Turán's graph theorem

Chapter 41

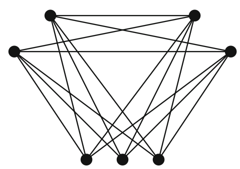
One of the fundamental results in graph theory is the theorem of Turán from 1941, which initiated extremal graph theory. Turán's theorem was rediscovered many times with various different proofs. We will discuss five of them and let the reader decide which one belongs in The Book.

Let us fix some notation. We consider simple graphs G on the vertex set $V = \{v_1, \dots, v_n\}$ and edge set E. If v_i and v_j are neighbors, then we write $v_i v_j \in E$. A *p-clique* in G is a complete subgraph of G on p vertices, denoted by K_p. Paul Turán posed the following question:

Paul Turán

> *Suppose G is a simple graph that does not contain a p-clique. What is the largest number of edges that G can have?*

We readily obtain examples of such graphs by dividing V into $p-1$ pairwise disjoint subsets $V = V_1 \cup \dots \cup V_{p-1}$, $|V_i| = n_i$, $n = n_1 + \dots + n_{p-1}$, joining two vertices if and only if they lie in distinct sets V_i, V_j. We denote the resulting graph by $K_{n_1,\dots,n_{p-1}}$; it has $\sum_{i<j} n_i n_j$ edges. We obtain a maximal number of edges among such graphs with given n if we divide the numbers n_i as evenly as possible, that is, if $|n_i - n_j| \leq 1$ for all i, j. Indeed, suppose $n_1 \geq n_2 + 2$. By shifting one vertex from V_1 to V_2, we obtain $K_{n_1-1,n_2+1,\dots,n_{p-1}}$ which contains $(n_1 - 1)(n_2 + 1) - n_1 n_2 = n_1 - n_2 - 1 \geq 1$ more edges than $K_{n_1,n_2,\dots,n_{p-1}}$. Let us call the graphs $K_{n_1,\dots,n_{p-1}}$ with $|n_i - n_j| \leq 1$ the *Turán graphs*. In particular, if $p-1$ divides n, then we may choose $n_i = \frac{n}{p-1}$ for all i, obtaining



The graph $K_{2,2,3}$

$$\binom{p-1}{2}\left(\frac{n}{p-1}\right)^2 = \left(1 - \frac{1}{p-1}\right)\frac{n^2}{2}$$

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

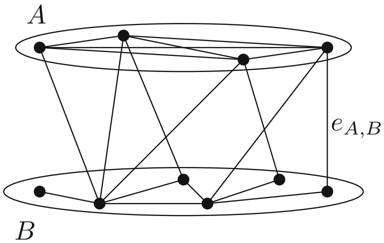
Theorem. *If a graph $G = (V, E)$ on n vertices has no p-clique, $p \geq 2$, then*

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

For $p = 2$ this is trivial. In the first interesting case $p = 3$ the theorem states that a triangle-free graph on n vertices contains at most $\frac{n^2}{4}$ edges. Proofs of this special case were known prior to Turán's result. Two elegant proofs using inequalities are contained in Chapter 20.

M. Aigner, G. M. Ziegler, *Proofs from THE BOOK*, https://doi.org/10.1007/978-3-662-57265-8_41

Let us turn to the general case. The first two proofs use induction and are due to Turán and to Erdős, respectively.

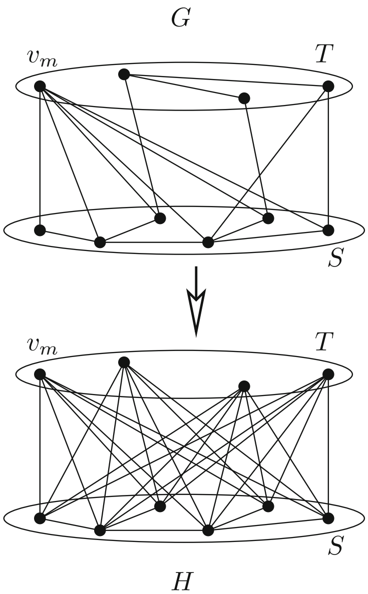


■ **First proof.** We use induction on n. One easily computes that (1) is true for $n < p$. Let G be a graph on $V = \{v_1, \ldots, v_n\}$ without p-cliques with a maximal number of edges, where $n \geq p$. G certainly contains $(p-1)$-cliques, since otherwise we could add edges. Let A be a $(p-1)$-clique, and set $B := V \backslash A$.

