A tight upper bound on the number of variables for average-case k-clique on ordered graphs

Benjamin Rossman

June 14, 2012

Abstract

A first-order sentence φ *defines k-clique in the average-case* if

$$\lim_{n\to\infty} \Pr_{G=G(n,p)} \big[G \models \varphi \Leftrightarrow G \text{ has a } k\text{-clique}\big] = 1$$

where $G = G(n,p)$ is the Erdős-Rényi random graph with $p = p(n)$ being the exact threshold such that $\Pr[G(n,p)$ has a k-clique$] = 1/2$. A question of interest is:

How many variables are required to define average-case k-clique in first-order logic?

Beyond just the usual language of graphs (with only an adjacency relation), we may consider this question for sentences which involve background relations on $\{1,\dots,n\}$ (e.g. the standard linear order). The following have been known:

- With arbitrary background relations, $k/4$ variables are necessary [6].
- With no background relations, $k/2$ variables are necessary and $k/2 + O(1)$ variables are sufficient (Ch. 6 of [7]).
- With arithmetic background relations ($<$, $+$ and $\times$), $k/4 + O(1)$ variables are sufficient (Amano [1]).

In this note, we tie up a loose end—strengthen this last lower bound—by showing that $k/4 + O(1)$ variables suffice with just a linear order in the background.

1 Introduction

The *number of variables* in a first-order formula φ refers to the number of distinct variable symbols (x, y, z, etc.) occurring in φ. This number includes both free and bound variables, and we allow variables to be quantified multiple times. For example, the following 2-variable sentence[1] expresses "the universe has $\geqslant 5$ elements" on the class of linear orders:

$$\exists x \exists y \big(x < y \wedge \exists x \big(y < x \wedge \exists y \big(x < y \wedge \exists x \big(y < x\big)\big)\big)\big).$$

[1] Recall that a *sentence* is a formula with no free variables.

The number of variables is an important measure of the complexity of a first-order formula. Under a well-known descriptive complexity characterization of first-order logic in terms of the complexity class AC^0, every s-variable formula has an equivalent AC^0 circuit of size $O(n^s)$ [3].

A well-studied question in model theory and finite model theory is: over which classes of structures does first-order logic increase in express power with respect to the number of variables? That is, when is the so-called *variable hierarchy* strict? For instance, 2 variables are enough to express every first-order property over the class of finite linear orders, whereas 3 variables are enough over the class of finite words [5]. On the other hand, the variable hierarchy is strict on the class of finite graphs. A longstanding open question was whether the variable hierarchy is strict on the class of finite ordered graphs (see [2]). Answering this question, in [6] we showed that the property "there exists a k-clique" requires $k/4$ variables on the class of finite ordered graphs. This lower bounds is in fact an average-case hardness result: in the first-order language of ordered graphs, $k/4$ variables are required even to express "there exists a k-clique" with high probability on a certain natural distribution (the Erdos-Renyi random graph $G(n,p)$ for $p = p(n)$ an appropriate threshold).

Following [6], Kazayuki Amano [1] gave uniform AC^0 circuits of size $n^{k/4+O(1)}$ which define k-clique in the same average-case sense. Under the descriptive complexity characterization of uniformity, Amano's circuits are equivalent to a sentence in the first-order language of graphs on $\{1,\dots,n\}$ with arithmetic background relations $<$, $+$ and $\times$. In the author's Ph.D. thesis [7], it was noted that the "$k/2$-extension axiom" (famous from the 0-1 law for first-order logic) implies a lower bound of $k/2$ variables for the average-case definability of k-clique in the absence of background relation; together with Joel Spencer, an upper bound of $k/2+O(1)$ was also shown. One question left open from all this work is whether $k/4+O(1)$ or $k/2 + O(1)$ (or something in-between) is the true number of variables required to define k-clique in the average-case with only a linear order in the background. Tying up this loose end, in this paper we show that $k/4 + O(1)$ variables suffice.

