Combinatorics, Probability and Computing (2010) 19, 121–131. © Cambridge University Press 2009
doi:10.1017/S0963548309990289

Finding an Unknown Acyclic Orientation of a Given Graph

OLEG PIKHURKO†

Department of Mathematical Sciences, Carnegie Mellon University,
Pittsburgh, PA 15213, USA
(web: http://www.math.cmu.edu/~pikhurko)

Received 26 November 2008; revised 7 April 2009; first published online 9 July 2009

Let $c(G)$ be the smallest number of edges we have to test in order to determine an unknown acyclic orientation of the given graph G in the worst case. For example, if G is the complete graph on n vertices, then $c(G)$ is the smallest number of comparisons needed to sort n numbers.

We prove that $c(G) \leqslant (1/4 + o(1))\, n^2$ for any graph G on n vertices, answering in the affirmative a question of Aigner, Triesch and Tuza [*Discrete Mathematics* **144** (1995) 3–10]. Also, we show that, for every $\varepsilon > 0$, it is NP-hard to approximate the parameter $c(G)$ within a multiplicative factor $74/73 - \varepsilon$.

1. Introduction

The *acyclic orientation game* is as follows. There are two players, *Algy* and *Strategist*, to whom we shall also refer as *him* and *her* correspondingly. Let G be a given graph, known to both players. At each step of the game, Algy selects any edge of G and Strategist has to orient this edge. The only restriction on Strategist's replies is that the revealed orientation has to be *acyclic*, that is, it does not contain directed cycles. The game ends when the current partial orientation extends to a *unique* acyclic orientation of the whole graph G. Algy tries to minimize the number of steps while Strategist aims at the opposite. Let $c(G)$ be the length of the game, assuming that both players play optimally.

In other words, Algy wants to discover a 'hidden' acyclic orientation of G by querying edges. The parameter $c(G)$ measures the worst-case complexity, that is, it is the smallest number such that Algy has a strategy that needs at most $c(G)$ steps for every acyclic orientation of G.

The special case when $G = K_n$ (the complete graph on n vertices) is equivalent to the well-known *minimum-comparison sorting* problem. While the asymptotic result $c(K_n) = (1 + o(1))\, n \log_2 n$ of Ford and Johnson [9] is not hard to prove, the exact computation of $c(K_n)$ seems

† Partially supported by the National Science Foundation, grants DMS-0457512 and DMS-0758057.

very difficult. For example, the problem of computing $c(K_{13})$ that appeared in Knuth's book [15, Chapter 5.3.1, Exercise 35] was solved only some 30 years later by Peczarski [21] (see also [22]).

One interpretation of $c(G)$ for a general order-n graph G is that Algy has to discover as much information as possible about the relative order of n elements given that certain pairs (namely, those corresponding to the edges of the complementary graph $\overline{G}$) cannot be queried. Manber and Tompa [17, 18] considered a related but different problem where the player can query *any* of the $\binom{n}{2}$ possible pairs but has to find the relative order for every edge of the given graph G.

Various results and bounds on $c(G)$ for general graphs were obtained by Aigner, Triesch and Tuza [2], who in particular studied graphs with $c(G) = e(G)$, calling them *exhaustive*. Even this property seems out of grasp. For example, the computational complexity of checking whether $c(G) = e(G)$ is not known; see Tuza [29, Problem 58]. Alon and Tuza [3] studied $c(G)$, where $G \in \mathcal{G}_{n,p}$ is a random graph of order n with edge probability p. They obtained, among other results, the correct order of magnitude when p is a non-zero constant: in this case $c(G) = \Theta(n \log n)$ almost surely.

The parameter $c(G)$ is not monotone with respect to the subgraph relation. For example, while $c(K_n) = (1 + o(1))n \log_2 n$, there are graphs G of order n with $c(G) \geqslant \lfloor n^2/4 \rfloor$. Indeed, let G be obtained from the *Turán graph* $T_2(n)$, the complete bipartite graph with vertex parts V_1 and V_2 of size $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$, by adding an arbitrary bipartite graph H inside one part of $T_2(n)$. Let $V(H) = U_1 \cup U_2$ be a bipartition of H. Suppose, for example, that $U_1, U_2 \subseteq V_1$. Strategist, in her replies, orients all edges from U_1 to $V(G) \setminus U_1$ and from V_2 to $V_1 \setminus U_1$. It is easy to see that Algy has to ask about the orientation of *every* edge of the original Turán graph $T_2(n)$, giving the claimed bound $c(G) \geqslant \lfloor n/2 \rfloor \lceil n/2 \rceil = \lfloor n^2/4 \rfloor$. We did not see any improvement over this bound in the literature; our Proposition 3.1 improves it by 1.