A contains $\binom{p-1}{2}$ 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 $e_B \leq \frac{1}{2}(1 - \frac{1}{p-1})(n-p+1)^2$. Since G has no p-clique, every $v_j \in B$ is adjacent to at most $p-2$ vertices in A, and we obtain $e_{A,B} \leq (p-2)(n-p+1)$. Altogether, this yields

$$|E| \leq \binom{p-1}{2} + \frac{1}{2}\Big(1 - \frac{1}{p-1}\Big)(n-p+1)^2 + (p-2)(n-p+1)\,,$$

which is precisely $(1 - \frac{1}{p-1})\frac{n^2}{2}$. □



■ **Second proof.** This proof makes use of the structure of the Turán graphs. Let $v_m \in V$ be a vertex of maximal degree $d_m = \max_{1 \leq j \leq n} d_j$. Denote by S the set of neighbors of v_m, $|S| = d_m$, and set $T := V \backslash S$. As G contains no p-clique, and v_m is adjacent to all vertices of S, we note that S contains no $(p-1)$-clique.

We now construct the following graph H on V (see the figure). 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 p-cliques. Let d_j' be the degree of v_j in H. If $v_j \in S$, then we certainly have $d_j' \geq d_j$ by the construction of H, and for $v_j \in T$, we see $d_j' = |S| = d_m \geq d_j$ by the choice of v_m. We infer $|E(H)| \geq |E|$, and find that among all graphs with a maximal number of edges, there must be one of the form of H. By induction, the graph induced by S has at most as many edges as a suitable graph $K_{n_1,\ldots,n_{p-2}}$ on S. So $|E| \leq |E(H)| \leq E(K_{n_1,\ldots,n_{p-1}})$ with $n_{p-1} = |T|$, which implies (1). □

The next two proofs are of a totally different nature, using a maximizing argument and ideas from probability theory. They are due to Motzkin and Straus and to Alon and Spencer, respectively.

■ **Third proof.** Consider a *probability distribution* $\boldsymbol{w} = (w_1, \ldots, w_n)$ on the vertices, that is, an assignment of values $w_i \geq 0$ to the vertices with $\sum_{i=1}^{n} w_i = 1$. Our goal is to maximize the function

$$f(\boldsymbol{w}) \;=\; \sum_{v_i v_j \in E} w_i w_j.$$

Suppose $\boldsymbol{w}$ is any distribution, and let v_i and v_j be a pair of nonadjacent vertices with positive weights w_i, w_j. Let s_i be the sum of the weights of all vertices adjacent to v_i, and define s_j similarly for v_j, where we may assume that $s_i \geq s_j$. Now we move the weight from v_j to v_i, that is, the new weight of v_i is $w_i + w_j$, while the weight of v_j drops to 0. For the new new distribution $\boldsymbol{w}'$ we find

$$f(\boldsymbol{w}') \;=\; f(\boldsymbol{w}) + w_j s_i - w_j s_j \;\geq\; f(\boldsymbol{w}).$$

"Moving weights"

