Clique is Hard on Average for Regular Resolution

Albert Atserias — Barcelona
Ilario Bonacina — Barcelona
Susanna F. de Rezende — Stockholm
Massimo Lauria — Rome
Jakob Nordström — Stockholm
Alexander Razborov — Chicago

1 The K-CLIQUE problem…

k-CLIQUE
INPUT: a graph $G = (V, E)$ with n vertices and $k \in \mathbb{N}$
OUTPUT: YES if G contains a k-clique as a subgraph; NO otherwise

- k-CLIQUE can be solved in time $n^{O(k)}$, by checking all $\binom{n}{k}$ subsets of G of size k
- k-CLIQUE is NP-complete and, assuming ETH, there is no $f(k)n^{o(k)}$-time algorithm for k-CLIQUE for any computable function f

GOAL: prove an $f(k)n^{\Omega(k)}$ lower bound for the k-CLIQUE problem.
(This implies $\mathsf{P} \neq \mathsf{NP}$…)

MORE DOABLE GOAL: prove an $f(k)n^{\Omega(k)}$ lower bound for the k-CLIQUE problem **in restricted models of computation** (powerful enough to capture algorithms used in practice).

Main Theorem (informal)
Let $\mathbf{G}$ be an Erdős-Rényi random graph with n vertices (and edge density s.t. $\mathbf{G}$ a.a.s. has no k-clique) and $k = o(n^{1/2})$. A.a.s. every regular resolution proof of the fact that $\mathbf{G}$ does not contain a k-clique has size $\geq n^{\Omega(k)}$

2 …and its Propositional Encoding

The $\mathsf{Clique}(G,k)$ formula is the conjunction of

$$\begin{aligned} &\bigvee_{v\in V} x_{v,i} && \text{for } i \in [k]\ , \text{ and} \\ &\neg x_{u,i} \vee \neg x_{v,j} && \text{for } u \neq v \in V,\ (u,v) \notin E \text{ and } i \neq j \in [k]\ , \text{ and} \\ &\neg x_{u,i} \vee \neg x_{v,i} && \text{for } i \in [k],\ u \neq v \in V\ , \end{aligned}$$

where $x_{v,j} \equiv$ "v is the j-th vertex of a k-clique in G".

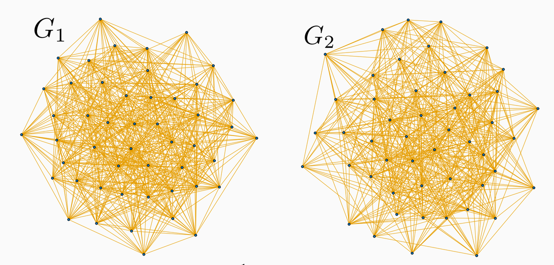
- If G does not contain a k-clique then $\mathsf{Clique}(G,k)$ is unsatisfiable.

4 Erdős–Rényi vs (k-1)-Colorable Graphs

A graph $\mathbf{G} = (V, E)$ sampled from the Erdős-Rényi distribution $\mathcal{G}(n,p)$ is such that $|V| = n$ and each edge $\{u,v\} \in E$ with prob. $p \in [0,1]$

- if $p \ll n^{-2/(k-1)}$ then a.a.s. there are no k-cliques in $\mathbf{G} \sim \mathcal{G}(n,p)$:
  By the fact that expected number of k-cliques in $\mathbf{G}$ is $p^{\binom{k}{2}}\binom{n}{k}$ and Markov's inequality
- $\mathbf{G} \sim \mathcal{G}(n, \frac{1}{2})$ a.a.s. has no clique of size $2\log n$
- there are poly-time algorithms solving k-CLIQUE on graphs $\mathbf{G} \sim \mathcal{G}(n, \frac{1}{2})$ and $k \approx \sqrt{n}$ [AKS98]. (Open problem for $\log n \ll k \ll \sqrt{n}$)

Example



Sampled from $\mathcal{G}(49, \frac{1}{2})$ not 7-colorable — A graph with 49 vertices 7-colorable

Both $\mathsf{Clique}(G_1, 8)$ and $\mathsf{Clique}(G_2, 8)$ are unsatisfiable.
But $\mathsf{Clique}(G_2, 8)$ is way easier to refute for regular resolution.

