Turán's Graph Theorem

Martin Aigner

One of the fundamental results in graph theory is the Theorem of Turán, proved in 1941, which initiated extremal graph theory. (See the book [2] by Bollobás as a standard reference.) Turán's theorem was rediscovered many times, and it is the purpose of this article to discuss some of the most beautiful older and more recent proofs.

Let us fix some notation. We consider graphs G on the vertex-set $V = \{1, 2, \ldots, n\}$ and edge-set $E \subseteq \binom{V}{2}$. If i and j are neighbors, then we write $ij \in E$. A *k-clique* in G is a complete subgraph of G with k vertices, denoted by K_k. Turán posed the following question: Suppose G does not contain a k-clique, how many edges can G maximally have? Let us denote this number by $t(n, k)$. We have $t(n, 2) = 0$, and $t(n, k)$ is clearly an increasing function in k.

We readily obtain examples of such graphs by dividing V into $k - 1$ pairwise disjoint subsets, $V = V_1 \dot{\cup} \cdots \dot{\cup} V_{k-1}$, $|V_i| = n_i, n = n_1 + \cdots + n_{k-1}$, joining two vertices if and only if they lie in distinct V_i, V_j. Let us denote the resulting graph by $K_{n_1, \ldots, n_{k-1}}$. Figure 1 shows the graph $K_{2,2,3}$.

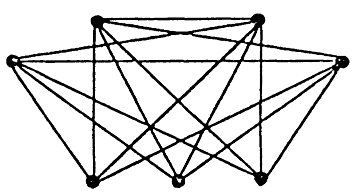

Figure 1.

The graph $K_{n_1, \ldots, n_{k-1}}$ contains $\sum_{i \neq j} n_i n_j$ edges, and it is clear that we obtain a maximal number of edges among these graphs if we divide the numbers n_i as evenly as possible, i.e. $|n_i - n_j| \leq 1$ for all i, j. If, in particular, $k - 1$ divides n, then we may choose $n_i = n/k - 1$ for all i, obtaining

$$\binom{k-1}{2} \frac{n^2}{(k-1)^2} = \frac{k-2}{k-1} \cdot \frac{n^2}{2}$$

edges. Turán's theorem now states that this number is an upper bound for the edge-number of *any* graph G on n vertices without k-cliques.

Theorem of Turán. *Let $G(V, E)$ be a graph on n vertices without a k-clique, then*

$$|E| \leq \frac{(k-2)n^2}{2(k-1)}. \tag{1}$$

More precisely, the theorem states that the graph $K_{n_1,\dots,n_{k-1}}$ with $|n_i - n_j| \le 1$ for $i \ne j$ is the *unique* graph without a k-clique with the maximal number $t(n,k)$ of edges. These graphs are therefore called *Turán graphs* $T(n,k)$. In the following, we will restrict ourselves to showing (1), but in some of the proofs we will demonstrate that the graphs $T(n,k)$ attain the maximum for arbitrary k. The uniqueness is then supplied by an easy argument.

As a warm-up let us look at the first interesting case $k = 3$: A triangle-free graph contains at most $n^2/4$ edges, and the unique extremal graph is $K_{n/2,n/2}$ if n is even, respectively $K_{(n-1)/2,(n+1)/2}$ if n is odd. For this special case, proofs were known before Turáns work. Before we look at two of them we need some more notation.

The *degree* d_i of vertex i is the number of edges incident with i. By counting in two ways we obtain

$$\sum_{i=1}^{n} d_i = 2|E|. \tag{2}$$

A set $A \subseteq V$ is called *independent*, if A contains no edges. As an example, all the defining vertex-sets V_i in the graph $K_{n_1,\dots,n_{k-1}}$ are independent. The number $\alpha(G) = \max(|U|: U \subseteq V \text{ independent})$ is called the *independence number* of G.

