Triangle-Freeness is Hard to Detect *

Stasys Jukna[1,2,3] Georg Schnitger[1,3]

February 7, 2002

Abstract

We show that recognizing the K_3-freeness and K_4-freeness of graphs is hard, respectively, for two-player nondeterministic communication protocols using exponentially many partitions and for nondeterministic syntactic read-r times branching programs.

The key ingredient is a generalization of a coloring lemma, due to Papadimitriou and Sipser, which says that for every balanced red-blue coloring of the edges of the complete n-vertex graph there is a set of ϵn^2 triangles, none of which is monochromatic and no triangle can be formed by picking edges from different triangles. We extend this lemma to *exponentially many* colorings and to *partial* colorings.

1 Introduction

Triangle-freeness is a major property of graphs and its communicational as well computational complexity deserves attention. One of the first results in this direction was obtained almost 20 years ago by Papadimitriou and Sipser [11] who proved that

[1]Universität Frankfurt, Institut für Informatik, Robert-Mayer-Str. 11-15, D-60054 Frankfurt, Germany.Email: {georg,jukna}@thi.informatik.uni-frankfurt.de

[2]Institute of Mathematics and Informatics, Akademijos 4, LT-2600 Vilnius, Lithuania

[3]Work partly supported by a DFG grant SCHN 503/2-1.

*The title of the submitted version was: *On Multi-Partition Communication Complexity of Triangle-Freeness* recognizing the triangle-freeness of graphs on n vertices requires $\Omega(n^2)$ bits of communication in nondeterministic best-partition two-party communication games. Since nondeterministically, graphs that contain a triangle can be recognized by communicating only $O(\log n)$ bits (just guess a potential triangle), this first showed that NP$\neq$co$-$NP in the context of best partition communication protocols.

The main step in Papadimitriou–Sipser's proof was a combinatorial lemma about the number of mixed (non-monochromatic) triangles under balanced red-blue colorings of edges of a complete n-vertex graph K_n. Given such a coloring χ, one looks for a large set Δ of triangles such that: (i) each triangle in Δ is *mixed* under χ (not all edges of the same color), and (ii) Δ is *collision-free* (no new triangle can be formed by picking edges from different triangles in Δ).

The first condition (i) is easy to ensure: if r_i is the number of red edges incident to the i-th vertex then $\frac{1}{2}\sum_{i=1}^{n} r_i(n-1-r_i)$ triangles will be mixed under χ. If χ is *strongly balanced* in that the number of red edges is equal to the number of blue edges ± 1 then at least $\Omega(n^3)$ of the triangles will be mixed. More interesting (and important for applications) is condition (ii). An easy argument shows (see, for example, the first paragraph of the next section) that no collision-free set can have more that $\binom{n}{2}$ triangles. What Papadimitriou and Sipser proved is that for any strongly balanced coloring χ of K_n there exists a collision-free set Δ_χ of $\Omega(n^2)$ triangles, all of which are mixed under this coloring.

A natural question is whether a similar result holds for more than one coloring. That is, given a set $\mathcal{C}$ of balanced colorings of K_n, the problem is to find a large collision-free set Δ of triangles, a constant fraction of which is mixed under *each* $\chi \in \mathcal{C}$. For different colorings χ the sets Δ_χ of mixed triangles, guaranteed by Papadimitriou and Sipser, may be (and, actually, are) rather different, and taking just their union might produce a lot of collisions. So, in the case of more than one coloring, the choice of a large collision-free set of triangles is a far more subtle task.

In [8] this task was solved under the additional requirement that all the colorings in $\mathcal{C}$ are balanced on a *fixed* bipartition $K_{n,n}$ of K_{2n}. Although the number of allowed colorings in [8] is $2^{\epsilon n^2}$, the requirement itself is crucial and forbids a lot of colorings: for every bipartition $K_{n,n}$, almost all balanced colorings of K_{2n} are *not* balanced on it.

In this paper we solve this problem in the case of *arbitrary* balanced colorings of K_n, as well as in the case of *partial* colorings; moreover, we allow the colorings to be only weakly balanced (Lemmas 2.1, 2.2 and 2.3). This is the main combinatorial contribution of this paper. Then we present several applications of the extended coloring lemmas: we show that detecting the absence of t-cliques for $t = 3, 4$ is hard for nondeterministic communication protocols with exponentially many partitions and for nondeterministic syntactic read-s times branching programs.

The model of nondeterministic *multi-partition* communication protocols is a strengthening of Papadimitriou–Sipser's model where instead of just one partition of input variables we allow the players to use different partitions for different inputs. The cost of such an extended protocol is the maximum over all inputs of the number of communicated bits *plus* the number of binary bits required to specify a particular partition used for this input. The *multi-partition communication complexity* of a boolean function f is the minimum cost of a multi-partition protocol for f; hence, as in the case of one partition, the communication complexity of any function does not exceed the number of its variables (see Section 3.2 for more precise definitions). As shown in [8], using more partitions may drastically decrease the communication complexity: for every $k = k(N)$ there exist (rather artificial but explicit) boolean functions f in N variables whose multi-partition communication complexity with k partitions drops from $\Omega(N)$ to $O(\log k)$ by taking just one more partition. On the other hand, in this paper we show that for some natural graph-theoretic functions, using many partitions does not help much.

Next, we consider the usual model of *branching programs*. This model captures in a natural way the deterministic space whereas nondeterministic branching programs do the same for the nondeterministic mode of computation (see, e.g., [13] for more information). The model of nondeterministic syntactic read-r times branching programs (r-n.b.p.) was introduced in [6]. These are the usual nondeterministic branching programs with the restriction that along each path (be it consistent or not) each variable can be tested at most r times. (In non-syntactic read-r times branching programs this restriction concerns only *consistent* paths.) The size of such a program is the number of edges in its underlying graph.

Given a set Δ of triangles in K_n, let $\Delta\text{-FREE}_n$ be a boolean function in $\binom{n}{2}$ variables (corresponding to edges of K_n) which, given a subgraph G of K_n, outputs 1 if and only if none of the triangles from Δ is present in G. Let also $K_t\text{-FREE}_n$ be a boolean function which, given a subgraph of K_n, outputs 1 if and only if it contains no clique on t vertices. Our main results are the following.

1. The multi-partition communication complexity of K_3-FREE$_n$ is $\Omega(n^{3/2})$, and is $\Theta(n^2)$ if the number of partitions does not exceed $k = 2^{O(n)}$ (Theorem 3.2).
2. There exists a set Δ of triangles in K_n such that the multi-partition communication complexity of Δ-FREE$_n$ is $\Theta(n^2)$ (Theorem 3.1).
3. If $r = o(\log n/\log\log n)$, then the 4-clique-freeness function K_4-FREE$_n$ requires r-n.b.p. of size exponential[1] in $(n/r)^2$ (Theorem 3.3).

The first result extends the lower bound of Papadimitriou and Sipser to the case of exponentially many partitions. The second gives a *truly linear* (in the number of variables) lower bound on the nondeterministic multi-partition communication complexity. The third result also gives the first *truly exponential* (in the number of variables) lower bounds for nondeterministic syntactic read-r times branching programs computing a natural combinatorial function. In the case of *deterministic* read-once branching programs ($r = 1$) such a (truly exponential) lower bound was earlier obtained in [2] for the $\oplus$ CLIQUE$_{n,3}$ function which, given a graph G, outputs the parity of triangles in G. In the case of *nondeterministic* read-once branching programs such lower bounds for $\oplus$ CLIQUE$_{n,3}$, as well as for K_3-FREE$_n$, were proved in [8]. In the case of larger values of r, the only known *truly exponential* lower bounds were obtained in [6, 4, 1, 5] for boolean functions based on some special quadratic forms; the proofs employ non-trivial probabilistic and algebraic arguments. Our method for the K_4-FREE$_n$ function is different, and requires only simple probabilistic reasoning. Of course, being syntactic is a severe restriction on the computational power of r-n.b.p. On the other hand, this model is *nondeterministic* and, so far, no lower bounds are known for non-syntactic r-n.b.p. even for constant $r \geq 2$. Recent lower bounds for the non-syntactic model, proved by Ajtai [1], hold only for *deterministic* branching programs. As shown in [5], Ajtai's method can be extended to yield lower bounds also for *randomized* branching programs if the error probability is small enough. But the case of *nondeterministic* non-syntactic r-n.b.p.'s remains open, even for $r = 2$.

The paper is organized as follows. In Section 2 we state our main combinatorial results—the coloring lemmas for triangles and 4-cliques. In Section 3 we use them to prove lower bounds for multi-partition communication complexity and for nondeterministic read-r times branching programs of the corresponding boolean functions.

[1] A function $f(x)$ is *exponential* in $g(x)$ if $\log f(x) = \Omega(g(x))$.

The rest is devoted to the proof of the coloring lemmas. We conclude the paper with several remarks and open problems.

