The Fundamentals of Extremal Graph Theory

Cheng, Feng

Introduction:

The history of graph theory, a subset of combinatorics, can be traced back to the days of Leonard Euler, when he found out the solution to the Seven Bridges of Königsberg problem in 1736. It was not until the 1930s and 40s that extremal graph theory, a branch of graph theory that deals with certain properties of graphs over maximality and minimality, emerged and initiated its development. Pál Turán, a Hungarian mathematician, was the first to concern these extremal properties of graphs and became the founder of extremal graph theory for his Turán's theorem. Many mathematicians, especially Hungarian ones, such as P. Erdös, B. Bollobás, V. Sós, and E. Szemerédi, have been following his steps, pushing this field forward since Turán.

In this paper, we will focus on the earliest yet most fundamental theory in extremal graph theory, Turán's theorem, proven and published in 1941. First, we will take a look at Mantel's theorem, a special case of Turán's theorem. Then we will shift our attention to Turán's theorem itself. The paper will produce several different proofs for both theorems. We will concentrate on some extensions of these two theorems through the paper as well, which emphasize the degree sequence, subgraph, or other graph properties beyond what Mantel's theorem and Turán's theorem have already displayed. All properties shown and theorems proven in this paper will help us understand how the extremal graph theory was founded and some major thoughts in this field.

Preliminaries:

In this paper, all graphs are taken as simple graphs. If the letter G is not explicitly defined in the proofs and the propositions themselves, then we recognize it as the graph G that contains all the vertices and edges. *Triangles* are another way of expressing K_3 or C_3.

Let G be a graph, and v be a vertex in G. Denote the vertex set of G by $V(G)$, and the edge set of G by $E(G)$ (sometimes abbreviated as V and E). Denote the number of edges in G by $e(G) = |E(G)|$. Denote $N(v)$ as the set of vertices that are adjacent to v. Denote a graph G of order n by G_n. The *complement graph* of G, denoted by $\bar{G}$, is the graph with the same vertex set as G, but with the edge set that is the complement of the edge set of G.

A *clique* is a subset of the vertices of a graph such that its induced subgraph is complete. A k-*clique* means that the clique has k vertices.

The *degree sequence* of a graph is the sequence of the degree of each vertex (usually written in nonincreasing order) in the graph.

An *independent set* of vertices is defined as a set of vertices that are pairwise nonadjacent.

Mantel's theorem. A simple graph G with order n and free of any triangle has at most $\left\lfloor\frac{n^2}{4}\right\rfloor$ edges.

In the following pages, we will give four different proofs (modified upon [1]) of Mantel's theorem. All of these proofs are worth noticing because they employ commonly used methods of solving problems in graph theory, or combinatorics.

For the extremal problems in combinatorics, it is often the case that we pick a single element with extremal properties out of all the elements. Since we choose to analyze an element linked to some maximal or minimal properties, it is very possible that we can obtain some other extremal properties for a larger context.

Proof I (from maximality).

Suppose v_1 is the vertex in G with maximum degree d, and let the vertices that are the neighbors of v_1 be

$$v_{n-d+1}, v_{n-d+2}, \dots, v_n.$$

Because G doesn't contain any K_3, meaning that any two of the vertices listed above are not neighbors, the number of edges in G satisfies

$$e(G) \le \sum_{i=1}^{n-d} d(v_i) \le (n-d)d \le \frac{(n-d+d)^2}{4} = \frac{n^2}{4}.$$

Since $e(G)$ must be an integer, G has at most $\left\lfloor \frac{n^2}{4} \right\rfloor$ edges.

To satisfy the equality, v_i $(i = 1, 2, 3, \dots, n-d)$ must not contain any pairwise neighbor, and all of these v_i must be of degree d, which satisfies the equation $n - d = d$ (indicating that d can be either $\left\lfloor \frac{n}{2} \right\rfloor$ or $\left\lceil \frac{n}{2} \right\rceil$ as n might not be even). Since $v_{n-d+1}, v_{n-d+2}, \dots, v_n$ do not contain any pairwise neighbor as well, G is a bipartite graph $K_{\left\lfloor \frac{n}{2} \right\rfloor, \left\lceil \frac{n}{2} \right\rceil}$. □

As we can see from Proof I, it is possible to open up the proof if we focus on one single vertex, it is worth the attempt to focus on a pair of vertices that are adjacent to each other and form an edge.

