SIAM J. DISCRETE MATH.
Vol. 12, No. 1, pp. 48–63

CONSTRUCTIVE QUASI-RAMSEY NUMBERS AND TOURNAMENT RANKING*

A. CZYGRINOW†, S. POLJAK†, AND V. RÖDL†

Abstract. A constructive lower bound on the quasi-Ramsey numbers and the tournament ranking function was obtained in [S. Poljak, V. Rödl, and J. Spencer, *SIAM J. Discrete Math.*, (1) 1988, pp. 372–376]. We consider the weighted versions of both problems. Our method yields a polynomial time heuristic with guaranteed lower bound for the linear ordering problem.

Key words. discrepancy, linear ordering problem, derandomization, regularity lemma

AMS subject classifications. 68R05, 68R10, 05D99

PII. S0895480197318301

1. Introduction.

The *quasi-Ramsey number* $g(n)$ is defined as the maximum discrepancy between the number of edges and nonedges that appears on some induced subgraph of any graph of order n, i.e.,

$$g(n) = \min_{f} \max_{S \subseteq [n]} |f(S)|,$$

where $[n] = 1, \dots, n$, f is a function from $[n]^2$ into $\{-1, 1\}$ and $f(S) = \sum_{e \in S^2} f(e)$. It is well known (Erdös and Spencer [4]) that for some positive, absolute constants c_1, c_2

$$c_1 n^{3/2} \le g(n) \le c_2 n^{3/2}.$$

The *tournament ranking function* $h(n)$ is defined as the maximum size of an acyclic (undirected) subgraph that appears in any tournament of order n. More precisely, let T_n be a tournament on n vertices, P_n a transitive tournament on n vertices, and let $|T_n \cap P_n|$ denote the number of common oriented arcs of T_n and P_n; then

$$h(n) = \min_{T_n} \max_{P_n} |T_n \cap P_n|.$$

It was shown by Spencer ([14], [15]) that

$$\frac{1}{2}\binom{n}{2} + c_1 n^{3/2} \le h(n) \le \frac{1}{2}\binom{n}{2} + c_2 n^{3/2}$$

where c_1 and c_2 are positive absolute constants. The proof of the upper bound has been simplified by Fernandez de La Vega [5]. Using the method of Spencer, the lower bound on $h(n)$ can be obtained by an algorithmic argument from the lower bound on $g(n)$.

Poljak, Rödl, and Spencer [12] proposed a fast $O(n^3 \log n)$ time algorithm that finds a set S with discrepancy at least $\frac{\pi^{-1/2}}{24} n^{3/2}$, the corresponding result for the tournament ranking function $h(n)$ is also presented in [12]. We will consider the

*Received by the editors March 12, 1997; accepted for publication (in revised form) March 5, 1998; published electronically January 29, 1999.

http://www.siam.org/journals/sidma/12-1/31830.html

†Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322 (aczygri@mathcs.emory.edu, rodl@mathcs.emory.edu).

weighted version of both problems. Our algorithm uses the Erdös–Selfridge method of conditional expectations that was also applied in [12]. For the lower bound on the quasi-Ramsey number $g(n)$ we prove the following result.

THEOREM 1. *Let $f : [n]^2 \to \mathbb{R}$ be a weight function on the edges of a complete graph K_n. Then there is a subset $S \subset [n]$ such that*

$$|f(S)| \geq \frac{1}{12\sqrt{\pi}} n^{-1/2} \sum_{e \in [n]^2} |f(e)|.$$

Moreover, S can be found in $O(n^3 \lg (nd) \lg n)$ time, provided the weights are integers from $\{-d, \ldots, d\}$.

The weighted version of the tournament ranking problem is also known as the *linear ordering problem* (see Grötschel, Jünger, and Reinelt [10]). The problem can be formulated in the following way: For a given tournament T with weight $c(i, j)$ on the arc $(i, j) \in T$, find the ordering σ of vertices for which the sum

$$\sum_{(i,j) \in T, \sigma(i) < \sigma(j)} c(i, j)$$

is a maximum. The list of applications of the linear ordering problem can be found in Lenstra [11]. It includes applications from different areas of econometrics (input-output matrix analysis), sociology (social choice), psychology, machine scheduling, and even archaeology. The problem is NP-complete (see Garey and Johnson [8]), but there were several methods developed for solving small instances, e.g., up to order of 50 by Grötschel, Jünger, and Reinelt [10]. Using the algorithm from Theorem 1, we will get a polynomial time heuristic with a guaranteed lower bound.

THEOREM 2. *Let T be a tournament on n vertices with nonnegative weights $w(e)$ on edges. Then there is an ordering σ such that the sum of weights on edges that agree with the ordering is at least*

$$\left(\frac{1}{2} + \frac{1}{4\sqrt{\pi}} n^{-1/2} \right) K,$$

where K is the total sum of weights. The ordering σ can be constructed by a $O(n^3 \lg (nd) \lg n)$ time algorithm, provided weights are integers from $\{0, \ldots, d\}$.

From the upper bound on $h(n)$, we conclude that there exists weight function for which the heuristic is best possible (up to a constant factor).

Given a real number ρ, $0 < \rho < 1$ a polynomial time approximation scheme (PTAS) for an optimization problem is an algorithm which when given an instance of size n, finds in polynomial time (in n) a solution of value at least $(1 - \rho)OPT$, where OPT is the optimal value. Using the regularity lemma and its constructive version of Alon et al. [1], we design a PTAS for the "dense" quasi-Ramsey problem and for tournament ranking. For the quasi-Ramsey number we have the following theorem. Let $f : E(K_n) \to \{-1, 1\}$ and $OPT(f) = \max_{S \subseteq [n]} |f(S)|$.

THEOREM 3. *Let $c > 0$ be fixed. If $OPT(f) \geq cn^2$, then for every ρ, $0 < \rho < 1$, there is a $O(n^{2.4})$ time algorithm that constructs set S such that*

$$|f(S)| \geq (1 - \rho) OPT(f).$$

For the tournament ranking we prove the following theorem for the case when $OPT(T_n) = \max_{P_n} |T_n \cap P_n|$ for a tournament T_n.

THEOREM 4. *For $0 < \rho < 1$ there is a polynomial time algorithm that constructs an ordering σ of vertices of T_n so that at least $(1-\rho)OPT(T_n)$ of arcs agree with σ.*

Note that Theorem 3 and Theorem 4 are in some sense counterparts to Theorem 1 and Theorem 2. For example, Theorem 1 provides the existence of a polynomial time algorithm to find the set S with $|f(S)|$ being the guaranteed minimum; Theorem 3 gives for every ρ the const$(\rho)n^{2.4}$ algorithm that finds a set S with $f(S)$ being a $(1-\rho)$ multiple of the optimal. Theorem 3 is based on the algorithmic version of the regularity lemma which "approximates the graph with error of ϵn^2". Therefore, it can be applied only to instances with $OPT(f) \geq cn^2$. On the other hand, in case of Theorem 4, clearly $OPT(T_n) \geq \frac{1}{2}\binom{n}{2}$ and, therefore, a PTAS for the linear ordering problem can be obtained with no additional assumptions. Independently, very recently Frieze and Kannan [6] and [7] applied a version of the regularity lemma to the maximum subgraph problem, an equivalent to tournament ranking. Our arguments differ from those in [7]. The rest of the paper is organized as follows: In section 2, for a given $\overrightarrow{v_1}, \ldots, \overrightarrow{v_n} \in \mathbb{R}^k$, we will show how to construct sign vector $\overrightarrow{X} = (X_1, \ldots, X_n)$ such that

$$\| X_1\overrightarrow{v_1} + \cdots + X_n\overrightarrow{v_n} \| \geq cn^{-1/2}\sum_{i=1}^{n} \| \overrightarrow{v_i} \|,$$

where $\|\overrightarrow{u}\| = \sum_{j=1}^{k} |u_j|$. The algorithm is later applied to quasi-Ramsey numbers and to the linear ordering problem. Section 3 includes the applications of the regularity lemma. We conclude with an open problem in section 4.

2. Constructing sign vectors. Set $\overrightarrow{1} = (1, \ldots, 1)$ and $\overrightarrow{0} = (0, \ldots, 0)$, and for $\overrightarrow{u}$ and $\overrightarrow{v}$ from $\mathbb{R}^k$, let $\langle \overrightarrow{v}, \overrightarrow{u} \rangle$ denote the dot product of $\overrightarrow{v}$ and $\overrightarrow{u}$, and $\|\overrightarrow{u}\|$ its l_1- norm, i.e., $\|\overrightarrow{u}\| = \sum_{j=1}^{k} |u_j|$. We first establish two auxiliary facts.