2 Coloring lemmas

Given a (partial) red-blue coloring of some set of points, we will say that it is λ-*balanced* if at least a λ-fraction of points are colored red and at least λ-fraction of points are colored blue. If not stated otherwise, the balance parameter $\lambda = \lambda(n)$ may be an arbitrary function such that $0 < \lambda(n) \leq 1/2$. A coloring is *balanced* if it is γ-*balanced* for some (arbitrary small, but fixed through the paper) constant $0 < \gamma \leq 1/2$. We will assume that the number n of vertices in the considered graphs is sufficiently large.

A triangle is just a set $T = \{u, v, w\}$ of three mutually adjacent vertices. A set Δ of triangles is *collision-free* if no triangle outside Δ can be formed by taking edges from three triangles in Δ. The reason why the collision-freeness property is important is roughly as follows. Distinguish one edge in each of the triangles from Δ, and construct a set $\mathcal{G}$ of graphs by taking from each triangle its distinguished edge and *precisely one* of the remaining two edges. Since Δ is collision-free, we obtain $|\mathcal{G}| = 2^{|\Delta|}$ graphs, none of which contains a triangle, but the union of any two of them already has a triangle. This, in particular, implies that no collision-free set can have more that $\binom{n}{2}$ triangles. On the other hand, it is easy to construct a collision-free set Δ of $\Omega(n^2)$ triangles in K_n by taking a matching on $n/2$ vertices, and joining the endpoints of its edges with all the remaining vertices. However, we need the triangles in Δ to be mixed under given colorings of K_n, which requires extra efforts.

For the ease of counting, it will be convenient to specify a triangle $T = \{u, w, v\}$ by a pair (e, v) where $e = uw$ is the *fixed edge* and v the *top vertex* of the triangle; the two edges uv and wv joining v with the endpoints of e are the *free* edges. A triangle (e, v) is *mixed* under a given coloring if its free edges receive different colors.

For a set E of edges in K_n, let Δ_E be the set of all $|E|(n-2)$ triangles whose fixed edges belong to E, that is,

$$\Delta_E := \{(e, v) \;:\; e \in E,\, v \notin e\}\,.$$

Given a set of edges E, we say that a pair of triangles in Δ_E *locally collide* if either they share a free edge, or they share an edge which is free in one of them and fixed in the other, or these two triangles together with an edge from E produce a new triangle (see Fig. 1).

Say that a set E of edges in K_n is *sparse* if $|E| = \Theta(n)$, the edges in E form no triangles, and at most $O(n)$ paths of length two or three in E. If not stated otherwise, $c > 0$ will stand for a sufficiently small constant depending only on the balance parameter.

Lemma 2.1 *There exists a sparse set E of edges in K_n with the following property. For every set of at most 2^{cn^2} balanced colorings of K_n there is a subset $\Delta \subseteq \Delta_E$ of $|\Delta| = \Theta(n^2)$ triangles such that Δ has no local collisions, and a constant fraction of triangles in Δ is mixed under each of the given colorings.*

In this lemma the sets Δ are large, but they may be not collision-free. That is, they may have some *global* collisions: a triangle can be formed by taking edges from some *three* triangles in this set. The next lemma gives us collision-free sets.

Lemma 2.2 *There exists a sparse set E of edges in K_n with the following property. For every set of k balanced colorings of K_n there exists a collision-free set $\Delta \subseteq \Delta_E$ of triangles such that a constant fraction of them is mixed under each of the given colorings, and*

(i) $|\Delta| = \Omega(n^{3/2})$ *if* $k \leq 2^{cn^{3/2}}$;

(ii) $|\Delta| = \Omega(n^2)$ *if* $k \leq 2^{cn}$.

We postpone the proof of these lemmas to Sections 4.3–4.4.

In the lemmas above the colorings are "total"—each edge receives one of the two colors. In applications, however, we often have colorings which are only "partial"—some edges may be left uncolored. To obtain a similar result also in the case of partial colorings, we will consider 4-cliques instead of triangles.

Fix a partition $V = V_1 \cup V_2$ of the vertex set V of K_n into two disjoint parts of the same size ± 1. By a *square* in K_n we will mean a 4-clique with one edge e_1 drawn in V_1, and the second edge e_2 drawn in V_2. These two edges are the *fixed* edges of the square; the four remaining edges joining the endpoints of e_1 and e_2 lie in $V_1 \times V_2$, and we call them *bipartite*. For each square (e_1, e_2) we fix two of its disjoint bipartite edges and call them *free* edges of the square. A square is *mixed* under a coloring of K_n if its free edges receive different colors.

A set $\mathcal{S}$ of squares is *collision-free* if no two of them share a common bipartite edge. As in the case of triangles, the reason why this property is important for applications is the following. If $\mathcal{S}$ is collision-free then we can form a set of $2^{|\mathcal{S}|}$ graphs, by picking from each of the squares all its edges, except precisely one of the two free ones. None of these graphs contains a 4-clique (because of their bipartite structure), but the union of any two of them already contains at least one 4-clique.

Lemma 2.3 *If $n^{-1/6} \ll \lambda \leq 1/2$ then there exists an absolute constant $c > 0$ and a collision-free set $\mathcal{S}$ of squares in K_n with the following property. For every set of at most $2^{c\lambda^8 n^2}$ λ-balanced partial colorings of $V_1 \times V_2$, at least $\Omega(\lambda^8 n^2)$ squares in $\mathcal{S}$ are mixed under each of them.*

We postpone the proof of this lemma to Section 4.5. Here we only mention that the lemma does not hold if we take triangles instead of squares: if we consider triangles whose free edges belong to $V_1 \times V_2$, then there exists a partial coloring χ of $V_1 \times V_2$ which is λ-balanced for $\lambda = 1/4$ but *none* of the triangles is mixed. To see this, just split both sets V_1 and V_2 into subsets V_1', V_1'' and V_2', V_2'' of size $n/4$, and color all edges of $V_1' \times V_2'$ in red and all edges of $V_1'' \times V_2''$ in blue.

3 Applications

In this section we apply the coloring lemmas to prove lower bound on the nondeterministic multi-partition communication complexity of the K_t-FREE$_n$ functions for $t = 3, 4$. As a consequence we derive lower bounds for r-n.b.p. recognizing the K_4-freeness of graphs.

3.1 Multi-partition communication

Perhaps, the best way to view a *nondeterministic* communication protocol between two parties, Alice and Bob, wishing to compute a boolean function f, is a scheme by which a third party, Carole (a "superior being"), knowing the whole input a, can convince Alice and Bob what the value of $f(a)$ is (see, e.g., Sect. 2.1 in [10]). Hence, we have three players, Alice, Bob and Carole. Before the game starts, Carole chooses some partition of the set X of variables into disjoint blocks X_A and X_B; the partition must be strongly balanced in that both blocks have the same size ± 1. After that the first two players have only partial information about the input: Alice can see only the bits in X_A, and Bob can see only the bits in X_B. Given an input $a \in f^{-1}(1)$, Carole's goal is to convince Alice and Bob that $f(a) = 1$. For this purpose, she announces to both players some binary string W_a, a *certificate* for (or a *proof* of) the fact that "$f(a) = 1$." Having this certificate, Alice and Bob verify it *independently* and respond with either Yes or No. Alice and Bob agree that $f(a) = 1$ (and accept the input a) if and only if they both replied with Yes. If $f(a) = 0$ then Alice and Bob must be able to detect that the certificate is wrong no matter what Carole says. The protocol is correct if, for every input a, Alice and Bob accept it if and only if $f(a) = 1$. The communication complexity of this game is the length of the certificate W_a in the worst case.

For example, Carole can easily convince Alice and Bob that a graph G *has* a triangle: using only $3\lceil\log_2 n\rceil$ bits she announces the binary code of a triangle in G; Alice and Bob can locally check whether the edges of this triangle she/he should see are indeed present. On the other hand, Papadimitriou and Sipser [11] show that to convince the players that a graph has *no* triangles, Carole must announce almost entire graph. Let us stress that in this game Carole can choose an *arbitrary* balanced partition of the variables X, but after that she must use this partition for *all* inputs.

In this paper we consider the generalization of this game where Carole is allowed to change her opinion and use "most appropriate" partitions for different inputs. Such a strengthening of Papadimitriou–Sipser's model was (more or less explicitly) used by several authors as a tool of proving lower bounds on different types of branching programs (see, e.g., [6]).

More formally, in the *multi-partition* communication game the players act as follows. Given an input $a \in f^{-1}(1)$, Carole announces a pair (W_a, P_a) of binary strings where, as before, W_a is a certificate for the input a, and P_a is the binary code of a partition of input variables to be used by Alice and Bob on this input. The partition does not need to be strongly balanced—we only require that each block contains a γ-fraction of all variables where $0 < \gamma \leq 1/2$ may be an arbitrarily small (but fixed) constant; for ease of notation we don't show γ explicitly. The *multi-partition communication complexity* $\mathsf{C}(f)$ of f is the sum $|W_a| + |P_a|$ of the lengths of strings W_a and P_a on the worst case input a. In the case when Carole can use at most k different partitions, the corresponding measure is denoted by k-$\mathsf{C}(f)$. In these terms, the result from [11] says that $1\text{-}\mathsf{C}(K_3\text{-FREE}_n) = \Omega(n^2)$ in the case of strongly balanced partitions. In the next section we will prove that k-$\mathsf{C}\left(K_3\text{-FREE}_n\right) = \Omega(n^2)$ as long as $k \le 2^{cn}$, and $\mathsf{C}\left(K_3\text{-FREE}_n\right) = \Omega(n^{3/2})$.

