Ranking tournaments

N. Alon *

Abstract

A tournament is an oriented complete graph. The feedback arc set problem for tournaments is the optimization problem of determining the minimum possible number of edges of a given input tournament T whose reversal makes T acyclic. Ailon, Charikar and Newman showed that this problem is NP-hard under randomized reductions. Here we show that it is in fact NP-hard. This settles a conjecture of Bang-Jensen and Thomassen.

1 Introduction

A *tournament* is an oriented complete graph. A *feedback arc set* in a digraph is a collection of edges whose reversal (or removal) makes the digraph acyclic. The *feedback arc set problem* for tournaments is the optimization problem of determining the minimum possible cardinality of a feedback arc set in a given tournament. The problem for general digraphs is defined analogously. Bang-Jensen and Thomassen conjectured in [6] that this problem is NP-hard, and Ailon, Charikar and Newman proved in [1] that it is NP-hard under randomized reductions. Here we show how to derandomize a variant of the construction of [1] and prove that the problem is indeed NP-hard. This is based on the known fact that the minimum feedback arc-set problem for general digraphs is NP-hard, (see [7], p. 192), and on certain pseudo-random properties of the quadratic residue tournaments described in [4], pp. 134-137. Similar constructions can be given using any other family of antisymmetric matrices with $\{-1, 1\}$ entries whose rows are nearly orthogonal. We note that unlike the authors of [1], we do not apply the known fact that the minimum feedback arc set problem is APX-hard, and only need the simpler fact that it is NP-hard, proved more than thirty years ago. In fact, the proof in [1] can also be modified slightly so as to rely only on this fact (to get hardness of approximation under randomized reductions).

*Schools of Mathematics and Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel and IAS, Princeton, NJ 08540, USA. Email: nogaa@tau.ac.il. Research supported in part by the Israel Science Foundation, by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University and by the Von Neumann Fund.

2 Notation

For a digraph $G = (V, E)$ and a permutation π of its vertices, an oriented edge $(u, v) \in E$ is *consistent* with π if u precedes v in π. Let $FIT(G, \pi)$ denote the number of edges whose orientation is consistent with π minus the number of edges whose orientation is not consistent with π. Similarly, if the edges of G are weighted, we let $FIT(G, \pi)$ denote the total weight of the edges whose orientation is consistent with π minus the total weight of the edges whose orientation is not consistent with π. It is convenient to consider unweighted digraphs as weighted digraphs in which the weight of each edge is 1, and the weight of each non-edge is 0. Most of the weighted digraphs we use here have weights in $\{0, 1, -1\}$, but it is helpful to use weights that can be added and subtracted in order to simplify notation.

Returning to unweighted digraphs, let $FA(G)$ denote the minimum size of a feedback arc set of $G = (V, E)$. It is easy to see that $FA(G) = (|E| - max_\pi FIT(G, \pi))/2$, where the maximum is taken over all permutations π of V. This is because omitting a feedback arc set leaves the remaining graph acyclic, ensuring that there is a permutation π consistent with the orientation of all edges left, and similarly, for any π one can omit all edges not consistent with π and get an acyclic digraph.

If $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ are two (weighted) digraphs on the same set of vertices, the sum $G_1 + G_2$ is the digraph on V in which the weight of each edge is the sum of its weights in G_1 and in G_2. The difference $G_1 - G_2$ is defined in a similar manner. Note that for every permutation π on V, $FIT(G_1 + G_2, \pi) = FIT(G_1, \pi) + FIT(G_2, \pi)$ and $FIT(G_1 - G_2, \pi) = FIT(G_1, \pi) - FIT(G_2, \pi)$.

If G is a digraph, and $U \subset V$, then $G[U]$ denotes the induced subgraph of G on U. We consider this subgraph, however, as a digraph whose vertex set is V, where all vertices in $V - U$ are isolated. If U and W are two disjoint subsets of V, then $G[U, W]$ denotes the subgraph of G consisting of all edges of G with an end in U and an end in W. Here, too, the vertex set is the original set V. Let $e(U, W)$ denote the total number of edges of G that start at U and end at W. Thus, the total number of edges of $G[U, W]$ is $e(U, W) + e(W, U)$.

3 The quadratic residue tournaments

Let $p \equiv 3 \ mod \ 4$ be a prime, and let $T = T_p$ be the tournament whose vertices are all elements of the finite field $GF(p)$, in which (i, j) is a directed edge iff $i - j$ is a quadratic residue. In [4], pp. 134-137 it is shown that for every permutation π of the vertices of T_p, $FIT(T, \pi) \leq O(p^{3/2} \log p)$. Here we need a stronger result, providing a similar bound for certain subgraphs of T.

