Acta Informatica 25, 515–535 (1988)

124

Acta Informatica


124

# Graph Complexity ⋆

Pavel Pudlák [1], Vojtěch Rödl [2], and Petr Savický [3]

[1] Mathematical Institute ČSAV, Žitná 25, Praha 1, Czechoslovakia

[2] Department of Mathematics, FJFI ČVUT, Husova 5, Praha 1, Czechoslovakia

[3] Faculty of Mathematics and Physics, Charles University, Malostranské nám. 25, Praha 1, Czechoslovakia

**Summary.** We develop a complexity theory based on the concept of the graph instead of the Boolean function. We show its relation to the Boolean complexity and prove some lower bounds to the complexity of explicitly given graphs.

## 0. Introduction

The theory of Boolean functions deals mainly with two basic complexity measures: the formula size and the circuit size. The complexity of a Boolean function $f$ is the minimal size of a formula resp. circuit needed to compute $f$. We can consider the computation to be performed on the algebra of all Boolean functions (with a given number of variables) i.e. we have some initial functions and generate new functions using Boolean operations. The initial functions, let us call them *generators*, are just the projections on coordinates.

Given a formula or a circuit with variables $x_1, \ldots, x_n$ we can evaluate it on any Boolean algebra. We only need to assign some elements of the Boolean algebra to the variables $x_1, \ldots, x_n$, put otherwise we have to choose the generators. We consider two such Boolean algebras: (1) subgraphs of a complete graph, (2) subgraphs of a complete bipartite graph.

Our reason for studying the Boolean complexity of graphs instead of Boolean functions is that graphs have simpler structure and have been extensively studied for a long time. However we consider the graph complexity only as a tool for proving lower bounds for Boolean functions. Therefore we introduce systems of generators (stars in the first case, complete bipartite graphs in the second) so that lower bounds on the complexity of graphs could be transferred to the lower bounds on the complexity of some Boolean functions associated with them. We have not considered other sets of generators, but they might be interesting too.

⋆ The paper was written while the first author was visiting Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago

We would like to find some properties of graphs which imply high complexity. We prove two theorems in this direction (2.4, 2.5), which show that such nontrivial lower bounds for graph complexity are possible. However both theorems use techniques known in Boolean complexity and give only bounds which are too small, hence do not imply nontrivial lower bounds for the complexity of Boolean functions. Theorem 2.5 uses a condition which is very similar to the conditions in Ramsey theory. We had hoped that if the parameters of this condition were close to the parameters of a random graph, the complexity must be quite large, since no explicit constructions of such graphs are known. Surprisingly, it has turned out that there are such graphs which have formula size complexity only $O(n \log n)$ and linear circuit size complexity (Corollaries 5.4 and 5.8).

It is an open problem whether it is possible to construct a "Ramsey graph" which is a graph without large cliques and large independent sets, say, using a deterministic Turing machine and polynomial time. When defining the graphs mentioned above we shall also show the existence of a Ramsey graph with the same upper bounds to its complexity. Our proof uses probabilistic methods, but it is in a sense more explicit than the classical existence proof of P. Erdös. We shall show that these graphs occur as relatively large subgraphs of the graphs associated with Hadamard matrices. The question of defining Ramsey graphs and related combinatorial objects by small formulas was extensively studied by A.A. Razborov in his recent paper [8].

Questions related to the graph complexity as presented here have been considered in the literature. Especially, it has been well known that in communication complexity one works, in fact, with bipartite graphs instead of Boolean functions. For instance, deterministic communication complexity has a simple graph-theoretical characterization [15]. The graph complexity based on stars as generators has been considered by Bublitz [4].

## 1. Basic Definitions

We consider the following general situation. Let $V$ be a set of vertices, $\mathcal{G} \subseteq \mathcal{P}([V]^2)$ a set of graphs on $V$ and $B$ a set of Boolean (set) operations. Given a graph $H$ on $V$ we ask how difficult it is to express $H$ using graphs $\mathcal{G}$ and operations from $B$. We shall consider two basic ways of expressing $H$ using $\mathcal{G}$ and $B$, namely, formulas and circuits.

In the sequel the size of $V$ will be denoted $n$, if not stated otherwise; the graphs in $\mathcal{G}$ will be called *generators*; $B$ will be called the *basis*.

We shall consider also bipartite graphs. In this case $V = U \cup W$, $|U| = |W| = \frac{n}{2}$, ($n$ even), $\mathcal{G} \subseteq \mathcal{P}(U \times W)$.

We shall use two sets of generators, one for arbitrary graphs, one for the bipartite graphs described above. Both sets are motivated by the goal of obtaining lower bounds to Boolean complexity. However, one can consider other sets of generators. A natural requirement is that $\mathcal{G}$ should be closed under isomorphisms (in the case of bipartite graphs, under isomorphisms preserving $U$ and $W$).

We shall only use bases that are complete for the given set of generators, which means that any graph (bipartite graph resp.) can be expressed. Since the completeness of the basis depends on the set of generators, it is not surprising that this concept is different from the concept of completeness used in the theory of Boolean functions.

*Formulas.* In order to simplify the definition, suppose the basis $B$ contains only unary and commutative binary operations. A *formula* with variables $x_1, \ldots, x_m$ and basis $B$ is a binary (i.e., fan-in $\leqq 2$) rooted tree with leaves labelled by variables and other vertices labelled by operations from $B$ respecting the arity of operations. The *size of a formula* is the number of leaves or equivalently the number of occurrences of variables. Given a mapping from the set of variables $\{x_1, \ldots, x_m\}$ into the set of generators $\mathscr{G}$, a formula determines a graph (perform the operations on vertices and take the graph which appears on the root). Sometimes it is more convenient to assume that the leaves are just labelled by generators. The *formula size complexity* of a graph $H$, denoted by $L(H)$, is the smallest $k$ such that $H$ can be computed by a formula of size $k$ (assuming $\mathscr{G}$ and $B$ is fixed).

*Circuit.* A *circuit* with variables $x_1, \ldots, x_m$ in basis $B$ is a sequence of equations $e_1, \ldots, e_k$ where, for $i=1, \ldots, k$, $e_i$ has the form

$$y_i = x_j$$

or

$$y_i = \alpha(z)$$

or

$$y_i = \alpha(z, t)$$

where

$$\alpha \in B \quad \text{and} \quad z, t \in \{x_1, \ldots, x_m, y_1, \ldots, y_{i-1}\}.$$

Given a mapping $\{x_1, \ldots, x_m\} \to \mathscr{G}$ the circuit determines a graph (computed at $y_k$). The size of the circuit is $k$. The *circuit size complexity* of a graph $H$, denoted by $C(H)$ is the smallest $k$ such that $H$ can be computed using a circuit of size $k$ (assuming $\mathscr{G}$ and $B$ is fixed).

We shall use standard graph theory notation. In particular $K_p$ is a complete graph on $p$ vertices, $K_{p,q}$ a complete bipartite graph with blocks of size $p$ and $q$, $E_p$ an empty graph on $p$ vertices.