LEMMA 5.

$$\sum_{\overrightarrow{X} \in \{-1,1\}^n} |\langle \overrightarrow{1}, \overrightarrow{X} \rangle| = 2n\binom{n-1}{\lfloor \frac{n}{2} \rfloor}.$$

The proof can be found in [12]. For $1 \leq i \leq n$, let $\mathbf{X}_i$ be independent random variables with distribution $Pr(\mathbf{X}_i = 1) = Pr(\mathbf{X}_i = -1) = \frac{1}{2}$.

LEMMA 6. *Let $b_1, \ldots, b_n$ and a be real numbers and let u be the arithmetic mean of $|b_1|, \ldots, |b_n|$. Then we have the following inequality:*

$$E(|a + \mathbf{X}_1 b_1 + \cdots + \mathbf{X}_n b_n|) \geq E(|a + \mathbf{X}_1 u + \cdots + \mathbf{X}_n u|).$$

Proof. We may assume that all b_i's are nonnegative since the random variables $Z_i = \mathrm{sgn(b_i)X_i}$ have the same distribution as X_i, i.e., $E(|a + X_1 b_1 + \cdots + X_n b_n|) = E(|a + Z_1 b_1 + \cdots + Z_n b_n|) = E(|a + X_1|b_1| + \cdots + X_n|b_n||)$. Given a vector $\overrightarrow{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n$, let $\overrightarrow{w^{(l)}}$ be the vector obtained from $\overrightarrow{w}$ by cyclic shifting, with ith coordinate $w_i^{(l)} = w_{i+l \mod n}$ for $i = 1, 2, \ldots, n$. We have

$$E(|a+\mathbf{X}_1 b_1+\cdots+\mathbf{X}_n b_n|) = \frac{1}{2^n}\sum_{\overrightarrow{X} \in \{-1,1\}^n}\left|a + \sum_{i=1}^{n} X_i b_i\right| = \frac{1}{2^n}\sum_{\overrightarrow{X}}\frac{1}{n}\sum_{l=1}^{n}\left|a + \sum_{i=1}^{n} X_i^{(l)} b_i\right|$$

$$\geq \frac{1}{2^n}\sum_{\overrightarrow{X}}\frac{1}{n}\left|na+\sum_{i=1}^{n}\sum_{l=1}^{n}X_i^{(l)}b_i\right| = \frac{1}{2^n}\sum_{\overrightarrow{X}}\left|a+\sum_{i=1}^{n}X_iu\right| = E(|a+\mathbf{X}_1u+\cdots+\mathbf{X}_nu|). \quad \square$$

LEMMA 7. *Let* $\overrightarrow{v_1},\dots,\overrightarrow{v_n}\in\mathbb{R}^k$*. Then*

$$E(\|\mathbf{X}_1\overrightarrow{v_1}+\cdots+\mathbf{X}_n\overrightarrow{v_n}\|)\geq\sqrt{\frac{2}{\pi}}n^{-1/2}\sum_{i=1}^{n}\|\overrightarrow{v_i}\|.$$

Proof. From Lemma 5 and Stirling's formula, we obtain

$$E(|\mathbf{X}_1+\cdots+\mathbf{X}_n|)=\frac{1}{2^n}\sum_{\overrightarrow{X}\in\{-1,1\}^n}|\langle\overrightarrow{1},\overrightarrow{X}\rangle| = 2n2^{-n}\binom{n-1}{\lfloor\frac{n}{2}\rfloor}\geq\sqrt{\frac{2n}{\pi}}.$$

Let u_j be the arithmetic mean of absolute values of the jth components of $\overrightarrow{v_1},\dots,\overrightarrow{v_n}$, where $j=1,\dots,k$ and let $\overrightarrow{u}=(u_1,\dots,u_k)$. Using Lemma 6 with $a=0$ we have

$$E(\|\mathbf{X}_1\overrightarrow{v_1}+\cdots+\mathbf{X}_n\overrightarrow{v_n}\|)\geq E(\|\mathbf{X}_1\overrightarrow{u}+\cdots+\mathbf{X}_n\overrightarrow{u}\|)=\sum_{j=1}^{k}E(|\mathbf{X}_1u_j+\cdots+\mathbf{X}_nu_j|)$$

$$=\sum_{j=1}^{k}u_jE(|\mathbf{X}_1+\cdots+\mathbf{X}_n|)\geq\sqrt{\frac{2}{\pi}}n^{-1/2}\sum_{i=1}^{n}\|\overrightarrow{v_i}\|. \quad \square$$

COROLLARY 8. *For given* $\overrightarrow{v_1},\dots,\overrightarrow{v_n}\in\mathbb{R}^k$*, there is a choice of signs* $(X_1,\dots,X_n)\in\{-1,1\}^n$ *such that*

$$\|X_1\overrightarrow{v_1}+\cdots+X_n\overrightarrow{v_n}\|\geq\sqrt{\frac{2}{\pi}}n^{-1/2}\sum_{i=1}^{n}\|\overrightarrow{v_i}\|.$$

Next we will show that a vector $\overrightarrow{X}=(X_1,\dots,X_n)$ from Corollary 8 can be constructed by a polynomial time algorithm. The idea is as follows. We have $E(\|\mathbf{X}_1\overrightarrow{v_1}+\cdots+\mathbf{X}_n\overrightarrow{v_n}\|)\geq T$, where $T=cn^{-1/2}\sum\|\overrightarrow{v_i}\|$ in the beginning. (For later convenience, we write the vectors in the reverse order.) Let us assume that signs $X_n,X_{n-1},\dots,X_{i+1}$ are chosen, one in each step, such that

$$E(\|X_n\overrightarrow{v_n}+\cdots+X_{i+1}\overrightarrow{v_{i+1}}+\mathbf{X}_i\overrightarrow{v_i}+\cdots\mathbf{X}_1\overrightarrow{v_1}\|)\geq T.$$

At this moment there are two possible choices of X_i, and we take the better one (the one that maximizes the value of the expectation). As we cannot compute quickly the expected value $E(\|X_n\overrightarrow{v_n}+\cdots+X_{i+1}\overrightarrow{v_{i+1}}+\mathbf{X}_i\overrightarrow{v_i}+\cdots+\mathbf{X}_1\overrightarrow{v_1}\|)$ for general $\overrightarrow{v_i},\dots,\overrightarrow{v_1}$, we compute $E(\|X_n\overrightarrow{v_n}+\cdots+X_{i+1}\overrightarrow{v_{i+1}}+\mathbf{X}_i\overrightarrow{u}+\cdots+\mathbf{X}_1\overrightarrow{u}\|)$ instead, where $\overrightarrow{u}$ is the component-wise "average" of $\overrightarrow{v_1},\dots,\overrightarrow{v_n}$.

To describe the algorithm more precisely, we need to introduce some notation. For vectors $\overrightarrow{a}=(a_1,\dots,a_k)$, $\overrightarrow{b}=(b_1,\dots,b_k)\in\mathbb{R}^k$ we define the polynomials

$$W(b_j,i,a_j)=E(|b_j+\mathbf{X}_ia_j+\cdots+\mathbf{X}_1a_j|)=\sum_{l=0}^{i}\binom{i}{l}2^{-i}|b_j+a_j(i-2l)|,$$

$$W(\overrightarrow{b}, i, \overrightarrow{a}) = \sum_{j=1}^{k} W(b_j, i, a_j) = \sum_{j=1}^{k}\sum_{l=0}^{i}\binom{i}{l}2^{-i}|b_j + a_j(i-2l)|.$$

For given $\overrightarrow{v_i} = (v_{i1}, \ldots, v_{ik}) \in \mathbb{R}^k$, $i = 1, \ldots, n$, let u_{ij} denote the arithmetic mean of absolute values of the jth coordinates of $\overrightarrow{v_i}, \ldots, \overrightarrow{v_1}$, i.e., $u_{ij} = \frac{1}{i}(|v_{ij}| + \cdots + |v_{1j}|)$, and set $\overrightarrow{u_i} = (u_{i1}, \ldots, u_{ik})$. By $\overrightarrow{S_i}$ we denote the partial sums: let $\overrightarrow{S_n} = \overrightarrow{0}$ and $\overrightarrow{S_i} = X_n\overrightarrow{v_n} + \cdots + X_{i+1}\overrightarrow{v_{i+1}}$, where $X_n, \ldots, X_{i+1}$ have already been defined. (Observe that $E(\|\overrightarrow{S_i} + \mathbf{X}_i\overrightarrow{u_i} + \ldots \mathbf{X}_1\overrightarrow{u_i}\|) = W(\overrightarrow{S_i}, i, \overrightarrow{u_i})$.) Now we choose

$$X_i = \begin{cases} 1 & \text{if } W(\overrightarrow{S_i} + \overrightarrow{v_i}, i-1, \overrightarrow{u_{i-1}}) \geq W(\overrightarrow{S_i} - \overrightarrow{v_i}, i-1, \overrightarrow{u_{i-1}}), \\ -1 & \text{otherwise.} \end{cases}$$

We can formalize the algorithm in the following procedure.