Aigner, Triesch and Tuza [2, p. 10] asked whether the above bound is asymptotically sharp, that is, whether $c(G) \leqslant (1/4 + o(1))\, n^2$ for every graph G of order n. This open question is also mentioned by Alon and Tuza [3, p. 263] and by Tuza [29, Problem 55]. Here we answer it in the affirmative.

Theorem 1.1. *For every $\varepsilon > 0$ there exists n_0 such that $c(G) \leqslant (1/4 + \varepsilon)n^2$ for every graph G of order $n \geqslant n_0$.*

Aigner, Triesch and Tuza [2, p. 10] also asked if, furthermore, the upper bound can be improved to $n^2/4 + C$ for some absolute constant C. Unfortunately, we cannot prove this strengthening.

Our proof of Theorem 1.1 shows more. Namely, for every $\varepsilon > 0$ there exists C such that, for any graph G of order n, Algy can point (in one go) a set D of at most $Cn^{3/2}(\ln n)^{1/2}$ edges so that every acyclic orientation of D implies the orientation of all but at most $(1/4 + \varepsilon)n^2$ remaining edges of G. Strategies of this type (when Algy has to send his questions in a few *rounds*) are useful in situations where the main limitation is on the number of times that the players can exchange (large amounts of) information. The study of comparison sorting in rounds was initiated by Valiant [30]. We refer the reader to a survey by Gasarch, Golub and Kruskal [10] for more information on the topic.

Aigner, Triesch and Tuza [2, p. 10] also asked about the computational complexity of deciding whether $c(G) \leqslant k$ on the input (G, k); see also Tuza [29, Problem 59]. We obtain some progress on this question as follows.

Theorem 1.2. *For every* $\varepsilon > 0$ *it is NP-hard to approximate* $c(G)$ *within a multiplicative factor* $74/73 - \varepsilon$.

It is possible that the acyclic orientation game is PSPACE-complete, but the author could not show this.

Also, one would like to complement Theorem 1.2 by providing a polynomial-time algorithm that approximates $c(G)$ within a multiplicative factor $O(1)$. Unfortunately, the best approximability ratio in terms of $n = v(G)$ that the author could find is $O(n/\log n)$: output $e(G)$ as an upper bound on $c(G)$ and $e(G)\log_2 n/(Cn)$ as a lower bound, where C is the constant given below in Theorem 5.1. It is a remaining open problem to close this gap.

2. Notation

We will use the standard graph terminology that can be found, for example, in the books by Bollobás [4] or Diestel [7]. Some of the less common conventions are as follows.

For brevity, we usually abbreviate an unordered pair $\{x, y\}$ as xy. We write (x, y) to denote that an edge xy is oriented from x to y. Let $[n] = \{1, \ldots, n\}$.

For a graph G and disjoint sets of vertices $X, Y \subseteq V(G)$, $G[X, Y]$ denotes the bipartite graph on $X \cup Y$ consisting of all edges of G connecting X to Y. A *cut* of G is a partition $V(G) = V_1 \cup V_2$. Its *value* is $e(G[V_1, V_2])$, the number of edges connecting V_1 to V_2. If the graph G comes equipped with the *edge weight* function $w : E(G) \to \mathbb{R}$, then the *value* of the cut $\{V_1, V_2\}$ is $\sum_{x_1 \in V_1} \sum_{x_2 \in V_2} w(x_1 x_2)$. The *max-cut* parameter $\text{MAX-CUT}(G)$ is the maximum value of a cut of G.

A partial order $\prec$ on $V(G)$ and an acyclic orientation of $E(G)$ are *compatible* if, for every edge $xy \in E(G)$, the elements x and y are comparable in the $\prec$-ordering and, moreover, (x, y) if and only if $x \prec y$. In this case, the phrases and expressions '(x, y)', 'y is above x', 'x is smaller than y', '$y \succ x$', and so on, are all synonymous.

3. Bounding $c(G)$ for order-n graphs

Proposition 3.1. *For every* $n \geqslant 3$ *there is a graph G of order n with* $c(G) \geqslant \lfloor n^2/4 \rfloor + 1$.

Proof. Let G be the complete 3-partite graph with parts $X \cup Y \cup Z$ where $|X| = |Y| = \lfloor (n-1)/2 \rfloor$. (Thus, depending on the parity, $n = 2k+1$ or $n = 2k$, the part sizes are either $(k, k, 1)$ or $(k-1, k-1, 2)$.)

Strategist orients (x, y) for every $x \in X$ and $y \in Y$ and answers Algy's questions about these edges accordingly.

For every $z \in Z$, Strategist does the following. She waits until Algy queries an edge incident to z for the first time. If this is an edge xz with $x \in X$, then Strategist orients all edges from $X \cup Y$ to z (and answers all Algy's questions accordingly). Note that Algy has to query every edge yz with $y \in Y$ because neither of its orientations would create a directed cycle in Strategist's ordering. Thus, Algy has to query at least $|Y| + 1$ edges at z (including the first edge xz). Likewise, if the first queried edge was yz with $y \in Y$, then Strategist orients all edges from z to $X \cup Y$ and Algy has to query all edges xz with $x \in X$.