We need the following known fact, proved, for example, in [2] (see also [4], Lemma 9.1.2).

Lemma 3.1 *Let $T = T_p = (V, E)$ be the quadratic residue tournament defined above. Then, for every two disjoint sets U_1, U_2 of T,*

$$e(U_1, U_2) - e(U_2, U_1) \leq |U_1|^{1/2} |U_2|^{1/2} p^{1/2}.$$

Therefore, if $|U_1|$ and $|U_2|$ are large, then the number of edges of G oriented from U_1 to U_2 is roughly the number of edges oriented from U_2 to U_1, as the difference between these two numbers is a most $|U_1|^{1/2}|U_2|^{1/2}p^{1/2}$, whereas their sum is $|U_1||U_2|$. We next observe that this property implies that for every large set of vertices U of T, and for every permutation π, $FIT(T[U],\pi)$ is small.

Corollary 3.2 *Let $T = T_p = (V,E)$ be as above, let $U \subset V$ be a set of vertices of T, and let $T[U]$ denote the induced subgraph of T on U. Then, for every permutation π of V,*

$$|FIT(T[U],\pi)| \leq |U|\lceil \log_2 |U| \rceil p^{1/2} \leq |U| \log_2(2|U|)p^{1/2}.$$

Proof. We prove that for every set U of a most 2^r vertices, and for every permutation π

$$FIT(T[U],\pi) \leq r2^{r-1}p^{1/2}. \tag{1}$$

Note that if $\pi = \pi_1, \pi_2, \ldots, \pi_p$ and $\overline{\pi} = \pi_p, \pi_{p-1}, \ldots, \pi_1$, then $FIT(T[U],\overline{\pi}) = -FIT(T[U],\pi)$, and hence the validity of (1) implies the assertion of the corollary (including the absolute value). We prove (1) by induction on r. The result is trivial for $r = 1$. Assuming it holds for $r-1$ we prove it for r. Suppose $|U| \leq 2^r$. Given π, split U into two disjoint sets U_1, U_2, each of size at most 2^{r-1}, so that all the elements of U_1 precede all those of U_2 in the permutation π. Clearly

$$FIT(T[U],\pi) = e(U_1,U_2) - e(U_2,U_1) + FIT(T[U_1],\pi) + FIT(T[U_2],\pi).$$

By Lemma 3.1 and the induction hypothesis, the right hand side is at most

$$2^{r-1}p^{1/2} + 2(r-1)2^{r-2}p^{1/2} = r2^{r-1}p^{1/2}.$$

This completes the proof. $\Box$

Corollary 3.3 *Let $T = T_p = (V,E)$ be as above, let U, W be two disjoint subsets of vertices of T, and let $T[U,W]$ denote the bipartite subgraph of T consisting of all edges of T with an end in U and an end in W. Then, for every permutation π of V,*

$$|FIT(T[U,W],\pi)| \leq [\ (|U|+|W|)\lceil \log_2(|U|+|W|) \rceil + |U|\lceil \log_2 |U| \rceil + |W|\lceil \log_2 |W| \rceil\ ]p^{1/2}.$$

In particular, if $|U| \leq a$ and $|W| \leq a$, then $|FIT(T[U,W],\pi)| \leq 4a \log_2(4a)p^{1/2}$.

Proof. In the notation of Section 2, $T[U,W] = T[U \cup W] - T[U] - T[W]$. Therefore, for every π,

$$|FIT(T[U,W],\pi)| = |FIT(T[U \cup W],\pi) - FIT(T[U],\pi) - FIT(T[W],\pi)|,$$

and the desired result follows from the triangle inequality and three applications of the previous corollary. $\Box$

The *a-blow-up* of a digraph H, which we denote by $H(a)$, is the digraph obtained by replacing each vertex v of H by an independent set $I(v)$ of size a, and each directed edge (u,v) of H by a complete bipartite digraph containing all a^2 edges from the members of $I(u)$ to those of $I(v)$. It is easy to check that the minimum size of a feedback arc set of $H(a)$ satisfies $FA(H(a)) = a^2 FA(H)$. Indeed, this follows from the fact that if π is a permutation of the vertices of the blow up $H(a)$ that maximizes $FIT(H(a), \pi)$, and if x, y are two vertices of $H(a)$ that lie in the same $I(v)$, then one may always either shift x to lie right next to y in π or vice versa without decreasing the number of consistent edges.