2 Preliminaries

Let k be a fixed constant (independent of n). Let $p = p(n) = n^{-2/(k-1)}$, although everything we write holds for any $p(n) = \Theta(n^{-2/(k-1)})$ including the exact threshold for k-clique (see any standard text such as [4] for background on random graphs). Let G be the Erdős-Rényi random graph $G(n,p)$, viewed as a linearly ordered graph. That is, G is random structure with universe $[n] = \{1,\dots,n\}$ and binary relations E and $<$ where $<$ is the standard linear on $[n]$ and E is an anti-reflexive symmetric binary relation such that events $\{(u,v) \in E\}$ occur independently with probability p over pairs (u,v) such that $1 \leqslant u < v \leqslant n$. Throughout this note, "almost surely" means "with probability tending to 1 as $n \to \infty$". For vertices $u,v \in [n]$ such that $u \leqslant v$, we denote by $[u,v]$ the interval of vertices including and between u and v.

In this paper, we prove the following:

Theorem 1. *There is a sentence φ in the first-order language of ordered graphs with only $k/4+O(1)$ variables such that, almost surely, $G \models \varphi$ if and only if G has a k-clique.*

To prove Theorem 1, we first define a property $\mathcal{P}$ of finite ordered graphs (Definition 10) such that $\mathcal{P}$ implies the existence of a k-clique. We then show that $\mathcal{P}$ is first-order definable with $k/4+O(1)$ variables (Lemma 11). Finally, we show that almost surely, if G has a k-clique then G has property $\mathcal{P}$ (Lemma 19).

For simplicity, we treat the case where $k \geqslant 7$ and $k \equiv 3 \bmod 4$. The proof holds with minor modifications when $k \not\equiv 3 \bmod 4$. Let $t = (k-1)/2$ and $s = (k-7)/4$. Note that $s \geqslant 0$ and $t = 2s+3$ are integers and $p = n^{-1/t}$.

3 Proof sketch

Before defining property $\mathcal{P}$ in the next section, we give some basic intuition. We start by showing how to define k-CLIQUE almost surely with $k/2+O(1)$ variables. Suppose that G contains a k-clique $\{v_1,\dots,v_k\}$ (i.e. condition on this event). Then almost surely vertices $v_{t+2},\dots,v_k$ are the only common neighbors of $v_1,\dots,v_{t+1}$. This is seen by the following union bound:

$$\begin{aligned}
&\Pr[v_1,\dots,v_{t+1} \text{ have a common neighbor beside } v_{t+2},\dots,v_k \mid \{v_1,\dots,v_k\} \text{ is a } k\text{-clique in } G]\\
&\quad\leqslant \sum_{w\in[n]\setminus\{v_1,\dots,v_k\}} \Pr[w \text{ is a common neighbor of } v_1,\dots,v_{t+1} \mid \{v_1,\dots,v_k\} \text{ is a } k\text{-clique in } G]\\
&\quad= (n-k)p^{t+1} < p = o(1).
\end{aligned}$$

Denote by $\mathcal{Q}$ the following property: *there exist distinct vertices $x_1,\dots,x_{t+1}$ such that $x_1,\dots,x_{t+1}$ form a clique and have $\geqslant t$ common neighbors and every two common neighbors of $x_1,\dots,x_{t+1}$ are adjacent.* Note that property $\mathcal{Q}$ implies the existence of a k-clique (as $k = 2t+1$). The above inequality also shows that, almost surely, if G has a k-clique then G has property $\mathcal{Q}$; hence $\mathcal{Q}$ is almost surely equivalent to k-CLIQUE with respect to the random graph G.

We claim that $\mathcal{Q}$ is definable with only $t+3 = k/2+O(1)$ variables on the class of finite ordered graphs. (Here the linear order is indispensable: $\mathcal{Q}$ is not definable with fewer than k variables on the class of finite graphs.) The key observation is that saying "$x_1,\dots,x_{t+1}$ have $\geqslant t$ common neighbors" can be achieved with only 2 bound variables in addition to free variables $x_1,\dots,x_{t+1}$: letting $\nu(\vec{x},y) \equiv \bigwedge_{i\in\{1,\dots,t+1\}} \mathrm{Edge}(x_i,y)$, we have

$$\begin{aligned}
&\text{“}x_1,\dots,x_{t+1} \text{ have } \geqslant t \text{ common neighbors”} \equiv\\
&\quad \exists y,\ \nu(\vec{x},y) \wedge \Big(\exists z,\ y<z \wedge \nu(\vec{x},z) \wedge \Big(\exists y,\ z<y \wedge \nu(\vec{x},y) \wedge \Big(\exists z,\ z<y \wedge \nu(\vec{x},z) \wedge \dots\Big)\Big)\Big)
\end{aligned}$$

where there are t existential quantifiers in total. Hence, property $\mathcal{Q}$ can be expressed with