Also, independently of the game scenario on the edges adjacent to Z, Algy has to query all edges between X and Y. Thus $c(G) \geqslant |X| \times |Y| + |Z|(|X|+1)$, which is easily seen to be the required bound. $\square$

Next, we prove Theorem 1.1. Its proof, where Algy queries random edges in the first round, is somewhat similar to the methods of Bollobás and Rosenfeld [5] (see also Häggkvist and Hell [11]), who studied how much information about the unknown linear order can be obtained in just one round with the given number of queries.

In order to prove Theorem 1.1 we will need the following auxiliary result.

Theorem 3.2 (Ruzsa and Szemerédi [23]). *For every $\varepsilon > 0$ there exists $\delta > 0$ such that if a graph G of order n has at most δn^3 triangles then we can remove at most εn^2 edges from G, making it triangle-free.*

Proof of Theorem 1.1. Given $\varepsilon > 0$, let $\delta = \delta(\varepsilon/2) > 0$ be the constant returned by Theorem 3.2 on the input $\varepsilon/2$. Fix an arbitrary positive constant C such that $C^2 > 2(1+\delta)/\delta^2$. Let n be sufficiently large. Let G be an arbitrary graph of order n. Let $V = V(G)$, $p = C\sqrt{\ln n/n}$, and $w = \lfloor \delta n/2 \rfloor$.

Algy selects a set D of edges of G by including each element of $E(G)$ into D with probability p, independently of the other choices. Let the acronym w.h.p. stand for 'with high probability', meaning with probability $1 - o(1)$ as $n \to \infty$.

Claim 1. With high probability, the following holds for every linear ordering $L = (V, \prec)$ of V and every $2w$ pairwise distinct vertices $x_1, \dots, x_w, y_1, \dots, y_w \in V$ with $x_i \prec y_i$ for $i \in [w]$. For $i \in [w]$, define

$$Z_i = \{z \in V : x_i \prec z \prec y_i,\ x_i z, z y_i \in E(G)\} \setminus \{x_1, \dots, x_{i-1}, y_1, \dots, y_{i-1}\}. \tag{3.1}$$

If each Z_i has at least δn elements, then there are $i \in [w]$ and $z \in Z_i$ such that $x_i z$ and $z y_i$ belong to D.

Proof of claim. Fix any linear order $\prec$ on V and any $2w$ pairwise distinct vertices $x_1, \dots, x_w$, $y_1, \dots, y_w$ such that $x_i \prec y_i$ and $|Z_i| \geqslant \delta n$ for each $i \in [w]$. Clearly, there are at most $n!\, n^{2w}$ choices of such a configuration.

The probability that this configuration violates the claim is at most $(1-p^2)^{\delta n w}$ because there are at least $w \times \delta n$ choices of (i, z) with $i \in [w]$ and $z \in Z_i$, the probability that at least one of the edges $x_i z$ and $z y_i$ of G is not in D is $1 - p^2$, while these probabilities are independent over distinct pairs (i, z). (Indeed, the events for different pairs (i, z) involve disjoint sets of edges; this was the reason for excluding any vertex in $\{x_1, \dots, x_{i-1}, y_1, \dots, y_{i-1}\}$ from Z_i in (3.1).)

The union bound shows that the total probability of failure is at most

$$n!\, n^{2w}(1-p^2)^{\delta n w} < \mathrm{e}^{n \ln n + 2w \ln n - p^2 \delta n w} \leqslant \mathrm{e}^{(1+\delta - C^2\delta^2/2 + o(1))\, n \ln n}.$$

This is $o(1)$ by the choice of C. The claim is proved. $\square$

Also, w.h.p. $|D| \leqslant pn^2$ by the Chernoff bound [6]. Hence, there is a set D that satisfies the conclusion of Claim 1 and has at most pn^2 elements. Fix such a set D.

During the first round, Algy asks about the orientation of all edges in D. After we have received Strategist's answers, let H be the spanning subgraph of G that consists of those edges of G whose orientation is still undetermined from the revealed orientation of D.

We claim that H has at most δn^3 triangles. Suppose on the contrary that this is false. Fix an arbitrary linear ordering $\prec$ of V that is compatible with the orientation of D. Let us define x_i and y_i inductively on i. Suppose that $i \in [w]$ and we have already defined $x_1, \dots, x_{i-1}$ and $y_1, \dots, y_{i-1}$.

Let $U = \{x_1, \dots, x_{i-1}, y_1, \dots, y_{i-1}\}$. The vertices in U belong to at most $2(i-1)\binom{n}{2} < wn^2$ triangles of H. So, the graph $H' = H - U$ has at least $\delta n^3 - wn^2 \geqslant \delta n^3/2$ triangles. By averaging, H' contains a pair of vertices x_i and y_i such that there are at least $(\delta n^3/2)/\binom{n}{2} > \delta n$ vertices $z \in V(H') = V \setminus U$ for which $x_i \prec z \prec y_i$ and $\{x_i, y_i, z\}$ spans a triangle in H'. Now, increase i by 1 and iterate the above step if the new index i is still at most w.

