On Graphs That Do Not Contain The Cube And Related Problems*

Rom Pinchasi† Micha Sharir‡

October 27, 2002

1 Introduction

Let Q denote the edge graph of the 3-dimensional cube (it has 8 vertices and 12 edges). The *Turan number* of Q is the maximum number of edges in a graph on n vertices that does not contain Q. Back in 1969, Erdős and Simonovits [1] have shown that the Turan number of Q is $O(n^{8/5})$. In this paper, we provide an alternative simpler proof of this result. The original proof in [1] was based on the assumption that the given graph G is regular, and required a nontrivial technical lemma that reduces the general case to the case of a regular graph. Our proof does not need to assume that G is regular (and so does not use this lemma), and follows a different approach that appears to be more powerful than the one in [1]. We demonstrate this by applying the technique to obtain Turan numbers for more general graphs than Q, which the method of [1] seems incapable of achieving.

2 Graphs That Do Not Contain the 3-Dimensional Cube

Certain applications of extremal graph theory involve bipartite graphs where the sizes of the two vertex sets are different from each other. For this reason, we state our main result for the bipartite case; the non-bipartite case is then an immediate corollary.

Theorem 2.1 *A bipartite graph $G \subseteq A \times B$, with $|A| = m$, $|B| = n$, that does not contain Q has $O(m^{4/5}n^{4/5} + mn^{1/2} + nm^{1/2})$ edges.*

Proof: Let E denote the number of edges of G. We define a *configuration* to be a 6-tuple $\kappa = (u, v, a, b, c, d)$ of distinct vertices of G, such that $a, c, v \in A$, $u, b, d \in B$, and (v, u),

*Work on this paper by Micha Sharir has been supported by NSF Grants CCR-97-32101 and CCR-00-98246, by a joint grant from the U.S.–Israel Binational Science Foundation, by a grant from the Israeli Academy of Sciences for a Center of Excellence in Geometric Computing at Tel Aviv University, and by the Hermann Minkowski–MINERVA Center for Geometry at Tel Aviv University.

†Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; room@math.mit.edu

‡School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel, and Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA; sharir@cs.tau.ac.il

(a,u), (a,b), (v,b), (c,u), (c,d), (v,d) are all edges of G; see Figure 1. In other words, κ consists of two C_4's (cycles of length 4) with a common edge and with no other common vertex. We say that a pair (a,b), (c,d) of edges of G is *tame* if all four endpoints are distinct, and there exist at most two pairs (u,v) such that (u,v,a,b,c,d) is a configuration. We say that $\kappa = (u,v,a,b,c,d)$ is a *tame* configuration if (a,b), (c,d) form a tame pair.

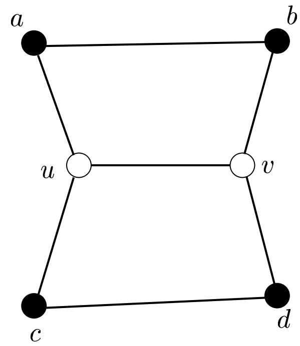

Figure 1: A configuration in the proof of Theorem 2.1.

Let K be the set of all tame configurations. A trivial upper bound for $|K|$ is $O(E^2)$, because the number of pairs $(a,b),(c,d)$ of edges of G is $O(E^2)$, and each of them, if tame, gives rise to only two configurations in K.

We next obtain a lower bound for $|K|$, as follows. Fix an edge $e = (v,u)$ of G, with $v \in A$, $u \in B$, and let G_e denote the graph whose vertices are the neighbors of either u or v in G, and whose edges are the edges of G that connect pairs of these neighbors. Let A_e, B_e, E_e denote the number of vertices in A, of vertices in B, and of edges of G_e, respectively. Put $V_e = A_e + B_e$.

We note that the (tame or untame) configurations of the form (u,v,a,b,c,d), for the fixed pair of vertices u,v, correspond in a 1-1 manner to the vertex-disjoint pairs of edges of G_e. For a vertex a of G_e, let $\delta_e(a)$ denote the degree of a in G_e.

