Shotgun threshold for sparse Erdős–Rényi graphs

Jian Ding, Yiyang Jiang, Heng Ma

Peking University

Shotgun Assembly Problems

Motivited by DNA shotgun sequencing and Reconstructing big neural networks in practice, **Mossel-Ross'19** proposed the following framework for shotgun problems.

- Model: $\mathcal{G}$ is a (fixed or random) graph, (possibly with random labeling of the vertices).
- Observation: For each vertex v, its r-neighborhood $\mathsf{N}_r(v)$ are given, which is the subgraphs induced by the vertices (forgetting their names) at distance no greater than r from v.
- Question: Can we reconstruct $\mathcal{G}$ from these r-neighborhoods profile $\{\mathsf{N}_r(v) : v \in \mathcal{G}\}$?

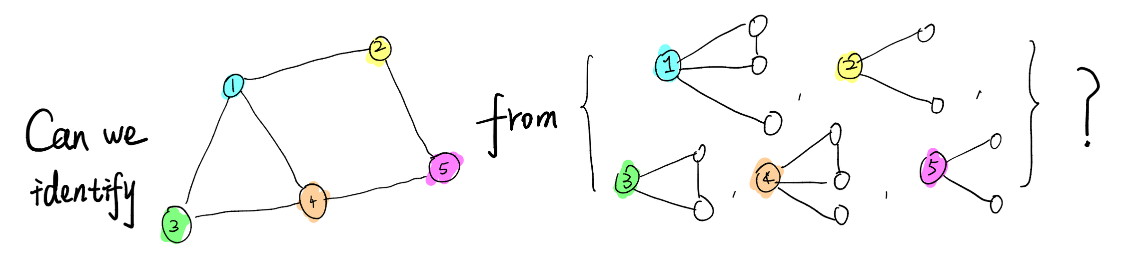


Figure 1: A sepical case of shotgun problem for $r = 1$.

Prior Works

We are mainly interested in the case the graph $\mathcal{G}$ is a Erdős–Rényi graph $\mathcal{G}_{n,p}$.

Mossel-Ross'19: with probability tending to 1,

- for $\lambda \neq 1$, there is a constant C_λ (with explicit formula) such that $\mathcal{G}_{n,\lambda/n}$ is r-identifiable for $r \geq C_\lambda \log n$.
- $\mathcal{G}_{n,\lambda/n}$ is r-nonidentifiable for $r \leq \frac{1}{2(\lambda - \log \lambda)} \log n$.

Remark: The assumption $\lambda \neq 1$ comes from the fact that each connected comopnent of $\mathcal{G}_{n,\frac{\lambda}{n}}$ has diameter less than $C_\lambda \log n$ (**Łuczak'98**, **Riordan–Wormald'10**).

However by **Nachmias-Peres'08**: the diameter is of order $n^{1/3}$ for $\lambda = 1$.

Our Results

Theorem. Take $\lambda \in (0, \infty)$. Let $\mathbf{T}, \mathbf{T}'$ be two independent Poisson(λ) Galton–Waston trees. Define