For $i \in [w]$, let Z_i be defined by (3.1); we have $|Z_i| \geqslant \delta n$. The obtained vertices $x_1, \dots, x_w$, $y_1, \dots, y_w$ satisfy all assumptions of Claim 1 with respect to the linear order $\prec$. By the definition of D (which was chosen to satisfy the conclusion of Claim 1), there are $i \in [w]$ and $z \in Z_i$ such that $x_i z, z y_i \in D$. By the definition of Z_i, we have $x_i \prec z \prec y_i$. Since $\prec$ was chosen to be compatible with Strategist's replies, the edges $x_i z, y_i z \in D$ are oriented as (x_i, z) and (z, y_i). Note that x_i and y_i are adjacent in H because these two vertices belong to at least $\delta n \geqslant 1$ triangles of H by the definition of x_i and y_i. But then the orientation of the edge $x_i y_i \in E(G)$ is determined after the first round, contradicting the fact that $x_i y_i \in E(H)$. Thus the graph H of order n has at most δn^3 triangles.

By the choice of δ (that is, by Theorem 3.2) there is a set F of at most $\varepsilon n^2/2$ edges such that $E(H) \setminus F$ contains no triangles. By the Turán theorem [28] (or rather the special case which was earlier proved by Mantel [19]) we have $|E(H) \setminus F| \leqslant n^2/4$. Thus $e(H) \leqslant n^2/4 + \varepsilon n^2/2$.

In the second round, Algy asks about the orientation of all edges of H. By the definition of H, this completely determines the orientation of all edges of G. Assuming that Strategist plays optimally, we have

$$c(G) \leqslant |D| + e(H) \leqslant pn^2 + \left(\frac{n^2}{4} + \frac{\varepsilon n^2}{2}\right) \leqslant \frac{n^2}{4} + \varepsilon n^2,$$

finishing the proof of Theorem 1.1. □

Remark. All known proofs of Theorem 3.2 use some version of Szemerédi's Regularity Lemma [25] and therefore return a function $\delta(\varepsilon)$ that approaches 0 extremely slowly and is of little practical value. Tao Jiang [14] observed that, instead of Theorem 3.2, one can use the result of Moon and Moser [20] that a graph of order n and size m contains at least $(m/3n)(4m - n^2)$ triangles. His calculations [14] based on this idea show that $c(G) \leqslant n^2/4 + 2n^{7/4}(\ln n)^{1/4}$ for any order-n graph G with n large. On the other hand, if we use Theorem 3.2, then we can deduce some structural information about almost extremal graphs. Namely, if an order-n graph G satisfies $c(G) = (\frac{1}{4} + o(1))n^2$, then by the Stability Theorem of Erdős [8] and Simonovits [24] applied to the triangle-free graph $H \setminus F$, there is a partition $V(G) = V_1 \cup V_2$ with $(\frac{1}{4} + o(1))n^2$ edges going across (which is asymptotically largest possible). Unfortunately, neither of these two approaches has led us to the complete answer so far.

4. Inapproximability results

In order to prove Theorem 1.2 we will need the following auxiliary result.

Lemma 4.1. *For every $\delta > 0$, it is NP-hard to approximate the graph parameter*

$$3e(G) + \text{MAX-CUT}(G)$$

within a multiplicative factor $74/73 - \delta$.

Proof. We will use the construction of Håstad [12, 13] that demonstrates that MAX-CUT is NP-hard to approximate within a factor $17/16 - \varepsilon$. Since we are interested in a parameter somewhat different from just MAX-CUT, we have to unfold Håstad's construction.

First, Håstad proves [12, Theorem 2.3] that it is NP-hard to approximate E3-LIN-2 within a factor less than 2. That is, for every $\varepsilon > 0$ it is NP-hard to distinguish, for an input system $\mathcal{S}$ of s equations over $\mathbb{Z}_2$ each of the form $x + y + z = 0$ or $x + y + z = 1$, between the cases when some assignment of variables satisfies at least $(1 - \varepsilon)s$ equations and when every assignment satisfies at most $(1/2 + \varepsilon)s$ equations.

Next, Håstad constructs [12, Theorem 4.2] a graph G from a given instance $\mathcal{S}$ of E3-LIN-2 with s equations as follows. We can assume that $s_0 \geqslant s/2$ equations of $\mathcal{S}$ are of the form $x + y + z = 0$. (If $s_0 < s/2$, we can simply replace each variable x by $1 - x$.) Let $s_1 = s - s_0$ be the number of equations of the form $x + y + z = 1$.