## 2. Star Complexity

Let $V$ be the set of vertices $|V|=n$. Let $\mathscr{S} = \{G \subseteq [V]^2 \mid G \cong K_{1,n-1}\}$ be the set of generators. The associated formula and circuit complexity will be called *star complexity*. In this section we assume that the basis $B$ is any basis containing $\cap$ and $\cup$, if not stated otherwise.

We shall explain how this complexity is related to the Boolean complexity. Let $f$: $\{0, 1\}^n \to \{0, 1\}$ be a Boolean function. Each $a \in \{0, 1\}^n$ can be associated with a subset $X_a \subseteq \{1, \ldots, n\}$ by defining

$$i \in X_a \leftrightarrow a_i = 1.$$

Then $f$ corresponds to the set system $\Sigma_f \subseteq \mathscr{P}(\{1, \ldots, n\})$ defined by

$$\Sigma_f = \{X_a: f(a) = 1\}.$$

It is obvious that Boolean operations on Boolean functions correspond to particular set operations on set systems; e.g., the conjunction $\wedge$ corresponds to the intersection $\cap$. The initial functions, the "generators", are the *projection functions*

$$f(x_1, \ldots, x_n) = x_i,$$

$i = 1, \ldots, n$. The set system corresponding to such a function is a "star"

$$\Sigma_i = \{X \subseteq \{1, \ldots, n\} \mid i \in X\}.$$

Now, given a formula or a circuit we can transform all the set systems occurring in such a computation into graphs just by intersecting them with set of all two element subsets of $\{1, \ldots, n\}$. Clearly, the initial graphs, the generators,

$$\Sigma_i \cap [\{1, \ldots, n\}]^2$$

will be isomorphic to $K_{1,n-1}$. Thus we can think of this kind of graph complexity as Boolean complexity where we use only a special part of each Boolean function. As a consequence we have that a lower bound to the circuit resp. formula size complexity of a graph $H$ is a lower bound to the circuit resp. formula size of any Boolean function $f$ which extends $H$, i.e.,

$$\Sigma_f \cap [\{1, \ldots, n\}]^2 = H.$$

We shall show a better trade-off in the next section.

First we shall estimate the maximal star complexity of graphs on $n$ vertices. This is a result of Bublitz [4].

**Theorem 2.1.** *For both circuit and formula size complexity the maximal complexity of graphs on n vertices is at least* $\Omega\left(\frac{n^2}{\log n}\right)$ *and at most* $O\left(\frac{n^2}{\log n}\right)$.

The lower bound follows by the standard counting argument. For the upper bound we need a lemma.

**Lemma 2.2.** $L(K_{p,q} \cup E_{n-p-q}) \leq p + q$.

*Proof.* Let $G$ be such a graph with sets of vertices $A$, $B$, $|A| = p$, $|B| = q$ on which it is a complete bipartite graph. Let $S_i$ be the star with the root $i$. Then

$$G = (\bigcup_{i \in A} S_i) \cap (\bigcup_{i \in B} S_i),$$

which proves the inequality. □

The idea of the proof of the upper bound in Theorem 2.1 is the following. Let $G$ be an arbitrary graph on $n$ vertices. We cover the edges of $G$ by complete bipartite graphs $G_1, \ldots, G_l$ so that, for $i = 1, \ldots, l$, $G_i$ is the largest subgraph of $H_i = G - \bigcup_{j<i} G_j$ which is of the form $K_{p,p} \cup E_{n-2p}$. If $H_i$ has $m$ edges and $t$ satisfies

$$n\binom{m/n}{t} \geqq t \cdot \binom{n}{t},$$

then there is such a graph with $p \geqq t$, see Erdös, Spencer [5]. Using the estimate and the lemma above one can compute that the size of the formula for $G = \bigcup_{i=1}^{l} G_i$ will be $O\left(\frac{n^2}{\log n}\right)$. Independently and in purely graph theoretical terms such an estimate was proved by Tuza [9]. □

*Lower Bounds*

**Theorem 2.3.** *For the basis* $B = \{\cap, \cup\}$

$$L(K_{n-1} \cup E_1) \geqq (n-1) \cdot \log_2(n-1) - 1.$$

*Proof.* Let $G$ be the graph, $V = V' \cup \{v\}$ the vertex set, where $V'$ is the $n-1$ element clique of $G$.

*Claim 1.* Any minimal formula for $G$ does not use $S_v$, the star with the root $v$. Suppose some formula does use $S_v$. If we replace the occurrences of $S_v$ by the empty graph we obtain some $G' \subseteq G$, because the operations are monotone. On the other hand no edges can be missing in $G'$, since $S_v$ does not contain any edges from $G$. Now the empty graphs are obviously superfluous and we can reduce the size of the formula by eliminating them; which proves the claim.

*Claim 2.* If $\varphi$ is a formula computing $G$ which does not use $S_v$, then $\varphi$ is a formula for the threshold function $T_2^{n-1}: \{0, 1\}^{n-1} \to \{0, 1\}$ such that

$$T_2^{n-1}(a) = 1 \leftrightarrow \Sigma a(i) \geqq 2.$$

So $\varphi(x_1, \ldots, x_{n-1})$ is a formula in which we substitute different $S_i$'s, $i \in V'$ for different $x_i$'s, in order to compute $G$. Since $\varphi$ must output 1 for every edge of the clique, we have $\varphi(a) = 1$ whenever $\Sigma a(i) = 2$. By monotonicity this must be true also for $\Sigma a(i) > 2$. If $\varphi(a) = 1$ for some $a$ such that $\Sigma a(i) = 1$ then $\varphi$ would output 1 for some edge $(u, v)$, $u \in V'$, which is not possible. By monotonicity we have $\varphi(\bar{0}) = 0$ too.

To prove the theorem it is now sufficient to apply a result of Kričevskij, see [9], which says that the formula size of $T_2^{n-1}$ even in a larger basis $\{\vee, \wedge, \neg\}$ is at least $(n-1)\log_2(n-1) - 1$. □

Observe that $K_{n-1} \cup E_1$ is the complement of a star. Hence the theorem shows that the complexity of the complement may be substantially larger than the complexity of the graph itself.

The next theorem uses a technique similar to the well-known Nečiporuk lower bound.

**Theorem 2.4.** *There exists $\varepsilon>0$ such that for any graph $G$ on $V$, $|V|=n$, if for every $X\subseteq V$, $|X|\geqq\frac{n}{2}$ there exist $K$ nonisomorphic induced graphs on $G|X$ each with $l$ vertices, then*

$$L(G)\geqq\varepsilon\cdot n\cdot\frac{\log K}{l\cdot\log l}.$$

*Remark.* We can assume that the vertices $V$ are linearly ordered and that two induced subgraphs are isomorphic only if there is an isomorphism which preserves the ordering. Thus we can obtain $K$ slightly larger.