3.2 Communication complexity of triangle-freeness

Recall that, given a set Δ of triangles in K_n, Δ-FREE is a boolean function which, given a graph G, outputs 1 if and only if none of the triangles from Δ is present in G. In the communication game for $\Delta\text{-FREE}_n$ the set Δ of triangles is known to all three players, and Carole's goal is to convince Alice and Bob that none of the triangles from Δ is present in a given graph G.

Theorem 3.1 *There exists a set Δ_0 of triangles in K_n such that $\mathsf{C}\left(\Delta_0\text{-FREE}\right) = \Theta(n^2)$.*

Proof. Let E be a set of edges guaranteed by Lemma 2.1. The set Δ_E has $\Theta(n^2)$ triangles. Say that a triangle is *chordal* if all its three vertices belong to some path of length three in E. Remove from Δ_E all such triangles, and let Δ_0 be the resulting set. Since the set of edges E is sparse, we only have a linear number (in n) paths of length three, and hence, we have removed at most $O(n)$ triangles. Our goal is to show that Δ_0 is the desired set of "hard" triangles.

Consider the communication game for $f = \Delta_0\text{-FREE}_n$. If Carole uses $k = 2^{\Omega(n^2)}$ partitions, we are done. So, assume that she uses $k \le 2^{o(n^2)}$ partitions. Our goal is to show that then Carole must use a certificate of length $\Omega(n^2)$.

To show this, let $\mathcal{C}$ be the set of $|\mathcal{C}| = k$ balanced colorings of K_n, corresponding to the partitions used by Carole. Since Δ_0 was obtained from Δ_E by removing a negligible number of triangles, Lemma 2.1 gives us a subset $\Delta \subseteq \Delta_0$ of $t := |\Delta| = \Theta(n^2)$ triangles such that Δ has no local collisions, has no chordal triangles, and for every coloring $\chi \in \mathcal{C}$ there exists a subset $\Delta_\chi \subseteq \Delta$ of $h := |\Delta_\chi| = \Omega(n^2)$ triangles, all of which are mixed under χ.

Let $x_i y_i z_i$ be the triple of variables where $x_i y_i$ correspond to the free edges and z_i to the fixed edge of the i-th triangle in Δ, $i = 1, \ldots, t$. Since Δ has no local collisions, no two triangles from Δ share a free edge, implying that all the variables $x_1, y_1, x_2, y_2, \ldots, x_t, y_t$ are distinct. Moreover, no two triangles in Δ share an edge which is free in one of them and fixed in the other, implying that these variables are different from the variables $z_1, \ldots, z_t$ (although z_i's themselves may be not distinct).

Hence, we can form a set $\mathcal{G}$ of $|\mathcal{G}| = 2^t$ graphs by picking from each of the triangles in Δ its fixed edge and *precisely one* of its free edges. That is, in the binary code of every graph in $\mathcal{G}$ each of the triples $x_iy_iz_i$ has one of the two values 011 or 101; all the remaining variables are set to zero.

We claim that none of the graphs in $\mathcal{G}$ has a triangle from Δ_0, and hence, must be accepted. To see this, take a graph in $\mathcal{G}$, and suppose that it contains a triangle $T = \{u, v, w\}$. If this triangle does not belong to Δ_E, there is nothing to prove. If it belongs to Δ_E then it must have at least one edge from E and, since E is triangle-free, at most two such edges. If only one edge of T would belong to E, then the remaining two edges of T would be free edges of some two triangles from Δ, and we would have a local collision between these two triangles (cf. the last three situations in Fig. 1(C)). So, the only possibility is that some two edges of T, say, uv and vw belong to E (cf. the first two situations in Fig. 1(C)). In this case the edge uw must be a free edge of some triangle (e, w') from Δ, implying that w' must be an endpoint of uw, say, $w' = w$. But then the edges e, uv and vw form a path of length three in E, meaning that the triangle T is chordal, and hence, cannot belong to Δ_0.

Thus, all the graphs from $\mathcal{G}$ must be accepted. Since we only have k colorings (partitions), Carole must use some one coloring χ for a set $\mathcal{G}' \subseteq \mathcal{G}$ of $|\mathcal{G}'| \geq |\mathcal{G}|/k \geq 2^t/k$ graphs from $\mathcal{G}$. We know that there is a subset $\Delta_\chi \subseteq \Delta$ of $h = \Omega(n^2)$ triangles, all of which are mixed under χ. Assume w.l.o.g. that these are the first h triangles $x_1y_1z_1, \ldots, x_hy_hz_h$, hence, $\chi(x_i) \neq \chi(y_i)$ for all $i = 1, \ldots, h$. That is, for each of the first h triangles $x_iy_iz_i$, each of its two free edges x_i and y_i is seen by precisely one of the players, Alice and Bob. Since the graphs in $\mathcal{G}'$ can take at most 2^{t-h} different values on the variables $x_{h+1}, y_{h+1}, \ldots, x_t, y_t$, there exists a subset $\mathcal{G}_\chi \subseteq \mathcal{G}'$ of

$$|\mathcal{G}_\chi| \geq |\mathcal{G}'|/2^{t-h} \geq 2^t/k2^{t-h} = 2^h/k$$

graphs having the same values on these variables.

We claim that for every graph from $\mathcal{G}_\chi$, Carole must use a different certificate. Since $h = \Omega(n^2)$ and $k = 2^{o(n^2)}$, this will imply that the binary length of a certificate must be at least $\log_2 |\Delta_\chi| = \Omega(n^2)$.

To show this, assume that Carole uses the same certificate for two different graphs G_1 and G_2 in $\mathcal{G}_\chi$. By the construction of $\mathcal{G}_\chi$, the union $G = G_1 \cup G_2$ of these graphs contains at least one triangle $T \in \Delta_\chi \subseteq \Delta_0$. Let x_i, y_i be the free edges of T. We may assume w.l.o.g. that x_i is seen by Alice, y_i is seen by Bob, and that x_i is present ($x_i = 1$) in the first graph G_1; hence, y_i is present ($y_i = 1$) in G_2. Let G be the union of the part of G_1 seen by Alice with the part of G_2 seen by Bob. Both players replied with Yes on both G_1 and G_2. Since the players have to verify the certificate *independently*, the players are forced to reply with Yes also on G. But this is impossible because G contains the triangle T and this triangle belongs to Δ_0. □

In Theorem 3.1 the set Δ of triangles is not specified. However, using the second coloring lemma, we can obtain the following lower bounds for the "pure" version K_3-FREE$_n$ of the triangle-freeness property.

Theorem 3.2 *Let $f = K_3\text{-FREE}_n$. Then $\mathsf{C}(f) = \Omega(n^{3/2})$. Moreover, there is a constant $c > 0$ such that $k\text{-}\mathsf{C}(f) = \Theta(n^2)$ as long as $k \leq 2^{cn}$.*

Proof. The proof is precisely the same as that of Theorem 3.1 with only one difference: this time we take the set Δ of triangles, guaranteed by Lemma 2.2. The fact that this set is free not only from local collisions but also from global ones, implies that the graphs in the constructed set $\mathcal{G}$ have no triangles at all, and hence, must be accepted by K_3-FREE$_n$. □

In the proof of Theorem 3.1 it was sufficient to work with local collisions only. The absence of such collisions in $\Delta \subseteq \Delta_0$ guarantees that no new triangle from Δ_0 can be formed by triangles in Δ. However, triangles outside of Δ_0 may be formed, because of possible global collisions. Demanding the absence of such collisions is a severe requirement and reduces the lower bound from $\Omega(n^2)$ to $\Omega(n^{3/2})$. At the moment we don't know whether such a drastic jump is an inherent property of the triangle-freeness function K_3-FREE$_n$ itself, or it is just a weakness of our argument. Although we cannot refute the second, it may well be that the former is true. Triangle-free graphs have many specific structural properties which Carole could try to encode in her certificates. In particular, in every triangle-free graph the neighborhoods of their vertices span at least $n^2/4$ non-edges, and Carole could try, say, to encode a large fraction of non-edges using much fewer than n^2 bits (see Section 5 for a discussion).

3.3 Branching programs for K_4-freeness

The model of nondeterministic syntactic read-r times branching programs (r-n.b.p.) was introduced in [6]. These are the usual nondeterministic branching programs with the restriction that along each path (be it consistent or not) each variable can be tested at most r times. The size of a branching program is the number of edges in the underlying graph. Using the coloring lemma for 4-cliques one can prove the following lower bound on the size of r-n.b.p. recognizing the K_4-freeness of graphs.

Theorem 3.3 *If* $r = o(\log n/ \log\log n)$ *then the* 4*-clique-freeness function* K_4-FREE$_n$ *requires* r*-n.b.p. of size exponential in* $(n/r)^2$.