Algorithm

input: vectors $\overrightarrow{v_1}, \ldots, \overrightarrow{v_n} \in \mathbb{R}^k$
output: sign vector $(X_1, \ldots, X_n)$
$\overrightarrow{S_n} = \overrightarrow{0}$
for i=n downto 1
  begin
    if $i < n$ then $\overrightarrow{S_i} = X_n\overrightarrow{v_n} + \cdots + X_{i+1}\overrightarrow{v_{i+1}}$
    compute $W_+ = W(\overrightarrow{S_i} + \overrightarrow{v_i}, i-1, \overrightarrow{u_{i-1}})$ and $W_- = W(\overrightarrow{S_i} - \overrightarrow{v_i}, i-1, \overrightarrow{u_{i-1}})$
    if $W_+ \geq W_-$ then $X_i = 1$
    else $X_i = -1$
  end
return $(X_1, \ldots, X_n)$

Proposition 9. *The above algorithm returns a vector* $(X_1, \ldots, X_n)$ *such that*

$$\|X_1\overrightarrow{v_1} + \cdots + X_n\overrightarrow{v_n}\| \geq \sqrt{\frac{2}{\pi}}n^{-1/2}\sum_{i=1}^{n}\|\overrightarrow{v_i}\|.$$

Proof. Since $E(\|\overrightarrow{S_i} + \mathbf{X}_i\overrightarrow{v_i} + \mathbf{X}_{i-1}\overrightarrow{u_{i-1}} + \cdots \mathbf{X}_1\overrightarrow{u_{i-1}}\|) = \frac{1}{2}W(\overrightarrow{S_i} + \overrightarrow{v_i}, i-1, \overrightarrow{u_{i-1}}) + \frac{1}{2}W(\overrightarrow{S_i} - \overrightarrow{v_i}, i-1, \overrightarrow{u_{i-1}})$, we have

$$\begin{aligned} W(\overrightarrow{S_{i-1}}, i-1, \overrightarrow{u_{i-1}}) &= W(\overrightarrow{S_i} + X_i\overrightarrow{v_i}, i-1, \overrightarrow{u_{i-1}}) \geq E(\|\overrightarrow{S_i} + \mathbf{X}_i\overrightarrow{v_i} + \mathbf{X}_{i-1}\overrightarrow{u_{i-1}} \\ &\quad + \cdots + \mathbf{X}_1\overrightarrow{u_{i-1}}\|) \\ &\geq E(\|\overrightarrow{S_i} + \mathbf{X}_i\overrightarrow{u_i} + \cdots \mathbf{X}_1\overrightarrow{u_i}\|) = W(\overrightarrow{S_i}, i, \overrightarrow{u_i}). \end{aligned}$$

The first inequality holds by the choice of X_i, the second one by Lemma 6, and the (obvious) fact that u_{ij} is an arithmetic mean of v_{ij} and $i-1$ copies of u_{i-1j}. Hence

$$\|X_n\overrightarrow{v_n} + \cdots + X_1\overrightarrow{v_1}\| \geq W(\overrightarrow{S_1}, 1, \overrightarrow{u_1}) \geq \cdots \geq W(\overrightarrow{S_n}, n, \overrightarrow{u_n})$$

and

$$\begin{aligned} W(\overrightarrow{S_n}, n, \overrightarrow{u_n}) &= \sum_{j=0}^{k}\sum_{l=0}^{n}\binom{n}{l}2^{-n}|u_{nj}(n-2l)| = \sum_{j=0}^{k}u_{nj}2^{-n}\sum_{l=0}^{n}\binom{n}{l}|n-2l| \\ &\geq \sqrt{\frac{2}{\pi}}n^{1/2}\sum_{j=0}^{k}u_{nj} = \sqrt{\frac{2}{\pi}}n^{-1/2}\sum_{i=1}^{n}\|\overrightarrow{v_i}\|. \qquad \square \end{aligned}$$

PROPOSITION 10. *For $k = O(n)$, the time complexity of the above algorithm is $O(n^3 \lg(nd) \lg n)$ provided the vectors $\overrightarrow{v_1}, \ldots, \overrightarrow{v_n} \in \mathbb{R}^k$ are integral and $|v_{ij}| \leq d$.*

Proof. The procedure consists of n iterations for computing $X_n, \ldots, X_1$. At each step we evaluate the expression $W(\overrightarrow{S_i}, i, \overrightarrow{u_i})$. To keep the computation in integers we replace it by

$$i2^i W(\overrightarrow{S_i}, i, \overrightarrow{u_i}) = \sum_{l=0}^{i} \binom{i}{l} \bigg(\sum_{j=1}^{k} |iS_{ij} + (i-2l)iu_{ij}| \bigg),$$

where $\overrightarrow{S_i} = (S_{i1}, \ldots, S_{ik})$. The $O(n^2)$ combinatorial coefficients $\binom{i}{l}$ can be evaluated in advance using the identity $\binom{i}{l} = \binom{i-1}{l} + \binom{i-1}{l-1}$. Since i is of size at most n and the terms S_{ij}, iu_{ij} are of size nd, we can compute $|iS_{ij} + (i-2l)iu_{ij}|$ in $O(\lg n \lg(nd))$ steps. The sum $\sum_{j=1}^{k} |iS_{ij} + (i-2l)iu_{ij}|$ can be evaluated in $O(k \lg n \lg(nd))$ steps. The number $\binom{i}{l}$ is less than 2^n and so the multiplication $\binom{i}{l} \cdot (\sum_{j=1}^{k} |iS_{ij} + (i-2l)iu_{ij}|)$ can be computed in $O(\lg(2^n) \lg(ndk))$ steps. The total complexity of the procedure is $O(n^2(k \lg n \lg(nd) + n \lg(ndk)))$, which when $k = O(n)$ becomes $O(n^3 \lg(nd) \lg n)$. ∎

Using the divide and conquer technique, one can design a slightly faster algorithm that gives a little worse results (for details consult [2]).

We will now apply the algorithm to quasi-Ramsey numbers and to the linear ordering problem. Let us start with the proof of Theorem 1.

Proof of the Theorem 1. We use the same technique that was applied in [12]. Let $K = \sum_{e \in [n]^2} |f(e)|$. First we need to find a large cut of K_n with edge weights $|f(e)|$. Obviously, by a greedy procedure we can construct disjoint sets X and Y such that $X \cup Y = [n]$ and

$$\sum_{x \in X, y \in Y} |f(x,y)| \geq \frac{K}{2}.$$

Indeed, assume that sets $X^i \cup Y^i = [i]$ have been constructed. Let $W_X^i = \sum_{j \in X^i} f(j, i+1)$ and $W_Y^i = \sum_{j \in Y^i} f(j, i+1)$. If $W_X^i \leq W_Y^i$ then set $X^{i+1} = X^i \cup \{i+1\}$ and $Y^{i+1} = Y^i$; otherwise, set $X^{i+1} = X^i$ and $Y^{i+1} = Y^i \cup \{i+1\}$. (Using the Goemans–Williamson algorithm from [9], one can possibly improve a constant in our theorem. However, since the result in [9] provides .878 approximation of the optimal cut, it does not guarantee that the produced cut is bigger than $\frac{K}{2}$. For slightly better cut algorithms consult [13].)