*Proof.* Let $\varphi$ be a formula for $G$, let $n\cdot\gamma$ be its size. There are at least $\frac{n}{2}$ variables which occur in $\varphi$ at most $2\gamma$ times. Let $X$ be the corresponding set of vertices. To each $l$ element subset $Y$ of $X$ there corresponds an $l$ element subset of variables of $\varphi$ such that if we substitute zeros into the other variables the resulting formula computes the subgraph induced by $Y$. Thus for each such an induced subgraph we get a formula such that each variable occurs in it at most $2\gamma$ times. W.l.o.g. assume that the basis $B$ contains all at most binary operations. Then each formula can be reduced either to a constant or to a formula not containing constants. The number of such formulas with $l$ variables and size $\leqq 2\gamma l$ can be easily estimated by

$$\leqq(c\cdot l)^{2\gamma l}$$

where $c$ is a suitable constant. This number must be larger or equal to the number of nonisomorphic graphs induced by $l$ element subsets of $X$, hence we get

$$K\leqq(c\cdot l)^{2\gamma l}\Rightarrow\gamma\geqq\frac{\log K}{2l(\log l+\log c)}.$$

In the estimate above we did not identify the formulas which are identical only after renaming variables. Hence the theorem is true also if we use ordered graphs as stated in the remark above. □

*Remarks.* 1. An easy computation shows that the largest bound which can be obtained using the above theorem cannot exceed

$$c\cdot n\frac{\log n}{\log\log n},\qquad c \text{ a constant.}$$

2. An easy counterexample shows that in the theorem it is not sufficient to count the number of $l$ elements induced subgraphs *in the whole* graph $G$. Let $G'$ have $K$ nonisomorphic $l$ element induced subgraphs and suppose $G'$ has $\frac{1}{2}\sqrt{n}$ vertices. Let $G$ be $G'$ plus $n-\frac{1}{2}\sqrt{n}$ isolated vertices. Then $L(G)\leqq n$. If the theorem could be strengthened, then $G$ should have complexity

$$\geqq\varepsilon\cdot n\cdot\frac{\log K}{l\cdot\log l}.$$

But we shall see later, that for suitable graphs the factor

$$\frac{\log K}{l \cdot \log l}$$

is not constant.

**Theorem 2.5.** *There exists an* $\varepsilon > 0$ *such that for every graph* $G$ *on* $V$, $|V| = n$,

$$L(G) \geqq \varepsilon \cdot n \cdot \log \frac{n}{k},$$

*where* $k$ *is maximal such that, for some* $A, B \subseteq V$ *disjoint,* $k = |A| = |B|$, $G$ *is monochromatic on* $A \times B$ *(i.e., either contains each edge* $(u, v)$, $u \in A$, $v \in B$ *or none).*

*Proof.* To prove the theorem we shall use one of the results of [2], Theorem 1. This theorem says that functions computed by small formulas are, in a sense, locally simple. Let $G$ be a graph computed by a formula $\varphi$ of size $n \cdot \gamma$. Let $f$ be the Boolean function computed by $\varphi$. Thus $G$ is determined by the restriction of $f$ to the inputs with exactly two l's. We can identify the variables of $f$ with the vertices of $G$. Now we shall apply the result mentioned above to $f$. By this theorem there exist disjoint subsets of variables $A_0$ and $B_0$, $|A_0| = |B_0| \geqq nb^{-\gamma}$ such that $f | A_0 \cup B_0$ ($f$ restricted to variables of $A_0 \cup B_0$ with other variables set to 0), satisfies the following property called $\lceil a\gamma \rceil$ simple in [2], ($a$, $b$ are constants): There are partitions

$$X_1 \cup \ldots \cup X_{l_1} = \{0, 1\}^{A_0}$$
$$Y_1 \cup \ldots \cup Y_{l_2} = \{0, 1\}^{B_0},$$

with $l_1, l_2 \leqq \lceil a\gamma \rceil$ such that the function $f | A_0 \cup B_0$ is constant on each set

$$X_i \times Y_j, \quad i = 1, \ldots, l_1, \ j = 1, \ldots, l_2.$$

Denote by

$$\bar{X}_i = \{a \in X_i \,|\, \Sigma\, a(j) = 1\},$$
$$\bar{Y}_i = \{a \in Y_i \,|\, \Sigma\, a(j) = 1\}.$$

Let $\bar{X}_i$ be a largest one among $\bar{X}_1, \ldots, \bar{X}_{l_1}$ and $\bar{Y}_j$ be a largest one among $\bar{Y}_1, \ldots, \bar{Y}_{l_2}$. The cardinality of $\bar{X}_i$ is at least

$$\frac{|A_0|}{\lceil a\gamma \rceil} \geqq \frac{n}{b^\gamma \lceil a\gamma \rceil} \geqq \frac{n}{c^\gamma}$$

for a suitable constant $c$, and the same bound holds for $\bar{Y}_j$. The sets $\bar{X}_i$ and $\bar{Y}_j$ correspond to subsets $A \subseteq A_0$ and $B \subseteq B_0$. It is not hard to see that the graph computed by $\varphi$ is monochromatic on $A \times B$. Hence

$$k \geqq \min(|A|, |B|) = \min(|\bar{X}_i|, |\bar{Y}_i|) \geqq \frac{n}{c^\gamma},$$

from which we get $\gamma \geqq \log_c \frac{n}{k}$. $\square$

*Remarks.* 1. The largest bound obtained from this theorem cannot exceed $c \cdot n \log n$, $c$ constant.

2. A slightly stronger theorem can be proved. Namely, if $V = U \cup W$, $|U| = |W|$, it suffices to maximize $k$ over subsets $A \subseteq U$ and $B \subseteq W$.

*Circuit Complexity*

The only lower bound which we know of is due to Jiři Sgall, [10]

$$C_{\{\cap,\cup\}}(K_{n-1} \cup E_1) \geqq 2n - O(1).$$

An interesting fact shown by Wegener [10] (also for other "slices" of Boolean functions) is that the circuit complexity in the monotone basis $\{\cap, \cup\}$ can be at most linearly larger than the complexity in the basis $\{\cap, \cup, \neg\}$. An explicit calculation gives

$$C_{\{\cap,\cup\}} \leqq 2 \cdot C_{\{\cap,\cup,\neg\}}(G) + 9n - 23$$

for graphs with $n$ vertices, $n \geqq 3$.

*Bounded Depth Formulas*

A $\cup\cap$-formula is a formula in basis $\{\cup, \cap\}$ in which all intersections precede unions. The dual concept is an $\cap\cup$-formula. The corresponding measures will be denoted by $L_{\cup\cap}$ and $L_{\cap\cup}$.

The measure $L_{\cup\cap}$ is trivial since $L_{\cup\cap}(G)$ equals the number of stars contained in $G$ plus two times the number of remaining edges.

