Discrete Mathematics 88 (1991) 179–192
North-Holland

Some results on the complexity of families of sets

Daniel Grieser

Department of Mathematics, MIT, Cambridge, MA 02139, USA
Current address: Department of Mathematics, University of California, Los Angeles, CA 90024, USA

Received 16 September 1988

Abstract

Grieser, D., Some results on the complexity of families of sets, Discrete Mathematics 88 (1991) 179–192.

Let $\mathscr{P}$ be a property of graphs on a fixed n-element vertex set V. The *complexity* $c(\mathscr{P})$ is the minimal number of edges whose existence in a previously unknown graph H has to be tested such that it becomes possible to decide whether H has property $\mathscr{P}$ or not. We investigate properties (in the broader sense of families of sets) whose complexity is much less than the maximum (which is $\binom{n}{2}$ for graph properties). It is well known that such properties must be representable as a disjoint union of long intervals (in the boolean lattice of graphs on V). We show that, if the number of intervals is not too large, the converse is true as well. We also show that there are graph properties whose complexities differ by at most $4n$ from any given number between $2n-4$ and $\binom{n}{2}$. Finally, we give estimates on the complexity of the scorpion graph property.

1. Introduction

Let $\mathscr{P}$ be a family of subsets of a set T. We consider a measure of complexity that was first introduced by Holt and Reingold ([4]) and Rosenberg ([5]). This measure is defined as follows:

Suppose two players, $\mathscr{A}$ (Algy or Seeker) and $\mathscr{S}$ (Strategist or Hider) play the following game: $\mathscr{S}$ thinks of some subset H of T. $\mathscr{A}$ wants to determine if H is in $\mathscr{P}$. For this purpose $\mathscr{A}$ chooses an element $x \in T$, and $\mathscr{S}$ tells $\mathscr{A}$ whether x is in H or not. Then $\mathscr{A}$ tries another element and so on until $\mathscr{A}$ is able to decide if H is in $\mathscr{P}$. The goal of $\mathscr{A}$ is to make this decision as soon as possible, while $\mathscr{S}$ tries to force $\mathscr{A}$ to ask many questions. We allow $\mathscr{S}$ to change the set H in the course of the game as long as the answers given so far remain correct. The *complexity* $c(\mathscr{P})$ is defined as the number of questions when both players play optimally.

Obviously $c(\mathscr{P})$ is bounded above by $t = |T|$, and is zero if and only if $\mathscr{P} = \emptyset$ or 2^T in which cases we call $\mathscr{P}$ *trivial*. $\mathscr{P}$ is *elusive* if $c(\mathscr{P}) = t$. We will use the word *algorithm* for the way $\mathscr{A}$ plays, and *strategy* for the way $\mathscr{S}$ plays. We think of $\mathscr{P}$ as a property of subsets of T. Especially, when T is the set of all two-element subsets of $V = \{1, 2, \ldots, n\}$, we may interpret subsets of T as graphs on the vertex set

V, and we call $\mathcal{P}$ a *graph property* if $\mathcal{P}$ is invariant under all permutations of the vertices.

Up to now research has concentrated on:

(i) showing that large classes of properties and some special ('natural') graph properties have high complexity, or even are elusive, and
(ii) giving general lower bounds for the complexity of graph properties, and constructing graph properties of low complexity.

We shortly state some results which will be of interest in the sequel.

Theorem 1 ([6, 2]). *If $c(\mathcal{P}) \leq k$, then $\mathcal{P}$ is the disjoint union of intervals of length $t-k$.*

Here an *interval* $[A, B]$ is a class of subsets of T of the form

$$[A, B] = \{X \mid A \subseteq X \subseteq B\}$$

where $A \subseteq B$. Its *length* is $l([A, B]) = |B| - |A|$. This theorem can be used to prove lower bounds or even elusiveness for many properties. Concerning general lower bounds for the complexity of graph properties, Bollobás and Eldrigde proved the following theorem.

Theorem 2 ([3]). *If $\mathcal{P}$ is a nontrivial graph property, then*

$$c(\mathcal{P}) \geq 2n - 4.$$

Probably this bound is not best possible. The graph property of lowest complexity known today is the property $\mathcal{S}c$ of being a scorpion graph. A graph on $n \geq 5$ vertices is a *scorpion graph* if it has a vertex t of degree 1 (tail), a vertex b of degree $n-2$ (body) and a vertex l of degree 2 incident to t and b (link). The remaining $n-3$ vertices can be connected in an arbitrary manner (see Fig. 1).

The property $\mathcal{S}c$ was investigated first in [2]. The algorithm was slightly improved in [1] yielding the following theorem.