$t+3$ variables as follows:

$$\mathcal{Q} \equiv \exists x_1 \dots \exists x_{t+1},\ \bigwedge_{1 \leqslant i<j \leqslant t+1} \text{Edge}(x_i,x_j)$$
$$\wedge \text{ “}x_1,\dots,x_{t+1} \text{ have} \geqslant t \text{ common neighbors”}$$
$$\wedge\ \forall y \forall z,\ (\nu(\vec{x},y) \wedge \nu(\vec{x},z) \wedge y \neq z) \to \text{Edge}(y,z).$$

Property $\mathcal{P}$ is similar to property $\mathcal{Q}$, except that we must use $k/4+O(1)$ variables to *isolate* the $k/2+O(1)$ vertices $x_1,\dots,x_{t+1}$ that make up the first half of a possible k-clique in the graph G. (As with property $\mathcal{Q}$, once we isolate these $t+1$ vertices, it will be easy to say that they belong to a k-clique using just $O(1)$ additional free variables.) What do we mean by *isolate*? Well, with only $k/4+O(1)$ parameters, there is no hope of defining the set $\{x_1,\dots,x_{t+1}\}$ exactly. But we can define a sequence of *intervals* $I_1,\dots,I_{t+1} \subseteq [n]$ where I_j contains x_j and is not too large; in fact, I_j has size roughly $n^{j/t}$. This sequence will *isolate* $x_1,\dots,x_{t+1}$ in the sense that for all $j \in \{1,\dots,t\}$, x_{j+1} is the unique common neighbor of $x_1,\dots,x_j$ in the interval I_j. This property allows us to efficiently define x_j (with $O(1)$ extra variables) given formulas defining $I_1,\dots,I_{t+1}$. As to defining intervals $I_1,\dots,I_{t+1}$ using just $k/4+O(1)$ variables, this is accomplished by using a single variable for each of $I_1,\dots,I_4$ and a single variable for each pair $(I_5,I_{t+1}),(I_6,I_t),(I_7,I_{t-1}),\dots,(I_{s+4},I_{s+5})$; that is, $s+4=k/4+O(1)$ total variables.

4 Property $\mathcal{P}$

The following definitions refer to a fixed but arbitrary finite ordered graph. Without loss of generality, we assume this graph has vertex set $[n]$ under the standard ordering. For a vertex $v \in [n]$, we denote by $v+1$ and $v-1$ the successor and predecessor of v (when defined).

Definition 2. A sequence $I_1,\dots,I_\ell$ of subsets of $[n]$ is an *ℓ-clique isolator* if $|I_1|=1$ and there exists $(u_1,\dots,u_\ell) \in I_1 \times \dots \times I_\ell$ such that for every $i \in \{2,\dots,\ell\}$, u_i is the unique common neighbor of $u_1,\dots,u_{i-1}$ in the set I_i.

Remark 3. The notion of an ℓ-clique isolator will be useful for the following reason. Suppose $I_1,\dots,I_\ell$ are given by unary relation symbols. Then the statement “$I_1,\dots,I_\ell$ is an ℓ-clique isolator” can be expressed in first-order logic using only 2 variables. To see this, we inductively define formulas $\psi_i(x)$ such that $\psi_i(x)$ is true iff $I_1,\dots,I_i$ is an i-clique isolator and $x=u_i$ (i.e., for the unique i-clique $\{u_1,\dots,u_i\}$ with $(u_1,\dots,u_i) \in I_1 \times \dots \times I_i$). In the base case,

$$\psi_1(x) \equiv I_1(x) \wedge \Big(\forall y,\ y \neq x \to \neg I_1(y)\Big).$$