Let $X = \{x_1, \ldots, x_{n_1}\}$, $Y = \{y_1, \ldots, y_{n_2}\}$. We assume $n_1 \leq n/2$. Assign a vector $\overrightarrow{v_i} = (v_{i1}, \ldots, v_{in_2})$ to each vertex x_i, where $v_{ij} = f(x_i, y_j)$, $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$. Using the algorithm from section 2, we construct a sign vector $(X_1, \ldots, X_{n_1})$ such that

$$\|X_1 \overrightarrow{v_1} + \cdots + X_{n_1} \overrightarrow{v_{n_1}}\| \geq \sqrt{\frac{2}{\pi}} {n_1}^{-1/2} \frac{K}{2} \geq \sqrt{\frac{2}{\pi}} (\frac{n}{2})^{-1/2} \frac{K}{2} \geq \frac{1}{\sqrt{\pi}} n^{-1/2} K.$$

We partition sets $X = X^+ \cup X^-$ and $Y = Y^+ \cup Y^-$ by $X^+ = \{x_i, X_i = 1\}$,

$Y^+ = \{y_j, \sum_{i=1}^{n_1} X_i f(x_i, y_j) \geq 0\}$ and $X^- = X - X^+$, $Y^- = Y - Y^+$. Then

$$\|X_1 \overrightarrow{v_1} + \cdots + X_{n_1} \overrightarrow{v_{n_1}}\| = \sum_{j=1}^{n_2} |\sum_{i=1}^{n_1} X_i f(x_i, y_j)| = \sum_{y \in Y^+, x \in X^+} f(x,y)$$
$$+ \sum_{y \in Y^+, x \in X^-} -f(x,y) + \sum_{y \in Y^-, x \in X^+} -f(x,y) + \sum_{y \in Y^-, x \in X^-} f(x,y).$$

Hence, we can choose $X^* \in \{X^+, X^-\}$ and $Y^* \in \{Y^+, Y^-\}$ such that

$$|f(X^*, Y^*)| = \left| \sum_{y \in Y^*, x \in X^*} f(x,y) \right| \geq \frac{1}{4\sqrt{\pi}} n^{-1/2} K.$$

We also have $f(X^*, Y^*) = f(X^* \cup Y^*) - f(X^*) - f(Y^*)$. Let S be one of X^*, Y^*, $X^* \cup Y^*$ for which $|f(S)| \geq \frac{1}{3}|f(X^*, Y^*)|$. We see that S is such that $|f(S)| \geq \frac{1}{12\sqrt{\pi}} n^{-1/2} K$. □

Taking $K = \binom{n}{2}$ we obtain a lower bound on the quasi-Ramsey numbers.

Corollary 11.

$$g(n) \geq \frac{1}{24\sqrt{\pi}} n^{3/2}.$$

We can now apply the result of Theorem 1 to the linear ordering problem. Since the proof resembles the reasoning for the corresponding result in [12], we omit the details.

Proof of Theorem 2. Let w_{ij} be the weight of the pair $\{i, j\}$. Define $f : [n]^2 \to \mathbb{Z}$ as follows. For $i < j$,

$$f(i,j) = \begin{cases} w_{ij} & \text{if } (i,j) \in T, \\ -w_{ij} & \text{if } (j,i) \in T. \end{cases}$$

Let X^*, Y^* be the sets constructed in the proof of Theorem 1 and let $R = [n] - X^* - Y^*$. Construct $\prec$ in the following way. Construct ranking on X^* such that at least half of the arcs with both endpoints in X^* are consistent with the ranking. (This can be obtained by considering an arbitrary ordering and its inverse.) Similarly construct rankings of Y^* and R. Assume that $f(X^*, Y^*) \geq 0$; then for $x \in X^*$ and $y \in Y^*$ let $x \prec y$. Suppose that $f(X^* \cup Y^*, R) \geq 0$; then for $r \in R$ and $z \in X^* \cup Y^*$ let $z \prec r$. □

3. Applications of the regularity lemma. In this section we present the applications of the regularity lemma to both quasi-Ramsey and tournament ranking functions. A variant of the regularity lemma was applied for max-cut, graph bisection, and a quadratic assignment problem in Frieze and Kannan [6] and for computing frequencies in graphs in Duke, Lefmann, and Rödl [3]. For simplicity, we restrict our discussion to the unweighted case, but similar results can be obtained for weighted graphs and tournaments. Let (V, E) be a graph on n vertices, for $V_1, V_2 \subset V$, $V_1 \cap V_2 = \emptyset$, the density $d(V_1, V_2)$ is defined as $d(V_1, V_2) = \frac{e(V_1, V_2)}{|V_1||V_2|}$, where $e(V_1, V_2)$ denotes the number of edges between V_1 and V_2.

Definition 12. *A pair of subsets (V_1, V_2) is called ϵ-regular if for every $W_1 \subset V_1$, with $|W_1| \geq \epsilon|V_1|$ and for every $W_2 \subset V_2$, with $|W_2| \geq \epsilon|V_2|$*

$$|d(W_1, W_2) - d(V_1, V_2)| < \epsilon.$$

Definition 13. *A partition* $V_1 \cup V_1 \cup \cdots \cup V_k$ *of* V *is* ϵ*-regular if*

(i) $||V_i| - |V_j|| \le 1$ *for all* i, j *and*

(ii) *all except at most* $\epsilon\binom{k}{2}$ *pairs* (V_i, V_j) *are* ϵ*-regular*

Let us now state the powerful regularity lemma of Szemerédi [16].

Lemma 14. *For every* $\epsilon > 0$ *and every integer* l *there exist* N *and* L *such that any graph with at least* N *vertices admits an* ϵ*-regular partition* $V_1 \cup \cdots \cup V_k$ *with* $l \le k \le L$.

The following version can be easily concluded from the original proof [16].

Lemma 15. *For every* $\epsilon > 0$ *and every integer* l*, there exists an* N *such that for any graph with at least* $N = N(l, \epsilon)$ *vertices and any partition* P *of the graph into* m *subsets, there exists* $L = L(l, \epsilon, m)$ *and an* ϵ*-regular partition* $V_1 \cup \cdots \cup V_k$ *with* $l \le k \le L$ *which is a refinement of* P.

The partition postulated in both lemmas can be found in $O(n^{2.4})$ time using the algorithm described in Alon et al. [1].

Proof of Theorem 3. The algorithm is as follows: Let $\epsilon = \frac{\rho c}{7}$.

1. Find an ϵ-regular partition $V_1 \cup \cdots \cup V_k$ with $k \ge \frac{1}{\epsilon}$ of the graph $G_1 = (V, f^{-1}(1))$.
2. Check all 2^k subsets of V of the form $S = \bigcup_{l \in L} V_l$, where $L \subset [k]$ and choose S that maximizes $|\sum_{1 \le i < j \le k}(2d_{ij} - 1)|V_i \cap S||V_j \cap S||$.

Note that if (V_i, V_j) is ϵ-regular with density d_{ij} in $G_1 = (V, f^{-1}(1))$, then (V_i, V_j) is ϵ-regular with density $1 - d_{ij}$ in $G_{-1} = (V, f^{-1}(-1))$. Given the partition $V_1 \cup \cdots \cup V_k$, we define $f^* : 2^{[n]} \to \mathbb{R}$ in the following way. For $T \subset [n]$, $f^*(T) = \sum_{1 \le i < j \le k}(2d_{ij} - 1)|V_i \cap T||V_j \cap T|$, where $d_{ij} = d(V_i, V_j)$.

Fact 16. *Let* T^* *be a minimal set that maximizes* f^**. Then for every* l *such that* $V_l \cap T^* \ne \emptyset$ *the sum* $\sum_{j \ne l}(2d_{lj} - 1)|V_j \cap T^*| > 0$.

Proof. We use proof by contradiction. Assume that there exists l such that $V_l \cap T^* \ne \emptyset$ and $\sum_{j \ne l}(2d_{lj} - 1)|V_j \cap T^*| \le 0$. Then

$$f^*(T^*) = \sum_{1 \le i < j \le k}(2d_{ij} - 1)|V_i \cap T^*||V_j \cap T^*| = \sum_{j \ne l}(2d_{lj} - 1)|V_l \cap T^*||V_j \cap T^*|$$

$$+ \sum_{i,j \ne l, i<j}(2d_{ij} - 1)|V_i \cap T^*||V_j \cap T^*| = |V_l \cap T^*| \sum_{j \ne l}(2d_{lj} - 1)|V_j \cap T^*|$$

$$+ \sum_{i,j \ne l, i<j}(2d_{ij} - 1)|V_i \cap T^*||V_j \cap T^*| \le \sum_{i,j \ne l, i<j}(2d_{ij} - 1)|V_i \cap T^*||V_j \cap T^*|$$

$$= \sum_{1 \le i < j \le k}(2d_{ij} - 1)|V_i \cap (T^* \setminus V_l)||V_j \cap (T^* \setminus V_l)| = f^*(T^* \setminus V_l)$$

and we get the contradiction with minimality of T^*. □

Fact 17. *Let* T^* *be a minimal set that maximizes* f^**. If* $T^* \cap V_l \ne \emptyset$*, then* $V_l \subset T^*$.

