Learning Erdős-Rényi Random Graphs via Edge Detecting Queries

Zihan Li[1], Matthias Fresacher[2], Jonathan Scarlett[1]

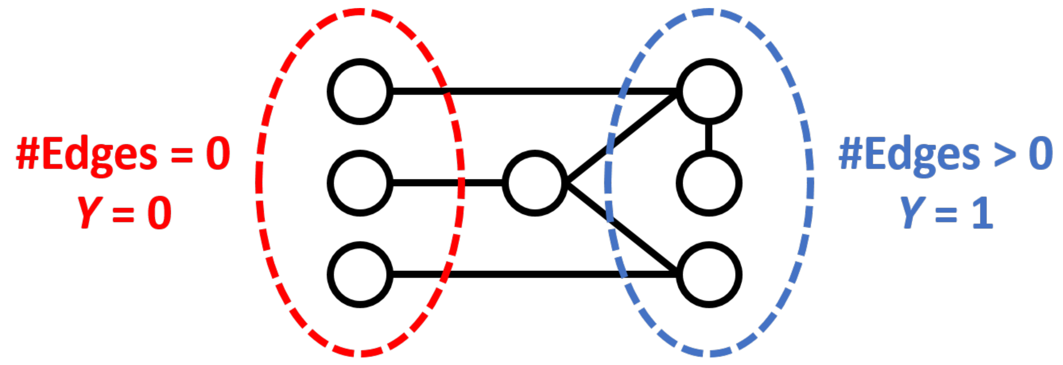
[1]National University of Singapore, [2]University of Adelaide

Problem Setup

- **Goal.** Learn a graph $G = (V, E)$ via *edge detecting queries*:

$$Y = \bigvee_{(i,j)\in E} \{\text{nodes i and j included in the test}\}$$

An illustration:



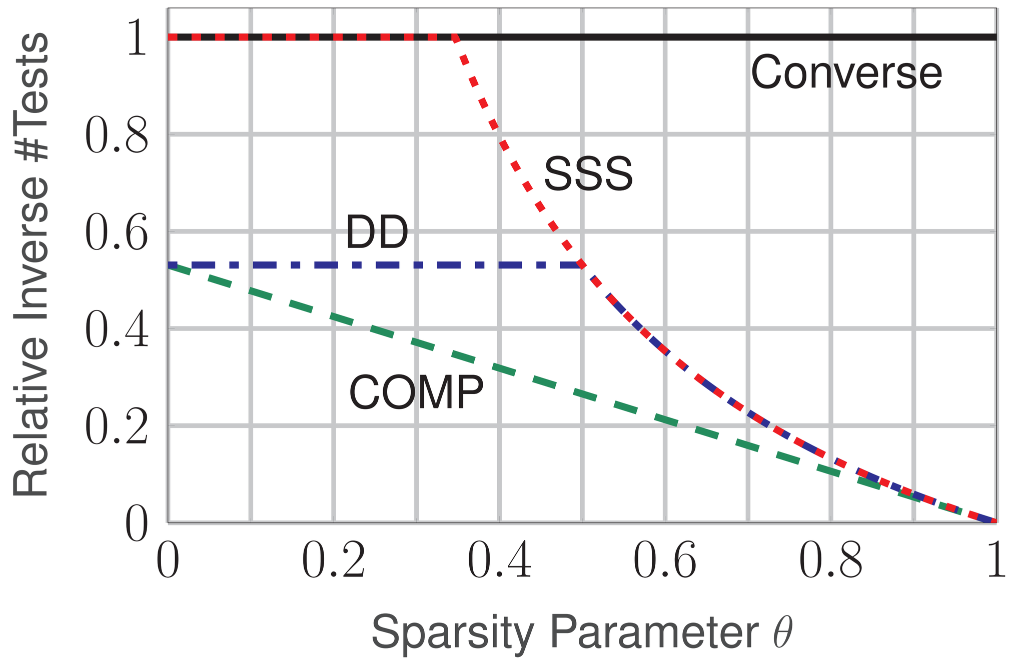
  - Potential applications: Learning chemical interactions, wireless connectivity, social networks with privacy constraints, etc.
- **Erdős-Rényi model:** Each edge appears independently with probability q (in this work, $q \ll 1$)
- **Error probability**: If the estimated graph is $\widehat{G}$, then