For $i \in \{2,\dots,\ell\}$, define

$$\psi_i(x) \equiv \theta_i(x) \wedge \Big(\forall y,\ y \neq x \to \neg\theta_i(y)\Big)$$
$$\text{where } \theta_i(x) \equiv I_i(x) \wedge \bigwedge_{j \in \{1,\dots,i-1\}} \Big(\exists y,\ \psi_j(y) \wedge \text{Edge}(x,y)\Big).$$

The statement "$I_1, \ldots, I_\ell$ is an ℓ-clique isolator" is equivalent to the 2-variable formula $\exists x,\ \psi_\ell(x)$. A corollary of this observation is that if each set I_i is definable by an m-variable formula, then the statement "$I_1, \ldots, I_\ell$ is an ℓ-clique isolator" is equivalent to a formula with $m+2$ variables.

Definition 4. A vertex $v \in [n]$ is a *pointer* if $v \geqslant t+1$ and $v, v-1, \ldots, v-t$ (i.e., v and its t predecessors) have a unique common neighbor. If v is a pointer, we denote by v^* the unique common neighbor of $v, v-1, \ldots, v-t$.

Remark 5. The predicate "x is a pointer and $x^* = y$" is definable with 3 variables (i.e., 1 variable in addition to x and y).

$$\text{“}x \text{ is a pointer and } x^* = y\text{”} \ \equiv\ \text{“}x \text{ has} \geqslant t \text{ predecessors”} \wedge \gamma(x,y) \wedge \Big(\forall z,\ z \neq y \to \neg\gamma(x,z)\Big)$$
$$\text{where } \gamma(x,y) \equiv \bigwedge_{j \in \{0,\ldots,t\}} \Big(\exists z,\ \text{“}z = x - j\text{”} \wedge \mathrm{Edge}(y,z)\Big).$$

We leave it as an exercise to show that "x has $\geqslant t$ predecessors" is definable with 1 variable in addition to x and "$z = x - j$" (for fixed j) is definable with 1 variable in addition to x and z.

Remark 6. For any $v \in [n]$ such that $v \geqslant t+1$, the probability that v is a pointer in G is roughly p. Conditioning on v being a pointer, v^* is uniformly distributed in $[n] \setminus \{v, v-1, \ldots, v-t\}$. (These facts come up in the proof of Lemma 14.)

Definition 7. For $j \geqslant 1$ and $v \in [n]$, denote by $f_j(v)$ the minimal $w \in [n]$ such that $w > v$ and w is a common neighbor of $v+1, \ldots, v+j$ (i.e., the j successors of v); in cases where $f_j(v)$ would be undefined (either because $v > n-j$ or because $v+1, \ldots, v+j$ have no common neighbor greater than v), we set $f_j(v) = n$.

Remark 8. For fixed $j \geqslant 1$, the predicate "$f_j(x) = y$" is definable with 3 variables (cf. Remark 5).

Remark 9. For any $j \in \{1, \ldots, t-1\}$ and $v \in [n]$ such that $v < n - n^{1-\varepsilon}$, we expect $f_j(v)$ to be around $v + p^{-j} = v + n^{j/t}$ in the random graph G. Indeed, for any constant $\varepsilon > 0$, it holds almost surely that $v + n^{(j/t)-\varepsilon} < f_j(v) < v + n^{(j/t)+\varepsilon}$. Moreover, this is true even if we condition on arbitrary events in G depending only on edges outside of the interval $[v + n^{(j/t)-\varepsilon}, v + n^{(j/t)+\varepsilon}]$.