We repeat this (reducing the number of vertices with a positive weight by one in each step) until there are no nonadjacent vertices of positive weight anymore. Thus we conclude that there is an optimal distribution whose nonzero weights are concentrated on a clique, say on a k-clique. Now if, say, $w_1 > w_2 > 0$, then choose ε with $0 < \varepsilon < w_1 - w_2$ and change w_1 to $w_1 - \varepsilon$ and w_2 to $w_2 + \varepsilon$. The new distribution $\boldsymbol{w}'$ satisfies $f(\boldsymbol{w}') = f(\boldsymbol{w}) + \varepsilon(w_1 - w_2) - \varepsilon^2 > f(\boldsymbol{w})$, and we infer that the maximal value of $f(\boldsymbol{w})$ is attained for $w_i = \frac{1}{k}$ on a k-clique and $w_i = 0$ otherwise. Since a k-clique contains $\frac{k(k-1)}{2}$ edges, we obtain

$$f(\boldsymbol{w}) \;=\; \frac{k(k-1)}{2}\frac{1}{k^2} \;=\; \frac{1}{2}\Big(1 - \frac{1}{k}\Big).$$

Since this expression is increasing in k, the best we can do is to set $k = p-1$ (since G has no p-cliques). So we conclude

$$f(\boldsymbol{w}) \;\leq\; \frac{1}{2}\Big(1 - \frac{1}{p-1}\Big)$$

for *any* distribution $\boldsymbol{w}$. In particular, this inequality holds for the *uniform* distribution given by $w_i = \frac{1}{n}$ for all i. Thus we find

$$\frac{|E|}{n^2} \;=\; f\Big(w_i = \frac{1}{n}\Big) \;\leq\; \frac{1}{2}\Big(1 - \frac{1}{p-1}\Big),$$

which is precisely (1). □

■ **Fourth proof.** This time we use some concepts from probability theory. Let G be an arbitrary graph on the vertex set $V = \{v_1, \ldots, v_n\}$. Denote the degree of v_i by d_i, and write $\omega(G)$ for the number of vertices in a largest clique, called the *clique number* of G.

Claim. *We have* $\omega(G) \;\geq\; \displaystyle\sum_{i=1}^{n} \frac{1}{n - d_i}.$

We choose a random permutation $\pi = v_1 v_2 \ldots v_n$ of the vertex set V, where each permutation is supposed to appear with the same probability $\frac{1}{n!}$, and then consider the following set C_π. We put v_i into C_π if and only if v_i is adjacent to all v_j $(j < i)$ preceding v_i. By definition, C_π is a clique in G. Let $X = |C_\pi|$ be the corresponding random variable. We have $X = \sum_{i=1}^{n} X_i$, where X_i is the indicator random variable of the vertex v_i, that is, $X_i = 1$ or $X_i = 0$ depending on whether $v_i \in C_\pi$ or $v_i \notin C_\pi$. Note that v_i belongs to C_π with respect to the permutation $v_1 v_2 \ldots v_n$ if and only if v_i appears *before* all $n - 1 - d_i$ vertices which are not adjacent to v_i, or in other words, if v_i is the *first* among v_i and its $n - 1 - d_i$ non-neighbors. The probability that this happens is $\frac{1}{n-d_i}$, hence $EX_i = \frac{1}{n-d_i}$.

Thus by linearity of expectation (see page 116) we obtain

$$E(|C_\pi|) \;=\; EX \;=\; \sum_{i=1}^{n} EX_i \;=\; \sum_{i=1}^{n} \frac{1}{n-d_i}.$$

Consequently, there must be a clique of at least that size, and this was our claim. To deduce Turán's theorem from the claim we use the Cauchy–Schwarz inequality from Chapter 20,

$$\Big(\sum_{i=1}^{n} a_i b_i\Big)^2 \;\le\; \Big(\sum_{i=1}^{n} a_i^2\Big)\Big(\sum_{n=1}^{n} b_i^2\Big).$$

Set $a_i = \sqrt{n-d_i}$, $b_i = \frac{1}{\sqrt{n-d_i}}$, then $a_i b_i = 1$, and so

$$n^2 \;\le\; (\sum_{i=1}^{n}(n-d_i))(\sum_{i=1}^{n}\frac{1}{n-d_i}) \;\le\; \omega(G)\sum_{i=1}^{n}(n-d_i). \qquad (2)$$