Theorem 3. $c(\mathcal{S}c) \leq 6n - 10.$

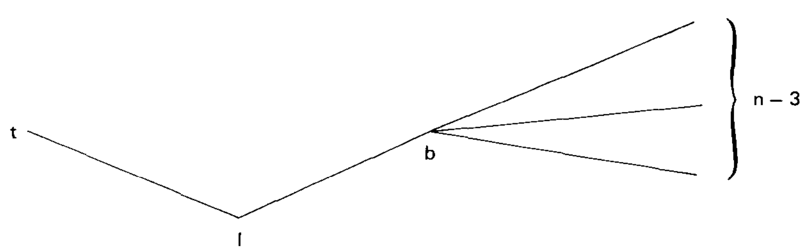

Fig. 1. Scorpion.

In Section 2 we will prove a result which can be viewed as a weak inversion of Theorem 1: If $\mathcal{P}$ is the disjoint union of $r > 1$ intervals of length at least $t - k$, then $c(\mathcal{P}) \leq 2k \log r$. This yields a nontrivial upper bound in the case where k is small compared with t (the intervals are 'long') and r is not too large, e.g. if $\mathcal{P}$ is a graph property (i.e. $t = \binom{n}{2}$), k is linear and r is polynomial in n (which is the case for scorpions).

In Section 3 we will investigate scorpion graphs. We will define what a scorpion with a tumour of size p is (a 'normal' scorpion has $p = 0$), and we will generalize the well-known algorithm for scorpions to the case of scorpions with tumours. We will also improve the algorithm to establish an upper bound of $6n - \sqrt{2n} - 6$ for $c(\mathcal{S}c)$. Turning to lower bounds in 3.2, we will construct a strategy to show $c(\mathcal{S}c) \geq (3 + \alpha)n - C_\alpha$ for any α, $0 < \alpha < 1.5$, and sufficiently large n, with some constant C_α only depending on α. This is considerably better than the easily obtained lower bound $3n - 6$, and possibly some ideas in the strategy might be applicable to more general classes of graph properties. Finally, in 3.3 we show how scorpions with tumours can be used to construct graph properties whose complexities are close to any given number between $2n - 4$ and $\binom{n}{2}$.

In the description of an algorithm that decides $\mathcal{P}$ (i.e. that decides if the set H is in $\mathcal{P}$) we will often use phrases like "$\mathcal{A}$ determines that H is of some specified form" and mean: "$\mathcal{A}$ decides that either H has this form or else H is not in $\mathcal{P}$". A question of Algy will sometimes be called a probe or a test.

For later use and as an example we now prove that the complexity of an interval $[A, B]$, $A \subseteq B \subseteq T$, is $t - l([A, B])$:

If $\mathcal{A}$ probes all elements of $A \cup B^c$ then $\mathcal{A}$ clearly is able to decide $\mathcal{P}$. On the other hand it is necessary to probe all these elements because $\mathcal{S}$ may choose the strategy of giving all elements of A (i.e., answering "$x \in H$" for all $x \in A$) and refusing all elements of B^c. Hence

$$c([A, B]) = |A \cup B^c| = t - (|B| - |A|).$$

2. An upper bound for $c(\mathcal{P})$

Theorem 4. *Let $\mathcal{P}$ be a disjoint union of r intervals of length $\geq t - k$. Then for the complexity of $\mathcal{P}$ we have the upper bound*:

$$c(\mathcal{P}) \leq \begin{cases} 2k \log r & \text{if } r \geq 2, \\ k & \text{if } r = 1, \end{cases}$$

where log *is the natural logarithm.*

Proof. We construct an algorithm which decides $\mathcal{P}$ in at most that many steps. The idea comes from the scorpion graph algorithm: In order to decide if a graph is a scorpion, player $\mathcal{A}$ determines tail, body and link and then checks the requirements on their incidences (see Section 3.1). Now in the class $\mathcal{S}c$ of scorpion graphs on n vertices the graphs with fixed t, b and l form an interval, and

$\mathcal{S}c$ is the disjoint union of all such intervals (one for each choice of t, l, b). Therefore, it is natural to generalize this proceeding as follows. If $\mathcal{P}$ is the disjoint union of the intervals $\mathcal{I}_j$, $j \in J$, then at first $\mathcal{A}$ determines the interval $\mathcal{I}_j$ which contains H as an element (see the convention at the end of Section 1) and then decides if H is actually in $\mathcal{I}_j$. Because every $\mathcal{I}_j$ has length $\geq t-k$, the second part takes at most k steps. Therefore we have to show that for $r=|J|\geq 2$ the first part can be accomplished in at most $k(2\log r-1)$ steps. We will do this recursively using the following observation.