The measure $L_{\cap\cup}$ can be characterized as follows.

$$L_{\cap\cup}(G) = \min \{\sum_{i \in I} (n - |X_i|) \mid \{X_i\}_{i \in I}, \text{ a system of independent sets of } G \text{ which cover the complement of } G\}$$

where $n$ is the number of vertices. That is $\{X_i\}_{i \in I}$ is a system of cliques for the complement of $G$ which covers all the edges of the complement of $G$. Such systems have been studied, see Berge [3]. The following is an easy corollary of a well-known theorem of Erdös, Goodmann and Pósa, see [3]

**Corollary 2.6.** *For every graph $G$ on $n$ vertices*

$$L_{\cap\cup}(G) \leqq \left\lceil \frac{n^2}{4} \right\rceil (n-2),$$

*where we have equality for*

$$G \cong K_{\lfloor \frac{n}{2} \rfloor} \cup K_{\lceil \frac{n}{2} \rceil}.$$

Thus the maximal complexity is attained by very simple graphs. Another characterization of this measure is the following. The representation graph of a hypergraph $(X, \{X_i\}_{i\in I})$, $X_i \subseteq X$, is the graph with vertices $I$ and $(i, j)$ is an edge if and only if $X_i \cap X_j \neq 0$.

$$L_{\cap\cup}(G) = \min\{\sum_{i\in I} |X - X_i| \,|\, \text{the representation graph of } (X, \{X_i\}_{i\in I}) \text{ is isomorphic to } G\}.$$

Things become much more complicated if we allow more alternations. For the measure $L_{\cup\cap\cup}$ we know that almost all graphs have $L(G)$ close to $L_{\cup\cap\cup}(G)$. This follows from the proof of Theorem 2.1, where we actually showed that $L_{\cap\cup\cap}(G) = O\left(\frac{n^2}{\log n}\right)$ for graphs on $n$ vertices. It seems very unlikely that for every graph $L_{\cup\cap\cup}(G)$ is close to $L(G)$, but we cannot disprove it. We can only prove that for some graphs the representation from Theorem 2.1 (covering by complete bipartite graphs which is a special kind of $\cup\cap\cup$ formulas) is much worse than representation by an unrestricted formula, cf. Sect. 5, Remark 3.

It is possible that in some special cases $L_{\cup\cap\cup}$ can be estimated easily. This would be the case if e.g., the following problem had an affirmative answer.

*Problem*

Is for every $G$ which does not contain a star or $K_{2,2}$ $L_{\cup\cap\cup}(G) \geqq$ the number of edges in $G$? (This is equivalent to the question if the same holds for $L_{\cap\cup}(G)$.)

## 3. Bipartite Complexity

In this section we assume that the set of vertices is $V = U \cup W$, $|U| = |W| = \frac{n}{2}$, $n$ even. The set of generators will be

$$\mathscr{B} = \{A \times W \,|\, A \subseteq U\} \cup \{U \times B \,|\, B \subseteq W\}.$$

The associated formula and circuit complexity will be called *bipartite complexity*. Again we assume that the basis contains at least $\cap$ and $\cup$. Thus the set of generators and the basis is complete w.r.t. bipartite graphs on $U$ and $W$. As we shall compare the two graph complexities, we shall use subscripts $\mathscr{S}$ and $\mathscr{B}$ to distinguish them.

Let us consider the relation to the Boolean complexity. Let $f$: $\{0, 1\}^m \to \{0, 1\}$ be a Boolean function, let $m$ be even. Then we can write $f$: $\{0, 1\}^{\frac{m}{2}} \times \{0, 1\}^{\frac{m}{2}} \to \{0, 1\}$, which shows that $f$ is essentially a $2^{\frac{m}{2}} \times 2^{\frac{m}{2}}$ 0,1-matrix or a bipartite graph with blocks $U = \{0, 1\}^{\frac{m}{2}}$ and $W = \{0, 1\}^{\frac{m}{2}}$. Now, as in the case of the star complexity, Boolean operations correspond to set operations on graphs. Thus it only remains to find out what the generators are, i.e., graphs corresponding to the projection functions. These are easily seen to be graphs isomorphic to $K_{2^{\frac{m}{2}-1},2^{\frac{m}{2}}}$ and $K_{2^{\frac{m}{2}},2^{\frac{m}{2}-1}}$. The system of graphs which correspond to the projection functions is very special, in particular it contains only $m$ graphs, while $|\mathscr{B}|=2^{\frac{n}{2}+1}=2^{2^{\frac{m}{2}+1}}$. As a result the measure determined by such a system is not invariant under isomorphisms. E.g., it is possible to show that there is a matching (between $U$ and $W$) which has complexity linear in $m$ and there is another one which has complexity exponential in $m$ (using a counting argument). Therefore we prefer a larger system such as $\mathscr{B}$. The motivation for the particular choice of $\mathscr{B}$ is in communication complexity. The basic idea of communication complexity is that there are two "computers", each having access to half of the input and each one can compute arbitrary function on its part of the input just for a unit cost. In graphs it means that the graphs $A\times W$ and $U\times B$ have complexity one, for arbitrary $A\subseteq U$ and $B\subseteq W$.

Since the number of vertices of the graph corresponding to a Boolean function is exponential in the number of variables of the function, we get the following proposition.

**Proposition 3.1.** *A lower bound $F(\log n)$ for bipartite complexity of a graph on $n$ vertices, with $n=2^{\frac{m}{2}+1}$, $m$ even, gives a lower bound $F\left(\frac{m}{2}+1\right)$ for an $m$ variable Boolean function. This is true for any system $\mathscr{G}$ which contains the graphs corresponding to the projection functions (and formula size as well as circuit size complexity).* □

We shall show that suitable lower bounds to the star complexity may imply exponential lower bounds to the Boolean complexity too. This is because of the following relation between the star complexity and the bipartite complexity.

**Proposition 3.2.** *Let $G$ be a bipartite graph with blocks $U$, $W$, $|U|=|W|=\frac{n}{2}$. Then*

$$L_{\mathscr{S}}(G)\leqq L_{\mathscr{B}}(G)\cdot\frac{n}{2}+n,$$

*where on both sides we take the same basis containing $\cap$, $\cup$. Moreover the same inequality holds for $C_{\mathscr{S}}$ and $C_{\mathscr{B}}$.*

*Proof.* Let $\varphi$ be a formula computing $G$ which uses the generators $\mathscr{B}$. Replace each graph $A\times W$ resp. $U\times B$ by

$$\bigcup_{i\in A} S_i \text{ resp. } \bigcup_{i\in B} S_i,$$

where $S_i$ is a star with root $i$. Let $\psi_1$ be the resulting formula. Let $\psi_2$ be a formula computing the complete bipartite graph $U\times W$ using the stars. Then $\psi_1\cap\psi_2$ is a formula for $G$. The size of $\psi_1$ is $\leqq$ size of $\varphi$ times $\frac{n}{2}$. The size of $\psi_2$ is $n$ (Lemma 2.2). The same construction works with circuits. □

For circuits we can obtain a better bound if we consider the measure $C_{\mathscr{G}}$ for some $\mathscr{G}\subseteq\mathscr{B}$ small. The proof is exactly the same.

**Proposition 3.3.** *Let $\mathscr{G} \subseteq \mathscr{B}$, let G be a graph as in Proposition* 3.2. *Then*

$$C_{\mathscr{S}}(G) \leqq C_{\mathscr{G}}(G) + |\mathscr{G}| \cdot \frac{n}{2} + n. \quad \square$$

**Corollary 3.4.** *1. A lower bound $\Omega(n \cdot F(\log n))$ for $L_{\mathscr{S}}(G)$, where G is as in Proposition* 3.2, *gives a lower bound $(F(m))$ to the formula size of an m variable Boolean function.*

*2. A lower bound $\alpha n \log_2 n$, with $\alpha > 1$, for $C_{\mathscr{S}}(G)$, G as above, gives an exponential lower bound to the circuit size of an m variable Boolean function.*

*Proof.* 1. Follows from Proposition 3.2.