To prove Theorem 3.3, we first relate the size of an r-n.b.p. computing a boolean function f with the so-called *overlapping* multi-partition communication complexity $\mathsf{C}_\lambda(f)$ of f. This measure is an extension of $\mathsf{C}(f)$ where the blocks of input variables X_A and X_B Carole gives to Alice and Bob need not be disjoint: we only require that both $|X_A - X_B|$ and $|X_B - X_A|$ are at least $\lambda \cdot |X|$. Put otherwise, instead of partitions of X into disjoint blocks, Carole chooses a balanced *partial* coloring of X in red and blue. Having such a coloring, the restriction is that Alice cannot see red variables, and Bob cannot see blue variables; the uncolored variables are seen by both players.

The following lemma gives a general lower bound on the size of an r-n.b.p. for a boolean function f in terms of the overlapping multi-partition communication complexity of f.

Lemma 3.4 *If* $r = o(\log n/ \log\log n)$ *then every boolean function* f *on* n *variables requires an* r*-n.b.p. of size exponential in* $\mathsf{C}_\lambda(f)/r^2$ *where* $\lambda = r^{-O(r)}$.

We give the proof of this lemma in Appendix A; here we only mention that the proof is based on a general result from [6] saying that if a boolean function can be computed by a small r-n.b.p. then it can be represented as an OR of a small number of boolean functions of a very special form.

Proof of Theorem 3.3. By Lemma 3.4, it is enough to show that there exists a subfunction f of K_4-FREE$_n$ such that $\mathsf{C}_\lambda(f) = \Omega(\lambda^d n^2)$ for some constant $d \geq 0$. To define the desired subfunction, fix a partition of the vertex set of K_n into two parts V_1, V_2 of size $|V_1| = |V_2| = n/2$, and let $\mathcal{S}$ be a collision-free set of squares, guaranteed by Lemma 2.3. Set all fixed edges of these squares (i.e., the variables corresponding to these edges) to the constant 1, and set to 0 all the remaining edges in V_1 and in V_2. The obtained subfunction f of K_4-FREE$_n$ depends only on $n^2/4$ variables, corresponding to bipartite edges in $V_1 \times V_2$.

Consider a λ-overlapping multi-partition communication game for this subfunction, and let $\mathcal{C}$ be the set of λ-balanced partial colorings of $V_1 \times V_2$ used by Carole. If

$|\mathcal{C}| \geq 2^{\Omega(\lambda^8 n^2)}$ then $\mathsf{C}_\lambda(f) = \Omega(\lambda^8 n^2)$, and we are done. So assume that $|\mathcal{C}| \leq 2^{o(\lambda^8 n^2)}$. Since $r = o(\log n/\log\log n)$, the balance parameter $\lambda = r^{-O(r)}$ is much larger than $n^{-1/6}$, and Lemma 2.3 gives us a collision-free set $\mathcal{S}$ of squares, at least $\Omega(\lambda^8 n^2)$ of which are mixed under each of the given colorings. Since no two squares in $\mathcal{S}$ share a free edge, we can construct a set $\mathcal{G}$ of $2^{|\mathcal{S}|}$ graphs by picking from each of the squares all its edges, except precisely one of the two free ones. Since none of squares in $\mathcal{S}$ share a bipartite edge, none of the graphs in $\mathcal{G}$ contains a square. But the union of any two of them already has a square. Arguing as in the proof of Theorem 3.2 we conclude that Carole must use a certificate of length at least $\log_2|\mathcal{S}| = \Omega(\lambda^8 n^2)$. □

4 Proofs of coloring lemmas

The proofs of all three coloring lemmas follow the same general frame: we first use so-called "joining lemma" to produce a large set of triangles (or 4-cliques) so that a large fraction of them is mixed under *every* balanced coloring. After that we use so-called "collision lemma" to remove possible collisions between triangles (or 4-cliques). We first prove these two simple lemmas. In their proofs we use the following simplest version of Chernoff's inequality: if X is the sum of n independent Bernoulli random variables each with success probability p then $X \leq np/2$ with probability at most $e^{-pn/8}$, and $X \geq 2pn$ with probability at most $e^{-pn/12}$.

4.1 Joining and collision lemmas

Given two subsets of vertices A and B, we say that an edge *joins* this pair if one its endpoint belongs to A and the other to B; in particular, both endpoints may belong to $A \cap B$. We say that a set E of edges in K_n is an *ϵ-expander with expansion D* if every two sets of at least ϵn vertices are joined by at least Dn edges from E; an *ϵ-expander* is an expander with expansion $D \geq 1$. (Note that our definition of an expander is slightly different from the standard notion where instead of pairs of sets one is interested in the number of edges joining a set with its complement.) An ϵ-expander E is *sparse* if $|E| = \Theta(n/\epsilon^2)$, the edges in E form no triangles and at most $O(n/\epsilon^{2l})$ paths of length $l = 2, 3$.

Lemma 4.1 **(Joining Lemma)** *If $n^{-1/6} \ll \epsilon \leq 1$ then sparse ϵ-expanders with arbitrary constant expansion exist.*

Proof. Let $D \geq 1$ be an arbitrary constant, and n be sufficiently large, $n \geq 40D/\epsilon^2$ is enough. Consider a random subset $\mathbf{E}$ of edges in K_n, each edge in which appears independently and with equal probability $p := 40D/\epsilon^2 n$. Chernoff's and Markov's inequalities imply that, with probability $> 1/2$, the set $\mathbf{E}$ has $\Theta(n/\epsilon^2)$ edges, at most $p^l n^{l+1} = O(n/\epsilon^{2l})$ paths of any constant length l, and at most $p^3 n^3 = O(1/\epsilon^6)$ triangles. We also claim that $\mathrm{Prob}\,[\mathbf{E}$ is an ϵ-expander with expansion $2D] > 1/2$.

To show this, let A and B be two sets of at least $m = \epsilon n$ vertices, and F be the set of all edges joining these two sets. If $|A \cap B| \leq 2m/3$ then we have at least $(m/3)^2 > 0.1m^2$ bipartite edges in $(A - B) \times (B - A)$, and if $|A \cap B| \geq 2m/3$ then at least least $0.2m^2$ edges join the vertices in $A \cap B$. Hence, $|F| \geq 0.1\epsilon^2 n^2$ and the expected number of edges in $\mathbf{E} \cap F$ is $p \cdot |F| \geq 4Dn$. By Chernoff's inequality, the actual number of edges in this intersection is smaller than $2Dn$ with probability at most $e^{-p|F|/8} \leq e^{-4Dn} \leq e^{-4n}$. Since the total number of large pairs A, B does not exceed 2^{2n} we conclude that, with probability at least $1 - 2^{2n} \cdot e^{-4n} > 1/2$, $\mathbf{E}$ is an ϵ-expander with expansion $2D$.

Thus, there exists a set E of edges with both properties. That is, E is an ϵ-expander with expansion $2D$, has $|E| = \Theta(n/\epsilon^2)$ edges, at most $O(n/\epsilon^{2l})$ paths of length $l = 2, 3$, and at most $O(1/\epsilon^6)$ triangles. Since $\epsilon \gg n^{-1/6}$, the number of triangles is $o(n)$ and we can safely remove one edge from each of them without destroying any of the remaining properties. In particular, the expansion of the obtained set of edges is still at least $2D - o(1) \geq D$. □

To remove the possible collisions between cliques we need the following property of independent sets in sparse hypergraphs. Recall that a *hypergraph* over a set V of vertices is a family $\mathcal{F}$ of subsets of V, called *hyperedges*. The *rank* of the hypergraph is the minimum cardinality of its hyperedge. As in the case of graphs, a set of vertices is *independent* if it contains no hyperedge of $\mathcal{F}$. We will consider hypergraph whose vertices are some configurations (edges, triangles, squares, etc.) and each hyperedge corresponds to a "collision" between these configurations. Hence, being independent in such a hypergraph is equivalent to being collision-free.

A hypergraph is *sparse* if the number of its hyperedges is linear in the number of vertices.

Lemma 4.2 **(Collision Lemma)** *Let $\mathcal{F}$ be a sparse hypergraph of rank $r \geq 2$ over a set V of N vertices. For every constant $c_1 > 0$ there exists a constant $c_2 > 0$ such* *that for every family of up to $2^{c_2 N}$ subsets of V of size at least $c_1 N$ there exists an independent set which contains at least $c_2 N$ vertices in each of these sets.*

This lemma is a very special (but handy) version of the following more general fact.

Lemma 4.3 *Let $\mathcal{F}$ be a hypergraph of rank $r \geq 2$ over a set V of N vertices. Let $\mu > 0$ and*

$$p = p(\mathcal{F}) := \left(\frac{\mu N}{8|\mathcal{F}|}\right)^{1/(r-1)}.$$

Then, for any family $V_1, \ldots, V_k$ of $k \leq 2^{\mu p N/8}$ subsets of V of size at least μN there exists an independent set $S \subseteq V$ such that $|V_i \cap S| \geq \mu p N/4$ for all $i = 1, \ldots, k$.