$k = 3$: **First Proof** (Mantel 1906). Let $ij \in E$. Since G contains no triangles we have $(d_i - 1) + (d_j - 1) \le n - 2$ (see Figure 2), hence $d_i + d_j \le n$. Summing over the edges we obtain

$$\sum_{ij \in E} (d_i + d_j) \le n|E| \tag{3}$$

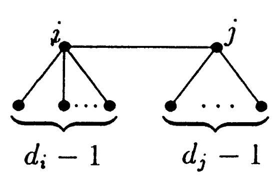

Figure 2.

The number d_i clearly appears d_i times in the sum of (3), and we conclude

$$\sum_{ij \in E} (d_i + d_j) = \sum_{i=1}^{n} d_i^2 \le n|E|. \tag{4}$$

By the Cauchy-Schwarz inequality $(\Sigma x_i y_i)^2 \le \Sigma x_i^2 \cdot \Sigma y_i^2$ applied to $x_i = d_i$, $y_i = 1$ we obtain by (2) and (4)

$$n^2|E| \ge \sum_{i=1}^{n} d_i^2 \cdot n \ge \left(\sum_{i=1}^{n} d_i\right)^2 = 4|E|^2, \tag{5}$$

and thus $|E| \le n^2/4$. □

Let us demonstrate how the uniqueness of the extremal graph $K_{n/2,n/2}$ is established for n even. (The case n odd is analogous.) If $|E| = n^2/4$, then we must have equality in (5). Now, we have equality in the Cauchy-Schwarz inequality iff the vectors are multiples of each other. For the vector (d_i) this means $d_i = d$ for all i, and we conclude $n^2|E| = n^2 d^2$ and hence $d = n/2$ because of $|E| = n^2/4$. But this immediately implies $G = K_{n/2,n/2}$.

$k = 3$: **Second proof** (Folklore). Let A be a largest independent set, $|A| = \alpha$. Since G is triangle-free, we have $d_i \leq \alpha$ for all i. The set $B = V \setminus A$ meets every edge of G, whence we obtain $|E| \leq \sum_{i \in B} d_i$ by counting in two ways. Setting $|B| = \beta = n - \alpha$ we obtain by the inequality of the arithmetic-geometric mean

$$|E| \leq \sum_{i \in B} d_i \leq \alpha \cdot \beta \leq \left(\frac{\alpha + \beta}{2} \right)^2 = \frac{n^2}{4}. \qquad \square$$

Now we turn to the proofs of the general case (1).

First proof (Turán 1941). We use induction on n. (1) is trivially true for small n. Let G be a graph on $V = \{1, \ldots, n\}$ without k-cliques with a maximal number of edges. G certainly contains $(k-1)$-cliques, since otherwise we could add edges. Let A be a $(k-1)$-clique, $B = V \setminus A, |B| = n - k + 1$ (Figure 3).

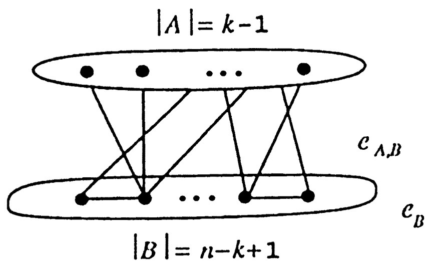

Figure 3.

A contains $\binom{k-1}{2}$ edges, and we now estimate the edge-number e_B in B and the edge-number $e_{A,B}$ between A and B. By induction, we have $e_B \leq (k - 2/2(k-1))(n - k + 1)^2$. Since G has no k-clique, every $j \in B$ is adjacent to at most $k - 2$ vertices in A, and we obtain $e_{A,B} \leq (k-2)(n-k+1)$. Altogether, this yields

$$|E| \leq \binom{k-1}{2} + \frac{k-2}{2(k-1)}(n-k+1)^2 + (k-2)(n-k+1), \tag{6}$$

which is precisely $(k - 2/2(k-1))n^2$. $\square$

Second proof (Erdös 1970). This proof makes use of the structure of the Turán graphs. Let $m \in V$ with $d_m = \max_{1 \leq j \leq n} d_j$. We denote by S the neighbors of $m, |S| = d_m$, and set $T = V \setminus S$. As G contains no k-clique, and m is adjacent to all of S, we note that S contains no $(k-1)$-clique. We now construct the following graph H on V (see Figure 4). H corresponds to G on S and contains all edges between S and T, but no edges within T.