Note that Fact 17 implies that if S is a set found by the algorithm, then $|f^*(S)| \ge f^*(T^*)$ as the algorithm checks all the possible unions of V_i's to maximize $|f^*|$. In the same way, one can show that $|f^*(S)| \ge -f^*(L^*)$ where L^* maximizes $-f^*$.

Proof.

$$f^*(T^*) = |V_l \cap T^*| \sum_{j \neq l} (2d_{lj} - 1)|V_j \cap T^*| + \sum_{i,j \neq l, i<j} (2d_{ij} - 1)|V_i \cap T^*||V_j \cap T^*|$$

$$\leq |V_l| \sum_{j \neq l} (2d_{lj} - 1)|V_j \cap T^*| + \sum_{i,j \neq l, i<j} (2d_{ij} - 1)|V_i \cap T^*||V_j \cap T^*|$$

$$= \sum_{1 \leq i < j \leq k} (2d_{ij} - 1)|V_i \cap (T^* \cup V_l)||V_j \cap (T^* \cup V_l)| = f^*(T^* \cup V_l).$$

Hence, $f^*(T^*) \leq f^*(T^* \cup V_l)$ and the equality holds only if $|V_l \cap T^*| = |V_l|$ as $\sum_{j \neq l}(2d_{lj} - 1)|V_j \cap T^*| > 0$ by the previous fact. ∎

It will be convenient to use the following notation. For two functions $A(n)$ and $B(n)$, we write $A(n) =_\epsilon B(n)$ if $|A(n) - B(n)| \leq \epsilon n^2$ for n large enough.

Our main lemma shows that f^* is a "good" approximation for the discrepancy function f.

LEMMA 18. *For every $U \subset V$ $|f^*(U) - f(U)| < \frac{7}{2}\epsilon n^2$.*

Proof. We divide the proof into five claims.

CLAIM 19. $f(U) =_{\frac{\epsilon}{2}} \sum_{\{i,j\} \in [k]^2} f(V_i \cap U, V_j \cap U)$.

Indeed, since $|V_i| \leq \frac{n}{k}$ and also $|V_i \cap U| \leq \frac{n}{k}$, we infer that $|f(V_i \cap U)| \leq \binom{\frac{n}{k}}{2} \leq \frac{n^2}{2k^2}$. Therefore,

$$|f(U) - \sum_{\{i,j\} \in [k]^2} f(V_i \cap U, V_j \cap U)| = \left| \sum_{i=1}^{k} f(V_i \cap U) \right| \leq \sum_{i=1}^{k} |f(V_i \cap U)| \leq \frac{n^2}{2k} \leq \frac{\epsilon}{2} n^2$$

which proves Claim 19.

We partition $[k]^2 = S \cup I \cup R$ as follows: $\{i,j\} \in S$ if and only if either $|V_i \cap U| < \epsilon |V_i|$ or $|V_j \cap U| < \epsilon |V_j|$, $\{i,j\} \in I$ if and only if the pair (V_i, V_j) is not ϵ-regular, $R = [k]^2 \setminus (S \cup I)$.

CLAIM 20. $f(U) =_\epsilon \sum_{R \cup I} f(V_i \cap U, V_j \cap U)$.

$$\left| \sum_{[k]^2} f(V_i \cap U, V_j \cap U) - \sum_{R \cup I} f(V_i \cap U, V_j \cap U) \right| = \left| \sum_{S} f(V_i \cap U, V_j \cap U) \right|$$

$$\leq \sum_{S} |f(V_i \cap U, V_j \cap U)| \leq \sum_{S} |V_i \cap U||V_j \cap U| \leq \binom{k}{2} \epsilon \frac{n^2}{k^2} = \frac{\epsilon}{2} n^2.$$

Since $|f(U) - \sum_{[k]^2} f(V_i \cap U, V_j \cap U)| \leq \frac{\epsilon}{2}$ by Claim 19, we infer that $|f(U) - \sum_{R \cup I} f(V_i \cap U, V_j \cap U)| \leq \epsilon$.

CLAIM 21. $f(U) =_{\frac{3\epsilon}{2}} \sum_{R} f(V_i \cap U, V_j \cap U)$.

Indeed, there are at most $\epsilon \frac{k^2}{2}$ irregular pairs and for each of them $|f(V_i \cap U, V_j \cap U)| \leq (\frac{n}{k})^2$, which implies

$$\left| \sum_{R \cup I} f(V_i \cap U, V_j \cap U) - \sum_{R} f(V_i \cap U, V_j \cap U) \right| = \left| \sum_{I} f(V_i \cap U, V_j \cap U) \right|$$

$$\leq \sum_I |f(V_i \cap U, V_j \cap U)| \leq \epsilon \frac{k^2}{2}\left(\frac{n}{k}\right)^2 = \frac{\epsilon}{2} n^2.$$

Together with Claim 20, this shows that $f(U) =_{\frac{3\epsilon}{2}} \sum_R f(V_i \cap U, V_j \cap U)$.

Claim 22. $f(U) =_{\frac{5\epsilon}{2}} \sum_R (2d_{ij} - 1)|U \cap V_i||U \cap V_j|$.

From Claim 21 we know that $f(U) =_{\frac{3\epsilon}{2}} \sum_R f(U \cap V_i, U \cap V_j)$. For $\{i, j\} \in R$ we can approximate $f(U \cap V_i, U \cap V_j)$ by $(2d_{ij} - 1)|U \cap V_i||U \cap V_j|$ with $2\epsilon(\frac{n}{k})^2$ error, namely,

$$\begin{aligned}
&|f(U \cap V_i, U \cap V_j) - (2d_{ij} - 1)|U \cap V_i||U \cap V_j||\\
&= |d(U \cap V_i, U \cap V_j)|U \cap V_i||U \cap V_j| - (1 - d(U \cap V_i, U \cap V_j))|U \cap V_i||U \cap V_j|\\
&\quad -(2d_{ij} - 1)|U \cap V_i||U \cap V_j||\\
&= 2|d(U \cap V_i, U \cap V_j) - d_{ij}||U \cap V_i||U \cap V_j| \leq 2\epsilon\left(\frac{n}{k}\right)^2.
\end{aligned}$$

Thus,

$$\left|\sum_R f(U \cap V_i, U \cap V_j) - \sum_R (2d_{ij} - 1)U \cap V_i||U \cap V_j||\right| \leq \frac{k^2}{2} 2\epsilon\left(\frac{n}{k}\right)^2 = \epsilon n^2$$

which proves the claim.

Claim 23. $f(U) =_{\frac{7\epsilon}{2}} f^*(U)$.

By definition, $f^*(U) = \sum_{[k]^2} (2d_{ij} - 1)|U \cap V_i||U \cap V_j|$ and by Claim 22 we have $f(U) =_{\frac{5\epsilon}{2}} \sum_R (2d_{ij} - 1)|U \cap V_i||U \cap V_j|$. Similar computations show

$$\left|\sum_{[k]^2} (2d_{ij} - 1)|U \cap V_i||U \cap V_j| - \sum_R (2d_{ij} - 1)|U \cap V_i|||U \cap V_j||\right|$$

$$\leq \sum_{I \cup S} |(2d_{ij} - 1)||U \cap V_i||U \cap V_j| \leq \binom{k}{2}\left(\epsilon\left(\frac{n}{k}\right)^2 + \epsilon\left(\frac{n}{k}\right)^2\right) \leq \epsilon n^2. \qquad \square$$

From Lemma 18 we can easily conclude that the set S found by the algorithm has discrepancy $|f(S)| \geq (1 - \rho)OPT(f)$. Indeed, let T be such that $|f(T)| = OPT(f)$ and S be the set chosen by the algorithm. From the note after Fact 17 we know that $|f^*(S)| \geq |f^*(T)|$ and Lemma 18 implies

$$|f(S) - f^*(S)| \leq \frac{7}{2}\epsilon n^2, |f(T) - f^*(T)| \leq \frac{7}{2}\epsilon n^2.$$

Thus,

$$|f(S)| = |f^*(S) + f(S) - f^*(S)| \geq |f^*(S)| - |f(S) - f^*(S)| \geq |f^*(T)| - \frac{7}{2}\epsilon n^2$$

$$= |f(T) + f^*(T) - f(T)| - \frac{7}{2}\epsilon n^2 \geq |f(T)| - |f^*(T) - f(T)| - \frac{7}{2}\epsilon n^2$$

$$\geq |f(T)| - 7\epsilon n^2.$$

Since $|f(T)| \geq cn^2$ and $\epsilon = \frac{\rho c}{7}$ we get $|f(S)| \geq (1-\rho)|f(T)|$. □

We will now turn our attention to the linear ordering problem. Let $T_n = (V, A)$ be a tournament with $V = [n]$. We denote by $OPT(T_n) = \max_{P_n} |T_n \cap P_n|$, where max is taken over all transitive tournaments of order n. For a pair of subsets (V_1, V_2) with $V_1 \cap V_2 = \emptyset$ we define the tournament density $d_T(V_1, V_2)$ as follows: $d_T(V_1, V_2) = \frac{arcs(V_1,V_2)}{|V_1||V_2|}$, where $arcs(V_1, V_2)$ is the number of arcs that start at V_1 and end at V_2. Note that $d_T(V_1, V_2) = 1 - d_T(V_2, V_1)$.