2. Let $\mathscr{G}$ be the set of bipartite graphs corresponding to the projection functions. Then $|\mathscr{G}| = m = 2 \log n - 2$. Hence by Proposition 3.3

$$\begin{aligned} C_{\mathscr{G}}(G) &\geqq C_{\mathscr{S}}(G) - \frac{n}{2}(2 \log n - 2) - n \\ &\geqq \alpha n \log n - n \log n \\ &= (\alpha - 1)\, 2^{\log n} \log n. \end{aligned}$$

By Proposition 3.1 this gives a bound

$$(\alpha - 1) \cdot 2^{\frac{m}{2}+1} \cdot \left(\frac{m}{2} + 1\right) = (\alpha - 1) \cdot 2^{\frac{m}{2}} \cdot (m + 2)$$

for an $m$ variable Boolean function. $\square$

For general graphs we have only a slightly weaker relation because of the following proposition which reduces the complexity of a general graph to the complexity of a bipartite graph.

**Proposition 3.5.** *Let G be a graph on V, $|V| = n$, n even. Then there exists a partition $V = U \cup W$ into two equal size blocks such that*

$$L_{\mathscr{S}}(G \cap U \times W) \geqq \frac{L_{\mathscr{S}}(G)}{\lceil \log_2 n \rceil};$$

$$C_{\mathscr{S}}(G \cap U \times W) \geqq \frac{C_{\mathscr{S}}(G)}{\lceil \log_2 n \rceil} - 1 + \frac{1}{\lceil \log_2 n \rceil}.$$

*Proof.* Clearly the first inequality is equivalent to

$$\max_{U, W} L_{\mathscr{S}}(G \cap U \times W) \cdot \lceil \log_2 n \rceil \geqq L_{\mathscr{S}(\mathscr{G})},$$

where we maximize over all partitions into equal size blocks $U$, $W$. The latter inequality is a consequence of the fact that $G$ is the union $\lceil \log_2 n \rceil$ graphs of the form $G \cap U \times W$. The proof for $C_{\mathscr{S}}$ is similar. $\square$

*Maximal Complexity*

Let $C^n_{\mathscr{B}}$ resp. $L^n_{\mathscr{B}}$ denote the maximal circuit resp. formula size complexity of graphs on $n$ vertices.

**Proposition 3.6.**

$$\frac{n}{2}-\log_2 n-O(1)\leqq C^n_{\mathscr{B}},\quad \frac{n}{2}-O(1)\leqq L^n_{\mathscr{B}},\quad C^n_{\mathscr{B}}\leqq L^n_{\mathscr{B}}\leqq n.$$

*Proof.* 1. Lower bounds. There are $2^{\frac{n^2}{4}}$ bipartite graphs and there are at most

$$\binom{|\mathscr{B}|}{k+1}(\alpha_1\, k^2)^k\leqq 2^{(\frac{n}{2}+1)k}(\alpha_1\, k^2)^k,\qquad \alpha_1 \text{ constant},$$

circuits of size $\leqq k$, since a circuit of size $k$ can have at most $k+1$ inputs. There are at most

$$(\alpha_2\cdot|\mathscr{B}|)^k=(\alpha_2\cdot 2^{\frac{n}{2}+1})^k$$

formulas of size $\leqq k$, $\alpha_1$, $\alpha_2$ constants. Comparing the number of graphs with the number of circuits and the number of formulas we get the lower bounds.

2. Upper bound. Let $G\subseteq U\times W$. Then

$$G=\bigcup_{i\in U} G_i,$$

where

$$G_i=\{(i,j)\,|\,(i,j)\in G\}=\{(i,j)\,|\,j\in W\}\cap\{(l,j)\,|\,l\in U,(i,j)\in G\}.$$

Hence $L_{\mathscr{B}}(G)\leqq 2|U|=n$. □

*Lower bounds.* The only lower bound that we know of is the following trivial one.

**Proposition 3.7.** *Let $G\subseteq U\times W$ be a graph computed by a formula or a circuit which uses only $l$ generators from $\mathscr{B}$ (this is true in particular if $L_{\mathscr{B}}(G)\leqq l$ or $C_{\mathscr{B}}(G)\leqq l-1$). Then there exist partitions $U=U_1\cup\ldots\cup U_{l_1}$, $W=W_1\cup\ldots\cup W_{l_2}$, $l_1, l_2\leqq 2^l$, such that $G$ is the union of some complete bipartite graphs $U_i\times W_j$.* □

*Remarks.* 1. The largest bound that one can obtain here is only $\log_2 n-1$. There are simple graphs for which this value is attained:

(a) matching;

(b) $G\subseteq U\times W$ where

$$U=\left\{-\frac{n}{2},\ldots,-1\right\},\qquad W=\left\{1,\ldots,\frac{n}{2}\right\},$$

$(i,j)\in G$ iff $i+j\geqq 0$.

2. This simple idea has been used implicitly many times. Weaker conditions have been used often, e.g.,

(a) there exists a partition

$$U = U_1 \cup \ldots \cup U_{l_1}, \quad l_1 \leqq 2^l$$

such that $G$ is a union of some complete bipartite graphs $U_i \times B$, $B \subseteq W$ *arbitrary*; implicit in Nečiporuk [7], see also [9];

(b) $G$ is the disjoint union of $\leqq 2^l$ complete bipartite graphs $A \times B$, $A \subseteq U$, $B \subseteq W$; see Yao [12];

(c) there exist sets $A$, $B$, $A \subseteq U$, $B \subseteq W$ such that

$$|A| = |B| \geqq \frac{n}{2^{l+1}}$$

and $G$ is monochromatic on $A \times B$.

Bipartite complexity is very much related to some recent research in communication complexity. We should at least briefly comment on this connection. The classical concept of deterministic communication complexity is essentially bipartite complexity with formulas, respectively circuits replaced by *decision trees*, and where the measure is the *depth* of the tree. This model is much weaker even than formulas, hence large lower bounds for this complexity *do not* imply significant lower bound for formula size and circuit size of Boolean functions. In a recent paper [1], analogues of $\Sigma_i$ and $\Pi_i$ classes of the polynomial time hierarchy and PSPACE have been defined for communication complexity. The classes $\Sigma_i$ and $\Pi_i$ correspond to the graphs whose bipartite bounded depth formula size complexity is $\leqq 2^{(\log n)^c}$, $c > 0$ constant (with just a small proviso). The analogue of PSPACE corresponds to the graphs with $L_{\mathscr{B}}(G) \leqq 2^{(\log n)^c}$ (without any restriction to the depth of formulas). We are indebted to G. Turán for pointing out the connection between [1] and the present paper.

## 4. Hadamard Matrices

We shall use the graphs associated with Hadamard matrices to determine the limits for our lower bound theorems. We shall use results proved in [6]. In this section we shall investigate *combinatorial* properties of these graphs.