Proof. If $p \geq 1$ then $|\mathcal{F}| \leq \mu N/8$, and the desired independent set S can be obtained by deleting one vertex from each hyperedge. So, assume that $p < 1$, and let $\mathbf{S}$ be a random set of vertices where each vertex is picked independently and with equal probability p. For every $i = 1, \ldots, k$, the expected number of vertices in $V_i \cap \mathbf{S}$ is $p \cdot |V_i| \geq \mu p N$. By Chernoff's inequality, the probability that $|V_i \cap \mathbf{S}| \geq \mu p N/2$ for all $i = 1, \ldots, k$, is at least $1 - k \cdot e^{-\mu p N/8}$, which is $> 1/2$ because $k \leq 2^{\mu p N/8}$. On the other hand, the expected number of hyperedges of $\mathcal{F}$ lying entirely in $\mathbf{S}$ does not exceed $p^r|\mathcal{F}|$, and by Markov's inequality, the actual number of such hyperedges does not exceed $2p^r|\mathcal{F}|$ with probability greater than $1/2$.

Fix a set satisfying both these conditions, and remove one vertex from each hyperedge lying within this set. The resulting set S is independent and

$$|V_i \cap S| \geq \mu p N/2 - 2p^r|\mathcal{F}| = pN\left(\frac{\mu}{2} - 2p^{r-1}\frac{|\mathcal{F}|}{N}\right) = \mu p N/4$$

for every $i = 1, \ldots, k$. □

4.2 Many mixed triangles under each coloring

Let $0 < \gamma \leq 1/2$ be the balance parameter of the considered colorings, and apply the Joining Lemma with $\epsilon := \gamma^2/5$. This gives us a sparse ϵ-expander E. Being an expander means that at least n edges of E join each pair of vertex sets of size at least ϵn. We first use this property to show that a constant fraction of triangles in $\Delta_E = \{(e, v) : e \in E, v \notin e\}$ is mixed under *every* γ-balanced coloring of K_n.

Claim 4.4 *At least ϵn^2 of triangles in Δ_E are mixed under every balanced coloring of K_n.*

Proof. Take an arbitrary such coloring, and call a vertex *red* (*blue*) if more that $c_1 n$ of its incident edges are red (blue) where $c_1 := 1-\gamma/3$ and γ is the balance parameter of χ. A vertex is *mixed* if it is neither red nor blue.

Our first goal is to prove that at least ϵn vertices must be mixed (a similar fact was proved in [11] for the case $\gamma = 1/2$). To show this, let R, B and M be the sets of red, blue and mixed vertices, respectively. Let r be the sum over all vertices $v \in R$, of the number of red edges incident to v. By Euler's theorem, r is at most two times the total number of red edges, implying that $r \leq 2(1-\gamma)\binom{n}{2} \leq (1-\gamma)n^2$. Since $r \geq c_1 n \cdot |R|$, this implies that $|R| \leq (1-\gamma)n^2/c_1 n = c_2 n$ where $c_2 := (1-\gamma)/c_1$. The same argument yields that $|B| \leq c_2 n$. Now suppose to the contrary that we have *fewer* than dn mixed vertices where $d := 2c_1 - 1 - c_2 = 2\gamma^2/(9-3\gamma) \geq \gamma^2/5 = \epsilon$. As $c_1 > 1/2$, all three sets B, R and M are disjoint, and cover all n vertices. Since $|B|, |R| \leq c_2 n$ this, together with our assumption $|M| < dn$, implies that $|B|, |R| > c_3 n$ where $c_3 := 1-(d+c_2) \geq 2(1-c_1)$. Hence, for every vertex $v \in R$ we have $|\{v\} \times B| = |B| > c_3 n$ edges going to the vertices in B. As v is red, fewer than $(1-c_1)n$ of these edges can be blue; hence, more than $|B| - (1-c_1)n \geq |B|/2$ of these edges must be red. Thus, more than half of edges in $R \times B$ must be red. Symmetrically, more than half of edges in $R \times B$ must be blue, a contradiction.

Thus, at least ϵn vertices are mixed. For each such vertex, at least $(1-c_1) = \gamma n/3$ of its incident edges are blue and at least $(1-c_1) = \gamma n/3$ of its incident edges are red. Select $\lfloor \epsilon n \rfloor$ of these mixed vertices and call them *top* vertices. As $2\epsilon \leq \gamma/6$, each top vertex v has at least ϵn red edges and ϵn blue edges to *bottom* (non-top) vertices. Since E is an ϵ-expander, the pair of sets of bottom vertices, connected to v by red and blued edges, are large enough to be joined by at least n edges from E. Together with v, every such edge $e \in E$ produces the triangle (e, v) in Δ_E which is mixed under χ. Since we have ϵn top vertices, the total number of mixed triangles in Δ_E is at least ϵn^2. □

4.3 Removing the collisions: proof of Lemma 2.1 and 2.2(i)

To prove Lemma 2.1 we have to remove all possible *local* collisions between the triangles in the set Δ_E given by Claim 4.5. Recall that a local collision between two

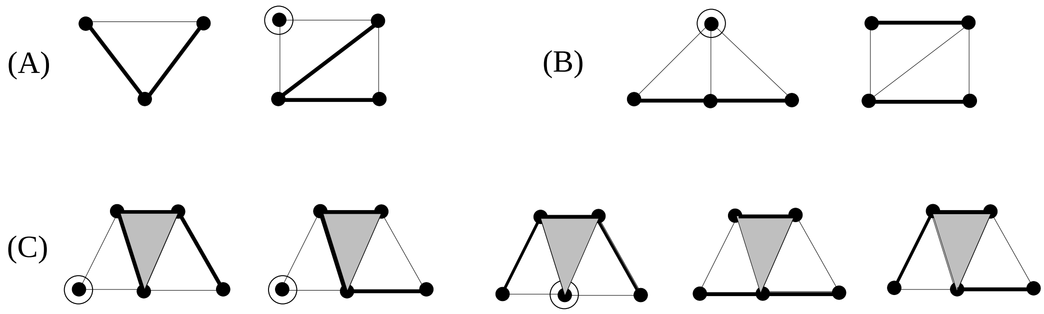

Figure 1: All types of local collisions; bold lines correspond to fixed edges

triangles occurs if either:

(A) the triangles share an edge which is free in one of them and fixed in the other, or

(B) the triangles share a common free edge, or

(C) a triangle can be formed by taking a free edge from the first triangle, an edge from the second triangle and an edge from E.

To eliminate such collisions, we first estimate their number. Let P_l be the number of paths of length l in E. Since the expander E is sparse, $P_l = O(n)$ for $l = 1, 2, 3$.

Claim 4.5 *There are at most $O(n^2)$ local collisions in Δ_E.*

Proof. Let (e_1, v_1) and (e_2, v_2) be two triangles in Δ_E. Fig. 1 depicts all possible local collisions between these triangles. For a collision of type (A) to occur, the fixed edges e_1, e_2 must form a path of length two, and the top vertex of at least one triangle must be an endpoint of the fixed edge of the other one. In this case we have at most $O(n \cdot P_2) = O(n^2)$ possibilities.

If the collision is of type (B) but not of type (A) then either the fixed edges form a path of length two and the top vertex is the same, or the fixed edges are disjoint and the top vertex v_i of the triangle (e_i, v_i) must be an endpoint of the second edge e_{3-i}. In the first case we have at most $O(n \cdot P_2) = O(n^2)$ possibilities, and in the second at most $2P_1^2 = O(n^2)$ possibilities.

In the case of type (C) but neither (A) nor (B) collision, a triangle T is formed by an edge $e \in E$, and two edges of the colliding triangles. Since E has no triangles, at least one of the edges e_1 or e_2 does not belong to T; say, $e_2 \notin T$. If the second edge e_1 belongs to T then the edges e, e_1 and e_2 form a path of length three, and the top vertex v_2 of the second triangle (e_2, v_2) belongs to this path. Hence, in this case we have at most $n \cdot P_3$ possibilities. If neither of the edges e_1, e_2 belongs to T then either the edges e_1, e, e_2 for a path of length three and $v_1 = v_2$ (at most $n \cdot P_3$ possibilities), or each of the top vertices v_1, v_2 is an endpoint of some of the edges e, e_1, e_2, some two of whom form a path of length two (at most $O(P_1 \cdot P_2)$ possibilities). Hence, the number of type (C) collisions is also at most $O(n^2)$. □

Now we are ready to finish the proof of Lemma 2.1. Fix an arbitrary set $\mathcal{C}$ of at most 2^{cn^2} balanced colorings of K_n where $c > 0$ is sufficiently small constant. Claim 4.4 gives a family $\{\Delta_\chi : \chi \in \mathcal{C}\}$ of large subsets of Δ_E such that all the triangles in Δ_χ are mixed under χ. Consider a "collision graph" graph $\mathcal{F}$ whose vertices are triangles from Δ_E, and edges are pairs of triangles forming a local collision. Since, by Claim 4.5, $\mathcal{F}$ is sparse, we can apply the Collision Lemma and obtain an independent set $\Delta \subseteq \Delta_E$ such that all the intersections $\Delta \cap \Delta_\chi$, $\chi \in \mathcal{C}$, are still large. Since independence of Δ in $\mathcal{F}$ is equivalent to having no local collisions, we are done.