$$\Pr[\text{error}] := \mathbb{P}[\widehat{G} \neq G].$$

- **Sparsity level**
  - Let n denote the number of nodes
  - We consider the sparse regime $q = \Theta(n^{-2(1-\theta)})$ with $\theta \in (0, 1)$
  - Average number of edges: $\bar{k} = \binom{n}{2} q = \Theta(n^{2\theta})$
- **Bernoulli random testing:** Each node is independently placed in each test with probability $p = \sqrt{\frac{\nu}{\bar{k}}}$ for some $\nu > 0$

Contributions

- **Existing hardness result [1]**: Learning an arbitrary graph with n nodes and k edges requires $\Omega(\min\{k^2 \log n, n^2\})$ tests
- **This work**: Learning an Erdős-Rényi random graph with an average of $\bar{k}$ edges only requires $O(\bar{k} \log n)$ tests
  - *Explicit near-optimal constant factors*:



Algorithm-Independent Lower Bound

Theorem 1. If the number of tests satisfies

$$\#\text{Tests} \leq \left(\bar{k} \log_2 \frac{1}{q}\right)(1 - o(1))$$

as $n \to \infty$, then it is impossible for any algorithm (and test design) to achieve $\Pr[\text{error}] \to 0$.

Algorithmic Upper Bounds

- **COMP Algorithm**:

(i) Initialize $\widehat{E}$ to contain all $\binom{n}{2}$ edges
(ii) For each negative test, remove all edges from $\widehat{E}$ whose nodes are both included in the test
(iii) Output $\widehat{G} = (V, \widehat{E})$

Theorem 2. The COMP algorithm achieves $\Pr[\text{error}] \to 0$ as $n \to \infty$ as long as

$$\#\text{Tests} \geq \left(2e \cdot \bar{k} \log n\right)(1 + o(1))$$

- **DD Algorithm**:

(i) Initialize $\widehat{E} = \emptyset$, and initialize PE to contain all $\binom{n}{2}$ edges
(ii) For each negative test, remove all edges from PE whose nodes are both included in the test
(iii) For each positive test, if the test covers exactly one edge in PE (the set of *possible edges*), add that edge to $\widehat{E}$
(iv) Output $\widehat{G} = (V, \widehat{E})$

Theorem 3. The DD algorithm achieves $\Pr[\text{error}] \to 0$ as $n \to \infty$ as long as

$$\#\text{Tests} \geq \left(2 \max\{\theta, 1 - \theta\} e \cdot \bar{k} \log n\right)(1 + o(1))$$

SSS Algorithm Lower Bound

SSS Algorithm: Using integer programming, find the graph with the fewest edges that is consistent with the test results

Theorem 4. The (essentially optimal) SSS algorithm yields $\Pr[\text{error}] \to 1$ as $n \to \infty$ as long as