Lemma 1. *Let $\mathcal{P}$ be a property of subsets of T. If $\mathcal{P}_1$, $\mathcal{P}_2 \subseteq \mathcal{P}$ and $x \in T$ is contained in every member of $\mathcal{P}_1$ and in no member of $\mathcal{P}_2$, then by probing x $\mathcal{A}$ can determine that $H \in \mathcal{P}-\mathcal{P}_1$ (if x is refused) or that $H \in \mathcal{P}-\mathcal{P}_2$ (if x is given).*

Proof. Clear by the definitions. □

Let's apply this to our situation:

$$\mathcal{P}=\bigcup_{j\in J}\mathcal{I}_j,\qquad \mathcal{I}_j=[A_j, B_j^c]\quad\text{with } A_j, B_j\subseteq T,\ A_j\cap B_j=\emptyset\ \forall j$$

and

$$\mathcal{I}_i\cap\mathcal{I}_j=\emptyset\quad \forall i\neq j.$$

We write $\mathcal{P}_{J'}=\bigcup_{j\in J'}\mathcal{I}_j$ for $J'\subseteq J$. Choose $x\in T$ and let $\mathcal{P}_1=\mathcal{P}_{J_1}$, $\mathcal{P}_2=\mathcal{P}_{J_2}$ with

$$J_1=\{j \mid x\in A_j\},\qquad J_2=\{j\mid x\in B_j\}.$$

By the lemma after the probe x $\mathcal{A}$ knows that $H\in\mathcal{P}_{J-J_1}$ (if x is refused) or that $H\in\mathcal{P}_{J-J_2}$ (otherwise). If, say, x is refused, then in the case $|J-J_1|\leq 1$ we are done, and in the case $|J-J_1|>1$ we can use the same procedure with J replaced by $J-J_1$ and thereby reduce the number of possible intervals step by step until we arrive at one interval. One such step will be the more effective the larger $\min\{|J_1|, |J_2|\}$ is (regarding the worst case for the answer to the probe x), and we may use our freedom in the choice of x to make this number large. Therefore we have to solve the following problem:

Given a family of $r>1$ pairwise disjoint intervals $[A_j, B_j^c]$ of length $\geq t-k$, find $x\in T$ which is in many A's and in many B's!

Observing that for intervals $\mathcal{I}=[A, B^c]$, $\mathcal{I}'=[C, D^c]$ the condition $\mathcal{I}\cap\mathcal{I}'=\emptyset$ is equivalent to ($A\cap D\neq\emptyset$ or $B\cap C\neq\emptyset$), we can formulate this as a problem on families of pairs of sets; we will prove the following lemma.

Lemma 2. *Let $\{(A_j, B_j)\}_{j\in J}$, $r=|J|>1$, be a family of pairs of subsets of a set T having the properties*:

(a) $A_j\cap B_j=\emptyset\quad\forall j$,
(b) $A_i\cap B_j\neq\emptyset$ *or* $A_j\cap B_i\neq\emptyset\quad\forall i\neq j$,
(c) $|A_j\cup B_j|\leq k\quad\forall j$.

Then there is an $x \in T$ *which is contained in at least* $\lceil r/2k \rceil$ *A's and at least* $\lceil r/2k \rceil$ *B's.*

Before we prove this lemma, we show how the theorem follows. By the lemma with every question the number of possible intervals can be reduced by the factor $1 - 1/2k$ as long as there are at least 2 intervals, hence after l questions there will be at most $\max\{r(1-1/2k)^l, 1\}$ intervals left. Using $(1-1/2k)^{2k} < e^{-1}$, it is easy to see that for $l \geq k(2\log r - 1)$ we have $r(1-1/2k)^l < \sqrt{e} < 2$, and the conclusion follows. □ (Theorem 4)

It remains to prove Lemma 2. We will first show that the problem of finding such an x is equivalent to a problem concerning coverings of a complete graph by bipartite graphs, and then solve this problem.

Let the dual family $\{(R_x, S_x)\}_{x\in T}$ of $\{(A_j, B_j)\}_{j\in J}$ be the family of pairs of subsets of J defined by

$$R_x = \{j \in J \mid x \in A_j\}, \qquad S_x = \{j \in J \mid x \in B_j\}.$$

Then (a), (b), (c) are easily seen to be equivalent to:

(a′) $R_x \cap S_x = \emptyset \quad \forall x$,

(b′) $\forall i \neq j \ \exists x$ such that $i \in R_x$, $j \in S_x$ or $i \in S_x$, $j \in R_x$,

(c′) Every j is contained in at most k of the sets $R_x \cup S_x$,

and the lemma asserts that for some x, both R_x and S_x contain at least $\lceil r/2k \rceil$ elements.