**Definition.** 1. A *Hadamard matrix* is an $n \times n$ matrix $M$ with entries $\pm 1$ such that

$$M \cdot M^t = n \cdot I_n,$$

i.e., each two different rows (resp. columns) are orthogonal, see [14].

2. The *graph associated* with a symmetric Hadamard matrix $H$ has vertices $V = \{1 \ldots n\}$ and edges $\{i, j\}$ where $H_{ij} = 1$, for $1 \leqq i < j \leqq n$.

**Definition.** A graph $G$ is $n$-uniform if for every $A$, $B \subseteq V(G)$, $A$, $B$ disjoint,

$$\left| |G \cap A \times B| - \frac{|A| \cdot |B|}{2} \right| \leqq \frac{1}{2} \sqrt{n |A| \cdot |B|}.$$

**Lemma 4.1** [6]. *The graph associated with a symmetric $n \times n$ Hadamard matrix is n-uniform.* □

**Corollary 4.2.** *Let G be associated with a symmetric $n \times n$ Hadamard matrix. Then for any two disjoint sets of vertices A, B, $|A|=|B|>\sqrt{n}$, G is not monochromatic on $A \times B$.*

*Proof.* If $G$ is monochromatic on $A \times B$, then, by Lemma 4.1,

$$\left| |G \cap A \times B| - \frac{|A|\cdot|B|}{2} \right| = \frac{|A|\cdot|B|}{2} \leqq \frac{1}{2}\sqrt{n|A|\cdot|B|}$$

hence $|A|\cdot|B| \leqq n$. □

**Theorem 4.3.** $\forall \delta > 0\ \exists \delta' > 0 \forall n \forall l \leqq (\frac{1}{2}-\delta)\log_2 n$ *such that for every graph H on l vertices, every n-uniform graph G on n vertices and every induced subgraph G′ of G on p vertices with* $p^{1+\delta'} \geqq n$ *there are*

$$\binom{p}{l} \cdot 2^{-\binom{l}{2}} \cdot \frac{l!}{|\mathrm{Aut}\, H|} \cdot (1-o(1))$$

*induced subgraphs of G′ isomorphic to H. If we assume that the graphs are linearly ordered and we consider only isomorphisms which preserve also the ordering the same formula without the factor* $\frac{l!}{|\mathrm{Aut}\, H|}$ *is true.* □

($|\mathrm{Aut}\, H|$ is the size of the automorphism group of $H$.)

*Note* that here and in the next theorems $o(1)$ denotes a fixed function, which we shall not specify, that tends to zero as $l$ tends to infinity.

A slightly weaker version of this theorem with $p=n$ with unordered graphs was proved in [6]. The proof can be adapted to prove the theorem above.

**Theorem 4.4.** *Let G be an n-uniform graph on n vertices,* $\delta > 0$,

$$l \leqq \left(\frac{1}{2}-\delta\right)\log_2 n$$
$$m < \frac{1}{\sqrt{2}} \cdot \frac{l}{e} \cdot 2^{l/2} \cdot (1-o(1)).$$

*Then G contains an induced subgraph F with m vertices with the property that neither F nor its complement contain* $K_l$.

Accidentally, the theorem gives an alternative proof to a well-known theorem of P. Erdös, see [5] Theorem 5.1, on a lower bound for the Ramsey function.

*Proof.* Suppose the contrary – this means that every set of $m$ vertices spans a subgraph which contains either $K_l$ or $E_l$ as an induced subgraph. As every copy of $K_l$ (or $E_l$) is contained in $\binom{n-l}{m-l}$ $m$-sets we obtain the inequality

$$(h(E_l, G)+h(K_l, G))\binom{n-l}{m-l} \geqq \binom{n}{m},$$

where $h(H, G)$ denotes the number of induced subgraphs of $G$ isomorphic to $H$. Using Theorem 4.3 with $p=n$ this yields

$$\binom{n}{l} 2^{-\binom{l}{2}+1}(1-o(1)) \geqq \left(\frac{n}{m}\right)^l (1-o(1))$$

and thus

$$m \geqq \frac{1}{\sqrt{2}} \cdot \frac{l}{e} \cdot 2^{\frac{l}{2}}(1-o(1)),$$

which contradicts to our assumption on $m$. □

The same proof yields another result.

Let us define that a graph $H$ is *special* if its vertex set can be decomposed into two disjoint part $V(H)=A \cup B$, $|A|=|B|$ with the property that $H$ contains all edges of the type $\{u, v\}$, $u \in A$, $v \in B$. The statement that a graph $G$ is monochromatic on $A \times B$ for some $A$, $B$ disjoint, $|A|=|B|=l$, is equivalent to the statement that $G$ or its complement contains a special graph on $2l$ vertices, but the latter one will be more convenient for the next theorem.

**Theorem 4.5.** *Let $G$ be an $n$-uniform graph on $n$ vertices, $\delta > 0$,*

$$2l \leqq \left(\frac{1}{2} - \delta\right) \log_2 n$$

$$m < 2^{l/2} \frac{l}{e}(1+o(1)).$$

*Then $G$ contains an induced subgraph $F$ with $m$ vertices with the property that neither $F$ nor its complement contain a special induced subgraph with $2l$ vertices.*

*Proof.* The same consideration (with $K_l$ and $E_l$ replaced by special graphs and their complements) as in the proof above leads to the inequality

$$\sum_H h^*(H, G) \cdot \binom{n-2l}{m-2l} \geqq \binom{n}{m}, \tag{*}$$

where we sum over all special graphs and their complements and $h^*(H, G)$ denotes the number of nonisomorphic ordered induced subgraphs. As the number of special graphs on $2l$ vertices can be clearly bounded from above by

$$\frac{1}{2}\binom{2l}{l} 4^{\binom{l}{2}}$$

we infer (using (*) and Theorem 4.3 with $p=n$) that

$$\binom{n}{2l} \cdot 2^{-\binom{2l}{2}} \cdot (1-o(1)) \cdot \binom{2l}{l} \cdot 4^{\binom{l}{2}} \cdot \binom{n-2l}{m-2l} \geqq \binom{n}{m}.$$

Simple calculation yields now

$$m \geqq (1+o(1)) \cdot 2^{l/2} \cdot \frac{l}{e},$$

which similarly as in the proof of Theorem 4.4 gives a contradiction. □

In particular, for fixed $\delta > 0$ (say $\delta = \frac{1}{4}$), Theorem 4.4 shows that there is a graph $G$ with $m$ vertices which is an induced subgraph of the graph associated with a symmetric Hadamard matrix with $n$ vertices such that

$$m \sim \text{const} \cdot n^{1/4 - \delta/2} \cdot \log n$$

and the size of the largest clique and largest independent set in $G$ are similar as in a random graph on $m$ vertices. Theorem 4.5 implies a similar statement for special graphs instead of the cliques and independent sets.

There are several constructions of Hadamard matrices. We shall need the following one.