$$\#\text{Tests} \leq \left(2\theta e \cdot \bar{k} \log n\right)(1 - o(1)).$$

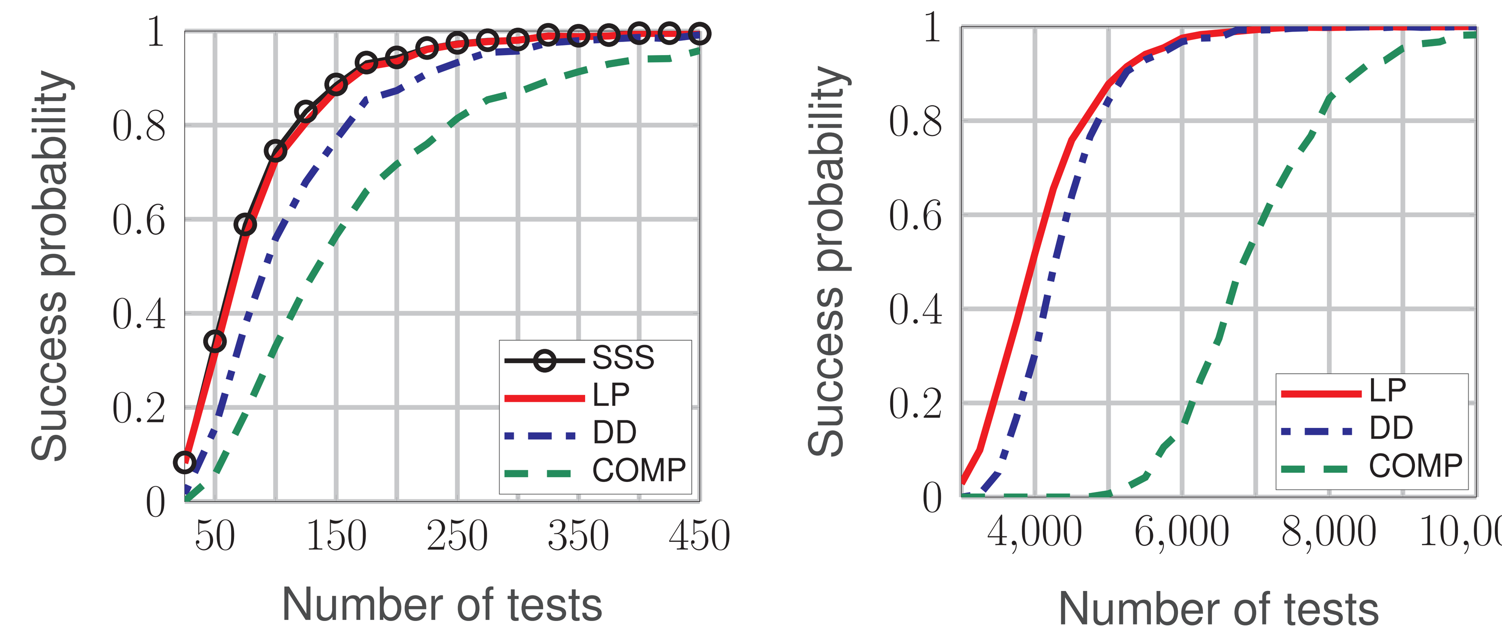
Sublinear-Time Decoding

Group Testing Quick and Efficient (GROTESQUE):

(i) Form $O(\bar{k} \log \bar{k})$ random bundles of nodes
(ii) Use *multiplicity tests* to find bundles with exactly one edge
(iii) Use *location tests* to determine those edges

Theorem 5. The GROTESQUE test design and decoding algorithm achieves $\Pr[\text{error}] \to 0$ with $t = O(\bar{k} \cdot \log \bar{k} \cdot \log^2 n)$ tests, and has $O(\bar{k} \log^2 \bar{k} + \bar{k} \log n)$ decoding time

Numerical Experiments



(i) $n = 50$, $\bar{k} = 5$ and $\nu = 1$

(ii) $n = 200$, $\bar{k} = 200$ and $\nu = 1$

References

[1] H. Abasi and N. H. Bshouty, “On learning graphs with edge-detecting queries,” 2018, https://arxiv.org/abs/1803.10639.

[2] M. Aldridge, L. Baldassini, and O. Johnson, “Group testing algorithms: Bounds and simulations,” *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3671–3687, June 2014.

[3] M. Aldridge, “The capacity of Bernoulli nonadaptive group testing,” *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7142–7148, 2017.

[4] M. Aldridge, O. Johnson, and J. Scarlett, “Group testing: An information theory perspective,” 2019, https://arxiv.org/abs/1902.06002.

[5] S. Cai, M. Jahangoshahi, M. Bakshi, and S. Jaggi, “Efficient algorithms for noisy group testing,” *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2113–2136, 2017.

Full Paper

https://arxiv.org/abs/1905.03410