If K_J is the complete graph on the vertex set J, then (a′), (b′) mean that the complete bipartite graphs G_x with vertex classes R_x, S_x cover the edges of K_J. We prove the following lemma.

Lemma 3. *If the bipartite graphs* G_x, $x \in T$, *cover all the edges of a complete graph* K_r, $r \geq 2$, *and if every vertex of the complete graph belongs to at most k of the* G_x, *then in one of the* G_x *both vertex classes have at least* $\lceil r/2k \rceil$ *elements.*

Proof. Loosely speaking, this comes from the fact that a bipartite graph with one small vertex class has many vertices compared with its number of edges, while the K_r has $(r-1)/2$ as many edges as vertices.

To be precise, we estimate for a bipartite graph G_x with vertex classes of size a and b, $a \leq b \leq r$, the ratio

$$\frac{|V(G_x)|}{|E(G_x)|} \geq \frac{a+b}{ab} = \frac{1}{a} + \frac{1}{b} \geq \frac{1}{a} + \frac{1}{r}.$$

If every G_x has one vertex class with at most v vertices (with some integer v), then $a \leq v$ and therefore

$$\sum_{x\in T} |V(G_x)| \geq \left(\frac{1}{v} + \frac{1}{r}\right) \sum_{x\in T} |E(G_x)| \geq \left(\frac{1}{v} + \frac{1}{r}\right) \frac{r(r-1)}{2} = r\frac{r-1}{2v} + \frac{r-1}{2}.$$

Therefore at least one of the vertices of the K_r is in more than $(r-1)/2v$ and hence in at least $r/2v$ of the G_x. Choosing v as the largest integer less than $r/2k$, $v = \lceil r/2k \rceil - 1$, yields the desired result. □ (Lemma 3 and Lemma 2)

As an example for Theorem 4 we consider the property $\mathcal{S}c$ of scorpion graphs. $\mathcal{S}c$ is the disjoint union of $n(n-1)(n-2)$ intervals of length $\binom{n}{2} - (3n-6)$, hence

$$c(\mathcal{S}c) \leq 2(3n-6)\log n(n-1)(n-2) < 18n \log n.$$

Even if this is much more than the known upper bound $6n$, it can be seen that $\mathcal{S}c$ is not elusive for large n.

3. Scorpions

We first define a scorpion with a tumour of size p.

Definition. Let n, p be natural numbers, $n \geq p+5$. The class $\mathcal{S}c^{(p)}$ consists of the Graphs G on n vertices with the property:

In G there are $p+2$ pairwise connected vertices; among them, there are p vertices which have no other incidences (the tumour vertices), and the vertices l (link) and b (body). l is incident to only one other vertex, t (tail), which has degree one, and b to all other vertices except t. The remaining $n-p-3$ vertices can be connected arbitrarily (see Fig. 2).

t, l, b and the tumour vertices are called *critical,* as well as all edges (of the complete graph K_n) incident with at least one of them. In order to prove upper and lower bounds for $c(\mathcal{S}c^{(p)})$ we will give algorithms and strategies. The following definitions will simplify the language in their description.

At any stage in the course of the questioning we call an edge *green* if it has been given, *red* if it has been refused, and *colourless* if it has not been probed yet. A vertex is colourless if all edges incident with it are colourless, otherwise it is green, red or many coloured with the obvious meanings.

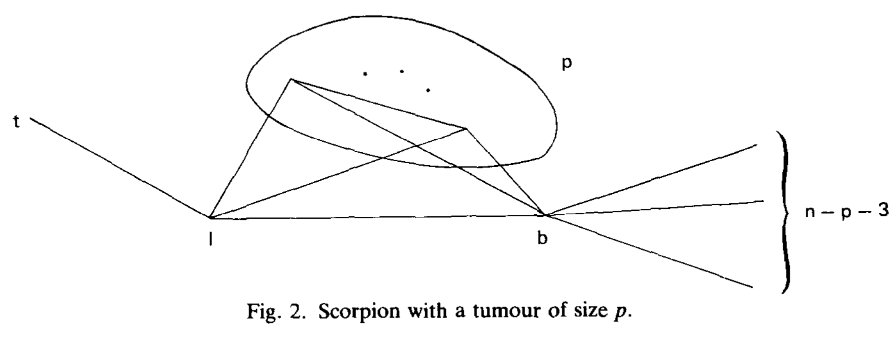

Fig. 2. Scorpion with a tumour of size p.

3.1. *Upper bounds for $c(\mathcal{S}c^{(p)})$*

We must slightly modify the well-known algorithm for scorpion graphs ([2, 1, 7]) to get a good algorithm for $\mathcal{S}c^{(p)}$.