Definition 10. A finite ordered graph has property $\mathcal{P}$ if there exist vertices v_1, v_2, v_3, v_4 and $w_1, \ldots, w_s$ such that

(i) $w_1, \ldots, w_s$ are pointers and

(ii) the following sequence of subsets of $[n]$ is a $(t+1)$-clique isolator:

$$\{v_1\}, \{v_2\}, \{v_3\}, \{v_4\}, \underbrace{[w_1, f_3(w_1)], \ldots, [w_s, f_{s+2}(w_s)]}_{[w_i, f_{i+2}(w_i)] \text{ for } i=1,\ldots,s}, \underbrace{[w_s^*, f_{t-s}(w_s^*)], \ldots, [w_1^*, f_{t-1}(w_1^*)]}_{[w_i^*, f_{t-i}(w_i^*)] \text{ for } i=s,\ldots,1},$$

(iii) for the unique $(t+1)$-clique $\{v_1, \ldots, v_{t+1}\}$ isolated by this sequence, $v_1, \ldots, v_{t+1}$ have exactly t common neighbors and these common neighbors form a t-clique.

Lemma 11. *There is a formula with $k/4 + O(1)$ variables that defines property $\mathcal{P}$ on the class of finite ordered graphs.*

Proof. The formula defining $\mathcal{P}$ begins with $\exists v_1, v_2, v_3, v_4, w_1, \ldots, w_s$. Each set $[w_i, f_{i+2}(w_i)]$ and $[w_i^*, f_{t-i}(w_i^*)]$ is definable with $C = O(1)$ variables in addition to parameter w_i (cf. Remarks 5 and 8). Therefore, the statement that

$$\{v_1\}, \{v_2\}, \{v_3\}, \{v_4\}, [w_1, f_3(w_1)], \ldots, [w_s, f_{s+2}(w_s)], [w_s^*, f_{t-s}(w_s^*)], \ldots, [w_1^*, f_{t-1}(w_1^*)]$$

is a $(t+1)$-clique isolator can be expressed with only $C+2$ variables in addition to parameters $v_1, v_2, v_3, v_4, w_1, \ldots, w_s$; moreover, the individual elements $v_1, \ldots, v_{t+1}$ of the unique $(t+1)$-clique isolated by this sequence are definable with the same $s + O(1)$ total variables (cf. Remark 3). Using the order, we can express that $v_1, \ldots, v_{t+1}$ have exactly t common neighbors with only 3 additional variables. To express that these common neighbors form a k-clique, we can say any two common neighbors are adjacent, using just 2 additional variables. So in total we require $s + O(1) = k/4 + O(1)$ variables (in fact, $k/4 + 10$ are sufficient). □

5 Almost surely, G has a k-clique iff G has property $\mathcal{P}$

Let $\varepsilon > 0$ be a sufficiently small constant ($\varepsilon = 1/k$ will do).

Definition 12. A tuple $(u_1, \ldots, u_\ell)$ of vertices in $[n]$ is *well-spaced* if

$$n^{1-\varepsilon} < u_1 < \cdots < u_\ell < n - n^{1-\varepsilon}$$

and $u_{i+1} - u_i > n^{1-\varepsilon}$ for $i \in \{1, \ldots, \ell - 1\}$.

Lemma 13. *Almost surely, if G contains a k-clique, then G contains a well-spaced k-clique.*

Proof. Condition on G containing a k-clique. Sample $\{v_1, \ldots, v_k\}$ uniformly from among the k-cliques of G where $v_1 < \cdots < v_k$. Notice that $(v_1, \ldots, v_k)$ is uniformly distributed among increasing k-tuples in $[n]^k$. The lemma follows from the observation that a uniform random increasing k-tuple in $[n]^k$ is well-spaced with high probability. □

Lemma 14. *Let $u, u' \in [n]$ be a fixed well-spaced pair of vertices and let $i \in \{1, \ldots, s\}$. Almost surely in G, there is a vertex w such that*

$$u - n^{\frac{i+2}{t} - \varepsilon} < w < u < f_{i+2}(w) < u + n^{\frac{i+2}{t} + \varepsilon},$$
$$u' - n^{\frac{t-i-1}{t} + 3\varepsilon} < w^* < u' < f_{t-i}(w^*) < u' + n^{\frac{t-i}{t} + \varepsilon}.$$