Lemma 2.2 *The number of vertex-disjoint tame pairs of edges in G_e is at least $\frac{1}{2}E_e(E_e - 4V_e - 1)$.*

Proof: Call any pair of edges, which is not tame, a *bad* pair. We wish to bound the number of bad pairs of edges in G_e. Suppose that (a,b), (c,d) is a bad pair. Consider any other pair (u',v'), such that (u',v',a,b,c,d) is also a configuration. If $u \neq u'$ and $v \neq v'$ then the 8-tuple (u,v,u',v',a,b,c,d) forms a forbidden copy of Q in G, contrary to assumption; see Figure 2. Hence, either all such pairs (u',v') satisfy $u' = u$, or all such pairs satisfy $v' = v$. In the former case we say that the pair (a,b), (c,d) is a *bad u-pair* and in the latter case we say that the pair (a,b), (c,d) is a *bad v-pair.*

By symmetry, it suffices to bound the number of bad u-pairs, and twice this bound will serve as a bound for the number of all bad pairs in G_e.

Let $a,c \in A$ be two distinct neighbors of u in G_e, and let $b \in B$ be a neighbor of a in G_e (so b is a neighbor of v in G). There is at most one edge of G_e incident to c (namely, (c,b)) that shares a vertex with (a,b).

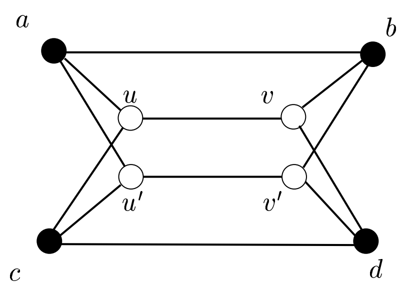

Figure 2: A double configuration forming a cube in G.

We claim that there is at most one vertex $d \in B$ (different from b) such that (a, b), (c, d) is a bad u-pair. Indeed, assume to the contrary that there are at least two such neighbors of c, say d_1, d_2. Since (a, b), (c, d_1) is a bad u-pair, there exist at least two distinct vertices $v_1', v_1'' \in A$, different from v, such that (u, v_1', a, b, c, d_1) and (u, v_1'', a, b, c, d_1) are configurations. Similarly, there exist at least two distinct vertices $v_2', v_2'' \in A$, different from v, such that (u, v_2', a, b, c, d_2) and (u, v_2'', a, b, c, d_2) are configurations. Clearly, there exist a pair of vertices $v_1 \in \{v_1', v_1''\}$, $v_2 \in \{v_2', v_2''\}$, such that $v_1 \neq v_2$. Then the eight vertices $(v, v_1, v_2, c, u, b, d_1, d_2)$ form a forbidden copy of Q in G (in fact, it is a copy of Q plus one main diagonal (u, v)); see Figure 3.

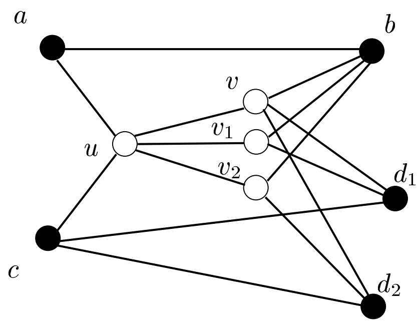

Figure 3: Two untame pairs $\{(a, b), (c, d_1)\}$ and $\{(a, b), (c, d_2)\}$, forming a cube in G.

This contradiction shows that there exist at most two 'bad' neighbors of c in G_e, with respect to the fixed edge (a, b) of that graph: one of them is b, and at most one other vertex d forms a bad u-pair (a, b), (c, d) in G_e. Let $N(u)$ (resp., $N(v)$) denote the set of neighbors of u (resp., of v); thus $A_e = |N(u)|$, $B_e = |N(v)|$. Then the number of bad u-pairs of edges (a, b), (c, d) is at most $2A_eE_e$. Similarly, the number of bad v-pairs of edges is at most $2B_eE_e$. Therefore the total number of bad pairs in G_e is at most $2E_e(A_e + B_e) = 2E_eV_e$.

