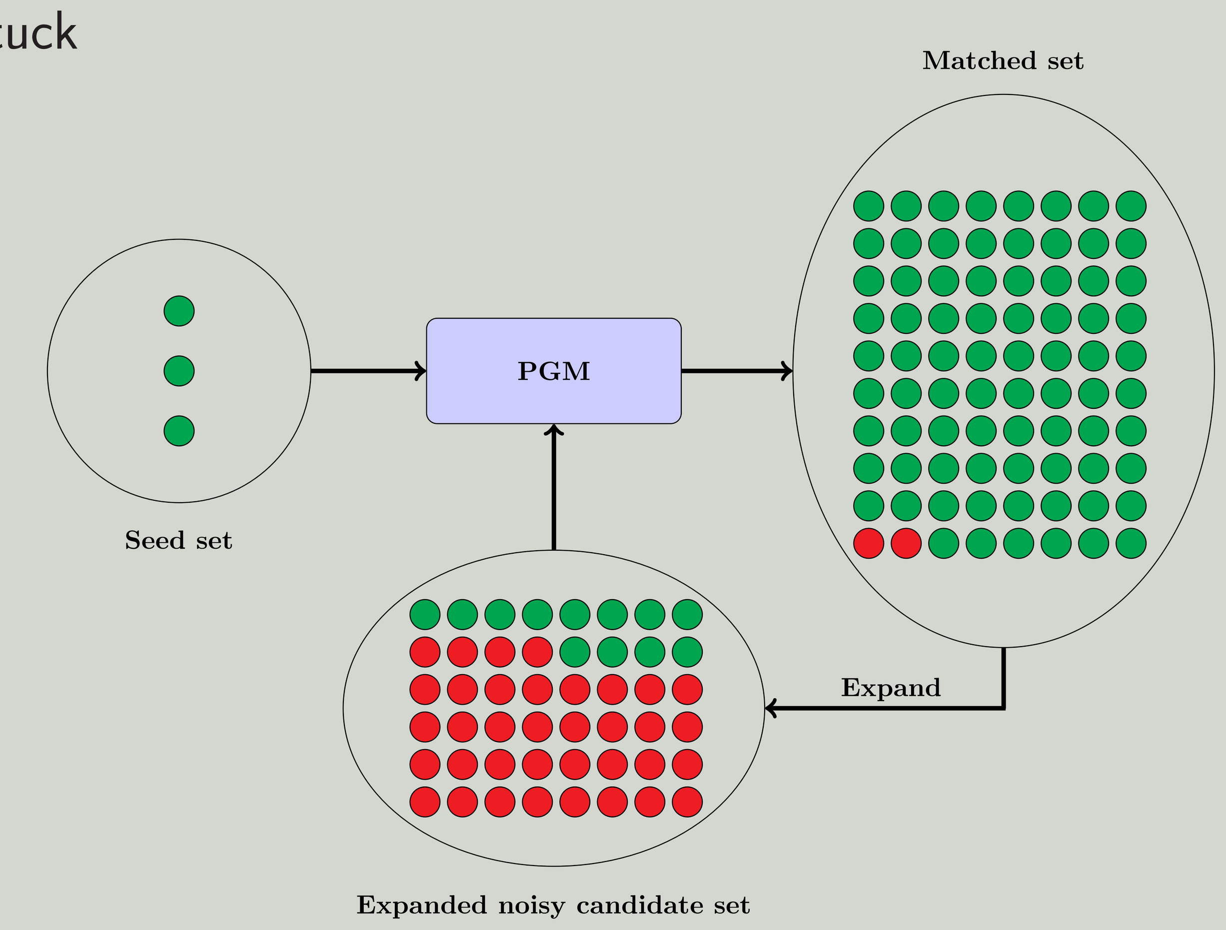
- **Theorem:** For $\text{Bi}(G(n,p);t,s)$ with fixed s and t, there exists a threshold $a_{t,s,r}$ for size of the correct pairs in the initial seed set such that:
  1. if $a/a_{t,s,r} < 1$, the percolation process **dies young**
  2. if $a/a_{t,s,r} \geq 1+\epsilon$, the percolation process matches almost all the nodes **correctly**

