

**Open Problem 2.2 (Erdős-Hajnal Conjecture [EH89])** *Prove or disprove the following:*

*For any finite graph  $H$ , there exists a constant  $\delta_H > 0$  such that any graph on  $n$  nodes that does not contain  $H$  as a subgraph (is a  $H$ -free graph) must have*

$$r(G) \gtrsim n^{\delta_H}.$$

It is known that  $r(G) \gtrsim \exp(c_H \sqrt{\log n})$ , for some constant  $c_H > 0$  (see [Chu13] for a survey on this conjecture). Note that this lower bound already shows that  $H$ -free graphs need to have considerably larger  $r(G)$ . This is an amazing local to global effect, where imposing a constraint on small groups of vertices are connected (being  $H$ -free is a local property) creates extremely large cliques or independence sets (much larger than  $\text{polylog}(n)$  as in random Erdős-Renyí graphs).

Since we do not know how to deterministically construct graphs with  $r(G) \leq \text{polylog}n$ , one approach could be to take  $G \sim G(n, \frac{1}{2})$  and check that indeed it has small clique and independence number. However, finding the largest clique on a graph is known to be NP-hard (meaning that there is no polynomial time algorithm to solve it, provided that the widely believed conjecture  $NP \neq P$  holds). That is a worst-case statement and thus it doesn't necessarily mean that it is difficult to find the clique number of random graphs. That being said, the next open problem suggests that this is indeed still difficult.

First let us describe a useful construct. Given  $n$  and  $\omega$ , let us consider a random graph  $G$  that consists of taking a graph drawn from  $G(n, \frac{1}{2})$ , picking  $\omega$  of its nodes (say at random) and adding an edge between every pair of those  $\omega$  nodes, thus "planting" a clique of size  $\omega$ . This will create a clique of size  $\omega$  in  $G$ . If  $\omega > 2 \log_2 n$  this clique is larger than any other clique that was in the graph before planting. This means that, if  $\omega > 2 \log_2 n$ , there is enough information in the graph to find the planted clique. In fact, one can simply look at all subsets of size  $2 \log_2 n + 1$  and check whether it is a clique: if it is a clique then it is very likely these vertices belong to the planted clique. However, checking all such subgraphs takes super-polynomial time  $\sim n^{\mathcal{O}(\log n)}$ . This motivates the natural question of whether this can be done in polynomial time.

Since the degrees of the nodes of a  $G(n, \frac{1}{2})$  have expected value  $\frac{n-1}{2}$  and standard deviation  $\sim \sqrt{n}$ , if  $\omega > c\sqrt{n}$  (for sufficiently large constant  $c$ ) then the degrees of the nodes involved in the planted clique will have larger degrees and it is easy to detect (and find) the planted clique. Remarkably, there is no known method to work for  $\omega$  significant smaller than this. There is a quasi-linear time algorithm [DM13] that finds the largest clique, with high probability, as long as  $\omega \geq \sqrt{\frac{n}{e}} + o(\sqrt{n})$ .<sup>11</sup>

## References

- [EH89] P. Erdos and A. Hajnal. Ramsey-type theorems. *Discrete Applied Mathematics*, 25, 1989.
- [DM13] Y. Deshpande and A. Montanari. Finding hidden cliques of size  $\sqrt{N/e}$  in nearly linear time. *Available online at arXiv:1304.7047 [math.PR]*, 2013.

---

<sup>11</sup>There is an amplification technique that allows one to find the largest clique for  $\omega \approx c\sqrt{n}$  for arbitrarily small  $c$  in polynomial time, where the exponent in the runtime depends on  $c$ . The rough idea is to consider all subsets of a certain finite size and checking whether the planted clique contains them.

MIT OpenCourseWare  
<http://ocw.mit.edu>

18.S096 Topics in Mathematics of Data Science  
Fall 2015

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.