Proof of Theorem 4. The ranking σ' can be constructed by the following procedure: Let $\epsilon = (\frac{\rho}{12})^2$.

1. Define an auxiliary graph G_T as $G_T = (V, E)$, where $E = \{\{v_i, v_j\} : i < j, (v_i, v_j) \in A\}$. Let $l = \frac{1}{\epsilon}$ and let $U_i = \{v_{\frac{n}{l}(i-1)}, \dots, v_{\frac{n}{l}i}\}$ where $i = 1, \dots, l$.
2. Apply Lemma 15 to obtain an ϵ-regular partition of V into $V_1 \cup \dots \cup V_k$, which is a refinement of $U_1 \cup \dots \cup U_l$.
3. Check all $k!$ permutations of the sets $\{V_1, \dots, V_k\}$ to find a permutation σ that maximizes $\sum_{1 \leq i_1 < i_2 \leq k} d_T(V_{\sigma(i_1)}, V_{\sigma(i_2)})|V_{\sigma(i_1)}||V_{\sigma(i_2)}|$.
4. Extend σ inside each of V_i in an arbitrary way to obtain the ranking σ' of V.

Let us first observe that in the first two steps of the algorithm we actually construct an ϵ-regular partition of the tournament T, where the regularity is defined as follows.

DEFINITION 24. *A pair of subsets (V_1, V_2) of V with $V_1 \cap V_2 = \emptyset$ is ϵ-regular in tournament (V, A) if for every $W_1 \subset V_1$ with $|W_1| \geq \epsilon|V_1|$, and every $W_2 \subset V_2$ with $|W_2| \geq \epsilon|V_2|$,*

$$|d_T(W_1, W_2) - d_T(V_1, V_2)| \leq \epsilon.$$

Then, since $\max U_i < \min U_j$ for $i < j$, the following fact holds.

FACT 25. *For $i < j$ let $V_i \subset U_i$ and $V_j \subset U_j$. If (V_i, V_j) is ϵ-regular in the graph G_T with density d_{ij}, then the pair (V_i, V_j) is ϵ-regular in tournament T with density $d_T(V_i, V_j) = d_{ij}$.*

Thus $V_1 \cup \dots \cup V_k$ is an ϵ-regular partition of a tournament T. Without loss of generality, we may assume that the optimal ordering of V is $1 < 2 < \dots < n$. For a subset $S \subset V$, define $h(S)$ as the number of arcs of T that agree with the optimal ordering, i.e., $h(S) = |\{(i, j) \in A : i < j, \text{ and } i, j \in S\}|$. For sets $S_1, S_2 \subset V$ with $S_1 \cap S_2 = \emptyset$ let $h(S_1, S_2)$ be the number of arcs of T between S_1 and S_2 that agree with the optimal ordering, i.e., $h(S_1, S_2) = |\{(i, j) \in A : i < j, i \in S_1, j \in S_2 \text{ or } i \in S_2, j \in S_1\}|$. Note that $h(S_1, S_2) = h(S_2, S_1)$. Define sets $Z_j = \{\frac{n}{s}(j-1), \dots, \frac{n}{s}j\}$, where $s = \frac{1}{\sqrt{\epsilon}}$ and $i = 1, \dots, s$. Simple computations show the following.

FACT 26.
1. $\sum_{j=1}^{s} h(Z_j) \leq \frac{\sqrt{\epsilon}}{2} n^2$;
2. $\sum_{i=1}^{k} h(V_i) \leq \frac{\epsilon}{2} n^2$.

Let $W_{ij} = V_i \cap Z_j$ where $i = 1, \dots, k$ and $j = 1, \dots, s$. We define

$$h^* = \sum_{1 \leq j_1 < j_2 \leq s} \sum_{i_1 \neq i_2} d_T(V_{i_1}, V_{i_2})|W_{i_1 j_1}||W_{i_2 j_2}|.$$

We will show that the number of arcs that agree with the optimal ordering cannot be much larger than h^*, namely, the following.

LEMMA 27. $h(V) \leq h^* + \frac{1}{2}(3\sqrt{\epsilon} + 5\epsilon)n^2$.

Before giving a proof we will establish some auxiliary facts.

CLAIM 28. $h(V) \leq \sum_{1 \leq j_1 < j_2 \leq s} \sum_{i_1 \neq i_2} h(W_{i_1 j_1}, W_{i_2 j_2}) + \frac{1}{2}(\epsilon + \sqrt{\epsilon})n^2$.

Indeed, since $\{W_{ij}\}$ is a partition of V we have

$$h(V) \leq \sum_{1\leq j_1<j_2\leq s} \sum_{i_1\neq i_2} h(W_{i_1j_1}, W_{i_2j_2}) + \sum_{i=1}^{k} h(V_i) + \sum_{j=1}^{s} h(Z_j)$$

$$\leq \sum_{1\leq j_1<j_2\leq s} \sum_{i_1\neq i_2} h(W_{i_1j_1}, W_{i_2j_2}) + \frac{\epsilon}{2}n^2 + \frac{\sqrt{\epsilon}}{2}n^2$$

by Fact 26.

We adopt the notation from the proof of Lemma 18. Let $[k]\times[k] = R\cup I$ where $(i_1, i_2) \in I$ if and only if (V_{i_1}, V_{i_2}) is not ϵ-regular in a tournament T. Note that if $(i_1, i_2) \in I$, then either

- $V_{i_1}, V_{i_2} \subset U_i$ for some $i \in [l]$ or
- (V_{i_1}, V_{i_2}) is irregular in the graph G_T.

Claim 29. $h(V) \leq \sum_{1\leq j_1<j_2\leq s} \sum_{(i_1,i_2)\in R} h(W_{i_1j_1}, W_{i_2j_2}) + \frac{1}{2}(3\epsilon + \sqrt{\epsilon})n^2$.

To prove the claim we bound $\sum_{1\leq j_1<j_2\leq s} \sum_{(i_1,i_2)\in I} h(W_{i_1j_1}, W_{i_2j_2})$ from above.

$$\sum_{1\leq j_1<j_2\leq s} \sum_{(i_1,i_2)\in I} h(W_{i_1j_1}, W_{i_2j_2}) = \sum_{(i_1,i_2)\in I} \sum_{j_1<j_2} h(W_{i_1j_1}, W_{i_2j_2})$$

$$\leq \sum_{i}^{l} h(U_i) + \epsilon\binom{k}{2}\frac{n^2}{k^2} \leq l\binom{\frac{n}{l}}{2} + \frac{\epsilon}{2}n^2 \leq \epsilon n^2.$$

Thus,

$$h(V) \leq \sum_{1\leq j_1<j_2\leq s} \sum_{(i_1,i_2)\in R} h(W_{i_1j_1}, W_{i_2j_2}) + \frac{1}{2}(3\epsilon + \sqrt{\epsilon})n^2.$$

Finally, let $[s]\times[k] = B\cup S$, where $S = \{(j,i), |W_{ij}| < \epsilon|V_i|\}$.

Claim 30. $h(V) \leq \sum_{1\leq j_1<j_2\leq s} \sum\{h(W_{i_1j_1}, W_{i_2j_2}), (i_1,i_2)\in R, (j_1,i_1), (j_2,i_2) \in B\} + \frac{1}{2}(3\epsilon + 3\sqrt{\epsilon})n^2$.