It follows that the number M_e of tame pairs of edges in G_e satisfies

$$M_e \geq \binom{E_e}{2} - 2E_eV_e = \frac{1}{2}E_e(E_e - 4V_e - 1),$$

as asserted. □

Put

$$G_1 = \{e \in G \mid E_e \geq 8V_e\},$$

$$G_2 = \{e \in G \mid E_e < 8V_e\}.$$

The total number $|K|$ of tame configurations thus satisfies

$$|K| = \sum_{e\in G} M_e \geq \sum_{e\in G_1} M_e \geq \sum_{e\in G_1} \frac{E_e(E_e - 4V_e - 1)}{2} \geq$$

$$\frac{1}{4}\sum_{e\in G_1} E_e^2 - \frac{1}{2}\sum_{e\in G} E_e \geq \frac{\left(\sum_{e\in G_1} E_e\right)^2}{4E} - \frac{\sum_{e\in G} E_e}{2}.$$

Assume for the time being that $\sum_{e\in G_2} E_e \leq \frac{1}{2}\sum_{e\in G} E_e$; the complementary case will be treated later. Then $\sum_{e\in G_1} E_e \geq \frac{1}{2}\sum_{e\in G} E_e$, and

$$|K| \geq \frac{\left(\sum_{e\in G} E_e\right)^2}{16E} - \frac{\sum_{e\in G} E_e}{2}.$$

Note that $\sum_e E_e = 4S$, where S is the number of C_4's in G. Hence,

$$|K| \geq \frac{S^2}{2E} - 2S.$$

For each pair of distinct vertices u, v of G (both in A or both in B), let $W_{u,v}$ denote the number of paths of length 2 that connect u and v in G. Note that $\sum_{u\neq v\in A} W_{u,v} = W^{(A)}$, where $W^{(A)}$ is the number of paths of length 2 in G whose extreme vertices are in A and whose middle vertex is in B, and that $\displaystyle\sum_{u\neq v\in A}\binom{W_{u,v}}{2} = S$. Similarly, $\sum_{u\neq v\in B} W_{u,v} = W^{(B)}$, where $W^{(B)}$ is the number of paths of length 2 in G whose extreme vertices are in B and whose middle vertex is in A, and $\displaystyle\sum_{u\neq v\in B}\binom{W_{u,v}}{2} = S$. Then we can lower bound S by

$$S = \sum_{u,v\in A}\binom{W_{u,v}}{2} = \sum_{u,v\in A}\left[\frac{W_{u,v}^2}{2} - \frac{W_{u,v}}{2}\right] \geq \frac{\left(\sum_{u,v\in A} W_{u,v}\right)^2}{2\binom{m}{2}} - \frac{\sum_{u,v\in A} W_{u,v}}{2} = \frac{(W^{(A)})^2}{2\binom{m}{2}} - \frac{W^{(A)}}{2}.$$

Finally, we have $W^{(A)} = \displaystyle\sum_{u\in B}\binom{\delta(u)}{2}$, where $\delta(u)$ is the degree of a vertex u in G. Hence,

$$W^{(A)} = \sum_{u\in B}\binom{\delta(u)}{2} = \sum_{u\in B}\left[\frac{\delta(u)^2}{2} - \frac{\delta(u)}{2}\right] \geq \frac{\left(\sum_{u\in B}\delta(u)\right)^2}{2n} - \frac{\sum_{u\in B}\delta(u)}{2} = \frac{2E^2}{n} - E.$$

We next assume that $W^{(A)} \geq 2\binom{m}{2}$, which implies that $S \geq \frac{(W^{(A)})^2}{4\binom{m}{2}}$. Finally, we assume that $E \geq n$, which implies that $W^{(A)} \geq \frac{E^2}{n}$. Putting it all together, we obtain

$$2S + |K| = \Omega\left(\frac{S^2}{E}\right) = \Omega\left(\frac{(W^{(A)})^4}{m^4 E}\right) = \Omega\left(\frac{E^7}{m^4 n^4}\right).$$

Combining this with the upper bound for $|K|$, which also holds trivially for S, we obtain

$$\frac{E^7}{m^4 n^4} = O(E^2),$$

or $E = O(m^{4/5}n^{4/5})$.

It remains to handle the cases that we have ignored so far. First, if $E \le n$ then clearly E satisfies the asserted bound. Next, suppose that $W^{(A)} \le 2\binom{m}{2}$. Since $W^{(A)} = \sum_{u\in B}\binom{\delta(u)}{2}$, we obtain

$$E = \sum_{u\in B}\delta(u) \le n + \sum_{u\in B,\ \delta(u)\ge 1}(\delta(u)-1) = O\left[n + \left(\sum_{u\in B}\binom{\delta(u)}{2}\right)^{1/2}\cdot n^{1/2}\right] = O(mn^{1/2}+n).$$

Note that by interchanging the roles of A and B, we may also assume that $E \ge m$ and that $W^{(B)} \ge 2\binom{n}{2}$. Otherwise we get, as above, $E = O(nm^{1/2} + m)$.