In other words, T is an independent set in H, and we conclude that H has again no k-cliques. Let d'_j be the degree of j in H. If $j \in S$, then we certainly have $d'_j \geq d_j$ by the construction of H, and for $j \in T$, we see $d'_j = |S| = d_m \geq d_j$ by the choice of m. We infer $|E(H)| \geq |E|$, and conclude that among all graphs with a maximal number of edges, there must be one of the form of H. Applying induction on S, we thus infer that among the graphs with a maximal number of edges there is a graph $K_{n_1, \ldots, n_{k-1}}$, which implies $|E| \leq \sum_{i \neq j} n_i n_j$ and therefore (1). $\square$

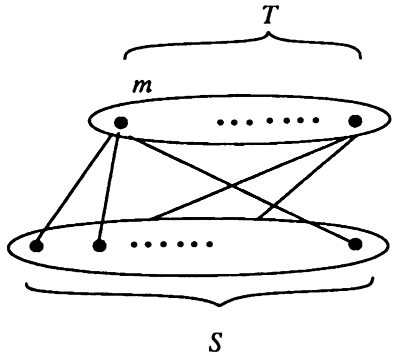

Figure 4.

We note that this proof yields the full statement $|E| \leq |E(H)|$, $H =$ Turán graph.

Third proof (Moon-Moser 1962). This proof generalizes the idea of the first proof for $k = 3$ and yields a quantitative estimate for the number of h-cliques. Let G be any graph on $V = \{1, \ldots, n\}$ and denote by $\mathscr{C}_h$ the set of h-cliques in G with $|\mathscr{C}_h| = C_h$. As examples we have $C_1 = n, C_2 = |E|, C_3 =$ number of triangles. For $A \in \mathscr{C}_h$ we denote by $d(A)$ the number of $(h+1)$-cliques containing A. Counting in two ways we obtain

$$\sum_{A \in \mathscr{C}_h} d(A) = (h+1)C_{h+1} (h \geq 1), \tag{7}$$

in generalization of (2). For $A \in \mathscr{C}_h (h \geq 2)$ let us denote by $A^{(1)}, \ldots, A^{(h)}$ the $(h-1)$-cliques contained in A.

Claim. *For any graph G*

$$\frac{C_{h+1}}{C_h} \geq \frac{1}{h^2 - 1}\left(h^2 \frac{C_h}{C_{h-1}} - n\right)(h \geq 2). \tag{8}$$

Consider $A \in \mathscr{C}_h, B = V \setminus A, |B| = n - h$. Among the vertices $j \in B$ there are precisely $d(A)$ vertices which are adjacent to all of A. Every other vertex in B is adjacent to at most one $(h-1)$-clique $A^{(i)}$, thereby forming an h-clique (Figure 5). We thus obtain (note -1 because of $A^{(i)} \subseteq A$)

$$\sum_{i=1}^{h} \left(d(A^{(i)}) - 1 - d(A)\right) + d(A) \leq n - h,$$

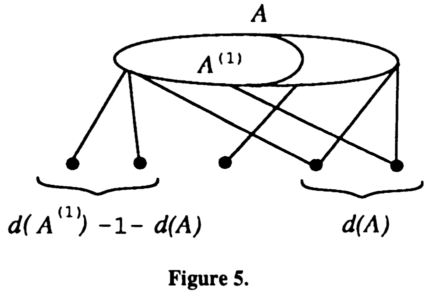

Figure 5.