Now we turn to the proof of Lemma 2.2(i). Let $\Delta \subseteq \Delta_E$ be the set of triangles guaranteed by Lemma 2.1. This set has $\Theta(n^2)$ triangles, and has no local collisions. Still, at least potentially, the set Δ may have some *global* collisions: it may happen that a triangle can be formed by taking edges from some *three* triangles in this set. To prove Lemma 2.2(i), we have to remove all possible global collisions. Let us first estimate their number.

Claim 4.6 *There are at most $O(n^3)$ global collisions in Δ.*

Proof. Suppose a triangle T can be formed by picking edges from some three triangles (e_i, v_i), $i = 1, 2, 3$ in Δ. Since E is triangle-free and Δ has no local collisions, we only have two possibilities: either T contains precisely two of the fixed edges e_i or none of them (see Fig. 2).

If T contains precisely two fixed edges, say, e_1 and e_2 then T is formed by e_1, e_2 and and a free edge of the third triangle (e_3, v_3). In this case the edges e_1, e_2, e_3 form a path of length three, and the top vertex v_3 must be an endpoint of e_1 or e_2. Since we only have $P_3 = O(n)$ paths of length three and at most n^2 possibilities for the

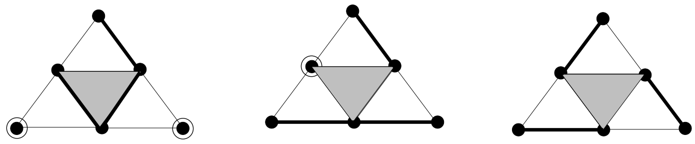

Figure 2: Global collisions; bold lines correspond to fixed edges.

choice of the top vertices v_1 and v_2, the total number of global collisions of this type is at most $O(P_3 n^2) = O(n^3)$.

If T contains no fixed edges then either: (a) some two of the edges e_1, e_2, e_3, say, e_1 and e_2, form a path of length two, the top vertex v_3 coincides with one of the top vertices v_1 or v_2, say, $v_3 = v_1$, and $v_2 \in e_3$, or (b) the edges e_1, e_2, e_3 are mutually disjoint and the top vertices v_1, v_2, v_3 of the corresponding triangles are the endpoints of these edges. In the first case (a) we have P_2 possibilities to choose the path $e_1 e_2$, P_1 possibilities for the edge e_3, and at most n possibilities for the top vertex $v_1 = v_3$. Since the top vertex v_2 must belong to e_3, we have at most $O(n \cdot P_1 \cdot P_2) = O(n^3)$ global collisions of this type. In the second case (b) each triple e_1, e_2, e_3 of edges in E can form only a constant number of triangles, since in this case for each edge e_i we have at most 4 possibilities to chose the top vertex v_i. Thus, in this case the number of global collisions does not exceed $O(P_1^3) = O(n^3)$. □

To destroy the possible global collisions between the triangles in Δ, we will use the assumption of Lemma 2.2(i) that the number k of given colorings $\chi_1, \ldots, \chi_k$ is at most $2^{cn^{3/2}}$ where $c > 0$ is a sufficiently small constant. By Lemma 2.1, we know that $|\Delta| = \Theta(n^2)$ and for each $i = 1, \ldots, k$ there is a set $\Delta_i \subseteq \Delta$ of $\Omega(n^2)$ triangles, all of which are mixed under χ.

Consider the "collision hypergraph" $(V, \mathcal{F})$ whose vertices are triangles in Δ, and hyperedges are triples of triangles forming global collisions. This hypergraph has $N = |\Delta| = \Theta(n^2)$ vertices, and by Claim 4.6, $|\mathcal{F}| = O(n^3)$ hyperedges. Moreover, the sets of vertices $V_i = \Delta_i$ are large enough, since $|\Delta_i| = \Omega(n^2) \geq \mu N$ for some constant $\mu > 0$ independent of k. Since $\mathcal{F}$ has rank $r = 3$, this implies that $p = p(\mathcal{F}) = (\mu N / 8|\mathcal{F}|)^{1/2} = \Omega(n^{-1/2})$. Hence, if the constant $c > 0$ is small enough to ensure the inequality $cn^{3/2} \leq \mu p N/8$ then, by Lemma 4.3, there exists a subset $\Delta' \subseteq \Delta$ such that there are no collisions between the triangles in Δ', and for each $i = 1, \ldots, k$, this set contains a subset $\Delta'_i = \Delta_i \cap \Delta'$ of $|\Delta'_i| = \Omega(pN) = \Omega(n^{3/2})$ triangles, all of which are mixed under χ_i. This completes the proof of Lemma 2.2(i)

4.4 Proof of Lemma 2.2(ii)

In the case of relatively small number of (at most 2^{cn}) colorings we can find a large set of triangles with an additional property that their fixed edges form a matching.

Claim 4.7 *For every set $\chi_1, \dots, \chi_k$ of $k \le 2^{cn}$ balanced colorings of K_n there exists a matching M of size $\Omega(n)$ such that a constant fraction of the triangles in Δ_M is mixed under each χ_i.*

Proof. Let $0 < \gamma \le 1/2$ be the balance parameter of the colorings considered, and $\epsilon := \gamma^2/5$. Take a sparse ϵ-expander E guaranteed by the Joining Lemma. When proving Claim 4.4 we have shown that for every coloring χ_i there exists a set of ϵn top vertices and a sequence of ϵn pairs of sets of bottom vertices of size at least ϵn such that every edge e joining any of these pairs, together with the corresponding top vertex v, produces a triangle (e, v) which is mixed under χ_i. Hence, if $E_{i,j}$ denotes the set of edges of E joining the j-th pair of sets of bottom vertices of the i-th coloring, it is enough to show that some matching contains $\Omega(n)$ edges in each of these sets. For this purpose, consider a "collision graph" $\mathcal{F}$ whose vertices are edges from E, and where $e_1, e_2 \in E$ are joined by an edge if and only if $e_1 \cap e_2 = \emptyset$. This graph has $|E| = \Theta(n)$ vertices and, since the expander E is sparse, at most $O(n)$ edges (at most so many paths of length two in E). Hence, $\mathcal{F}$ is sparse. On the other hand, since E is an ϵ-expander, the sets $E_{i,j}$ are large, $|E_{i,j}| \ge n$ for all $1 \le i \le k$ and $1 \le j \le \epsilon n$. Thus, if the constant $c > 0$ (in the upper bound $k \le 2^{cn}$) is sufficiently small, then the Collision Lemma gives us an independent set $M \subseteq E$ in $\mathcal{F}$ (a matching in K_n) such that all intersections $M \cap E_{i,j}$ have size $\Omega(n)$, as desired. $\square$

Now fix a matching, given by the previous claim, and consider the set Δ_M of $|\Delta_M| = \Theta(n^2)$ induced triangles. We know that a constant fraction of these triangles is mixed under each of the given 2^{cn} balanced colorings of K_n. To finish the proof of Lemma 2.2(ii), it remains to remove possible collisions from Δ_M.

Since M is a matching, the fixed edges of the triangles from Δ_M can form no path of length two. This immediately implies that we can only have two types of possible collisions between the triangles: type I depicted in Fig. 1(B) situation two, and type II depicted in Figure 2 situation three. Since fixed edges cannot form any path of length two, one triangle can participate in at most two collisions of type I. By the same reason, each of the remaining triangles can participate in at most one triple of triangles forming a collision of type II. Hence, the corresponding collision hypergraph has $|\Delta_M|$ vertices and at most $3|\Delta_M|$ hyperedges. By the Collision Lemma, there is a collision-free subset $\Delta \subseteq \Delta_M$ of $|\Delta| = \Omega(n^2)$ triangles, a constant fraction of which is mixed under each of the given colorings. This completes the proof of Lemma 2.2(ii).

4.5 Proof of Lemma 2.3

Let V_1, V_2 be a fixed bipartition of the vertices of K_n into two disjoint parts of equal size ± 1. Each pair (e_1, e_2) of edges where e_i is drawn between the vertices in V_i, defines a square. Given a set E of such edges, let $\mathcal{S}_E$ be the set of squares defined by the edges in E. Recall that two squares *collide* if they share an edge in $V_1 \times V_2$.

Claim 4.8 *Let $n^{-1/6} \ll \lambda \leq 1/2$. There exists a set of $\Theta(n^2/\lambda^4)$ squares, at most $O(n^2/\lambda^8)$ pairs of whom collide, and at least n^2 of whom are mixed under each λ-balanced partial coloring of $V_1 \times V_2$.*

Proof. Set $\epsilon = \lambda/4$, and let $E = E_1 \cup E_2$ where E_i is a sparse ϵ-expander in V_i given by the Joining Lemma. Take an arbitrary λ-balanced partial coloring of $V_1 \times V_2$. For a vertex $v \in V_1$ let its *red degree* (*blue degree*) be the number of red (blue) edges incident to it. Since the average red (blue) degree is at least $\lambda n/2$, there exists a set R (B) of at least $\lambda n/4 = \epsilon n$ vertices in V_1 of red (resp., blue) degree at least $\lambda n/4 = \epsilon n$. That is, for every two vertices $u \in R$ and $v \in B$ there exists a pair u_R, v_B of subsets in V_2 of size ϵn such that all edges joining u (resp., v) with the vertices in u_R (resp., in v_B) are red (resp., blue). Since E is an ϵ-expander, the pair R, B as well as each of the pairs v_R, v_B with $u \in R$ and $v \in B$, are joined by at least n edges from E. Since each edge between u and v, together with at least n edges of E joining the sets u_R and v_B, induces n squares, all of which are mixed under χ, the total number of mixed squares in $\mathcal{S}_E$ is at least n^2.