$$\gamma_\lambda = \mathbf{P}(\mathbf{T} \sim \mathbf{T}')$$

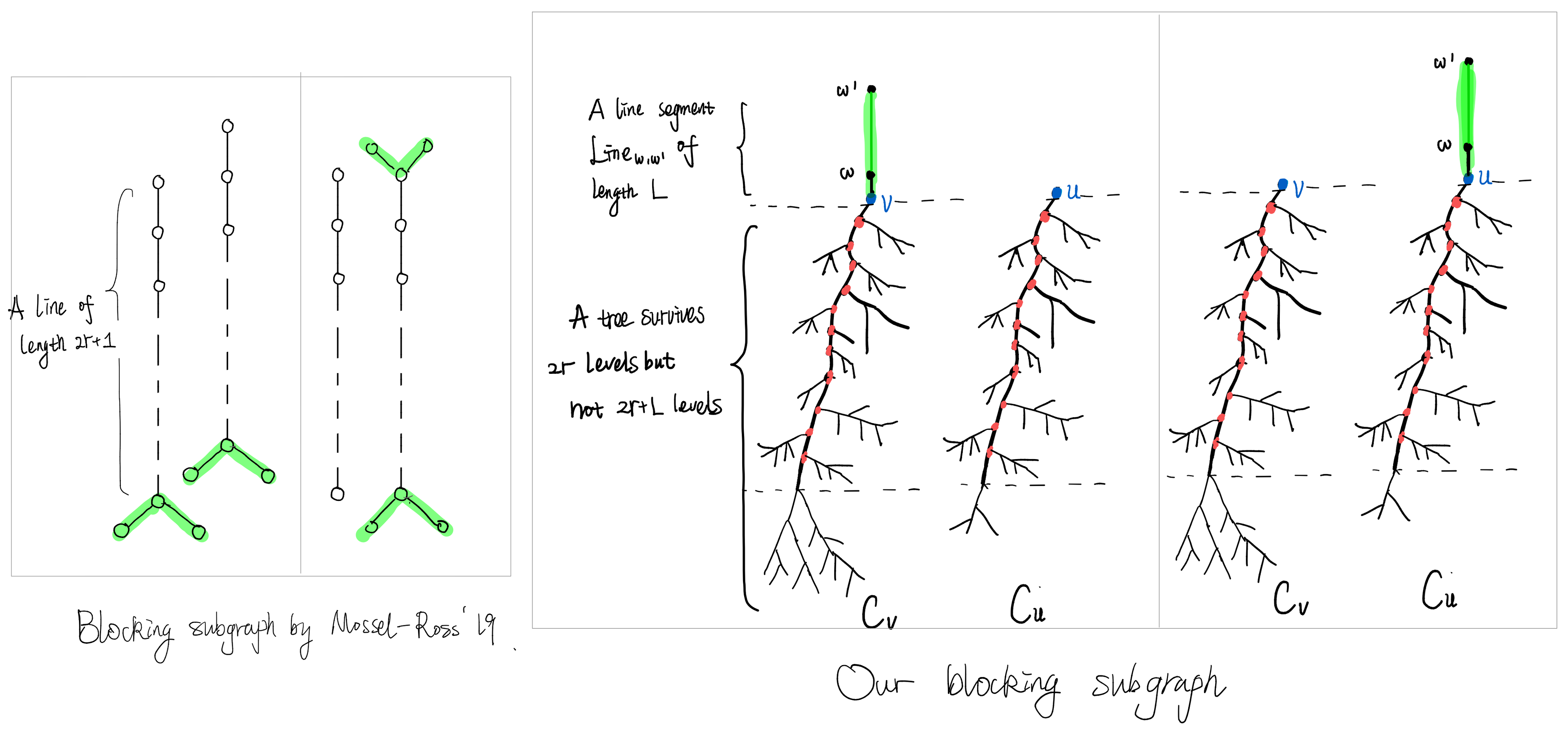
where $\mathbf{T} \sim \mathbf{T}'$ represent that there is an isomorphism from $\mathbf{T}$ onto $\mathbf{T}'$ keeping the root. Consider the Erdős–Rényi graph $\mathcal{G}\left(n, \frac{\lambda}{n}\right)$. Then the following hold for any $\epsilon_0 > 0$,

- for $r \leq (1 - \epsilon_0)\frac{1}{\log(\lambda^2\gamma_\lambda)^{-1}} \log n$, the shotgun problem is non-identifiable w.h.p. as $n \to \infty$;
- for $r \geq (1 + \epsilon_0)\frac{1}{\log(\lambda^2\gamma_\lambda)^{-1}} \log n$, the shotgun problem is identifiable w.h.p. as $n \to \infty$.