Our main technical lemma is the following.

Lemma 3.4 *Let $H = (U, F)$ be a digraph, let $p \equiv 3 \bmod 4$ be a prime, and let $T = T_p = (V, E)$ be the quadratic residue tournament described above. Let a be an integer and suppose that $a|U| \leq p$. For each $u \in U$, let $I(u)$ be an arbitrary subset of size a of V, where all $|U|$ sets $I(u)$ are pairwise disjoint, and let T' be the tournament obtained from T as follows: for each edge $(u, v) \in F$ of H, omit all edges of T that connect members of $I(u)$ with those of $I(v)$, and replace them with all the a^2 directed edges that start at a member of $I(u)$ and end at one of $I(v)$. Then, for every permutation π of V:*

$$|FIT(T', \pi) - FIT(H(a), \pi)| \leq p^{3/2} \log_2(2p) + 4|F|a \log_2(4a) p^{1/2}.$$

Proof. Consider $H(a)$ as a digraph on the sets of vertices $I(u),\ u \in U$. By construction,

$$T' = T - \sum_{(u,v) \in F} T[I(u), I(v)] + H(a).$$

Therefore, for every π,

$$FIT(T', \pi) = FIT(T, \pi) - \sum_{(u,v) \in F} FIT(T[I(u), I(v)], \pi) + FIT(H(a), \pi).$$

It follows that

$$|FIT(T', \pi) - FIT(H(a), \pi)| \leq |FIT(T, \pi)| + \sum_{(u,v) \in F} |FIT(T[I(u), I(v)], \pi)|,$$

and the desired result follows from Corollary 3.2 that implies that $|FIT(T, \pi)| \leq p^{3/2} \log_2(2p)$ and from Corollary 3.3 that implies that for each fixed $(u, v) \in F$, $|FIT(T[I(u), I(v)], \pi)| \leq 4a \log_2(4a) p^{1/2}$.

4 The main result

Theorem 4.1 *The minimum feedback arc set problem for tournaments is NP-hard.*

Proof. It is known (cf., e.g., [7], p. 192) that the minimum feedback arc set problem is NP hard, even for digraphs H in which all outdegrees and indegrees are at most 3 (this last point is not essential here, but we use it to make the computation explicit). Given a digraph $H = (U, F)$ as above, let $a = |U|^c$ where $c > 3$ is a fixed integer, and let $p \equiv 3 \; mod \; 4$ be a prime between $|U|a$ and, say, $2|U|a$. Such a prime always exists, by the known results on primes in arithmetic progressions, and it is easy to find such a prime in time polynomial in $|U|$, by exhaustive search. Let T' be the tournament constructed from T_p and the blow-up $H(a)$ of H as described in Lemma 3.4. Computing $FA(T')$ is equivalent to computing $max_\pi FIT(T', \pi)$, where the maximum is taken over all permutations π of V. However, by Lemma 3.4 it follows that the value of $max_\pi FIT(T', \pi)$ provides an approximation up to an additive error of $p^{3/2}\log_2(2p) + |F|4a\log_2(4a)p^{1/2} \le 13p^{3/2}\log_2(4p)$ for $max_\pi FIT(H(a), \pi)$, where here we used the fact that $|F| \le 3|U|$ and the fact that $|U|a \le p$. Since, as explained after the proof of Corollary 3.3, $max_\pi FIT(H(a), \pi) = a^2 max_\sigma FIT(H, \pi)$, where the last maximum is taken over all permutations σ of the vertices of H, we conclude that if $a^2 > 13p^{3/2}\log_2(4p)$, this approximation will enable us to determine $max_\sigma FIT(H, \pi)$ (and hence also $FA(H)$) precisely. Since $a = |U|^c$ and $p \le 2|U|a \le 2|U|^{c+1}$, this is the case provided $c \ge 4$, completing the proof. □

5 Remarks and problems

- By choosing c appropriately in the above proof it follows that for every fixed $\epsilon > 0$, it is NP-hard to approximate $FA(T)$ for a tournament on n vertices up to an additive error of $n^{2-\epsilon}$. Note that approximating it up to an additive error of ϵn^2 can be done in polynomial time using the algorithmic version of the regularity lemma (for digraphs), or the methods of [5].