Indeed, for $(j_1, i_1) \in S$ we have $h(W_{i_1j_1}, W_{i_2j_2}) < \epsilon|V_{i_1}||W_{i_2j_2}|$. Therefore,

$$\sum_{j_1<j_2} \sum\{h(W_{i_1j_1}, W_{i_2j_2}), (j_1,i_1)\in S, or (j_2,i_2)\in S\} \leq \sum_{[k]\times[k], i_1\neq i_2} \sum_{j_1<j_2} \epsilon|V_{i_1}||W_{i_2j_2}|$$

$$\leq \sum_{[k]\times[k], i_1\neq i_2} \sum_{j_1=1}^{s} \sum_{j_2=1}^{s} \epsilon|V_{i_1}||W_{i_2j_2}| \leq \epsilon s \sum_{[k]\times[k], i_1\neq i_2} |V_{i_1}||V_{i_2}| \leq \epsilon s k^2\frac{n^2}{k^2} = \sqrt{\epsilon}n^2$$

as $s = \frac{1}{\sqrt{\epsilon}}$.

Proof of Lemma 27. To show Lemma 27, we need to prove that $h(V) \leq h^* + \frac{1}{2}(7\epsilon + 3\sqrt{\epsilon})n^2$. For $j_1 < j_2$ we have

$$h(W_{i_1j_1}, W_{i_2j_2}) = arcs(W_{i_1j_1}, W_{i_2j_2}) = d_T(W_{i_1j_1}, W_{i_2j_2})|W_{i_1j_1}||W_{i_2j_2}|.$$

Since $|W_{i_1j_1}| \geq \epsilon|V_{i_1}|$, $|W_{i_2j_2}| \geq \epsilon|V_{i_2}|$, and (V_{i_1}, V_{i_2}) is ϵ-regular we can approximate $d_T(W_{i_1j_1}, W_{i_2j_2}) \leq d_T(V_{i_1}, V_{i_2}) + \epsilon$. Clearly,

$$\sum_{j_1<j_2} \sum_{i_1 \neq i_2} \epsilon|W_{i_1j_1}||W_{i_2j_2}| = \epsilon \sum_{i_1 \neq i_2} \sum_{j_1<j_2} |W_{i_1j_1}||W_{i_2j_2}| = \epsilon \sum_{i_1 \neq i_2} |V_{i_1}||V_{i_2}| \leq \epsilon n^2.$$

From Claim 30

$$h(V) \leq \sum_{1 \leq j_1 < j_2 \leq s} \sum \{h(W_{i_1j_1}, W_{i_2j_2}), (i_1, i_2) \in R, (j_1, i_1), (j_2, i_2) \in B\} + \frac{1}{2}(3\epsilon + 3\sqrt{\epsilon})n^2$$

$$\leq \sum_{j_1<j_2} \sum_{i_1 \neq i_2} d_T(V_{i_1}, V_{i_2})|W_{i_1j_1}||W_{i_2j_2}| + \frac{1}{2}(5\epsilon + 3\sqrt{\epsilon})n^2 = h^* + \frac{1}{2}(5\epsilon + 3\sqrt{\epsilon})n^2. \qquad \square$$

To complete the proof of Theorem 4, we first introduce an auxiliary digraph K with vertices corresponding to sets W_{ij}, weights on arcs corresponding to approximation of the number of arcs that are consistent with optimal ordering. More formally, let K be a complete k-partite, symmetric digraph with a vertex set $V(K) = \{y_{ij} : i \in [k], j \in [s]\}$ and with weights on arcs defined as follows: $w(y_{i_1j_1}, y_{i_2j_2}) = d_T(V_{i_1}, V_{i_2})|W_{i_1j_1}||W_{i_2j_2}|$ if $i_1 \neq i_2$, and $w(y_{i_1j_1}, y_{i_1j_2}) = 0$. Let $Y_i = \bigcup_{j \in [s]}\{y_{ij}\}$. Vertex $y_{ij} \in Y_i$ corresponds to the set $W_{ij} \subset V_i$ and Y_i corresponds to V_i, $\bigcup_{i \in [k]} y_{ij}$ to Z_j. We define the ordering $\prec$ of $V(K)$ in the following way: $y_{i_1j_1} \prec y_{i_2j_2}$ if and only if either $j_1 < j_2$ or $j_1 = j_2$ and $i_1 < i_2$. Then

$$h^* = \sum_{1 \leq j_1 < j_2 \leq s} \sum_{i_1 \neq i_2} w(y_{i_1j_1}, y_{i_2j_2}) \leq \sum_{y_{i_1j_1} \prec y_{i_2j_2}} w(y_{i_1j_1}, y_{i_2j_2}).$$

The final part of the proof is based on the following lemma.

LEMMA 31 (ordering lemma). *There exists a permutation $\sigma : [k] \to [k]$ such that for every ordering $\prec$ of $V(K)$*

$$\sum_{y_{i_1j_1} \prec y_{i_2j_2}} w(y_{i_1j_1}, y_{i_2j_2}) \leq \sum_{1 \leq i_1 < i_2 \leq k} \sum_{j_1, j_2 \in [s]} w(y_{\sigma(i_1)j_1}, y_{\sigma(i_2)j_2}).$$

In other words, the sum of weights of the arcs is maximized for an ordering $<$ in which $Y_{i_1} < Y_{i_2} < \cdots < Y_{i_k}$. We postpone the proof of Lemma 31 until the end of this section.

LEMMA 32. $h^* \leq \max_\sigma \sum_{1 \leq i_1 < i_2 \leq k} d_T(V_{\sigma(i_1)}, V_{\sigma(i_2)})|V_{\sigma(i_1)}||V_{\sigma(i_2)}|$

Proof. By the ordering lemma, there exists a permutation $\sigma : [k] \to [k]$ such that

$$h^* \leq \sum_{y_{i_1j_1} \prec y_{i_2j_2}} w(y_{i_1j_1}, y_{i_2j_2}) \leq \sum_{1 \leq i_1 < i_2 \leq k} \sum_{j_1, j_2 \in [s]} w(y_{\sigma(i_1)j_1}, y_{\sigma(i_2)j_2})$$

$$= \sum_{1 \leq i_1 < i_2 \leq k} \sum_{j_1, j_2 \in [s]} d_T(V_{\sigma(i_1)}, V_{\sigma(i_2)})|W_{i_1j_1}||W_{i_2j_2}|$$

$$= \sum_{1 \leq i_1 < i_2 \leq k} d_T(V_{\sigma(i_1)}, V_{\sigma(i_2)})|V_{\sigma(i_1)}||V_{\sigma(i_2)}|$$

$$\leq \max_{\sigma} \sum_{1\leq i_1<i_2\leq k} d_T(V_{\sigma(i_1)}, V_{\sigma(i_2)})|V_{\sigma(i_1)}||V_{\sigma(i_2)}|. \qquad \square$$

The number of arcs that are consistent with constructed ranking σ' is at least $\max_\sigma \sum_{1\leq i_1<i_2\leq k} d_T(V_{\sigma(i_1)}, V_{\sigma(i_2)})|V_{\sigma(i_1)}||V_{\sigma(i_2)}|$, which by Lemma 27 and Lemma 32 is at least $h(V) - \frac{1}{2}(5\epsilon + 3\sqrt{\epsilon})n^2$. When we combine it with the lower bound $h(V) = OPT(T_n) \geq \frac{1}{4}n^2$ mentioned in the introduction we conclude that the number of arcs that are consistent with constructed ordering is at least $(1-\rho)OPT(T_n)$ since $\rho \geq 10\epsilon + 6\sqrt{\epsilon}$. $\square$

We will now prove the ordering lemma.

Proof of the ordering lemma. To prove the lemma, it is sufficient to show that the sum of weights of arcs is maximized for an ordering in which every Y_i is an interval. Let $\prec$ be an ordering of $V(K)$. We denote by $h(\prec)$ the sum $\sum_{y_{i_1j_1}\prec y_{i_2j_2}} w(y_{i_1j_1}, y_{i_2j_2})$ and for every Y_i, where $i = 1, \dots, k$, we define a gap-number $g_i = gap_\prec(Y_i)$ as the minimum number of intervals I_{ij} such that $Y_i = \bigcup_{j=1}^{g_i+1} I_{ij}$. Note that the gap-numbers depend on the ordering of $V(K)$.

CLAIM 33. *If $gap_\prec(Y_{i_0}) > 0$ then there exists an ordering $\prec^*$ such that*

1. $h(\prec) \leq h(\prec^*)$,
2. $gap_{\prec^*}(Y_{i_0}) < gap_\prec(Y_{i_0})$, *and*
3. $gap_{\prec^*}(Y_i) = gap_{\prec^*}(Y_i)$ *for every* $i \neq i_0$.