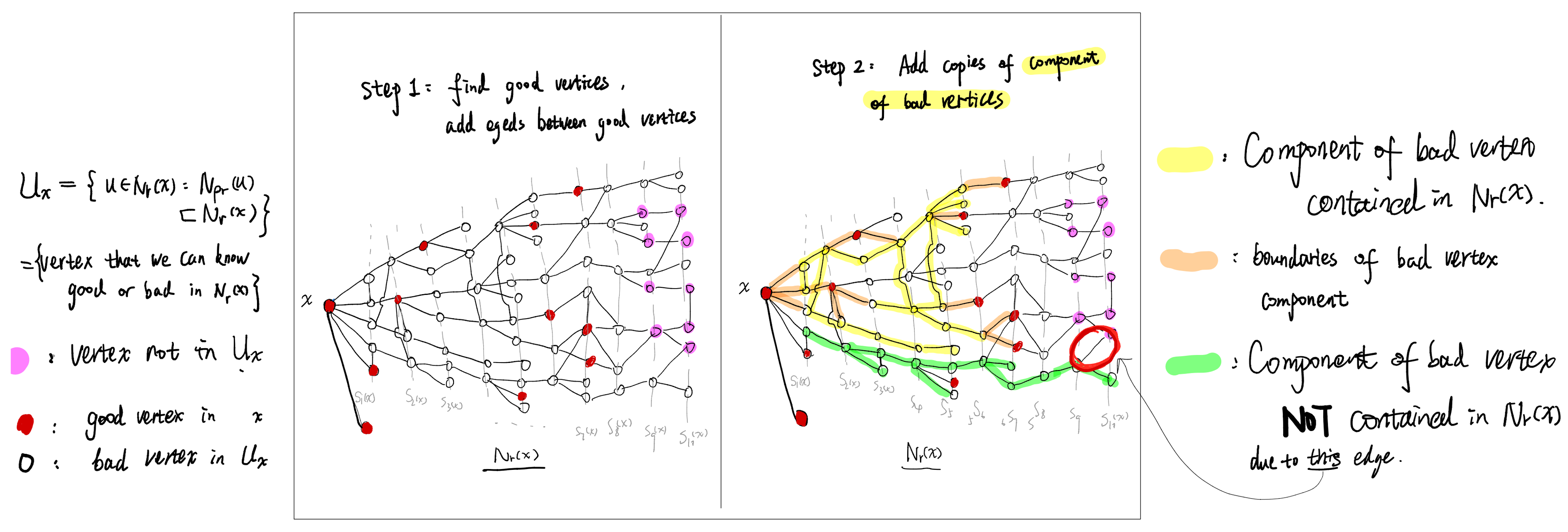
Remark 1. Indeed there is a power series A with non-negative coefficients such that $\lambda^2\gamma_\lambda = A(\lambda e^{-\lambda})$.
Remark 2. We also give an algorithm with polynomial running time for reconstructing the original graph.

Our Approach

- Appearance of the blocking subgraph: there is another graph has the same r-neighborhoods profile as this blocking subgraph, so one can not reconstruct. The expectated number of our blocking subgraph is roughly $n^2 \times \mathbf{P}(\mathbf{T} \sim_{2r} \mathbf{T}')$. We prove that $\mathbf{P}(\mathbf{T} \sim_{2r} \mathbf{T}') \asymp (\lambda^2\gamma_\lambda)^{2r}$. Letting $n^2 \times (\lambda^2\gamma_\lambda)^{2r} \geq 1$, we need $r \leq \frac{1}{\log((\lambda^2\gamma_\lambda)^{-1})} \log n$.