Finally, assume that neither of these inequalities hold but that

$$\sum_{e\in G_2} E_e > \frac{1}{2}\sum_{e\in G} E_e = 2S.$$

We have

$$\sum_{e\in G_2} E_e < 8\sum_{e\in G_2}(A_e + B_e) \le 8\sum_{e\in G}(A_e + B_e) = 16(W^{(A)} + W^{(B)}),$$

where the last equality is easily verified. Hence, $S < 8(W^{(A)} + W^{(B)})$. Suppose, without loss of generality, that $W^{(B)} \le W^{(A)}$, so $S < 16W^{(A)}$. This implies

$$S = \sum_{u,v\in A}\binom{W_{u,v}}{2} = \sum_{u,v\in A}\left[\frac{W_{u,v}^2}{2} - \frac{W_{u,v}}{2}\right] \le \sum_{u,v\in A} 16W_{u,v},$$

In other words, we have

$$\sum_{u,v\in A} W_{u,v}^2 = O(W^{(A)}),$$

which, using the Cauchy-Schwarz inequality, implies that

$$W^{(A)} = \sum_{u,v\in A} W_{u,v} \le \binom{m}{2}^{1/2}\cdot\left(\sum_{u,v\in A} W_{u,v}^2\right)^{1/2} = O(m(W^{(A)})^{1/2}),$$

or $W^{(A)} = O(m^2)$. Since we assume that $E \ge n$, we have $W^{(A)} \ge \frac{E^2}{n}$, implying that $E = O(mn^{1/2})$. The complementary case $W^{(A)} \le W^{(B)}$ yields, in a fully symmetric manner, $E = O(m^{1/2}n)$.

We have thus completed the proof of Theorem 2.1. □

The general case is now a straightforward corollary:

Corollary 2.3 *A graph with n vertices that does not contain Q has $O(n^{8/5})$ edges.*

3 A Generalization

In this section we generalize the method presented in Section 2 to bound the Turan number of more general families of graphs.

Let $k \le m$ be positive integers. Let A_1, A_2, B_1, B_2 be four pairwise disjoint sets, so that $|A_1| = |B_1| = k$ and $|A_2| = |B_2| = m$. Define $Q_{k,m}$ to be a bipartite graph whose set of edges is $(A_1 \times B_2) \cup (A_2 \times B_1) \cup M_1 \cup M_2$, where M_i is a perfect bipartite matching in $A_i \times B_i$, for $i = 1, 2$. It is easily checked that $Q_{2,2} = Q$. The following theorem bounds the Turan number of $Q_{k,m}$, but only for graphs that satisfy an additional assumption (for simplicity of presentation, we do not consider the bipartite version of this case):

Theorem 3.1 *Let $2 \le k \le m$ be positive integers, and let G be a graph on n vertices which does not contain a copy of $Q_{k,m}$, and also does not contain a copy of $K_{k+1,k+1}$. Then G has at most $O(n^{\frac{4k}{2k+1}})$ edges.*

Proof: Note first that the number of edges of a graph that satisfies only the second assumption of the theorem is $O(n^{2-\frac{1}{k+1}})$, and that this bound strictly dominates the bound asserted in the theorem, so the first assumption is non-redundant for the asserted bound.

Again, we assume without loss of generality that G is a bipartite graph. We define a *configuration* to be a $(2k+2)$-tuple $(u, v, a_1, \ldots, a_k, b_1, \ldots, b_k)$ of distinct vertices of G, so that (u, v), and (a_i, b_i), (a_i, u), (b_i, v), for $i = 1, \ldots, k$, are all edges of G,

We say that a $2k$-tuple $(a_1, \ldots, a_k, b_1, \ldots, b_k)$ of distinct vertices is *tame* if $(a_1, b_1), \ldots, (a_k, b_k)$ are all edges of G, and there are at most $2km$ edges (u, v) in G such that $(u, v, a_1, \ldots, a_k, b_1, \ldots, b_k)$ is a configuration. Every such configuration will be called a *tame* configuration. The proof proceeds along the same lines as in the proof of Theorem 2.1, but is actually simpler because of the second assumption of the theorem. It proceeds by effectively showing that all configurations are tame.