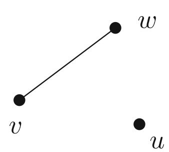
At this point we apply the hypothesis $\omega(G) \le p-1$ of Turán's theorem. Using also $\sum_{i=1}^{n} d_i = 2|E|$ from the chapter on double counting, inequality (2) leads to

$$n^2 \;\le\; (p-1)(n^2 - 2|E|),$$

and this is equivalent to Turán's inequality. □

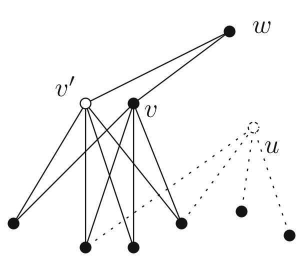
Now we are ready for the last proof, which may be the most beautiful of them all. Its origin is not clear; we got it from Stephan Brandt, who heard it in Oberwolfach. It may be "folklore" graph theory. It yields in one stroke that the Turán graph is in fact the unique example with a maximal number of edges. It may be noted that both proofs 1 and 2 also imply this stronger result.

■ **Fifth proof.** Let G be a graph on n vertices without a p-clique and with a maximal number of edges.



> **Claim.** *G does not contain three vertices u, v, w such that $vw \in E$, but $uv \notin E$, $uw \notin E$.*

Suppose otherwise, and consider the following cases.

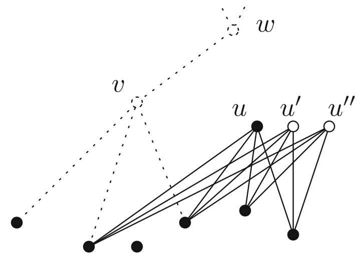


Case 1: $d(u) < d(v)$ or $d(u) < d(w)$.
We may suppose that $d(u) < d(v)$. Then we duplicate v, that is, we create a new vertex v' which has exactly the same neighbors as v (but vv' is not an edge), delete u, and keep the rest unchanged.
The new graph G' has again no p-clique, and for the number of edges we find

$$|E(G')| \;=\; |E(G)| + d(v) - d(u) \;>\; |E(G)|,$$

a contradiction.

Case 2: $d(u) \geq d(v)$ and $d(u) \geq d(w)$.
Duplicate u twice and delete v and w (as illustrated in the margin). Again, the new graph G' has no p-clique, and we compute (the -1 results from the edge vw):



$$|E(G')| \;=\; |E(G)| + 2d(u) - (d(v) + d(w) - 1) \;>\; |E(G)|.$$

So we have a contradiction once more.

A moment's thought shows that the claim we have proved is equivalent to the statement that

$$u \sim v \;\;:\Longleftrightarrow\;\; uv \notin E(G)$$

defines an equivalence relation. Thus G is a complete multipartite graph, $G = K_{n_1,\ldots,n_{p-1}}$, and we are finished. $\square$

References

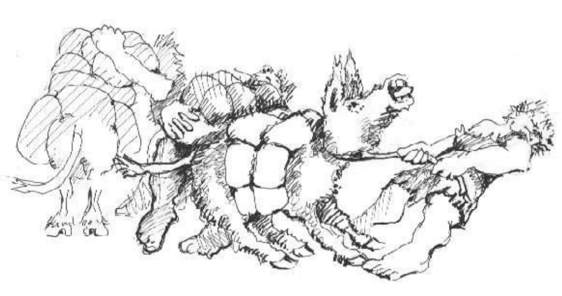
[1] M. AIGNER: *Turán's graph theorem,* Amer. Math. Monthly **102** (1995), 808-816.

[2] N. ALON & J. SPENCER: *The Probabilistic Method,* Third edition, Wiley-Interscience 2008.

[3] P. ERDŐS: *On the graph theorem of Turán (in Hungarian),* Math. Fiz. Lapok **21** (1970), 249-251.

[4] 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.

[5] P. TURÁN: *On an extremal problem in graph theory,* Math. Fiz. Lapok **48** (1941), 436-452.



"Larger weights to move"