Theorem 5.

$$c(\mathcal{S}c^{(p)}) \leq (6+p)n - 8 - \binom{p+4}{2}.$$

Proof. Player $\mathcal{A}$ determines t, l, b and the tumour vertices first and then checks all conditions on their incidences. Clearly, for the second part it is sufficient to test all edges incident with critical vertices. Some of these edges will have been probed before and need not be probed again. In order to simplify the counting of the number of steps, in the first part we don't count those probed edges which become critical in the end.

At any time during the questioning a *candidate tail* (*body*) is a vertex at most one of whose incident edges has been given (refused). T (B) denotes the set of candidate tails (bodies). The *weight* $w(x)$ of a candidate tail (body) x is 2 if none of its incident edges has been given (refused), otherwise it is 1. $w(x)$ is the least number of probes incident to x that can make x a noncandidate. In the beginning, $B = T = V$. The first part of the algorithm consists of three main steps: After the first step candidate bodies and candidate tails will be separated ($B \cap T = \emptyset$), after the second step B or T will contain only one element (i.e., body or tail is known), and after the third step all critical vertices will be known.

Step 1: $\mathcal{A}$ tests the n edges of a Hamiltonian circuit C. Then all vertices are coloured. Let $B_0 \subseteq B$ be the set of green vertices, $T_0 \subseteq T$ the set of red vertices, and $M = V - B_0 - T_0 = T \cap B$ the set of manycoloured vertices. Let $m = |M|$. Then m is even. We distinguish three cases:

Case a: $m = 0$.
Then T or B is empty, and H is not a scorpion.

Case b: $m = 2$, $M = \{x, y\}$ *and* x, y *are neighbours on* C.
Then $T_0 = \emptyset$ or $B_0 = \emptyset$. Choose the vertices x', y' such that x', x, y, y' are consecutive on C in this order. If $B_0 = \emptyset$ then $B = \{x, y\}$, $T = V - \{x, y\}$, and Step 1 is finished; if $T_0 = \emptyset$ then $B = V$, $T = \{x, y\}$, and after the tests $x'y$ and xy' we have $B \cap T = \emptyset$.

Case c: $m = 2$, $M = \{x, y\}$ *and* x, y *are not neighbours on* C, *or* $m \geq 4$.
Then there are $m/2$ colourless edges connecting the vertices of M in pairs, and after these $m/2$ tests we have $B \cap T = \emptyset$.

The sum of the weights of all candidates is at most $2(n-m) + m = 2n - m$ in any case, and the number of (noncritical) edges asked so far is at most $(n-4) + m/2$ because at least 4 edges of C become critical in the end and in Case b, $T_0 = \emptyset$ at least one of the additional edges becomes critical.

Step 2: $\mathscr{A}$ asks for colourless edges between T and B as long as there is any left and b or t is not determined uniquely. The sum of weights is reduced by 1 in every question, therefore at most $2n - m - 3$ questions are necessary. At the end of this questioning let $\tau = |T|$, $\beta = |B|$. If $\tau = 1$ or $\beta = 1$ then Step 2 is finished. Otherwise all $\tau\beta$ edges between T and B have been asked. At most β of them are red and at most τ are green, therefore $\tau\beta \leqslant \tau + \beta$, i.e. $(\tau - 1)(\beta - 1) \leqslant 1$, and we conclude $\tau = \beta = 2$.

In this case let $B = \{b_1, b_2\}$, $T = \{t_1, t_2\}$. There must be precisely two disjoint green edges joining B and T, or otherwise H is not a scorpion. Let the edges b_1t_1, b_2t_2 be green, the edges b_1t_2, b_2t_1 be red. Then either $b = b_1$, $t = t_2$, $l = b_2$ or $b = b_2$, $t = t_1$, $l = b_1$. By choosing $p + 1$ vertices $x_1, \ldots, x_{p+1}$ not in $B \cup T$ and asking for the edges b_1x_i, $i = 1, \ldots, p + 1$, $\mathscr{A}$ can easily decide if b_1 or b_2 is the body. Because all these additional edges become critical, the number of noncritical probes in this step is at most $2n - m - 3$.

Step 3: Now t or b is known, and by probing all edges incident to this vertex $\mathscr{A}$ gets l (resp. t) and then similarly the other critical vertices, only probing critical edges.

The number of critical edges is

$$\binom{n}{2} - \binom{n-p-3}{2} = (3+p)n - \binom{p+4}{2},$$

hence the total number of questions is at most

$$\left(n - 4 + \frac{m}{2}\right) + (2n - m - 3) + (3 + p)n - \binom{p+4}{2}$$

$$= (6+p)n - \frac{m}{2} - 7 - \binom{p+4}{2}. \qquad \square$$

By a good choice of the order in which $\mathscr{A}$ asks for the edges of the circuit C in Step 1 and for the edges between B and T in Step 2, we can improve this with the following theorem.