Each equation $x + y + z = 0$ and $x + y + z = 1$ is replaced respectively by the so-called 8-*gadget* and 9-*gadget* of Trevisan, Sorkin, Sudan and Williams [26, 27]. The definition of these gadgets can be found in the journal version [27, Lemmas 4.2 and 4.3]. For our purposes we need to know only that, for $\alpha = 8$ or 9, this particular α-gadget is an edge-weighted graph of total edge weight $\alpha + 1$ whose vertex set consists of the variables x, y, and z, the constant 0, and some new vertices so that:

- every $0/1$-assignment of x, y, and z that satisfies the equation can be extended to a cut of value at least α but not to a cut of a strictly larger value;
- no $0/1$-assignment of x, y, and z that violates the equation can be extended to a cut of value strictly larger than $\alpha - 1$.

(Here, a cut in a gadget H is encoded by an assignment $f : V(H) \to \{0, 1\}$ with $f(0) = 0$.) Also, the special vertices (the variables and the constant 0) form an independent set in both gadgets. Thus the constructed graph G has total edge weight $9s_0 + 10s_1$.

The above properties imply that if we can satisfy at least $(1 - \varepsilon)s$ equations of $\mathcal{S}$ then G has a cut of value at least $8s_0 + 9s_1 - 10\varepsilon s$. Also, if every assignment of variables violates at least $(1/2 - \varepsilon)s$ equations, then no cut of G can have value larger than $8s_0 + 9s_1 - (1/2 - \varepsilon)s$. Thus, if we cannot distinguish these two alternatives for E3-LIN-2 in polynomial time, then we cannot distinguish in polynomial time whether, for edge-weighted graphs, $3e(G) + \text{MAX-CUT}(G)$ is at least u_1 or at most u_2, where

$$u_1 = 3(9s_0 + 10s_1) + 8s_0 + 9s_1 - 10\varepsilon s = -4s_0 + 39s - 10\varepsilon s,$$
$$u_2 = 3(9s_0 + 10s_1) + 8s_0 + 9s_1 - (1/2 - \varepsilon)s = -4s_0 + 38.5s + \varepsilon s.$$

When $\varepsilon < 1/22$, then the ratio u_1/u_2 is minimized when $s_0 = s/2$ is as small as possible. Thus $u_1/u_2 \geqslant 74/73 - o(1)$ as $\varepsilon \to 0$, giving the inapproximability result for edge-weighted graphs.

Finally, we can get rid of edge weights by choosing a large integer l, say $l = s$, cloning each vertex of G l times, and replacing each edge of weight α by a pseudo-random bipartite graph of edge density α. (The edge weights in each gadget are real numbers lying between 0 and 1.) We refer the reader to a survey by Krivelevich and Sudakov [16] on the properties of pseudo-random graphs. Up to a multiplicative error $1 + o(1)$ as $l \to \infty$, any cut of the new graph G' corresponds to a *fractional cut* of G, where the vertices of G may be sliced between the two parts in some ratio and the value of the cut is defined in the obvious way. However, it is easy to see that, for an arbitrary edge-weighted (loopless) graph, there is an integer vertex cut which is at least as good as any fractional cut. Thus $\text{MAX-CUT}(G') = (1 + o(1))l^2\text{MAX-CUT}(G)$ (and $e(G') = (1 + o(1))l^2 e(G)$) as $l \to \infty$.

The obtained family of (unweighted) graphs G' establishes the lemma. □

Remark. The weaker result that it is NP-hard to approximate $3e(G) + \text{MAX-CUT}(G)$ within a factor $113/112 - \delta$ can be obtained from the statement of the 17/16-result of Håstad (without analysing the structure of his graphs) by observing that $\text{MAX-CUT}(G) \geqslant \frac{1}{2}\, e(G)$ for any G (and doing some easy calculations).

Proof of Theorem 1.2. Let l be a positive integer and let G be an arbitrary graph. Define $n = v(G)$, $m = e(G)$, and $t = \text{MAX-CUT}(G)$.

We construct a new graph $H = H(G, l)$ as follows. Let $V = V(G)$. For each $x \in V$, introduce a new vertex x'. Let $V' = \{x' : x \in V\}$ consist of all new vertices. For each edge $xy \in E(G)$, introduce a set U_{xy} of l new vertices. Let $U = \cup_{xy \in E(G)} U_{xy}$. The new graph H has $V \cup V' \cup U$ for the vertex set. Thus the total number of vertices is $v(H) = 2n + lm$. The edges of H are as follows. Let V span the complete graph. Connect x to x' for each $x \in V$. Put a complete bipartite graph between U_{xy} and $\{x, y, x', y'\}$ for every $xy \in E(G)$. These are all the edges (all other pairs of $V(H)$ are non-adjacent). Thus, for example, the size of H is $e(H) = \binom{n}{2} + n + 4lm$.

Claim 1. $c(H) \geqslant 3lm + lt$.