Let E denote the number of edges of G, and let N denote the number of tame configurations. An easy upper bound for N is $2kmE^k = O(E^k)$.

We next obtain a lower bound for N. We fix an edge (u, v) of G and define G_e exactly as in Section 2, namely, its vertices are the neighbors of u and the neighbors of v in G, and its edges are the edges of G that connect the neighbors of u to the neighbors of v. Define, as above, V_e and E_e to be the number of vertices and edges of G_e, respectively.

We claim that any matching of size k in G_e gives rise to a tame configuration. Indeed, let $(a_1, b_1), \ldots, (a_k, b_k)$ be such a matching, and consider all the edges (u_i, v_i) (including (u, v)), so that $(u_i, v_i, a_1, \ldots, a_k, b_1, \ldots, b_k)$ is a configuration; note that the edges (a_j, b_j) are distinct and vertex-disjoint.

Among the edges (u_i, v_i) one cannot find a matching of size m, because such a matching would have induced a copy of $Q_{k,m}$ in G. On the other hand, there can be at most k indices i with a common v_i, and at most k indices with a common u_i. Indeed, assume, without loss of generality, that $v_1 = \cdots = v_{k+1}$. Then $A = \{u_1, \ldots, u_{k+1}\}$ and $B = \{b_1, \ldots, b_k, v_1\}$ induce a copy of $K_{k+1,k+1}$ in G, contrary to assumption. Now take a maximum matching among the edges (u_i, v_i); its size is at most $m-1$, and we may write it as $(u_1, v_1), \ldots, (u_j, v_j)$, for some $j \le m-1$. Any other edge (u_i, v_i) must be incident to one of the $2j$ vertices

$u_1, \ldots, u_j, v_1, \ldots, v_j$, and each of these vertices is incident to at most $(k-1)$ such additional edges, for a total of at most $j + 2j(k-1) \leq (m-1)(2k-1) < 2km$. In other words, we have shown that, for any choice of a matching of size k from G_e, the corresponding configuration $(u, v, a_1, \ldots, a_k, b_1, \ldots, b_k)$ is tame.

The number of ways to pick k distinct and vertex-disjoint edges from G_e is at least

$$\frac{E_e(E_e - V_e)(E_e - 2V_e) \cdots (E_e - (k-1)V_e)}{k!} > \frac{(E_e - (k-1)V_e)^k}{k!},$$

assuming that $E_e \geq (k-1)V_e$.

The total number N of tame configurations satisfies, using Hölder's inequality,

$$N \geq \frac{1}{k!} \sum_{e | E_e > (k-1)V_e} (E_e - (k-1)V_e)^k \geq \frac{\left(\sum_{e|E_e>(k-1)V_e} (E_e - (k-1)V_e)\right)^k}{k! E^{k-1}}.$$

Arguing as before, one has $\sum_e E_e = 4S$, where S is the number of C_4's in G, and $\sum_e V_e = 2W$, where W is the number of paths of length 2 in G. Let us assume that $S \geq (k-1)W$. Then we have

$$\sum_{e|E_e>(k-1)V_e} (E_e - (k-1)V_e) \geq \sum_e (E_e - (k-1)V_e) = 4S - 2(k-1)W \geq 2S,$$

and thus

$$N \geq \frac{(2S)^k}{k! E^{k-1}}.$$

Using the analysis in the previous section we obtain, assuming $E \geq n$ and $W \geq 2\binom{n}{2}$, that $S = \Omega(E^4/n^4)$. Thus, $N = \Omega(E^{3k+1}/n^{4k})$. Combining this with the upper bound $O(E^k)$, we get $E = O(n^{\frac{4k}{2k+1}})$.

The remaining cases $E < n$, $W < 2\binom{n}{2}$, or $S < (k-1)W$, are analyzed in a manner similar to that in Section 2. (Recall that these cases yield the bound $E = O(n^{3/2})$, which is dominated by the bound asserted in the theorem, provided that $k \geq 2$.) $\Box$

Remarks: (1) We do not know whether Theorem 3.1 also holds without the assumption that G does not contain $K_{k+1,k+1}$.

(2) The approach of [1] seems incapable of obtaining this bound.

References

[1] P. Erdős and M. Simonovits, Some extremal problems in graph theory, *Combinatorial Theory and Its Applications* 1 (Proc. Colloq. Balatonfüred, 1969), North Holland, Amsterdam, 1970, pp. 377–390.