hence

$$\sum_{i=1}^{h} d(A^{(i)}) - (h-1)d(A) \leq n.$$

Summation over $A \in \mathscr{C}^{(h)}$ yields

$$\sum_{A \in \mathscr{C}_h} \sum_{i=1}^{h} d(A^{(i)}) - (h-1) \sum_{A \in \mathscr{C}_h} d(A) \leq nC_h. \tag{9}$$

As in (4) we conclude

$$\sum_{A \in \mathscr{C}_h} \sum_{i=1}^{h} d(A^{(i)}) = \sum_{B \in \mathscr{C}_{h-1}} d(B)^2, \tag{10}$$

and by (7) we have

$$(h-1) \sum_{A \in \mathscr{C}_h} d(A) = (h^2-1)C_{h+1}. \tag{11}$$

Substituting (10) and (11) into (9) gives us

$$\sum_{B \in \mathscr{C}_{h-1}} d(B)^2 \leq nC_h + (h^2-1)C_{h+1}. \tag{12}$$

By the Cauchy-Schwarz inequality applied to the vectors $(d(B))$, (1) of length C_{h-1}, we finally obtain

$$nC_h + (h^2-1)C_{h+1} \geq \sum_{B \in \mathscr{C}_{h-1}} d(B)^2 \geq \frac{1}{C_{h-1}} \left(\sum_{B \in \mathscr{C}_{h-1}} d(B) \right)^2 = \frac{h^2 C_h^2}{C_{h-1}},$$

which is precisely (8).

In order to prove (1) we must find a relationship between (8) and the edge-number $|E|$. Let us set

$$|E| = \left(1 - \frac{1}{\vartheta}\right) \frac{n^2}{2} (\vartheta \in \mathbb{R}). \tag{13}$$

Since the right-hand side of (13) is increasing in ϑ, we must thus prove $\vartheta \leq k-1$ for graphs without k-cliques.

Claim. *We have*

$$\frac{C_{h+1}}{C_h} \geq \frac{\vartheta - h}{\vartheta} \frac{n}{h+1} (h \geq 1). \tag{14}$$

For $n = 1$ we have $C_2 = |E|, C_1 = n$, and (14) is satisfied with equality by the definition of ϑ. Using (8) and induction on h we infer

$$\frac{C_{h+1}}{C_h} \geq \frac{1}{h^2-1} \left(h^2 \frac{\vartheta - h + 1}{\vartheta} \frac{n}{h} - n \right) = \frac{1}{h^2-1} \frac{(\vartheta - h)(h-1)n}{\vartheta}$$
$$= \frac{\vartheta - h}{\vartheta} \cdot \frac{n}{h+1},$$

as claimed.

Now, if G contains no k-clique, then $C_k = 0$, and we infer $\vartheta \leq k-1$ from (14) for $h + 1 = k$. □

EXAMPLE. Consider (8) for $h = 2$. In this case, the inequality states that any graph satisfies

$$C_3 \geq \frac{|E|}{3}\left(\frac{4|E|}{n} - n\right) = \frac{|E|}{3n}(4|E| - n^2).$$

We conclude that a graph G on an even number n of vertices with $|E| = n^2/4 + 1$ not only contains one triangle (as it must by Turán's Theorem), but more than $n/3$. If we add one edge to $K_{n/2, n/2}$, then we obtain $n/2$ triangles, and it can be easily shown that this holds for any graph with $n^2/4 + 1$ edges.

So far, the proofs have employed counting techniques, the following three proofs use entirely different ideas.

Fourth proof (Motzkin-Straus 1965). Let G be an arbitrary graph on $V = \{1, \dots, n\}$. By $\omega = \omega(G)$ we denote the number of vertices in a largest clique of G, $\omega(G)$ is called the *clique-number*. Now, we associate to each $i \in V$ a variable x_i (over $\mathbb{R}$) and consider the function $f(x_1, \dots, x_n) = 2\Sigma_{ij \in E} x_i x_j$.