Proof of claim. Let $V = X \cup Y$ be a maximum cut of G, that is, $e(G[X, Y]) = t$. Let $V' = X' \cup Y'$ be the corresponding partition of V'. Let $X = \{x_1, \ldots, x_a\}$ and $Y = \{y_1, \ldots, y_b\}$.

Let $P = (V(H), \preceq)$ be the partially ordered set on $V(H)$, where $\prec$ is the transitive closure of the digraph D that consists of the following ordered pairs:

- (x_i, x_{i+1}) and (x'_i, x'_{i+1}) for $i \in [a - 1]$,
- (y_i, y_{i+1}) and (y'_i, y'_{i+1}) for $i \in [b - 1]$,
- (x_i, x'_i) for $i \in [a]$,
- (y'_i, y_i) for $i \in [b]$,
- (x_a, y_1) and (y'_b, x'_1),
- (x_i, u), (u, x_j), and (u, x'_i) for $u \in U_{x_i x_j}$ and $x_i x_j \in E(G[X])$ with $i < j$,
- (y_i, u), (y'_j, u), and (u, y_j) for $u \in U_{y_i y_j}$ and $y_i y_j \in E(G[Y])$ with $i < j$,
- (x, u), (y', u), (u, x'), and (u, y) for $u \in U_{xy}$ and $xy \in E(G[X, Y])$ with $x \in X$.

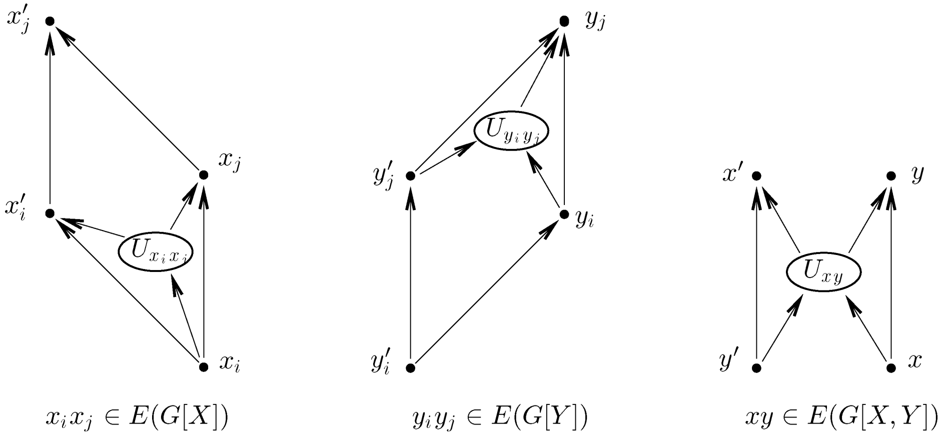

Figure 1. The placement of the set U relative to $V \cup V'$.

In other words, we take two chains, namely $x_1 \prec \cdots \prec x_a \prec y_1 \prec \cdots \prec y_b$ and $y'_1 \prec \cdots \prec y'_b \prec x'_1 \prec \cdots \prec x'_a$. We let $x \prec x'$ for $x \in X$ and $y \succ y'$ for $y \in Y$. For all $x_i \prec x_j$ that are adjacent in G, we insert the set $U_{x_ix_j}$ (as an antichain) above x_i but below x_j and x'_i. For all $y_i \prec y_j$ that are adjacent in G, we insert the set $U_{y_iy_j}$ (as an antichain) above y_i and y'_j but below y_j. For each $xy \in E(G[X, Y])$ with $x \in X$, we insert the set U_{xy} (as an antichain) above x and y' but below x' and y. Figure 1 shows the placement of the vertices of U relative to $V \cup V'$. Finally, we add those order relations that are implied by the above relations.

It is easy to check that D has no oriented cycles and that the obtained partial order P determines the orientation of every edge of H. Strategist chooses this orientation and answers all Algy's questions accordingly.

The digraph D defined above is not in general the Hasse diagram of the poset P: for example, if x_ix_{i+1} is an edge of G, then the relation $x_i \prec x_{i+1}$ can be determined from $x_i \prec u \prec x_{i+1}$ for some $u \in U_{x_ix_{i+1}}$. However (and this is the crucial property!) one can routinely check that every arc of D that connects $V \cup V'$ and U (in either direction) does belong to the Hasse diagram of P, that is, the orientation of this edge is not determined from the order relation of all other pairs of P.

Clearly, Algy has to query *every* edge that belongs to the Hasse diagram of P. Thus, Algy has to ask at least $3l$ (resp. $4l$) questions per edge of $G[X]$ and $G[Y]$ (resp. $G[X, Y]$). This shows that $c(H) \geqslant 3l(m - t) + 4lt = 3lm + lt$, as required. □

Claim 2. $c(H) \leqslant 3lm + lt + c(K_n) + n$.