Proof II (induction).

Mantel's theorem can be verified easily when $n = 3, 4$.

Suppose the case is true for $n < k$ $(k > 4)$. Then for $n = k$, suppose $v_i v_j$ is an edge in G. Since K_3 is not contained in the graph, v_i and v_j do not have a common neighbor. Thus, $d(v_i) + d(v_j) \le n$ and has at most $(n-1)$ edges that are incident with either of them. If we remove the two vertices v_i and v_j, we get a new graph of $(n-2)$ vertices, which according to the inductive hypothesis, should have at most $\frac{(n-2)^2}{4}$ edges; hence

$$e(G) \le \frac{(n-2)^2}{4} + n - 1 = \frac{n^2}{4}.$$

The equality is satisfied only when $d(v_i) + d(v_j) = n$, which means that $N(v_i) \cup N(v_j) = V(G)$. Also note that both $N(v_i)$ and $N(v_j)$ should be independent sets since there is no K_3 in G.

Since when $n = 3, 4$, G is a bipartite graph $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$, and for every $n \geq 5$, G will increase the cardinalities of two subsets by 1 if n increases by 2, respectively. Hence, the number of edges reaches its maximum if and only if G is a bipartite graph $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$ (for every n). □

The following proof uses the notion $d(v_i) + d(v_j) \leq n$ as well, while this proof expands this inequality to the whole context and take a more algebraic approach.

Proof III (summation and transformation).

As shown in Proof II, $d(v_i) + d(v_j) \leq n$ is true for every edge $v_i v_j$. Thus, we get the inequality

$$\sum_{v_i v_j \in E} \left(d(v_i) + d(v_j)\right) \leq en. \tag{1}$$

Note we have the following equation

$$\sum_{v_i v_j \in E} \left(d(v_i) + d(v_j)\right) = \sum_{v \in V} d^2(v)$$

as all the "$d(v)$"s are added for $d(v)$ times in this summation.
From the Cauchy-Schwarz Inequality, we get the inequality

$$\sum_{v \in V} d^2(v) \geq \frac{1}{n} \left(\sum_{v \in V} d(v)\right)^2 = \frac{1}{n} 4e^2. \tag{2}$$

Combining this with inequality (1), we get $e \leq \frac{n^2}{4}$, or $e \leq \left\lfloor \frac{n^2}{4} \right\rfloor$.

The equality is satisfied only when every pair of neighbors agrees with $d(v_i) + d(v_j) \leq n$, and all the "$d(v)$"s are equal to each other. Here we use $N(v_i) \cup N(v_j) = V(G)$ in the previous proof, and we get the same result that G reaches its maximum only when G is a $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$. □

Proof IV takes a more unique approach toward Mantel's theorem. It intuitively builds up a function and attempts to obtain the result through an inequality regarding the function. The core step in this proof is the use of *local adjustment*, a method applied to solve some complex inequalities. The method is about shifting some elements in the function once at a time, which gradually increases or decreases the value of the function. This proof shows the equal strength of the method in combinatorics, while we are solving extremal combinatorial problems.

Proof IV (weight and local adjustment).

Let $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ be a vector satisfying $\sum x_i = 1$ $(0 \leq x_i \leq 1)$. Define $f(x) = \sum_{ij \in E} x_i x_j$.

We first observe that if all the x_is are of weight $\frac{1}{n}$, $f(x) = \frac{e}{n^2}$. This means that there must exist an x that satisfies

$$f(x) \geq \frac{e}{n^2}.$$

Note that if $ij \notin E$, shifting the weight assigned to x_i to x_j (assuming the total weight of the neighbors of x_j is at least as large as the total weight assigned to the neighbors of x_i) will not decrease $f(x)$, as

$$x_i \sum_{pi \in E} x_p \leq x_j \sum_{qj \in E} x_q.$$

Following and repeating this step above, we find that $f(x)$ may always increase when the weights are not yet concentrated on a clique, since at this moment every ij belongs to E. However, there does not exist any K_3, meaning that $f(x)$ may only concentrate on two vertices.