Claim. *We have*

$$1 - \frac{1}{\omega} = \max\left(2 \sum_{ij \in E} x_i x_j \colon \sum_{i=1}^{n} x_i = 1, x_i \geq 0 \textit{ for all } i\right). \tag{15}$$

Since f is continuous on a compact set, there exists x with $f(x) = \max$. Among all such vectors x, we choose one with a maximal number of $x_i = 0$. Let $C = \{i \in V\colon x_i > 0\}$. We show first that C is a clique. Suppose this is false with $1, 2 \in C$ but $12 \notin E$. For any $t \in \mathbb{R}$ in the range $-x_1 \leq t \leq x_2$ the vector $x_t = (x_1 + t, x_2 - t, x_3, \dots, x_n)$ satisfies the conditions in (15), and furthermore, $f(x_t)$ is a *linear* function in t, since the product $(x_1 + t)(x_2 - t)$ does not appear in $f(x_t)$ because of $12 \notin E$. Since by the choice of x, $f(x_t)$ assumes the maximum at $t = 0$ (i.e. in the interior) we conclude that $f(x_t)$ is, in fact, constant for all t. For $t = x_2$, $\bar{x} = (x_1 + x_2, 0, x_3, \dots, x_n)$, we therefore obtain $f(\bar{x}) = f(x)$, contradicting the choice of x.

We can thus assume $f(x) = \max$ with $C = \{i\colon x_i > 0\}$ a clique. Since

$$1 = (x_1 + \dots + x_n)^2 = 2 \sum_{ij \in C} x_i x_j + \sum_{i \in C} x_i^2$$

we conclude that $f(x)$ is maximal if and only if $\Sigma_{i \in C} x_i^2$ is minimal. Under the assumption $\Sigma_{i \in C} x_i = 1$ this is clearly the case for $x_i = 1/|C|$, and we obtain

$$f(x) = 1 - \sum_{i \in C} x_i^2 = 1 - \frac{1}{|C|} \leq 1 - \frac{1}{\omega}$$

with equality for $|C| = \omega$, which is what we wanted to prove.

Inequality (1) is now an immediate consequence. Setting $x_i = 1/n$, we have $f(x) = 2|E|/n^2$ and therefore

$$\frac{2|E|}{n^2} = f(x) \leq 1 - \frac{1}{k-1} = \frac{k-2}{k-1},$$

since G contains no k-clique. □

Fifth proof (Li-Li 1981, Kleitman-Lovász 1994). The basis for this proof is again an algebraic structure. To every vertex $i \in V$ of the graph G we again associate a variable x_i and consider the polynomial

$$p_G(x_1,\dots,x_n) = \sum_{i<j,\, ij\notin E} (x_i - x_j). \tag{16}$$

The fundamental observation on the polynomial p_G is the following obvious fact:

$$\omega(G) \le k-1 \Leftrightarrow \text{the identification } x_{i_1} = \cdots = x_{i_k} \text{ of any } k \text{ variables in } p_G \text{ yields the zero-polynomial.} \tag{17}$$

Let $P(n,k)$ be the set of real polynomials in n variables which satisfy the right-hand side of (17). $P(n,k)$ is clearly an ideal in $\mathbb{R}[x_1,\dots,x_n]$. Let $\mathscr{H}(n,k)$ be the following family of graphs on $V = \{1,\dots,n\}$: H is in $\mathscr{H}(n,k)$ if and only if the vertex-set V can be partitioned into $k-1$ disjoint independent subsets (in the language of graph theory, this means H is $(k-1)$-partite or $(k-1)$-colorable). In particular, all our graphs $K_{n_1,\dots,n_{k-1}}$ are in $\mathscr{H}$ and therefore all Turán graphs. By our remarks on the graphs $K_{n_1,\dots,n_{k-1}}$ we can therefore state

$$|E(H)| \le \frac{(k-2)n^2}{2(k-1)} \text{ for all } H \in \mathscr{H}(n,k). \tag{18}$$

By $\hat{P}(n,k)$ we denote the ideal in $\mathbb{R}[x_1,\dots,x_n]$ generated by $\{p_H\colon H \in \mathscr{H}(n,k)\}$. Since we have $\omega(H) \le k-1$ for any such graph, we infer $\hat{P}(n,k) \subseteq P(n,k)$.