Proof of claim. Algy finds the orientation of all edges in the clique $H[V]$ by asking $c(K_n)$ questions. Then he asks about the orientation of every edge xx' with $x \in V$. Let X consist of those $x \in V$ for which we have (x, x'). Let $Y = V \setminus X$.

Take any $xy \in E(G[X])$. Suppose without loss of generality that $x \prec y$. For each $u \in U_{xy}$, Algy asks about the orientation of the edge uy. Whatever the answer is, it determines the orientation of ux or uy'. Hence, at most $3l$ questions are enough to determine the orientation of all edges incident to U_{xy}. The same applies to the case $xy \in E(G[Y])$. Finally, Algy asks about all edges incident to U_{xy} where $xy \in E(G[X, Y])$, posing $4l$ questions per edge of the cut $\{X, Y\}$. Thus the total number of questions is at most

$$c(K_n) + n + 3l\big(e(G[X]) + e(G[Y])\big) + 4le(G[X, Y]) = c(K_n) + n + 3le(G) + le(G[X, Y]),$$

giving the required bound. □

As was shown by Ford and Johnson [9], $c(K_n) = (1 + o(1))n \log_2 n$. Thus Claims 1 and 2 show that $c(H) = (1 + o(1))l(3e(G) + \text{MAX-CUT}(G))$ as $n \to \infty$, if we take $l \gg \ln n$, say $l = n$. (Note that, by removing isolated vertices from G, we can assume that $e(G) \geqslant v(G)/2$.) Since the order of H is bounded by a polynomial in $v(G)$, the desired inapproximability result for the parameter c follows from Lemma 4.1. □

5. A general lower bound on $c(G)$

Here is the lower bound on $c(G)$ that implies the approximability result mentioned at the end of the Introduction.

Theorem 5.1. *There is a constant $C > 0$ such that any graph G satisfies*

$$c(G) \geqslant \frac{e(G)\log_2(v(G))}{Cv(G)}. \tag{5.1}$$

Proof. Fix a sufficiently large C. Let G be an arbitrary graph of order n and size m.

Clearly, it is enough to prove the theorem under the assumption that G has no isolated vertices. Indeed, if we remove isolated vertices, then $c(G)$ remains the same while the right-hand side of (5.1) can only increase.

We have $c(G) \geqslant n/2$ because, for every vertex x of G, we have to query at least one edge incident to x. It follows that (5.1) holds unless

$$m > \frac{Cn^2}{2\log_2 n}. \tag{5.2}$$

So suppose that (5.2) holds. The average degree of G is $2m/n$. If we remove a vertex whose degree is less than m/n, then the average degree of G goes up. By iteratively repeating this step, we can find a non-empty set $X \subseteq V(G)$ such that the induced subgraph $H = G[X]$ has minimum degree at least $d = \lceil m/n \rceil$.

The graph H contains at least $|X|\, d! \geqslant (d + 1)!$ directed paths P of length d: there are $|X|$ choices for the first vertex and, inductively for $i = 2, \ldots, d + 1$, at least $d - i + 2$ choices for the ith vertex. For every choice of P choose an acyclic orientation of the whole graph G compatible with the orientation of P. Clearly, each orientation of G can appear this way for at most $\binom{n}{d+1} \leqslant 2^n$ different directed d-paths P. Hence, $a(G)$, the number of acyclic orientations of G, is at least $(d + 1)!/2^n$. The usual information-theoretic lower bound (see, for example, Aigner [1, p. 24]) implies that

$$c(G) \geqslant \log_2(a(G)) \geqslant \log_2\left(\frac{(d + 1)!}{2^n}\right).$$

If C is large, then also n is large by (5.2) and because $m \leqslant \binom{n}{2}$, namely $n > 2^C$. Again by (5.2), we have

$$d \geqslant \frac{m}{n} > \frac{Cn}{2\log_2 n} \quad \text{and} \quad \log_2 d > \frac{\log_2 n}{2}, \tag{5.3}$$

so d is forced to be large too. By Stirling's formula, $\log_2((d+1)!) > 0.9\, d\log_2 d$. We have by (5.3) that, for example, $d\log_2 d > (Cn/(2\log_2 n)) \times (\log_2 n)/2 > 2n$. Thus

$$\log_2\left(\frac{(d+1)!}{2^n}\right) > 0.9\, d\log_2 d - n > 0.4\, d\log_2 d \geqslant 0.4 \times \frac{m}{n} \times \frac{\log_2 n}{2},$$

as required. □

Remark. The inequality in (5.1) is sharp (up to an $O(1)$-factor) when G is the complete graph K_n or, more generally, when G is a typical graph in $\mathcal{G}_{n,p}$ with constant edge probability $p > 0$ by the result of Alon and Tuza [3].

Acknowledgements

The author thanks Alan Frieze, Tao Jiang and Oleg Verbitsky for helpful comments.