Proof. Let $M = \{1, \ldots, \lceil n^{\frac{i+2}{t} - 2\varepsilon} \rceil\}$ and for $m \in M$, let $x_m = u - 2tm$ and denote by Z_m the event that x_m is a pointer and $u' - n^{\frac{t-i-1}{t} + 3\varepsilon} < x_m^* < u'$. Note that events Z_m are mutually independent (using the fact that u, u' are well-spaced). We have

$$\Pr[Z_m] \sim n^{-\frac{i+2}{t} + 3\varepsilon}.$$

This is obtained from the following inequalities:

- $\Pr[Z_m] = \Pr[x_m \text{ is a pointer}] \Pr[u' - n^{\frac{t-i-1}{t} + 3\varepsilon} < x_m^* < u' \mid x_m \text{ is a pointer}]$,
- $\Pr[x_m \text{ is a pointer}] = \Pr[x_m, x_m - 1, \ldots, x_m - t \text{ have a unique common neighbor}]$
  $= \Pr[\geqslant 1 \text{ common neighbor}] - \Pr[\geqslant 2 \text{ common neighbors}]$,
- $\Pr[\geqslant 1 \text{ common neighbor}] = 1 - (1 - p^{t+1})^{n-t-1} \sim 1 - \exp(n^{1+(1/t)})^{n-t-1} \sim n^{-1/t}$,
- $\Pr[\geqslant 2 \text{ common neighbors}] \leqslant \binom{n-t-1}{2} (p^{t+1})^2 < n^{-2/t}$,
- $\Pr[u' - n^{\frac{t-i-1}{t} + 3\varepsilon} < x_m^* < u' \mid x_m \text{ is a pointer}] \sim n^{-\frac{i+1}{t} + 3\varepsilon}$

  since x_m^* is uniformly distributed in $[n] \setminus \{x, x-1, \ldots, x-t\}$ conditioned on x_m being a pointer.

By independence of Z_m's, we have

$$\Pr[\bigwedge_{m \in M} \neg Z_m] = \prod_{m \in M} \Pr[\neg Z_m] \leqslant (1 - n^{-\frac{i+2}{t} + 3\varepsilon} + o(n^{-\frac{i+2}{t} + 3\varepsilon}))^{n^{\frac{i+2}{t} - 2\varepsilon}} \sim \exp(n^{-\varepsilon}).$$

Therefore, almost surely at least one of the events Z_m holds in $G_{\vec{v}}$.

Now observe the following (cf. Remark 9)

$$\Pr\left[x_m + n^{\frac{i+2}{t} - \varepsilon} < f_{i+2}(x_m) < x_m + n^{\frac{i+2}{t} + \varepsilon} \;\middle|\; Z_m\right] = 1 - o(1),$$
$$\Pr\left[x_m^* + n^{\frac{t-i}{t} - \varepsilon} < f_{t-i}(x_m^*) < x_m^* + n^{\frac{t-i}{t} + \varepsilon} \;\middle|\; Z_m\right] = 1 - o(1).$$

It follows that for any $m \in M$ such that Z_m holds in $G_{\vec{v}}$, the vertex x_m is almost surely a suitable witness for w. □

We now fix an arbitrary well-spaced k-tuple of vertices $\vec{v} = (v_1, \ldots, v_k) \in [n]^k$. Denote by $G_{\vec{v}}$ the random graph G conditioned on $\vec{v}$ being a k-clique (that is, $G_{\vec{v}} = G \cup \{k\text{-clique}$ supported on $v_1, \ldots, v_k\})$.

Lemma 15. *The following hold almost surely in $G_{\vec{v}}$.*

1. *For all $j \in \{1, \ldots, t\}$, v_{j+1} is the unique common neighbor of $v_1, \ldots, v_j$ in the interval $[v_{j+1} - n^{\frac{j}{t} - \varepsilon}, v_{j+1} + n^{\frac{j}{t} - \varepsilon}]$. Hence, the sequence*

   $$\{v_1\}, [v_2 - n^{\frac{1}{t} - \varepsilon}, v_2 + n^{\frac{1}{t} - \varepsilon}], [v_3 - n^{\frac{2}{t} - \varepsilon}, v_3 + n^{\frac{2}{t} - \varepsilon}], \ldots, [v_{t+1} - n^{1-\varepsilon}, v_{t+1} + n^{1+\varepsilon}]$$

   is almost surely a $(t+1)$-clique isolator in $G_{\vec{v}}$.