Claim. *We have* $P(n,k) = \hat{P}(n,k)$.

Before proving this claim, let us see how Turán's theorem follows from it. Let G be a graph with $\omega(G) \le k-1$. Then $p_G \in P(n,k) = \hat{P}(n,k)$, i.e.

$$p_G = \sum_{i=1}^{m} q_i p_{H_i} \text{ with } H_i \in \mathscr{H}(n,k),\, q_i \in \mathbb{R}[x_1,\dots,x_n]. \tag{19}$$

By (16), p_G is a homogeneous polynomial of degree $(p_G) = \binom{n}{2} - |E(G)|$, and analogously degree $(p_{H_i}) = \binom{n}{2} - |E(H_i)|$. We thus infer from (19), $\binom{n}{2} - |E(G)| \ge \binom{n}{2} - |E(H_i)|$ for some i, and therefore (1) from (18).

Let $f \in P(n,k)$. To prove $f \in \hat{P}(n,k)$ we use induction on n. For $n = 2$ there is nothing to prove. For a subset $S \subseteq \{1,\dots,n-1\}$ we denote by f_S the polynomial which results from f by identifying $x_n = x_i$ for all $i \in S$. Clearly, $f_S \in P(n,k)$ and hence $f_S \in \hat{P}(n,k)$ for $S \ne \varnothing$ by induction (note $\hat{P}(n-1,k) \subseteq \hat{P}(n,k)$). Now consider the polynomial

$$g = \sum_{S\subseteq\{1,\dots,n-1\}} (-1)^{|S|} f_S. \tag{20}$$

Cancelling terms we see that every identification $x_n = x_i$ $(i = 1,\dots,n-1)$ in g yields the zero-polynomial. We conclude that $(x_1 - x_n)\dots(x_{n-1} - x_n)$ divides g, hence

$$g = (x_1 - x_n)\dots(x_{n-1} - x_n)h. \tag{21}$$

Since $f_S \in P(n,k)$ for all S, we have $g \in P(n,k)$ by (20), whence h becomes by (21) the zero-polynomial whenever we identify k of the variables $x_1,\dots,x_{n-1}$ in h. Expanding h with respect to x_n, we see that every coefficient polynomial p of a power x_n^l lies in $P(n-1,k)$ and hence in $\hat{P}(n-1,k)$ by induction. We conclude that the polynomial g is a sum of expressions

$$q(x_1 - x_n)\dots(x_{n-1} - x_n)p_{\bar{H}}, \tag{22}$$

$$\text{with } \bar{H} \in \mathscr{H}(n-1,k),\, q \in \mathbb{R}[x_1,\dots,x_n].$$

Adding the vertex n to each such $\overline{H}$ without edges from n to $\overline{H}$, we obtain $(x_1 - x_n)\dots(x_{n-1} - x_n)p_{\overline{H}} = p_H$ with $H \in \mathscr{H}(n, k)$. This now implies $g \in \hat{P}(n, k)$ by (22), and thus

$$f = g - \sum_{S \neq \emptyset} (-1)^{|S|} f_S \in \hat{P}(n, k),$$

as claimed. □

REMARK. We note that this proof again yields the full implication of (1), that the Turán graphs attain the maximal number of edges, and it can be shown that the polynomials p_H, $H =$ Turán graph, already generate the ideal $P(n, k)$.

Sixth proof (Alon-Spencer 1992). Our last and perhaps most elegant proof uses ideas from probability theory. Let G be an arbitrary graph on $V = \{1, \dots, n\}$.