References

[1] Aigner, M. (1988) *Combinatorial Search*, Wiley–Teubner Series in Computer Science, Wiley, Chichester.

[2] Aigner, M., Triesch, E. and Tuza, Z. (1995) Searching for acyclic orientations of graphs. *Discrete Math.* **144** 3–10.

[3] Alon, N. and Tuza, Z. (1995) The acyclic orientation game on random graphs. *Random Struct. Alg.* **6** 261–268.

[4] Bollobás, B. (1998) *Modern Graph Theory*, Springer, Berlin.

[5] Bollobás, B. and Rosenfeld, M. (1981) Sorting in one round. *Israel J. Math.* **38** 154–160.

[6] Chernoff, H. (1952) A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

[7] Diestel, R. (2006) *Graph Theory*, 3rd edn, Springer, Berlin.

[8] Erdős, P. (1967) Some recent results on extremal problems in graph theory: Results. In *Theory of Graphs* (Rome 1966), Gordon and Breach, New York; pp. 117–123 (English); pp. 124–130 (French).

[9] Ford, L. R. and Johnson, S. M. (1959) A tournament problem. *Amer. Math. Monthly* **66** 387–389.

[10] Gasarch, W., Golub, E. and Kruskal, C. (2003) Constant time parallel sorting: An empirical view. *J. Comput. Syst. Sci.* **67** 63–91.

[11] Häggkvist, R. and Hell, P. (1981) Parallel sorting with constant time for comparisons. *SIAM J. Comput.* **10** 465–472.

[12] Håstad, J. (1999) Some optimal inapproximability results. In *Proc. 29th ACM Symposium on Theory of Computing* (El Paso 1997), pp. 1–10. ACM, New York.

[13] Håstad, J. (2001) Some optimal inapproximability results. *J. Assoc. Comput. Mach.* **48** 798–859.

[14] Jiang, T. (2008) Personal communication.

[15] Knuth, D. E. (1973) *The Art of Computer Programming, Vol. 3: Sorting and Searching*, Addison-Wesley.

[16] Krivelevich, M. and Sudakov, B. (2006) Pseudo-random graphs. In *More Sets, Graphs and Numbers*, Vol. 15 of *Bolyai Society Mathematical Studies*, Springer, Berlin, pp. 199–262.

[17] Manber, U. and Tompa, M. (1981) The effect of number of Hamiltonian paths on the complexity of a vertex-coloring problem. In *Proc. 22nd Annual Symposium on Foundations of Computer Science* (Nashville 1981), IEEE Computer Society Press, pp. 220–227.

[18] Manber, U. and Tompa, M. (1984) The effect of number of Hamiltonian paths on the complexity of a vertex-coloring problem. *SIAM J. Comput.* **13** 109–115.

[19] Mantel, W. (1907) Problem 28. *Winkundige Opgaven* **10** 60–61.

[20] Moon, J. W. and Moser, L. (1962) On a problem of Turán. *Publ. Math. Inst. Hungar. Acad. Sci.* **7** 283–287.

[21] Peczarski, M. (2002) Sorting 13 elements requires 34 comparisons. In *Proc. 10th Annual European Symposium on Algorithms*, Vol. 2461 of *Lecture Notes in Computer Science*, Springer, Berlin, pp. 785–794.

[22] Peczarski, M. (2004) New results in minimum-comparison sorting. *Algorithmica* **40** 133–145.

[23] Ruzsa, I. Z. and Szemerédi, E. (1978) Triple systems with no six points carrying three triangles. In *Combinatorics II* (A. Hajnal and V. Sós, eds), North-Holland, Amsterdam, pp. 939–945.

[24] Simonovits, M. (1968) A method for solving extremal problems in graph theory, stability problems. In *Theory of Graphs* (Tihany 1966), Academic Press, pp. 279–319.

[25] Szemerédi, E. (1976) Regular partitions of graphs. In *Proc. Colloq. Int. CNRS*, Paris, pp. 309–401.

[26] Trevisan, L., Sorkin, G. B., Sudan, M. and Williamson, D. P. (1996) Gadgets, approximation, and linear programming. In *Proc. 37th Annual Symposium on Foundations of Computer Science* (Burlington 1996), IEEE Computer Society Press, Los Alamitos, CA, pp. 617–626.

[27] Trevisan, L., Sorkin, G. B., Sudan, M. and Williamson, D. P. (2000) Gadgets, approximation, and linear programming. *SIAM J. Comput.* **29** 2074–2097.

[28] Turán, P. (1941) On an extremal problem in graph theory (in Hungarian). *Mat. Fiz. Lapok* **48** 436–452.

[29] Tuza, Z. (2001) Unsolved combinatorial problems. BRICS Lecture Series, LS-01-1. Available from: `http://www.brics.dk/publications/`.

[30] Valiant, L. G. (1975) Parallelism in comparison problems. *SIAM J. Comput.* **4** 348–355.