Since both expanders E_1 and E_2 are sparse, each of them has $P_1 = \Theta(n/\lambda^2)$ edges and at most $P_2 = O(n/\lambda^4)$ paths of length 2. Hence, $|\mathcal{S}_E| = \Theta(n^2/\lambda^4)$, and it remains to show that at most $O(n^2/\lambda^8)$ pairs of squares in $\mathcal{S}_E$ can collide, i.e. share a bipartite edge. If two squares share a bipartite edge, then either they share two such edges, or they share only one such edge. In the first case we have at most $O(P_1 \cdot P_2)$

possibilities, whereas in the second we have at most $O(P_2 \cdot P_2)$ possibilities. Hence, at most $O(n^2/\lambda^8)$ pairs of squares can collide. $\Box$

Now we are ready to finish the proof of Lemma 2.3. Fix an arbitrary set $\mathcal{C}$ of at most $2^{c\lambda^8 n^2}$ balanced partial colorings of $V_1 \times V_2$ where $c > 0$ is sufficiently small constant. Claim 4.8 gives us a family $\mathcal{S}$ of $\Theta(n^2/\lambda^4)$ squares such that at most $O(n^2/\lambda^8)$ pairs of then collide and, for each coloring $\chi \in \mathcal{C}$, $\mathcal{S}$ contains a subset $\mathcal{S}_\chi \subseteq \mathcal{S}$ of n^2 squares all of whom are mixed under χ. It remains to destroy possible collisions between the squares in $\mathcal{S}$.

For this purpose, let us consider the "collision graph" $(V, \mathcal{F})$ whose vertices are squares in $\mathcal{S}$, and two squares are joined by an edge iff these squares share a common bipartite edge. By Claim 4.8 we know that the collision graph has $N = \Theta(n^2/\lambda^4)$ vertices and at most $|\mathcal{F}| = O(n^2/\lambda^8) = O(N/\lambda^4)$ edges. Moreover, the sets of vertices $V_\chi = \mathcal{S}_\chi$ are large enough: $|\mathcal{S}_\chi| \geq n^2 \geq \mu N$ for $\mu = \Omega(\lambda^4)$. Since $\mathcal{F}$ has rank $r = 2$, this implies that $p(\mathcal{F}) = \mu N/(8|\mathcal{F}|) = \Omega(\lambda^8)$. If $c > 0$ is sufficiently small to ensure $c\lambda^8 n^2 \leq \mu p N/8$, then $k \leq 2^{\mu p N/8}$, and the Collision Lemma gives us a subset $\mathcal{S}' \subseteq \mathcal{S}$ such that no two squares in $\mathcal{S}'$ share a bipartite edge, and for each coloring $\chi \in \mathcal{C}$, this set contains a subset $\mathcal{S}_\chi \cap \mathcal{S}'$ of at least $\mu p N/4 = \Omega(\lambda^8 n^2)$ squares, all of which are mixed under χ. This completes the proof of Lemma 2.3.

5 Concluding remarks

1.

Usually, a lower bound on the communication complexity of a given boolean function f is obtained by choosing a large enough set $F \subseteq f^{-1}(1)$ of inputs which is "hard" for *every* partition of input variables. That is, given a partition of the input variables, Carole cannot use one certificate for "too many" inputs from F without forcing Alice and Bob to (wrongly) accept an input from $f^{-1}(0)$. To our knowledge, all the lower bounds on the communication complexity of explicit boolean functions, including the highest ones proved in [6, 4, 1, 5], were obtained in this way. In the case of K_3-FREE$_n$ and K_4-FREE$_n$ functions this approach does not work (at least directly). We are not able to exhibit a large set F of graphs which is hard for every partition—the set of hard inputs F in our proof *depends* on the given set of partitions used by Carole: we are able to construct F only after Carole has fixed her opinion about the partitions she would like to use. This (dependency of F on the partitions) seems to be a new aspect in understanding the communication complexity.

2.

Let us note that, although simple, the arguments we used are quite general and may be also applied to other graph-theoretic problems. Just to mention an example, let $C_4\text{-FREE}_n$ be a boolean function which, given a graph G on n vertices, accepts it if and only if G has no cycle of length four. Kleitman and Winston [9]) proved that the number 4-cycle-free graphs is $2^{\Theta(n^{3/2})}$. This immediately implies that already the one-partition communication complexity of $C_4\text{-FREE}_n$ does not exceed $O(n^{3/2})$: given a C_4-free graph G, Carole just announces the entire graph G to both players. On the other hand, essentially the same argument as in the proof of the coloring lemma for triangles can be used to show that there is a set $\mathcal{C}$ of $\Theta(n^2)$ 4-cycles, a constant fraction of which is mixed under each balanced coloring of K_n. Moreover, due the "sparseness" of the underlying set of fixed edges given by the Joining Lemma, at most $O(n^4)$ quartets of 4-cycles in $\mathcal{C}$ collide, i.e., form a new 4-cycle. Applying Lemma 4.3 in this situation, we have a hypergraph of rank $r = 4$ on $|V| = \Omega(n^2)$ vertices, and with $|\mathcal{F}| = O(n^4)$ edges. In this case, $p(\mathcal{F}) = \Omega\left((N/|\mathcal{F}|)^{1/3}\right) = \Omega(n^{-2/3})$, and we obtain a collision-free subset $\mathcal{C}'$ of $|\mathcal{C}| = \Omega(p \cdot |V|) = \Omega(n^{4/3})$ 4-cycles. Arguing as in the proof of Theorem 3.2, this implies that the multi-party communication complexity of $C_4\text{-FREE}_n$ is $\Omega(n^{4/3})$.

3.

Next, we mention that the construction of mixed triangles can be used to give a lower bound on the number $f(n)$ of maximal triangle-free graphs on n vertices. A triangle-free graph is *maximal* if no edge can be added without forming a triangle. Erdős asked (see, e.g., [12], Problem 10.2 or [7], Problem 48) to determine or estimate $f(n)$. It is known (see [3]) that $f(n) \leq 2^{n^2/4}$. On the other hand, the following simple argument shows that $f(n) \geq 2^{n^2/8}$. Let $n = 4m$ and fix a partition V_1, V_2 of the vertex set into two parts of equal size. Let M be a maximal matching in V_2; $|M| = m$. Consider the family of $2^{|V_1|\cdot|M|} \geq 2^{2m^2} = 2^{n^2/8}$ graphs, each of which is obtained by joining every vertex in V_1 with *precisely one* endpoint of each of the edges in M. Add to each of these graphs all the edges from M. The obtained graphs are still triangle-free, and "maximal" in a sense that the addition of any new bipartite edge from $V_1 \times V_2$ creates a triangle. In each of these graphs draw edges between the vertices in V_1 and between the vertices in V_2 in an arbitrary way until the obtained graph becomes maximal triangle-free. Since each pair of the obtained graphs differ in at least one edge from $V_1 \times V_2$, we are done.

4.

In the context of this paper, the most interesting open question certainly is whether the lower bound $\mathsf{C}(K_3\text{-FREE}_n) = \Omega(n^{3/2})$ given in Theorem 3.2 is far from the optimum. Theorem 3.1 only says that $\mathsf{C}(\Delta\text{-FREE}_n) = \Theta(n^2)$ for *some* set Δ of triangles, and its proof fails if Δ is the set of *all* triangles. Triangle-free graphs have many interesting structural properties which (apparently) may help Carole to convince that the input graph has no triangles at all. In particular, by Mantel–Turán's theorem, in every triangle-free graph the neighborhoods of its vertices span at least $n^2/4$ non-edges. Hence, to improve the trivial upper bound $\mathsf{C}(K_3\text{-FREE}_n) \leq n^2$, one could try to encode a non-trivial fraction of non-edges using much fewer than n^2 bits. Let $D = D(n)$ be something much smaller than $\sqrt{n}$, say $D = n^\epsilon$ where $\epsilon < 1/2$ is a very small constant. In the case of λ-balanced partitions with $\lambda \to 0$ the argument used in the proof of Lemma 2.2(i) and Theorem 3.2 yields the lower bound $\mathsf{C}(K_3\text{-FREE}_n) = \Omega\left(\lambda^c n^{3/2}\right)$ where c is an absolute constant. Taking $\lambda = 1/D$ this yields $\mathsf{C}(K_3\text{-FREE}_n) = \Omega\left(n^{3/2}/D^c\right)$, which is close to $n^{3/2}$ if ϵ is sufficiently small.

Problem 5.1 Does $Dn^{3/2}$ bits are enough to encode n^2/D non-edges in every maximal triangle-free graph with more than $Dn^{3/2}$ edges?