Claim. *We have*

$$\omega(G) \geq \sum_{i=1}^{n} \frac{1}{n - d_i}. \tag{23}$$

We choose with equal probability $1/n!$ a permutation $\pi_1, \pi_2, \dots, \pi_n$ of V and construct the following set C. We put π_i into C if and only if π_i is adjacent to all π_j $(j < i)$. By definition C is a clique in G. Let $X = |C|$ be the corresponding random variable. We have $X = \sum_{i=1}^{n} X_i$, where X_i is the indicator random variable of i, i.e. $X_i = 1$ or 0 depending on $i \in C$ or $i \notin C$. Now we note $i \in C$ with respect to the permutation $(\pi_1, \dots, \pi_n)$ iff i appears *before* all $n - 1 - d_i$ non-neighbors of i, or in other words, if i is the *first* among i and its non-neighbors. We conclude $EX_i = 1/n - d_i$ for the expectation and hence

$$E(|C|) = EX = \sum_{i=1}^{n} EX_i = \sum_{i=1}^{n} \frac{1}{n - d_i}$$

by the linearity of expectation. Consequently, there must be a clique C with at least $E(|C|)$ vertices, and this is just our claim (23).

To deduce Turán's theorem from (23) we use the Cauchy-Schwarz inequality in the form

$$n^2 = \left(\sum \sqrt{x_i}\sqrt{x_i^{-1}}\right)^2 \leq \sum x_i \cdot \sum x_i^{-1}$$

with $x_i = n - d_i$. Indeed, (23) and (2) imply

$$\omega(G) \geq \frac{n^2}{\sum_{i=1}^{n} n - d_i} = \frac{n^2}{n^2 - 2|E|}. \tag{24}$$

If G has no k-clique, then $\omega(G) \leq k - 1$ and (24) reduces precisely to (1). □

REMARK. Inequality (23) was first proved in Wei **[10]** by successively removing vertices similar to the second proof.

REFERENCES

1. N. Alon-J. Spencer: *The Probabilistic Method.* Wiley-Interscience 1992.
2. B. Bollobás: *Extremal graph Theory.* Academic Press 1978.
3. P. Erdös: On the graph theorem of Turán (in Hungarian). *Math. Fiz. Lapok* 21 (1970), 249–251.

4. S. R. Li-W. W. Li: Independence number of graphs and generators of ideals. *Combinatorica* 1 (1981), 55–61.
5. L. Lovász: Stable sets and polynomials. *Discrete Math.* 124 (1994), 137–153.
6. W. Mantel: Problem 28. *Wiskundige Opgaven* 10 (1906), 60–61.
7. J. W. Moon-L. Moser: On a problem of Turán. *Publ. Math. Inst. Hungar. Acad. Sci.* 7 (1962), 283–286.
8. T. S. Motzkin-E. G. Straus: Maxima for graphs and a new proof of a theorem of Turán. *Canad. J. Math.* 17 (1965), 533–540.
9. P. Turán: On an extremal problem in graph theory (in Hungarian). *Math. Fiz. Lapok* 48 (1941), 436–452.
10. V. K. Wei: A lower bound on the stability number of a simple graph. *Bell Lab. Tech. Mem.* 81-11217-9 (1981).

II. Mathematisches Institut
Freie Universität Berlin
Arnimallee 3 D-14195
Berlin, GERMANY
aigner@math.fu-berlin.de

The fact is, although DNA testing may be as foolproof as fingerprinting, it doesn't cause excitement. It's difficult to respond to. It's like advanced math, brilliant but boring, astonishing but passionless. It made everyone eager to move on to the next phase of the trial, which consisted of autopsy pictures...

From *"If the Gloves Fit"* by Dominick Dunne, in Vanity Fair/August 1995.

Submitted by J. Foster
Weber State University

Answer to Picture Puzzle

(p. 797)

A. S. Bessicovitch.