Let $X$ be a set, let $V = \mathscr{P}(X)$ be the power set of $X$, $n = |V|$. An $n \times n$ symmetric Hadamard matrix $H$ with rows and columns indexed by the elements of $V$ is defined by

$$H_{uv} = (-1)^{|u \cup v|}.$$

To prove that $H$ is Hadamard we should show that, for $u \neq u'$, $u, u' \subseteq X$,

$$\sum_{v \subseteq X} H_{uv} \cdot H_{u'v} = 0.$$

Let $x$ belong to the symmetric difference of $u$ and $u'$. Then we can write the sum as follows

$$\sum_{v \subseteq X - \{x\}} (H_{uv} \cdot H_{u'v} + H_{u(v \cup \{x\})} \cdot H_{u'(v \cup \{x\})}).$$

Clearly each pair in the sum sums to zero, which proves that $H$ is Hadamard. The associated graph has $V$ as the vertex set and $(u, v)$ is an edge iff $|u \cup v|$ is even. The matrices described above are called Sylvester matrices, see [14]. The corresponding Boolean function has been studied in the communication complexity under the name "inner product mod 2".

## 5. The Complexity of Graphs Associated with Hadamard Matrices

*Lower Bounds.* We shall apply Theorems 2.4 and 2.5. Let $G$ denote a graph on $n$ vertices associated with a symmetric Hadamard matrix; in fact, it suffices to assume that $G$ is $n$-uniform. We consider only *the star complexity* in this section.

**Corollary 5.1.**

$$L(G) = \Omega\left(n \cdot \frac{\log n}{\log \log n}\right).$$

*Proof.* Let $G'$ be an arbitrary induced subgraph of $G$ with $m$ vertices, $m \geqq \frac{1}{2}n$, let $l = \lfloor \frac{1}{4} \log_2 n \rfloor$. By Theorem 4.3, if $n$ is sufficiently large, then each graph $H$ on $l$ vertices occurs as an induced subgraph of $G'$. This is because for $l$ large the factor

$$1 - o(1)$$

in the theorem is positive. To simplify the computation we consider ordered graphs. Thus there are

$$K = 2^{\binom{l}{2}}$$

such subgraphs of $G'$. By Theorem 2.4 we have the following lower bound,

$$L(G) \geqq \varepsilon \cdot n \cdot \frac{\binom{l}{2}}{l \cdot \log l} = \Omega\left(n \cdot \frac{\log n}{\log \log n}\right). \quad \square$$

We have proved this corollary only in order to show that Theorem 2.4 can give such a lower bound. The next theorem gives a better bound for $G$.

**Corollary 5.2**

$$L(G) = \Omega(n \cdot \log n).$$

*Proof.* By Corollary 4.2 the largest pair $A, B \subseteq V$, $|A| = |B|$ such that $G$ is monochromatic on $A \times B$ has size $\leqq \sqrt{n}$. Hence by Theorem 2.5

$$L(G) \geqq \varepsilon \cdot n \cdot \log \frac{n}{\sqrt{n}} = \frac{\varepsilon}{2} \cdot n \cdot \log n. \quad \square$$

*Upper Bounds.* In the rest of the section we consider the graph $G$ defined at the end of Sect. 4, i.e., a graph associated with a Sylvester matrix, $n = 2^k$ is the number of vertices of $G$.

**Proposition 5.3.** *For every $G'$ on $m$ vertices which is an induced subgraph of $G$,*

$$L(G') \leqq m \cdot \log_2 n,$$

*where the basis should contain $\cup$, $\oplus$ and 1.*

*Proof.* Suppose the vertices of $G$ are subsets of $\{1, \ldots, k\}$. then we have the following formula for $G'$

$$G' = 1 \oplus \bigoplus_{i=1}^{k} \bigcup_{i \in v \in V(G')} S_v,$$

where $S_v$ is the star with the root $v$. Clearly, an edge $(u, v)$ belongs to $G'$ iff $|u \cup v|$ is even iff the number of unions to which $u$ or $v$ belongs is even. Each union has size $\leqq |V(G')| = m$, there are $k = \log_2 n$ such unions. $\square$

**Corollary 5.4.** *For every $m$ there exists a graph $G'$ on $m$ vertices such that*

(i) $L(G') = O(m \log m)$,

*(ii) $G'$ does not contain $K_l$ nor its complement with $l \geqq (2-o(1)) \cdot \log_2 m$,*

*(iii) $G'$ does not contain a special graph on $2l$ vertices with $l \geqq (2-o(1)) \cdot \log_2 m$.*

*Proof.* By Theorem 4.4 there exists a subgraph $G'$ of $G$ whose size is

$$m \sim n^{1/4-\delta}$$

and which satisfies (ii). By Proposition 5.3

$$L(G') \leqq m \log n \sim \frac{1}{1/4-\delta} m \log m \sim O(m \log m),$$

for fixed $\delta$, $0<\delta<1/4$. Using Theorem 4.5 we can also show that there exists $G'$ satisfying (i) and (iii). Since the proofs of Theorems 4.4 and 4.5 show the conditions (ii) and (iii) for random subgraphs of $G$, they can be satisfied simultaneously. □

*Remarks.* 1. Corollary 5.4 shows that Ramsey type properties of graphs do not imply lower bounds to the formula size complexity larger than $\Omega(m \log m)$, where $m$ is the number of vertices.

2. A similar conclusion concerning Theorem 2.4 can be reached using Proposition 5.3 and Theorem 4.3. That is even the condition that graphs with $l$ vertices with $l \leqq (\frac{1}{2}-\delta) \log_2 n$, occur in each subgraph with $m$ vertices, $m^{1+\delta'} \geqq n$, with the frequencies similar as in random graphs on $m$ vertices cannot imply a larger lower bound than $\Omega(n \cdot \log n)$. We, however, cannot exclude that counting the number of nonisomorphic induced subgraphs on $l$ vertices with $l$ larger than $\log n$ could lead to larger lower bounds. Of course, if this were true, then the formula relating the number of such subgraphs to the complexity must be different from the formula of Theorem 2.4.

3. The graph $G'$, about which we spoke in the first remark, also shows that for some graphs any covering by complete bipartite graphs gives much larger formula than an optimal one. Namely, we know that $L(G') = O(m \cdot \log m)$, but the largest $K_{l,l}$ in $G'$ has $l = O(\log m)$. Hence any formula

$$\bigcup_j \left[\left(\bigcup_{i \in A_j} S_i\right) \cap \left(\bigcup_{i \in B_j} S_i\right)\right]$$

which computes $G'$ must have the size $\Omega\left(\frac{m^2}{\log m}\right)$. This is because $G'$ contains $\Omega(m^2)$ edges and each $K_{l,l}$ has the complexity $2l$ but covers only $l^2$ edges.

The last thing that we want to show is that the conditions considered above cannot imply nonlinear lower bounds to the circuit complexity. The following lemma and its corollary might be of an independent interest.