4. Aggressive Percolation Graph Matching

- State-of-the-art PGM algorithms needs **many seeds**
- Finding many seeds is **difficult** and **expensive**: how to grow a graph matching with a **handful of seeds?**
- Addition of **many wrong pairs** to the initial seed set have a **negligible** effect on the performance of PGM:
  - ▷ **Expand** the initial seed set to a larger noisy set
  - ▷ In the new noisy seed pairs there are **more correct pairs**

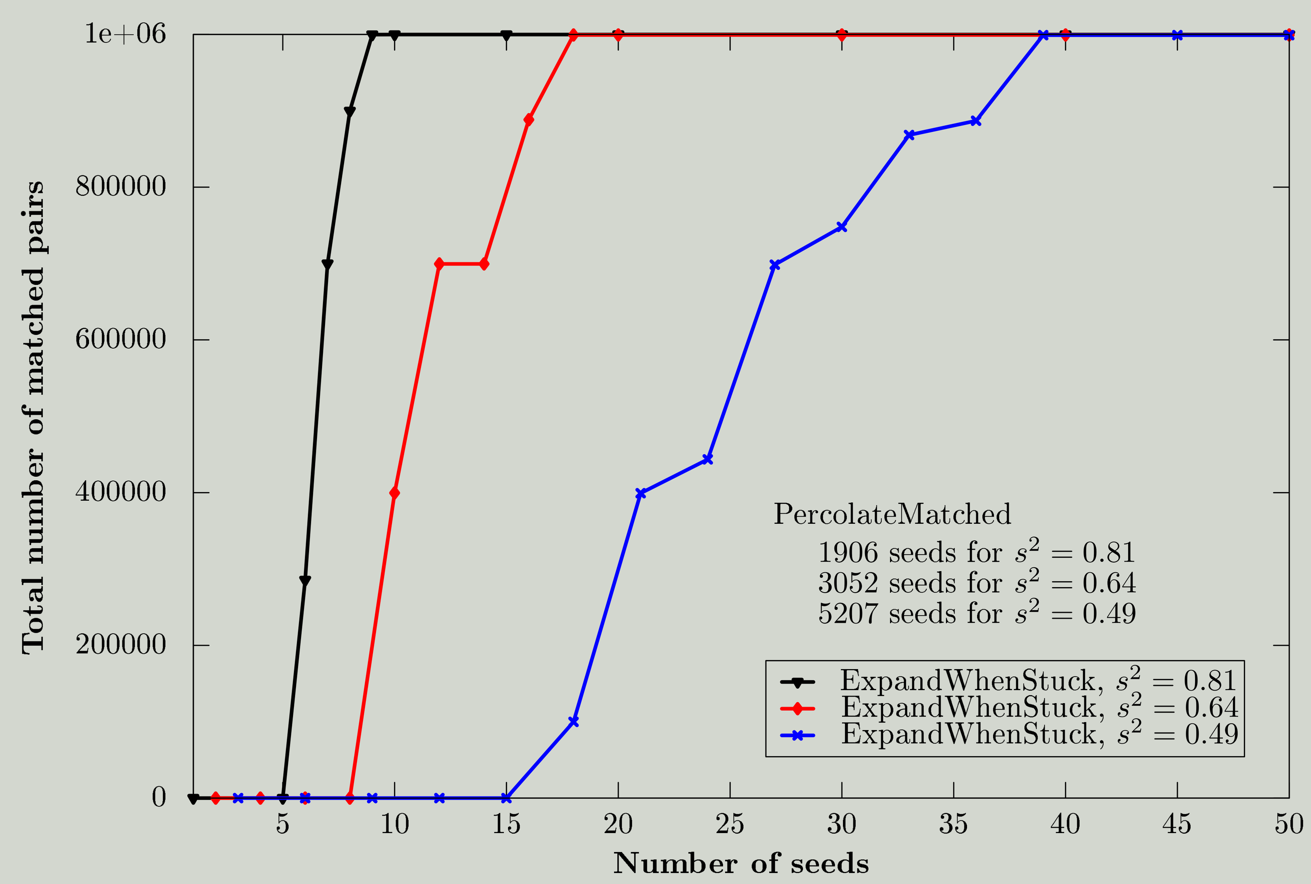
5. ExpandWhenStuck

- Expand the candidate pairs by many noisy pairs whenever the percolation process stuck