It is interesting that in the case of t-cycle-free graphs with $t \geq 4$, a similar question has a positive answer in a very strong sense. For example, for $t \in \{4, 6, 8\}$ it is possible to encode *all* t-cycle-free graphs using only $O(n^{1+2/t})$ bits (see [9, 14]).

A positive solution of Problem 5.1 would imply $\mathsf{C}(K_3\text{-FREE}_n)$ is at most $O(Dn^{3/2}\log n)$. Indeed, triangle-free graphs G with at most $Dn^{3/2}$ edges are "for free:" Carole can announce to both players the entire graph using only $O(Dn^{3/2}\log n)$ bits. If G has more edges then Carole can use the partitions X_A, X_B with X_A's being the sets of encoded non-edges. Given a triangle-free graph G with more than $Dn^{3/2}$ edges, Carole can choose a maximal triangle-free graph containing G, and use the corresponding partition X_A, X_B. After that, Alice replies with Yes if and only if she does not see any edge of G, and Bob replies with Yes if and only if the subgraph he can see is triangle-free.

Appendix A

Here we prove Lemma 3.4 relating the size of an r-n.b.p. computing a boolean function f with the overlapping multi-partition communication complexity of f.

First we recall from [6] the following result stating that functions computed by nondeterministic syntactic read-r times programs (r-n.b.p.) can be represented in some special form. Let $X = \{x_1, \dots, x_n\}$ be the set of boolean variables. Say that a boolean function $g(X)$ is an (r,t)-*rectangle* if it can be represented in the form

$$g_1(X_1) \wedge g_2(X_2) \wedge \cdots \wedge g_m(X_m)$$

where $m \leq rt$; g_i is a boolean function depending only on variables from $X_i \subseteq X$, $|X_i| \leq \lceil n/t \rceil$ and each variable belongs to at least one and to at most r of the sets $X_1, \dots, X_m$.

Theorem 5.2 **(Borodin–Razborov–Smolensky [6])** *Let f be a Boolean function and r, t be positive integers. If f can be computed by an r-n.b.p. of size L then f is an OR of at most $(2L)^{2rt}$ (r,t)-rectangles.*

Our next goal is to replace (r,t)-rectangles in this theorem by the ANDs of *two* functions whose sets of variables do not overlap too much. Say that a boolean function $R(X)$ in n variables $X = \{x_1, \dots, x_n\}$ is a λ-*overlapping rectangle* if it can be represented in the form $R(X) = R_1(X_1) \wedge R_2(X_2)$, where R_i is a Boolean function depending only on variables from $X_i \subseteq X$ and $|X_i - X_{3-i}| \geq \lambda n$ for both $i = 1, 2$.

Lemma 5.3 *If a boolean function f can be computed by an r-n.b.p. of size L then f is an OR of at most $L^{O(r^2)}$ λ-overlapping rectangles with $\lambda = r^{-O(r)}$.*

Proof. Let $f(X)$ be a boolean function in n variables, and suppose that it can be computed by an r-n.b.p. of size L. Set $t := 3r$ and apply Theorem 5.2. This gives us a covering of f by $(2L)^{6r^2}$ (r,t)-rectangles $g = g_1(X_1) \wedge g_2(X_2) \wedge \cdots \wedge g_m(X_m)$ with $m \leq rt = 3r^2$. Hence, it is enough to show that each such rectangle g is also a λ-overlapping rectangle with $\lambda = r^{-O(r)}$.

For this purpose, define the *trace* of a variable x in the sequence of sets $X_1, \ldots, X_m$ as the set $T(x) = \{i : x \in X_i\}$. All what we need is to find a pair Y_1, Y_2 of subsets of variables which is *good* in the sense that: (i) both Y_1 and Y_2 have size at least λn, (ii) for both $i = 1, 2$ all the variables in Y_i have the same trace T_i, and (iii) $T_1 \cap T_2 = \emptyset$. Having such a pair Y_1, Y_2 we can obtain a λ-overlapping rectangle $R = R_1 \wedge R_2$ by taking R_j $(j = 1, 2)$ to be the AND of all functions $g_i(X_i)$ with $i \notin T_{3-j}$.

To show that a good pair Y_1, Y_2 exists, partite the set of variables according to their traces. Since we have only $K = \sum_{j=0}^{r} \binom{m}{j} \leq \sum_{j=0}^{r} \binom{3r^2}{j} \leq r^{O(r)}$ possible traces, the average size of one block in this partition is at least n/K. Hence at least $n/2$ of the variables belong to blocks of size at least $n/(2K) \geq \lambda n$, where $\lambda = 1/(2K) = r^{-O(r)}$. Let Y be the union of these (large) blocks; hence, $|Y| \geq n/2$. Take any of large blocks Y_1. By the definition, $|Y_1| \geq \lambda n$ and all the variables in Y_1 have the same trace T_1. The union of the corresponding to this trace subsets X_i can contain at most $r \cdot \lceil n/t \rceil < n/2 \leq |Y|$ different variables. Hence, there must be a variable $y \in Y$ which belongs to none of these subsets. This means that the trace $T_2 = T(y)$ of y is disjoint from the trace T_1 of the variables in Y_1. Let Y_2 be the (unique) large block containing y. Then $|Y_2| \geq \lambda n$ and all the variables in Y_2 have the same trace T_2 which is disjoint from the trace T_1 of the variables in the first block Y_1. Hence, Y_1, Y_2 is a good pair, and we are done. $\square$

Now we can finish the proof of Lemma 3.4 as follows. Suppose that a boolean function f can be computed by an r-n.b.p. of size L. Then, by Lemma 5.3, f can be written as an OR of $k = L^{O(r^2)}$ λ-overlapping rectangles. All three players can see this set of rectangles. Given an input $a \in f^{-1}(1)$, Carole can take the first of the rectangles $R = R_1(X_1) \wedge R_2(X_2)$ which accepts a, and announce that on this input Alice must compute R_1 and Bob must compute R_2. That is, Carole has only to announce the binary code of the overlapping rectangle to be used; the certificate W_a is always empty. This implies $\mathsf{C}_\lambda(f) \leq \lceil \log(k+1) \rceil = O(r^2 \log L)$, and hence, the size L on the r-n.b.p. for f is exponential in $\mathsf{C}_\lambda(f)/r^2$.

References

[1] Ajtai, M. (1999) A non-linear time lower bound for boolean branching programs. In: *Proc. of 40th Ann. IEEE Symp. on Foundations of Comput. Sci.*, pp. 60–70.

[2] Ajtai, M., Babai, L., Hajnal, P., Komlos, J., Pudlák, P., Rödl, V., Szemeredi, E. and Turán, Gy. (1986) Two lower bounds for branching programs. In *Proc. of 18th Ann. ACM Symp. on the Theory of Computing*, pp. 30–38.

[3] Barefoot, C., Casey, K., Fisher, D., Fraughnaugh, K. and Harary, F. (1995) Size in maximal triangle-free graphs and minimal graphs of diameter 2. *Discrete Mathematics* **138**, no. 1-3, 93–99.

[4] Beame, P.W., Saks, M. and Thathachar, J.S. (1998) Time-space tradeoffs for branching programs. In *Proc. of 39th Ann. IEEE Symp. on Foundations of Comput. Sci.*, pp. 254–263. To appear in: *J. Comput. Syst. Sci.*

[5] Beame, P., Saks, M., Sun, X. and Vee, E. (2000) Super-linear time-space tradeoff lower bounds for randomized computation. In *Proc. of 41st Ann. IEEE Symp. on Foundations of Comput. Sci.*, 169–179.

[6] Borodin, A., Razborov, A. and R. Smolensky (1993) On lower bounds for read-k times branching programs. *Computational Complexity* **3** 1–18.

[7] Chung F.R.K. (1997) Open problems of Paul Erdős in graph theory. *Journal of Graph Theory* **25**:1 3–36.

[8] Ďuriš, P., Hromkovič, J., Jukna, S., Sauerfhoff, M. and Schnitger, G. (2001) On multipartition communication complexity. In *Proc. of 18th Int. Symp. on Theoretical Aspects of Computer Science* vol. 2010 of *Springer Lecture Notes in Computer Science*, pp. 206–217.

[9] Kleitman, D.J. and Winston, K.J. (1982) On the number of graphs without 4-cycles. *Discrete Mathematics* **41** 167–172.

[10] Kushilevitz, E. and Nisan, N. (1997) *Communication Complexity.* Cambridge University Press.

[11] Papadimitriou Ch. H. and Sipser M. (1982) Communication complexity. In *Proc. of 14th Ann. ACM Symp. on the Theory of Computing*, pp. 196–200. Journal version: Communication complexity, *J. Comput. Syst. Sci.*, **28**:2

[12] Simonovits, M. (1996) Paul Erdős' influence on extremal graph theory. In: *The Mathematics of Paul Erdős*, J. Nešetřil and R. Graham (eds.), Springer-Verlag, pp. 148–192.

[13] Wegener, I. (2000) *Branching programs and Binary Decision Diagrams: Theory and Applications.* SIAM Series in Discrete Mathematics and Applications.

[14] Wilson, D.B. and Kleitman, D.J. On the number of graphs which lack small cycles. *Discrete Mathematics* (to appear).