**Lemma 5.5.** *Let $f$ be a Boolean vector function, $f$: $\{0, 1\}^n \to \{0, 1\}^k$, determined by*

$$\langle x_1 \ldots x_n \rangle \mapsto \langle \bigvee_{i \in I_1} x_i \ldots \bigvee_{i \in I_k} x_i \rangle.$$

*Then there is a circuit of size $n + 2^{k+1} - k - 2$ for $f$. (This is true also for any commutative and associative operation instead of $\bigvee$.)*

*Proof.* For $w \in \{0, 1\}^{\leqq k}$ define sets $J_w \subseteq \{1, 2, \ldots, n\}$ by

$$i \in J_w \leftrightarrow \forall j \leqq |w| (i \in I_j \leftrightarrow w(j) = 1).$$

For $|w| = k$ the sets $J_w$ are disjoint since $i \in J_w$ iff $w$ is the characteristic function of the occurrence of $i$ in $I_1, \ldots, I_k$. We can compute all nonempty disjunctions

$$\bigvee_{i \in J_w} x_i, \qquad |w| = k$$

using only $\leqq n$ or's as each variable occurs in at most one. Next we compute nonempty disjunctions

$$\bigvee_{i \in J_w} x_i$$

for $|w| < k$ using the equality

$$\bigvee_{i \in J_w} x_i = \bigvee_{i \in J_{w0}} x_i \vee \bigvee_{i \in J_{w1}} x_i.$$

Thus we need $2^k - 1$ more or's. Now we can compute $\bigvee_{i \in I_j} x_i$ using equality

$$\bigvee_{i \in I_j} x_i = \bigvee_{|w| = j-1} \bigvee_{i \in J_{w1}} x_i,$$

which requires

$$|\{w \,|\, |w| = j-1\}| - 1 = 2^{j-1} - 1$$

new or's. Thus for $j = 1, \ldots, k$ all together

$$\sum_{j=1}^{k} (2^{j-1} - 1) = 2^k - 1 - k$$

or's. Hence the size of the circuit is

$$\leqq n + 2^k - 1 + 2^k - 1 - k = n + 2^{k+1} - k - 2. \quad \square$$

**Corollary 5.6.** *If $l$ divides k, then f can be computed by a circuit of size*

$$\leqq l(n-2) + 2^{\frac{k}{l}+1} - k.$$

*In particular if $k = O(\log n)$ then f has a linear size circuit.*

*Proof.* 1. Decompose $f$ into $l$ functions $\{0, 1\}^n \to \{0, 1\}^{\frac{k}{l}}$ and compute each of these functions separately using Lemma 5.4.

2. Suppose $k \leqq C \cdot \log n$, $C$ a constant. Let $l$ be an integer such that $l > C$. Let $k' = l \left\lceil \frac{C}{l} \log n \right\rceil \geqq k$. Extend $f$ onto an $f'$: $\{0, 1\}^n \to \{0, 1\}^{k'}$, say, by repeating

some coordinates, and apply the first part of the corollary. We get a circuit of size

$$\leqq l(n+2^{\frac{C}{l}\log_2 n+2}-2)-k' \leqq ln+O(n^{\frac{C}{l}}). \quad \square$$

**Theorem 5.7.** *Let $\varepsilon>0$ be fixed. Then the induced subgraphs of $G$ with $m$ vertices, where $m \geqq n^{\varepsilon}$ have the circuit complexity at most linear in $m$.*

*Proof.* We already know that if $G'$ is an induced subgraph of $G$ then

$$G' = 1 \oplus \bigoplus_{i=1}^{k} \bigvee_{i \in v \in V(G')} S_v,$$

where $k=\log_2 n$. Since $m \geqq n^{\varepsilon}$, we have $k \leqq \frac{1}{\varepsilon} \log m$. Thus the vector function

$$\langle \bigvee_{1 \in v \in V(G')} S_v \dots \bigvee_{k \in v \in V(G')} S_v \rangle$$

can be computed (by Corollary 5.5) using a circuit of size $O(m)$. Hence $G'$ has a circuit of size $O(m)$ too. $\square$

**Corollary 5.8.** *For every $m$ there exists a graph $G'$ on $m$ vertices such that*
*(i) $C(G')=O(m)$,*
*(ii) $G'$ does not contain $K_l$ nor its complement with $l \geqq (2-o(1)) \cdot \log_2 m$,*
*(iii) $G'$ does not contain a special graph on $2l$ vertices with $l \geqq (2-o(1)) \cdot \log_2 m$.*

*Proof.* The same as the proof of Corollary 5.4 with Theorem 5.7 used instead of Proposition 5.3. $\square$

## Appendix

In a recent paper [8] Razborov proved several results related to our approach. For instance he associates a graph $G$ on $2^m$ vertices with any Boolean function $f$ of $2m$ variables and proves that there exists a Boolean function $f$ whose formula size is polynomial in $m$ and such that the corresponding graph $G$ does not contain cliques and independent sets larger than $2m$. It is an open problem whether a sequence of such Boolean functions can be computed by a deterministic Turing machine in polynomial time.

## References

1. Babai, L, Frankl, P., Simon, J.: Complexity classes in communication complexity, 27th FOCS, pp 337–347 (1986)
2. Babai, L., Pudlák, P., Rödl, V., Szemeredi, E.: Lower bounds to the complexity of symmetric Boolean functions. Theor. Comput. Sci. (submitted for publication)
3. Berge, C.: Graphs and hypergraphs. Amsterdam Oxford New York: North-Holland, American Elsevier 1973

4. Bublitz, S.: Decomposition of graphs and monotone formula size of homogeneous functions. Acta Informatica **23**, 689–696 (1986)
5. Erdös, P., Spencer, J.: Probabilistic methods in combinatorics. Budapest: Akadémiai Kiadó 1979
6. Frankl, P., Rödl, V., Wilson, R.M.: The number of submatrices of given type in a Hadamard matrix and related results. J. Comb. Theor B (to appear)
7. Nečiporuk, E.T.: On a Boolean function. Sov Math Dokl **2**, **4**, 999–1000 (1966)
8. Razborov, A.A.: Formuly organičenoj glubiny v bazise $\{\&, \oplus\}$ i nekotorye kombinatornye zadači (Formulas of bounded depth in basis $\{\&, \oplus\}$ and some combinatorial problems), in Složnost' algoritmov i prikladnaja matematičeskaja logika, S.I. Adjan editor (1987)
9. Savage, J.E.: The complexity of computing. New York: Wiley 1976
10. Sgall, J.: Personal communication (1987)
11. Tuza, Z.: Covering of graphs by complete bipartite subgraphs; complexity of 0-1 matrices. Combinatorica **4**, 111–116 (1984)
12. Wegener, I.: On the complexity of slice functions. In: Chytil, M.P., Koubek, V. (eds.) Mathematical Foundations of Computer Science. Proceedings, 11th Symposium Praha, Czechoslovakia Sept. 3–7, 1984. (Lect. Notes Comput. Sci., vol. 176, pp. 553–561) Berlin Heidelberg New York: Springer 1984
13. Wegener, I.: The complexity of Boolean functions. Stuttgart: B.G. Teubner/New York: Wiley 1987
14. MacWilliams, F.J., Sloane, N.J.A.: The theory of error-correcting codes. Amsterdam New York: North-Holland 1978
15. Yao, A.C.-C.: Some complexity questions related to distributive computing. Proc. 11th ACM STOC, pp 209–213 (1979)

Received May 15, 1987 / March 4, 1988