5 Known Size Upper and Lower Bounds

- whenever G does not contain a k-clique, $\mathsf{Clique}(G,k)$ has resolution refutations of size $\leq n^{O(k)}$
- if G is $(k-1)$-colorable then $\mathsf{Clique}(G,k)$ has regular resolution refutations of size $\leq 2^k k^2 n^2$ [BGL13, this work]
- If G is the complete $(k-1)$-partite graph (or $G \sim \mathcal{G}(n,p)$ for suitable edge density p) then all *tree-like* resolution refutations of $\mathsf{Clique}(G,k)$ have size $\geq n^{\Omega(k)}$ [BGL13]
- for $n^{5/6} \ll k < \frac{n}{3}$ and $\boldsymbol{G} \sim \mathcal{G}(n,p)$ an Erdős-Rényi random graph(and suitable edge density p), all resolution refutations of $\mathsf{Clique}(\boldsymbol{G},k)$ a.a.s. have size $\geq 2^{n^{\Omega(1)}}$ [BIS07]
- if we encode k-CLIQUE using some other, specially tailored, Boolean encoding we get $n^{\Omega(k)}$ size lower bounds for *full* resolution [LPRT17]

MAIN DIFFICULTY: for $k < \sqrt{n}$ all the known techniques from proof complexity fail. Moreover, we need a new combinatorial property distinguishing graphs from $\mathcal{G}(n,p)$ from $(k-1)$-colorable graphs.

3 Regular Resolution………………………………. and Read-Once Branching Programs

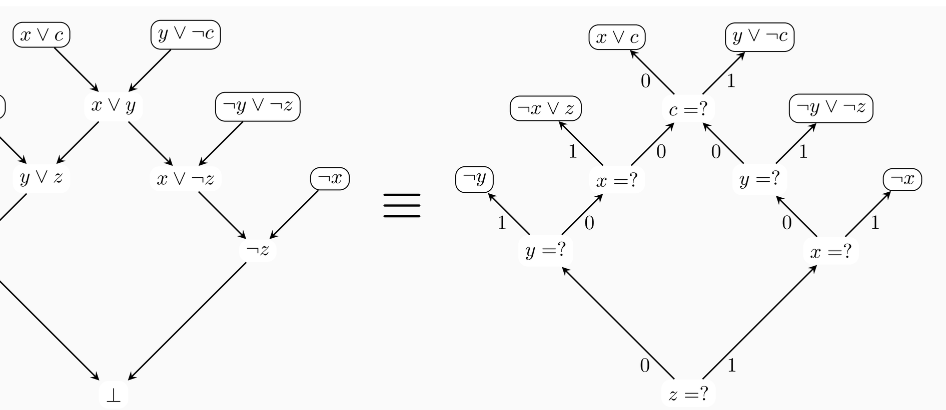
A *regular resolution* refutation of $C_1 \wedge \dots \wedge C_m$ is a labelled DAG where:

- input nodes are labeled with clauses from $\{C_1, \dots, C_m\}$;
- there is a unique sink labelled with $\perp$ (the empty clause);
- each internal node has label $C \vee D$ for some clauses C and D and has two parents, one labelled $C \vee x$ and one labelled $D \vee \neg x$ for some variable x (the variable x is *resolved* in that node). Moreover no variable is resolved twice along any path (*regularity*).

A *ROBP* for (the falsified search problem on) $C_1 \wedge \dots \wedge C_m$ is a labelled DAG with one source where:

- each internal node has label $x =?$ for some variable x and has exactly two outgoing edges labelled 0 and 1.
- along every path no variable is queried twice (i.e. paths define assignments of the variables);
- each sink v is labelled with a clause from $\{C_1, \dots, C_m\}$ falsified by every path from the source to v.

Example



- regular resolution refutations and ROBPs are two equivalent computational models [Kra95]
- The *size* of a resolution refutation (or ROBP) is the number of vertices in the associated DAG
- regular resolution captures the reasoning power of many state-of-the-art algorithms used in practice to solve the k-CLIQUE problem.

6 New Hardness Results

Main Theorem (more formal)

Let $\mathbf{G} \sim \mathcal{G}(n,p)$ be an Erdős-Rényi random graph with p somewhat close to the threshold for containing a k-clique (say for simplicity $p = n^{-4/(k-1)}$) and let $k \leq n^{1/2-\epsilon}$ for some arbitrary small ϵ.

Then, with exponentially high probability (in n), any regular resolution refutation of $\mathsf{Clique}(\mathbf{G},k)$ has size $\geq n^{\Omega(k)}$.

The full statement actually decrease smoothly with the edge density p (see paper)

In particular, for $\mathbf{G} \sim \mathcal{G}(n,1/2)$, regular resolution requires a.a.s. $\geq n^{\Omega(\log n)}$ size to refute $\mathsf{Clique}(\mathbf{G},k)$ for $k = O(\log n)$ and super-polynomial size for $k = o(\log^2 n)$.