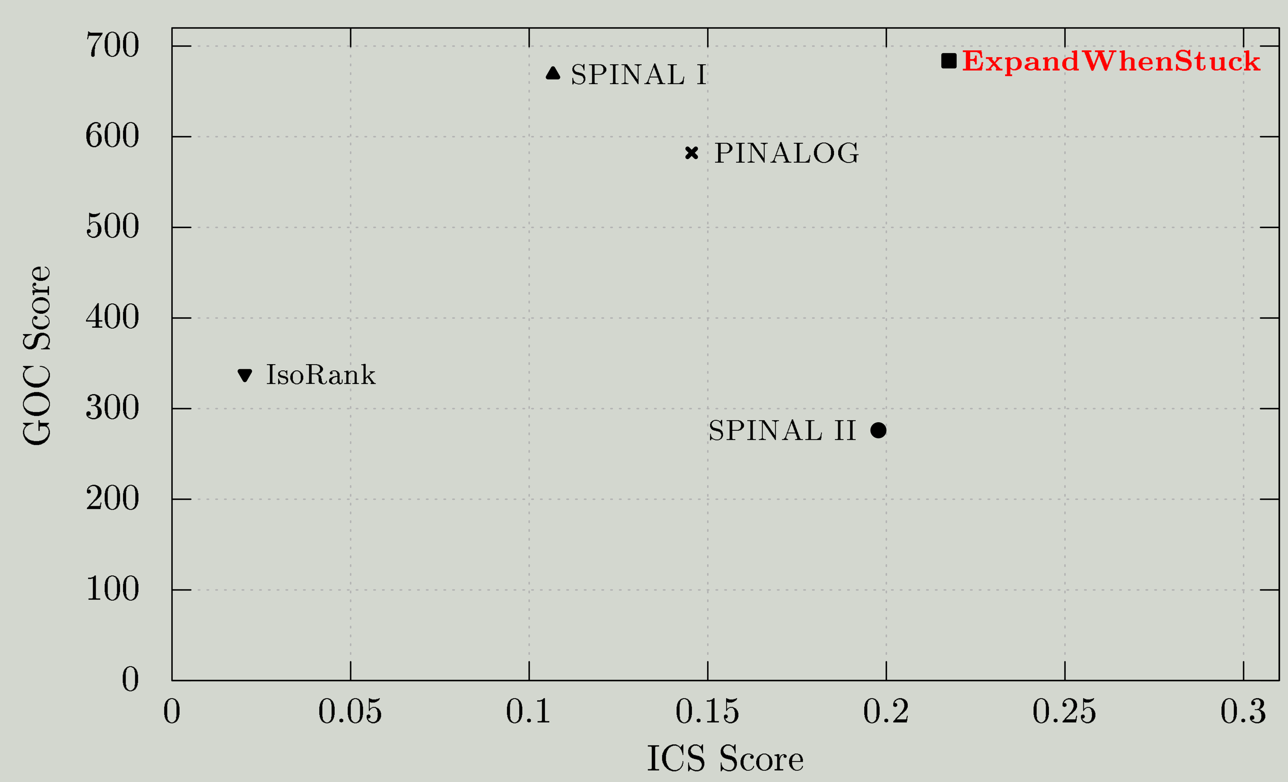
6. Experiments

- **ExpandWhenStuck** vs **PercolateMatched** [Yartseva and Grossglauser, 2013] over $\text{Bi}(G(n,p);t,s)$ with $n = 10^6$, $p = \frac{20}{n}$ and $t^2 = 1.0$



238 times improvement for $s^2 = 0.81$

- **ExpandWhenStuck** vs. state-of-the-art PPI network alignment algorithms



Access: `http://proper.epfl.ch`

7. Take Away Message

- Graph matching is a canonical operation in many fields
- We can have a dramatic reduction in the required size of the seed set with only a small increase in the matching error
- There are sharp phase transitions on the size of the final matching depending on the size of the initial seed set