- It will be interesting to decide if the minimum feedback arc set problem for tournaments is APX-hard. The authors of [1] describe a randomized algorithm that provides a constant approximation of this quantity.

- The assertion of Lemma 3.1 here follows from the fact that the absolute value of the sum of entries in any submatrix of the p by p matrix B in which $B_{ij} = \chi(i-j)$, where χ is the quadratic character, can be bounded as described in the lemma. If $G = (V, E)$ is a general directed graph, with weights on its edges, let $A = A_G$ be a matrix whose rows and columns are indexed by the vertices of G, in which for each $u, v \in V$, $A(u, v) = w(u, v) - w(v, u)$ is the difference between the weight of the directed edge from u to v and that from v to u (0 if both these edges are missing). Thus, the matrix B above is the matrix A_{T_p}, where T_p is the quadratic residue tournament described in Section 3.

  The *cutnorm* $||A||_C$ of a real matrix A is the maximum absolute value of the sum of entries in a submatrix of A. Note that if $A = A_G$, where G is a weighted directed graph, then for two subsets $U_1, U_2 \subset V$, the sum $\sum_{u_1 \in U_1, u_2 \in U_2} A(u_1, u_2)$ can be expressed as follows. Put $U_3 = U_1 \cap U_2$. For two disjoint subsets X, Y of V let $D(X, Y)$ denote the total weight of all edges oriented from X to Y minus the total weight of all edges oriented from Y to X. Then

$$\sum_{u_1 \in U_1, u_2 \in U_2} A(u_1, u_2) = D(U_1 - U_2, U_2) + D(U_3, U_2 - U_1). \quad (2)$$

The authors of [3] describe a polynomial time algorithm that finds, given a matrix A, two subsets U_1, U_2 such that $|\sum_{u_1 \in U_1, u_2 \in U_2} A(u_1, u_2)|$ is at least $\alpha||A||_C$, for some absolute constant $\alpha > 0$, (for randomized algorithms $\alpha > 0.56$). As in our case the matrix A is antisymmetric, the algorithm provides U_1, U_2 so that the above sum (with no absolute value) approximates the maximum cut norm. In view of the expression (2) this supplies an $\alpha/2$ approximation for the maximum possible value of $D(X, Y)$, as X and Y range over all pairs of disjoint subsets of V.

- The bound in Corollary 3.2 can be slightly improved, using the expression in (2) and the fact that for the matrix $A = A_{T_p}$ of the quadratic residue tournament, the absolute value of the sum of entries of any submatrix with s rows and t columns is at most $\sqrt{stp}$. Indeed, plugging this fact in a simple modified version of the proof of Corollary 3.2 one can prove the following:

  If U is a set of vertices of T_p, and $|U| \leq 3^r$ for some integer r, then for every permutation π of the vertices of T_p: $FIT(T[U], \pi) \leq 2r3^{r-1}p^{1/2}$.

- The basic approach of proving hardness results for dense instances of computational problems by reducing the task of solving precisely sparse instances to dense ones, adding a pseudo-random collection of edges to a blow-up of a sparse instance, can be applied to various additional similar problems. Several far reaching applications of this approach, combined with some additional ideas will appear in subsequent joint work with Ailon and with Shapira and Sudakov.

References

[1] N. Ailon, M. Charikar and A. Newman, Aggregating Inconsistent Information: Ranking and Clustering, Proc. of the 37^{th} ACM STOC, Baltimore, ACM Press (2005), to appear.

[2] N. Alon, Eigenvalues, geometric expanders, sorting in rounds and Ramsey Theory, Combinatorica 6 (1986), 207-219.

[3] N. Alon and A. Naor, Approximating the Cut-Norm via Grothendieck's Inequality, Proc. of the 36^{th} ACM STOC, Chicago, ACM Press (2004), 72-80. Also: SIAM J. Computing, to appear.

[4] N. Alon and J. H. Spencer, *The Probabilistic Method, Second Edition*, Wiley, 2000.

[5] S. Arora, D. Karger and M. Karpinski, Polynomial time approximation schemes for dense instances of NP-hard problems, Proceedings of the 27^{th} ACM STOC, ACM Press (1995), 284-293.

[6] J. Bang-Jensen and C. Thomassen, A polynomial algorithm for the 2-path problem for semi-complete digraphs, SIAM J. Discrete Math. 5 (1992), 366-376.

[7] M. R. Garey and D. S. Johnson, *Computers and Intractability, A Guide to the Theory of NP-Completeness*, W. H. Freeman and Company, New York, 1979.