2. $v_{t+2}, \ldots, v_k$ *are the only common neighbors of* $v_1, \ldots, v_{t+1}$.

Proof. Taking union bounds, we have

1. $\Pr\left[\begin{matrix} v_1, \ldots, v_j \text{ have a common neighbor beside} \\ v_{j+1} \text{ in } [v_{j+1} - n^{\frac{j}{t}-\varepsilon}, v_{j+1} + n^{\frac{j}{t}-\varepsilon}] \text{ in } G_{\vec{v}} \end{matrix}\right] \leqslant 2n^{\frac{j}{t}-\varepsilon}p^j = 2n^{-\varepsilon} = o(1),$

2. $\Pr\left[\begin{matrix} v_1, \ldots, v_{t+1} \text{ have a common neighbor} \\ \text{beside } v_{t+2}, \ldots, v_k \text{ in } G_{\vec{v}} \end{matrix}\right] \leqslant (n-k)p^{t+1} < p = o(1).$ □

For the next two lemmas, it will be convenient to relabel the first $t+1$ $(= 2s+4)$ vertices in $\vec{v}$ as follows. Let

$$v_1, \ldots, v_{t+1} = v_1, v_2, v_3, v_4, v'_1, \ldots, v'_s, v''_s, \ldots, v''_1.$$

That is, $v'_i = v_{i+4}$ and $v''_i = v_{t-i+2}$ for $i \in \{1, \ldots, s\}$.

Lemma 16. *Almost surely in* $G_{\vec{v}}$*, there exist vertices* $w_1, \ldots, w_s$ *such that*

$$v'_i - n^{\frac{i+2}{t}-\varepsilon} < w_i < v'_i < f_{i+2}(w_i) < v'_i + n^{\frac{i+2}{t}+\varepsilon},$$
$$v''_i - n^{\frac{t-i-1}{t}+3\varepsilon} < w^*_i < v''_i < f_{t-i}(w^*_i) < v''_i + n^{\frac{t-i}{t}+\varepsilon}.$$

Proof. This is pretty much a corollary of the argument in Lemma 14. Whereas Lemma 14 concerns a single well-separated pair (u, u') in the random graph G, we now consider s well-separated pairs $(v'_1, v''_1), \ldots, (v'_s, v''_s)$ in the random graph $G_{\vec{v}}$. However, we can apply the argument in Lemma 14 independently to each pair (v'_i, v''_i) using the fact that $(v'_1, \ldots, v'_s, v''_s, \ldots, v''_1)$ is well-separated; conditioning on $\{v_1, \ldots, v_k\}$ being a clique does not affect the argument. □

Lemma 17. *Almost surely in* $G_{\vec{v}}$*, there exist vertices* $w_1, \ldots, w_s$ *such that*

- $v'_i \in [w_i, f_{i+2}(w_i)]$ *and* $v''_i \in [w^*_i, f_{t-i}(w^*_i)]$ *for all* $i \in \{1, \ldots, s\}$,
- *the sequence*

  $$\{v_1\}, \{v_2\}, \{v_3\}, \{v_4\}, [w_1, f_3(w_1)], \ldots, [w_s, f_{s+2}(w_s)], [w^*_s, f_{t-s}(w^*_s)], \ldots, [w^*_1, f_{t-1}(w^*_1)]$$

  is a $(t+1)$*-clique isolator (and hence isolates the clique* $\{v_1, \ldots, v_{t+1}\}$*).*