Theorem 6.

$$c(\mathscr{S}c^{(p)}) \leqslant (6+p)n - \sqrt{2n} - \binom{p+4}{2}.$$

Proof. Let $\tau'(\beta')$ be the number of red (green) vertices after probing the n edges of C. $\mathscr{A}$ can force $\tau' \leqslant m + 1$ or $\beta' \leqslant m + 1$ (see below). Let $\tau' \leqslant m + 1$, the other case is analogous. Then $|T| \leqslant \tau' + m \leqslant 2m + 1$, hence either m is 'large' which leads to an improved bound by the last expression in the proof of Theorem 5, or $|T|$ is 'small'. If $\mathscr{A}$ asks the T-B-edges in Step 2 in such an order that their t-end goes cyclically through T, then after k questions every remaining t-candidate is incident to at least $k/(\tau' + m) - 1$ of all tested T-B-edges, and at least this number of T-B-edges tested in Steps 1 and 2 will become critical. Hence, if $k \leq 2n - m - 3$ is the number of edges tested in Step 2, the total number of steps is at most

$$n + \frac{m}{2} + k - \left(\frac{k}{\tau' + m} - 1\right) + \left((3+p)n - \binom{p+4}{2}\right)$$

$$\leq (4+p)n - \binom{p+4}{2} + \frac{m}{2} + 1 + (2n - m - 3)\left(1 - \frac{1}{2m+1}\right)$$

$$\leq (6+p)n - \sqrt{2n} - \binom{p+4}{2},$$

as can easily be shown by use of the arithmetic-geometric mean inequality.

We still have to show how to test the edges of C. Let n be even, the case n odd is similar.

First ask for the edges of a perfect matching. Let the number of red edges be less than or equal to the number of green edges. Now choose the circuit C in such a way that its edges alternately are matching- and nonmatching-edges and such that every red matching edge lies between green matching edges. Then every edge of C asked for in the sequel which generates a red vertex must generate a manycoloured vertex also, and hence in the end there are at least as many manycoloured as red vertices. □

3.2 Lower bounds

We obtain a simple lower bound for the complexity of $\mathcal{S}c^{(p)}$ and then strengthen this bound for regular scorpions ($p = 0$).

Theorem 7.

$$c(\mathcal{S}c^{(p)}) \geq \binom{n}{2} - \binom{n-p-3}{2} = (3+p)n - \binom{p+4}{2}.$$

Proof. If $\mathcal{P} \neq \emptyset$ is a property and $c(\mathcal{P}) \leq t - k$ then by Theorem 1 $\mathcal{P}$ must contain an interval $\mathcal{I}$ of length k. If $\mathcal{I} = [A, B]$, then

$$l(\mathcal{I}) = |B| - |A| \leq \max\{|X| \mid X \in \mathcal{P}\} - \min\{|X| \mid X \in \mathcal{P}\}.$$

For $\mathcal{S}c^{(p)}$ this last difference is just $\binom{n-p-3}{2}$, the maximal number of edges between noncritical vertices. The assertion follows. □

Theorem 8. *Let $0 < \alpha < 1.5$. Then there is a constant C_α such that for sufficiently large n:*

$$c(\mathcal{S}c) \geq (3 + \alpha)n - C_\alpha.$$

Proof. We will give a strategy for player $\mathcal{S}$. One probe and its answer will be called a *step*. $\mathcal{S}$ always answers in such a way that there is still a scorpion compatible with the given answers. The strategy consists of two parts. In the first part (αn steps) $\mathcal{S}$ gives answers according to rules to be described later. Then $\mathcal{S}$ fixes vertices b, t and l and answers in the second part accordingly. When b, t, l have been fixed, $\mathcal{A}$ must still ask all critical unprobed edges. Because the total number of critical edges is $3n-6$, $\mathcal{S}$ can force $\mathcal{A}$ to ask $3n - C_\alpha$ questions if it's possible to choose t, l, b in such a manner that the sum of their degrees in the graph of probed edges is less than some constant (and, of course, such that there are scorpions compatible with this choice). The following lemma shows that this is possible (for large n) if after the first part there are linearly many red vertices and linearly many green vertices.

Lemma 4. *Let α, λ, $\mu > 0$ be fixed real numbers. Then the following is true for sufficiently large n:*

If G is a graph with n vertices and αn edges, some of whose vertices are coloured green or red, and if at least λn vertices are coloured green and at least μn red, then there are a green vertex b and two nonadjacent red vertices l, t whose degrees sum up to at most $2\alpha/\lambda + 6\alpha/\mu$.