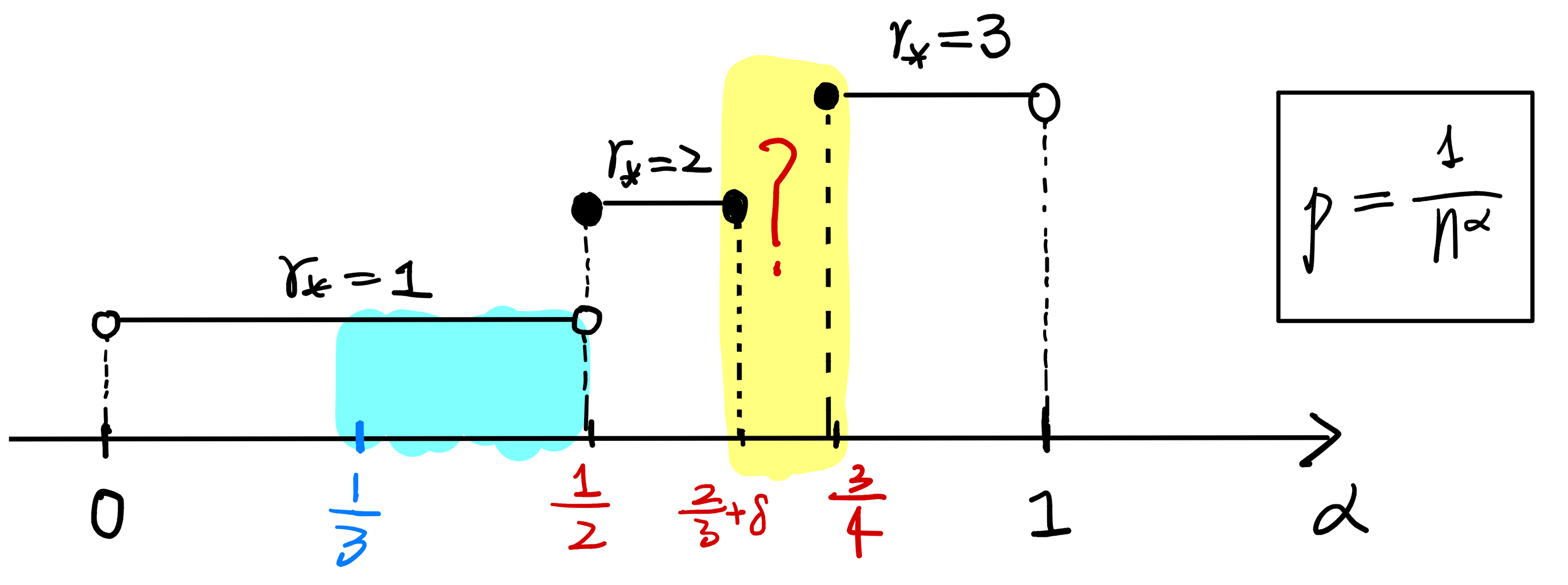


- The reconstructing algorithm for $r \geq (1 + \epsilon_0)\frac{\log n}{\log(\lambda^2\gamma_\lambda)^{-1}}$. We say a vertex is good if its ρr-beighborhood is unique, where $\rho \in (0, 1)$ depending on ϵ_0 is carefully choosen.
  1. For any pair of good vertices, whether there is an edge can be determined by the r-neighborhood for either of them and we then add an edge if there is one.
  2. For each good vertex x with unique ρr-neighborhood and each "bad components" contained in $N_r(x)$, we add a copy of it and each such added copy are disjoint.



Further Questions

Our approach depends heavily on the fact that $\mathcal{G}_{n,\lambda/n}$ looks locally like a Poisson GW tree, and hence can not be applied to dense Erdős–Rényi graphs $\mathcal{G}_{n,n^{-\alpha}}$ with $\alpha \in (0, 1)$. Results in this case were studied by **Mossel-Ross'19**, **Gaudio-Mossel'20**, **Huang-Tikhomirov'21+**, **Johnston-Kronenberg-Roberts-Scott'22+**; and the case $2/3 < \alpha < 3/4$ remains open.



References

[1] Elchanan Mossel and Nathan Ross. Shotgun assembly of labeled graphs. *IEEE Transactions on Network Science and Engineering*, 6(2):145–157, 2019.

[2] Jian Ding, Yiyang Jiang, and Heng Ma. Shotgun threshold for sparse Erdős-Rényi graphs. *IEEE Transactions on Information Theory*, To appear.