Proof. Condition on the almost sure properties of $G_{\vec{v}}$ given by Lemma 15(1) and 16. For vertices $w_1, \ldots, w_s$ as in Lemma 16, we have $v'_i \in [w_i, f_{i+2}(w_i)]$ and $v''_i \in [w^*_i, f_{t-i}(w^*_i)]$ for all $i \in \{1, \ldots, s\}$. The claim that the sequence

$$\{v_1\}, \{v_2\}, \{v_3\}, \{v_4\}, \underbrace{[w_1, f_3(w_1)]}_{\ni v_5}, \ldots, \underbrace{[w_s, f_{s+2}(w_s)]}_{\ni v_{s+2}}, \underbrace{[w^*_s, f_{t-s}(w^*_s)]}_{\ni v_{s+3} = v_{t-i+2}}, \ldots, \underbrace{[w^*_1, f_{t-1}(w^*_1)]}_{\ni v_{t+1}}$$

is a $(t+1)$-clique isolator follows from the fact that is subsumed by the $(t+1)$-clique isolator

$$\{v_1\}, [v_2 - n^{\frac{1}{t}-\varepsilon}, v_2 + n^{\frac{1}{t}-\varepsilon}], [v_3 - n^{\frac{2}{t}-\varepsilon}, v_3 + n^{\frac{2}{t}-\varepsilon}], \dots, [v_{t+1} - n^{1-\varepsilon}, v_{t+1} + n^{1-\varepsilon}].$$

That is, we have $\{v_{i_0}\} \subseteq [v_{i_0} - n^{\frac{i_0-1}{t}-\varepsilon}, v_{i_0} + n^{\frac{i_0-1}{t}-\varepsilon}]$ trivially for $i_0 \in \{2,3,4\}$, while for $i \in \{1,\dots,s\}$, we have

$$[w_i, f_{i+2}(w_i)] \subseteq [v'_i - n^{\frac{i+2}{t}-\varepsilon}, v'_i + n^{\frac{i+2}{t}+\varepsilon}] \subseteq [v_{i+4} - n^{\frac{i+3}{t}-\varepsilon}, v_{i+4} + n^{\frac{i+3}{t}-\varepsilon}],$$

$$[w^*_i, f_{t-i}(w^*_i)] \subseteq [v''_i - n^{\frac{t-i-1}{t}+3\varepsilon}, v''_i + n^{\frac{t-i}{t}+\varepsilon}] \subseteq [v_{t-i+2} - n^{\frac{t-i+1}{t}-\varepsilon}, v_{t-i+2} + n^{\frac{t-i+1}{t}-\varepsilon}]. \quad \square$$

Lemma 18. *Almost surely, $G_{\vec{v}}$ has property $\mathcal{P}$.*

Proof. Condition on the almost sure properties of $G_{\vec{v}}$ given by Lemmas 15(2) and 17. Vertices v_1, v_2, v_3, v_4 together with $w_1, \dots, w_s$ from Lemma 17 witness property $\mathcal{P}$. Lemma 17 takes care of conditions (i) and (ii) in Definition 10, while Lemma 15(2) takes care of condition (iii). □

Lemma 19. *Almost surely, G contains a k-clique iff G has property $\mathcal{P}$.*

Proof. Property $\mathcal{P}$ implies the existence of a k-clique (with probability 1). The other direction follows from Lemmas 13 and 18. Almost surely, if G contains a k-clique then it contains a well-spaced k-clique. But for any well-spaced k-clique $\vec{v} = (v_1, \dots, v_k)$ that we condition on, $G_{\vec{v}}$ has property $\mathcal{P}$ almost surely. Therefore, the existence of a k-clique in G implies that property $\mathcal{P}$ holds almost surely. □

References

[1] Kazuyuki Amano. k-Subgraph isomorphism on AC^0 circuits. *Computational Complexity*, 19(2):183–210, 2010.

[2] Anuj Dawar. How many first-order variables are needed on finite ordered structures? In *We Will Show Them: Essays in Honour of Dov Gabbay*, pages 489–520, 2005.

[3] Neil Immerman. *Descriptive Complexity*. Graduate Texts in Computer Science. Springer-Verlag, New York, 1999.

[4] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random Graphs*. John Wiley, 2000.

[5] Bruno Poizat. Deux ou trois choses que je sais de L_n. *Journal of Symbolic Logic*, 47(3):641–658, 1982.

[6] Benjamin Rossman. On the constant-depth complexity of k-clique. In *STOC '08: Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 721–730, 2008.

[7] Benjamin Rossman. *Average-Case Complexity of Detecting Cliques*. PhD thesis, MIT, 2010.