Proof. First we remark that in a graph with n vertices and k edges there are less than $2k/\rho$ vertices with degree greater than ρ; otherwise the sum of all degrees would be greater than $\rho 2k/\rho = 2k$ which is impossible. Now there are less than λn vertices of degree $>2\alpha n/\lambda n = 2\alpha/\lambda$, hence there is a green vertex with degree $\leqslant 2\alpha/\lambda$. There are less than $2\mu n/3$ vertices of degree $>3\alpha/\mu$; hence at least $\mu n/3$ red vertices have degree $\leqslant 3\alpha/\mu$. If $\mu n/3 > 3\alpha/\mu + 1$ then two of them are not connected, and this inequality holds for sufficiently large n. □

It remains to prescribe the answers in the first part such that there are linearly many red and green vertices in the end. $\mathcal{S}$ answers to the probe of the edge xy according to the colour of x and y and to the number R of red and G of green vertices existing at that time. Table 1 shows the strategy. The type number serves for reference (note that type numbers 4, 5 and 6 occur in two rows). c is some integral constant (only depending on α, $c \geqslant 3$) that will be determined later. The idea is as follows: $\mathcal{S}$ tries to make the number of one-coloured (i.e. red or green) or colourless vertices large while keeping the ratio G/R bounded (between $1/c$ and c, if possible). Questions of the form g–c or r–c (i.e., one vertex of the probed edge is green resp. red and the other is colourless) are more difficult to handle because these two objectives may become inconsistent. It is mainly this difficulty which forces us to choose $\alpha < 1.5$.

In the sequel we will prove several assertions which imply in the end that this strategy fulfills our requirements. Initially all vertices are colourless. Step 1 is of Type 1 and Step 2 of Type 1 or 5. Assertions (I) through (IV) are true after the

Table 1
Strategy for $\mathcal{S}c$ (g = green, r = red, m = manycoloured, c = colourless)

Type	Question	Answer	Condition
1	c–c		
2	m–c	red if $G \geqslant R$,	
3	r–g	green otherwise	
4	g–c	green	$G/R < c$ or $G = 1,\ R = 0$
5	g–c	red	$G = 2,\ R = 0$
6	g–c	red	else
4	r–c	red	$R/G < c$ or $R = 1,\ G = 0$
5	r–c	green	$R = 2,\ G = 0$
6	r–c	green	else
7	other cases (m–m, r–r, g–g, g–m, r–m)	such that $G + R$ is maximal afterwards	

first step:

(I). $R + G > 0$.

For $R + G$ is reduced only in a Type 3 step, and even this leaves one one-coloured vertex.

(II). $\quad R + G \geqslant 3 \;\Rightarrow\; \dfrac{1}{c} \leqslant \dfrac{G}{R} \leqslant c.$

This is true after the first step. Suppose it is true before Step k and let R', G' be the values of R, G after Step k, and suppose $R' + G' \geqslant 3$. If Step k has Type 1, 2, 3 or 7 then it is clear that the assertion is true for R', G' also. Suppose now that Step k has Type 4, 5 or 6 and the question was g–c. First, assume $R + G \geqslant 3$; then $G/R \leqslant c$ by assumption. If $G/R < c$ then we have Type 4 and

$$\frac{G'}{R'} = \frac{G+1}{R} \leqslant c.$$

If $G/R = c$ then we have Type 6 and

$$\frac{G'}{R'} = \frac{G-1}{R+1} = c - \frac{c+1}{R+1} \geqslant 1 > \frac{1}{c}.$$

Finally, if $R + G < 3$ we must have Type 4 (because $R' + G' > R + G$) and $R = G = 1$, hence the assertion is true in this case as well.

Assertions (III) and (IV) show that Type 5 and 6 cannot occur too often (they are bad in the sense that they destroy colourless vertices without increasing $G+R$):

(III). *Between any two steps of Type 6 there are at least* $(c-1)/2$ *steps not of Type* 6.

Proof. Let Step k be of Type 6, with the question g–c and numbers $G=G_0$, $R=R_0$. By (II) we have $G_0=cR_0$ and by (I) $R_0>0$, hence $G_0-R_0=(c-1)R_0\geqslant c-1$. Let $k<i\leqslant k+(c-1)/2$ and suppose Step i exists (i.e. $i\leqslant \alpha n$). Because $G-R$ is changed by at most 2 in every Step, we have $G-R\geqslant 0$ before Step i. From this we conclude that R is never reduced in Steps $k, k+1, \ldots, i-1$. Because Step k is of Type 6, R is increased by one in this step, and before Step i we have

$$G\leqslant G_0+2(i-k)\leqslant G_0+c-1=c(R_0+1)-1<cR.$$

Also, $R\leqslant G<cG$, and Step i cannot have Type 6. □

(IV). *Between any two steps of Type* 5 *there is at least one step of Type* 4.