7 Proof Ideas

The notation box

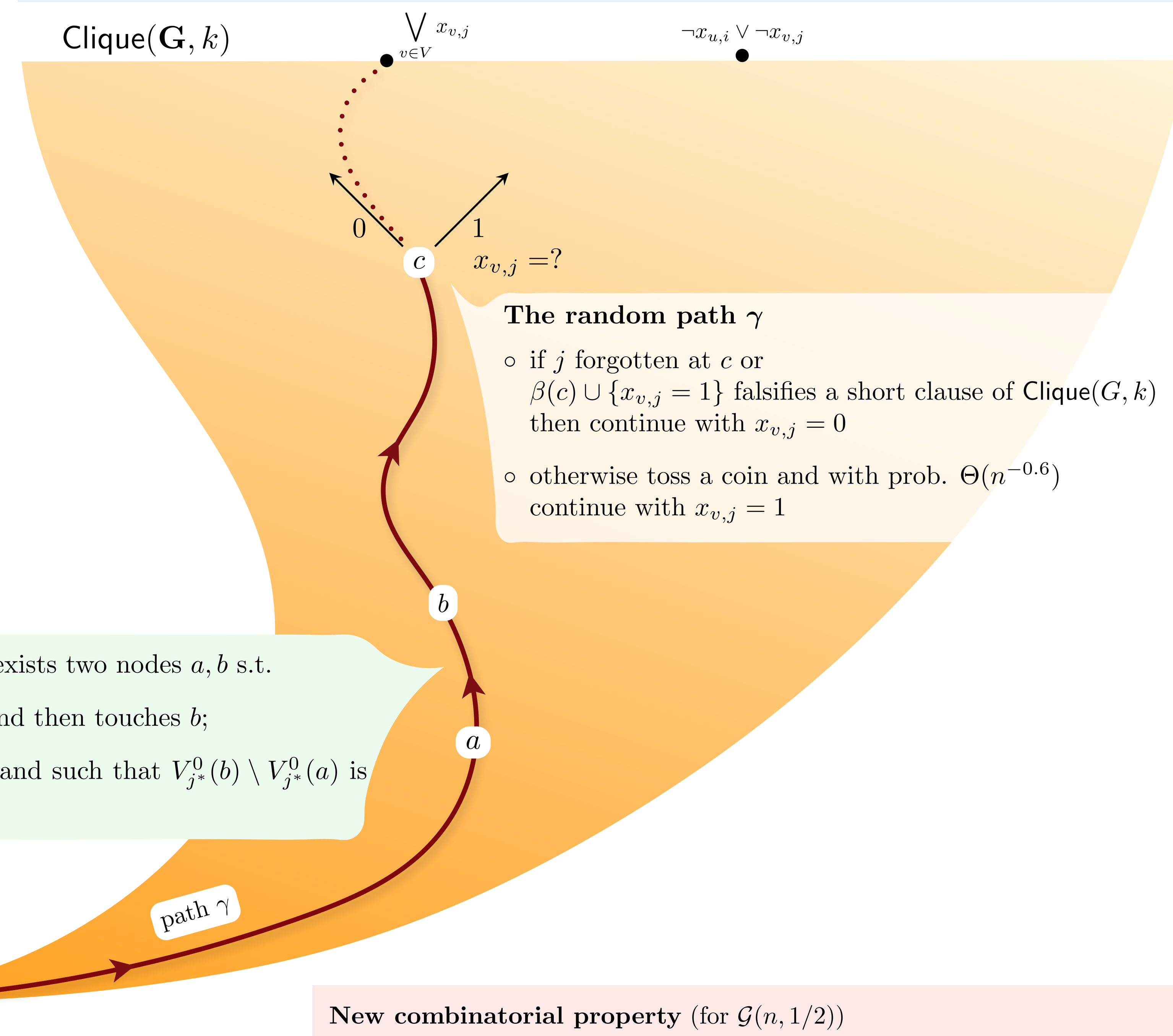
- $\beta(c) =$ max (partial) assignment contained in all paths from the source to c
- j is **forgotten** at c if no sink reachable from c has label $\bigvee_{v\in V} x_{x,j}$
- $\widehat{N}_W(R)$ is the set of **common neighbors** of R in W
- $V_j^0(a) = \{v \in V_j \ :\ \beta(a)(x_{v,j}) = 0\}$
- W is (r,q)**-dense** if for every subset $R \subseteq V$ of size $\leq r$ it holds $|\widehat{N}_W(R)| \geq q$

For simplicity just focus on $\mathbf{G} \sim \mathcal{G}(n, 1/2)$ with $k = 2\log n$.