Applying the claim to $Y_1, Y_2, \dots, Y_k$, we construct the ordering in which every $g_i = 0$, i.e., all Y_i are intervals. $\square$

Proof of the claim. Since $gap_\prec(Y_{i_0}) > 0$ there exist two intervals I_{i_01}, I_{i_02} such that $I_{i_01}, I_{i_02} \in Y_{i_0}$ and

$$I_{i_01} < I_{i_1j_1} < I_{i_2j_2} < \cdots < I_{i_tj_t} < I_{i_02}.$$

Without loss of generality we may assume that $d_T(V_{i_0}, V_{i_1})|W_{i_1j_1}| + \cdots + d_T(V_{i_0}, V_{i_t})$ $|W_{i_tj_t}| \geq d_T(V_{i_1}, V_{i_0})|W_{i_1j_1}| + \cdots + d_T(V_{i_t}, V_{i_0})|W_{i_tj_t}|$. Then the sum of the weights of arcs between intervals $I_{i_01} < I_{i_1j_1} < I_{i_2j_2} < \cdots < I_{i_tj_t} < I_{i_02}$ is

$$d_T(V_{i_0}, V_{i_1})|W_{i_1j_1}||I_{i_01}| + \cdots + d_T(V_{i_0}, V_{i_t})|W_{i_tj_t}||I_{i_01}|$$

$$+d_T(V_{i_1}, V_{i_0})|W_{i_1j_1}||I_{i_02}| + \cdots + d_T(V_{i_t}, V_{i_0})|W_{i_tj_t}||I_{i_02}|$$

$$\leq d_T(V_{i_0}, V_{i_1})|W_{i_1j_1}||I_{i_01}| + \cdots + d_T(V_{i_0}, V_{i_t})|W_{i_tj_t}||I_{i_01}|$$

$$+d_T(V_{i_0}, V_{i_1})|W_{i_1j_1}||I_{i_02}| + \cdots + d_T(V_{i_t0}, V_{i_t})|W_{i_tj_t}||I_{i_02}|,$$

which equals the sum of weights of arcs between intervals

$$I_{i_01} < I_{i_02} < I_{i_1j_1} < I_{i_2j_2} < \cdots < I_{i_tj_t}.$$

Therefore, we can reduce the number of gaps of Y_{i_0}. $\square$

4. Conclusions and an open problem. In this paper, we considered the weighted version of discrepancy and tournament ranking problems. In the first part of the paper we generalized the approach from [12] to weighted graphs. In the second part we presented algorithms for both problems which were based on the algorithmic regularity lemma. We want to conclude with the following open problem.

Open Problem 1. *For a given n, construct an $m \times n$ matrix $M = [m_{ij}]$ of $+1$'s and -1's with m small which has the following property. For every vector $\overrightarrow{u} \in \{-1,1\}^n$*

$$\frac{1}{m}\sum_{i=1}^{m}\left|\sum_{j=1}^{n} m_{ij}u_j\right| \geq c\sqrt{n}$$

for some constant c.

By probabilistic method one can show the existence of matrix M with $m = n$ and a constant $c = 0.0017$ sufficiently small. Note that Hadamard matrices do not possess the required property, taking $\overrightarrow{u}$ as one of the row vectors of M results in $\sum_{i=1}^{m}|\sum_{j=1}^{n} m_{ij}u_j| = n$.

Let us observe that we can use the solution matrix M to our initial problem of finding a sign vector. Namely, for given $\overrightarrow{v_1}, \ldots, \overrightarrow{v_n} \in \{-1,1\}^n$, there is an $O(n^2 m)$ algorithm that finds $\overrightarrow{X} = (X_1, \ldots, X_n) \in \{-1,1\}^n$ such that

$$||X_1\overrightarrow{v_1} + \cdots + X_n\overrightarrow{v_n}|| \geq cn^{3/2}.$$

Indeed, let $\overrightarrow{v_i} = (v_{i,1}, \ldots, v_{i,n})$ and $\overrightarrow{w_j} = (v_{1,j}, \ldots, v_{n,j})$. We can construct a sign vector in the following way: For every row vector $\overrightarrow{m_i}$ of matrix M we compute $\sum_{j=1}^{n}|\langle\overrightarrow{w_j}, \overrightarrow{m_i}\rangle|$ and we choose $\overrightarrow{m_i}$ such that the sum is the largest.

By the property of the matrix M, we know that for every vector $\overrightarrow{w_j}$, $\sum_{i=1}^{m}|\langle\overrightarrow{w_j}, \overrightarrow{m_i}\rangle| \geq cm\sqrt{n}$ and so $\sum_{j=1}^{n}\sum_{i=1}^{m}|\langle\overrightarrow{w_j}, \overrightarrow{m_i}\rangle| \geq cmn^{3/2}$. This implies that if a vector $\overrightarrow{m} = (m_1, \ldots, m_n)$ is chosen by the algorithm, then $\sum_{j=1}^{n}|\langle\overrightarrow{w_j}, \overrightarrow{m}\rangle| \geq cn^{3/2}$. We verify that

$$||\overrightarrow{v_1}m_1 + \cdots + \overrightarrow{v_n}m_n|| = \sum_{j=1}^{n}|v_{1,j}m_1 + \cdots + v_{n,j}m_n| = \sum_{j=1}^{n}|\langle\overrightarrow{w_j}, \overrightarrow{m}\rangle| \geq cn^{3/2}.$$

In computing the sum $\sum_{j=1}^{n}|\langle\overrightarrow{w_j}, \overrightarrow{m_i}\rangle|$ we add n^2 numbers of size $O(1)$. Note that the same argument can be repeated (resulting in different constant c) if $\overrightarrow{v_1}, \ldots, \overrightarrow{v_n} \in \{-1,1\}^k$, and $k = \Theta(n)$.

As long as m is smaller then $n \lg n$ this will improve the time complexity of results in [12]. A similar question can be asked for the weighted case.

Acknowledgments. We would like to thank referees for helpful comments and suggestions.

REFERENCES

[1] N. Alon, R.A. Duke, H. Lefmann, V. Rödl, and R. Yuster, *The algorithmic aspects of the regularity lemma*, J. Algorithms, 16 (1994), pp. 80–109.

[2] A. Czygrinow, S. Poljak, and V. Rödl, *On the Linear Ordering Problem*, Emory University Technical Report, Atlanta, GA, 1996.

[3] R.A. Duke, H. Lefmann, and V. Rödl, *A fast algorithm for computing frequencies in a given graph*, SIAM J. Comput., 24 (1995), pp. 598–620.

[4] P. Erdos and J. Spencer, *Probabilistic Methods in Combinatorics*, Academic Press, New York, 1974.

[5] W. Fernandez de la Vega, *On the maximal cardinality of a consistent set of arcs in a random tournament*, J. Combin. Theory Ser. B, 35 (1983), pp. 328–332.

[6] A. Frieze and R. Kannan, *The regularity lemma and approximation schemes for dense problems*, in Proc. 37th FOCS, 1996, pp. 12–20.

[7] A. Frieze and R. Kannan, *Quick Approximation to Matrices and Applications*, preliminary manuscript, February 1997.

[8] M. R. Garey and D. S. Johnson, *Computers and Intractability. A Guide to the Theory of NP-Completeness*, W.H. Freeman, San Francisco, CA, 1979.

[9] M. X. Goemans and D. P. Williamson, .878-*Approximation Algorithm for MAX CUT and MAX 2SAT*, in Proc. 26th STOC, 1994, pp. 422–431.

[10] M. Grötschel, M. Jünger, and G. Reinelt, *A cutting plane algorithm for the linear ordering problem*, Oper. Res., 32 (1984), pp. 1195–1220.

[11] J. K. Lenstra, *Sequencing by Enumerative Methods*, in Mathematical Center Tracts 69, Mathematisch Centrum, Amsterdam.

[12] S. Poljak, V. Rödl, and J. Spencer, *Tournament ranking with expected profit in polynomial time*, SIAM J. Discrete Math., 1 (1988), pp. 372–376.

[13] S. Poljak and D. Turzik, *A polynomial time heuristic for certain subgraph optimization problems with guaranteed lower bound*, Discrete Math., 58 (1986), pp. 99–138.

[14] J. Spencer, *Optimal ranking of tournaments*, Networks, 1 (1971), pp. 135–138.

[15] J. Spencer, *Nonconstructive Methods in Discrete Mathematics*, in Studies in Combinatorics, G. C. Rota, ed., Mathematical Association of America, Washington, D.C., 1978, pp. 142–178.

[16] E. Szemerédi, *Regular partitions of graphs*, in Proc. Colloque Internat. CNRS, J.C. Bermond, ed., Paris, 1978, pp. 399-401.