For $G=2$, $R=0$ can only originate from $G=1$, $R=0$ by a Type 4 step (except in the beginning), and analogously for $G=0$, $R=2$.

Now let $r=\lfloor \alpha n\rfloor$, and denote by a_i the number of steps of Type i, $i=1, 3, 5, 6$, and by s the number of all the other steps, counting only Steps 2 to r.

(V). $a_3\leqslant 2a_1+s+1$.

For $R+G$ equals 2 after Step 1, increases by 2 in Type 1, by at most 1 in the steps counted by s and is left unchanged in Type 5 and 6; Type 3 reduces $R+G$ by 1, and in the end we have $R+G\geqslant 1$ by (I).

Now the main point is following.

(VI). *After r steps there are at most* $\frac{2}{3}r(1+1/(c+1))+\frac{1}{3}$ *manycoloured vertices.*

Proof. Denote the number of manycoloured vertices after r steps by m. Such vertices are created only in Type 3, 5 and 6, hence

$$m=a_3+a_5+a_6.$$

By (III), (IV), (V) we have

$$a_6\leqslant\frac{2r}{c+1}+1,\qquad a_5\leqslant s+1,\qquad a_3\leqslant 2a_1+s+1.$$

Hence

$$r-1=a_1+a_3+a_5+a_6+s=\left(a_1+\frac{s}{2}\right)+a_3+a_5+a_6+\frac{s}{2}$$

$$\geqslant \frac{a_3-1}{2}+a_3+a_5+a_6+\frac{a_5-1}{2}=\frac{3}{2}m-\frac{1}{2}a_6-1$$

and

$$m\leqslant \frac{2}{3}r+\frac{1}{3}a_6\leqslant \frac{2}{3}r\left(1+\frac{1}{c+1}\right)+\frac{1}{3}. \qquad \square$$

Now it's easy to see that our strategy is good: Because $r=\lfloor \alpha n\rfloor$ and $\alpha<\frac{3}{2}$, we can choose c so large (only depending on α) that $m\leqslant(1-\delta)n$ for some $\delta>0$. Hence after r questions there are at least δn colourless or unicoloured vertices. If at least three of them are colourless, we can take them as t, l, b. Otherwise by (II) there must be linearly many green and linearly many red vertices which was to be shown. □ (Theorem 8)

3.3. Graph properties with given complexity

The example of an interval shows that for a fixed set T there are properties of subsets of T with any given complexity between 0 and $|T|$. By Theorem 2, an analogous statement is not true if we restrict ourselves to graph properties. Nevertheless a weaker statement can be proved.

Theorem 9. *If $n\geqslant 5$ and $2n-4\leqslant c\leqslant \binom{n}{2}$, then there is a graph property $\mathcal{P}$ with*

$$|c(\mathcal{P})-c|<4n.$$

Proof. Consider the graph properties $\mathcal{S}c^{(p)}$, $p=0,\dots,n-5$. By Theorems 5 and 7

$$(3+p)n-\binom{p+4}{2}\leqslant c(\mathcal{S}c^{(p)})\leqslant(6+p)n-\binom{p+4}{2}.$$

The upper and lower bound differ by $3n$, the lower bound equals $3n-6$ for $p=0$ and increases by less than n when p increases by 1. Finally, the upper bound exceeds $\binom{n}{2}$ for $p=n-7$. Hence there is a p such that $|c(\mathcal{S}c^{(p)})-c|<4n$. □

Acknowledgement

I thank Martin Aigner at the Freie Universität Berlin for inciting my interest in this subject and for many interesting conversations.

References

[1] M. Aigner, Combinatorial Search (Wiley, New York, 1988).
[2] M. R. Best, P. van Emde Boas and H. W. Lenstra Jr, A sharpened version of the Aanderaa–Rosenberg conjecture, Report ZW 30/74, Math. Centrum Amsterdam, 1974.
[3] B. Bollobás and S. E. Eldridge, Packings of graphs and applications to computational complexity, J. Combin. Theory Ser. B 25 (78) 105–124.
[4] R. L. Holt and E. M. Reingold, On the time required to detect cycles and connectivity in graphs, Math. Systems Theory 6 (1972) 103–106.
[5] A. L. Rosenberg, On the time required to recognize properties of graphs: a problem, SIGACT News 5 (1973) 15–16.
[6] R. L. Rivest and V. Vuillemin, A generalization and proof of the Aanderaa–Rosenberg conjecture, Proc. 7th Annual ACM symp. Theory of Computing (SIGACT comf.) (1975) 6–11.
[7] H. P. Yao, Some Topics in Graph Theory (Cambridge Univ. Press, Cambridge, 1986).