The ROBP size lower bound is by a non-trivial bottleneck counting.

Claim 1 (easy proof) and **Claim 2** imply that there must be $\geq n^{\Theta(k)}$ nodes in the ROBP.

Q.E.D.



The random path γ

- if j forgotten at c or $\beta(c) \cup \{x_{v,j} = 1\}$ falsifies a short clause of $\mathsf{Clique}(G,k)$ then continue with $x_{v,j} = 0$
- otherwise toss a coin and with prob. $\Theta(n^{-0.6})$ continue with $x_{v,j} = 1$

Claim 1: For every random path γ, there exists two nodes a, b s.t.

1. γ touches a, sets $\leq \lceil \frac{k}{200} \rceil$ variables to 1 and then touches b;
2. there exists a $j^* \in [k]$ not-forgotten at b and such that $V_{j^*}^0(b) \setminus V_{j^*}^0(a)$ is $(\frac{k}{10000}, \Theta(n^{0.9}))$-dense.

Claim 2: For every pair of nodes (a,b) satisfying point (2) of Claim 1,

$$\Pr_\gamma[\gamma \text{ touches } a, \text{ sets} \leq \lceil k/200 \rceil \text{ vars to 1 and then touches } b] \leq n^{-\Theta(k)}$$

New combinatorial property (for $\mathcal{G}(n,1/2)$)
A.a.s. $\mathbf{G} = (V,E) \sim \mathcal{G}(n,1/2)$ satisfies the following:

($\star$) For every $(\frac{k}{10000}, \Theta(n^{0.9}))$-dense $W \subseteq V$ there exists $S \subseteq V$, $|S| \leq \sqrt{n}$ s.t. for every $R \subseteq V$, with $|R| \leq \frac{k}{50}$ and $|\widehat{N}_W(R)| < \tilde{\Theta}(kn^{0.6})$ it holds that $|R \cap S| \geq \frac{k}{10000}$

Proof sketch of Claim 2:

Let $E =$ "γ touches a, sets $\leq \lceil k/200 \rceil$ vars to 1 and then touches b"
and let $W = V_{j^*}^0(b) \setminus V_{j^*}^0(a)$

CASE 1: $V^1(a) = \{v \in V \ :\ \exists I \in [k]\ \beta(a)(x_{v,i}) = 1\}$ has large size ($\geq k/20000$).
Then $\Pr[E] \leq n^{-\Theta(k)}$ because of the prob. of 1s in the random path γ and a Markov chain argument.

CASE 2.1: $V^1(a)$ is not large but many ($\geq \tilde{\Theta}(kn^{0.6})$) vertices in W are set to 0 by coin tosses. So $\Pr[E \wedge W$ has many coin tosses$] \leq n^{-\Theta(k)}$ again by a Markov chain argument as in CASE 1.

CASE 2.2: $V^1(a)$ is not large and not many vertices in W are set to 0 by coin tosses.
Then many of the 1s set by the random path γ between a and b must belong to a set of size at most $\sqrt{n}$, by the new combinatorial property ($\star$).
So $\Pr[E \wedge W$ has not many coin tosses$] \leq n^{-\Theta(k)}$.

STOC paper

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov (1998), Finding a large hidden clique in a random graph. *Random Struct. Alg.*

[BGL13] Olaf Beyersdorff, Nicola Galesi, and Massimo Lauria (2013), Parameterized complexity of DPLL search procedures. *ACM Trans. Comput. Logic*

[BIS07] Paul Beame, Russell Impagliazzo, and Anshish Sabharwal (2007), The resolution complexity of independent sets and vertex covers in random graphs. *Computational Complexity*

[BW01] Eli Ben-Sasson, and Avi Wigderson (2001), Short proofs are narrow—resolution made simple. *J. ACM*

[Kra95] Jan Krajíček (1995), Bounded Arithmetic, Propositional Logic, and Complexity Theory. *Cambridge University Press*

[LPRT17] Massimo Lauria, Pavel Pudlák, Vojtěch Rödl, and Neil Thapen (2017), The complexity of proving that a graph is Ramsey. *Combinatorica*

European Union's Horizon 2020 Research and Innovation Programme / ERC grant agreement no. 648276 AUTAR and Seventh Framework Programme (FP7/2007–2013) / ERC grant agreement no. 279611

Vetenskapsrådet

Swedish Research Council grants 621-2012-5645 and 2016-00782

Poster created by **Ilario Bonacina** (bonacina@cs.upc.edu) for the 50th Annual ACM Symposium on the Theory of Computing. June 25-29, 2018 in Los Angeles, CA