Finally, we get the inequality

$$\frac{e}{n^2} \leq f(x) \leq \frac{1}{4}. \tag{1}$$

Therefore, $e \leq \frac{n^2}{4}$ (or $\left\lfloor \frac{n^2}{4} \right\rfloor$).

The equality is satisfied only when the two equalities in inequality (1) are satisfied. Thus, the weight of all the vertices should be the same (or nearly the same) and the sum of the weight of each vertices' neighbors should also be the same (or nearly the same). Therefore, every vertex should have a similar number of neighbors. Since G doesn't include any K_3, we can easily determine that the maximum can be reached only when G is a $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$. □

Remember that in Mantel's theorem, we ask about the result when there is no triangle in the graph G. It is equally important in extremal graph theory to explore the result that involves the number of existing triangles.

Here we introduce **notation** that will be useful. Denote by $k_r(G)$ the number of "K_r"s in graph G. Denote by $(d_i)_1^n$ be the degree sequence of graph G with order n.

The degree sequence of $\bar{G}$ is $(\bar{d}_i)_1^n = (n - 1 - d_i)_n^1$. Thus, we get $\sum_{i=1}^n \binom{d_i}{2}$ pairs of adjacent edges in G and $\sum_{i=1}^n \binom{n-1-d_i}{2}$ pairs of adjacent edges in G, where adjacent edges mean that for the two edges, there is only one vertex that is shared by them. Each of $k_3(G) + k_3(\bar{G})$ triangles for the two graphs G and G' contains three such pairs of adjacent edges, while the remaining number of triple vertices is

$$\binom{n}{3} - k_3(G) - k_3(\bar{G}),$$

and all these triple vertices contains exactly one such pair of adjacent vertices. Hence, we get the equation

$$\sum_{i=1}^n \binom{d_i}{2} + \sum_{i=1}^n \binom{n-1-d_i}{2} = 3k_3(G) + 3k_3(\bar{G}) + \binom{n}{3} - k_3(G) - k_3(\bar{G}).$$

In turn we get the number of total triangles in G and $\bar{G}$ (Here we denote $e(G)$ by e, while we denote $e(\bar{G})$ by $\bar{e}$.)

$$\frac{1}{2}\left[\sum_{i=1}^{n}\binom{d_i}{2}+\sum_{i=1}^{n}\binom{n-1-d_i}{2}-\binom{n}{3}\right]$$
$$=\frac{1}{2}\sum_{i=1}^{n}\left[\frac{1}{2}d_i(d_i-1)+\frac{1}{2}(n-1-d_i)(n-2-d_i)\right]-\frac{1}{2}\binom{n}{3}$$
$$=\frac{1}{2}\sum_{i=1}^{n}\left[d_i^2-d_i-(n-2)d_i+\frac{1}{2}(n-1)(n-2)\right]-\frac{1}{2}\binom{n}{3}$$
$$=\sum_{i=1}^{n}\binom{d_i}{2}-(n-2)e+\left(\frac{1}{4}-\frac{1}{12}\right)n(n-1)(n-2)$$
$$=\sum_{i=1}^{n}\binom{d_i}{2}-(n-2)e+\binom{n}{3}$$

The number of total triangles is also equal to

$$\sum_{i=1}^{n}\binom{\bar{d}_i}{2}-(n-2)\bar{e}+\binom{n}{3}$$

since we can interchange G and $\bar{G}$, meaning that

$$\sum_{i=1}^{n}\binom{d_i}{2}-(n-2)e+\binom{n}{3}=\sum_{i=1}^{n}\binom{\bar{d}_i}{2}-(n-2)\bar{e}+\binom{n}{3}.$$

This formula [2] above provides a rather simple form to represent the number of triangles in graphs. Such simplicity gives rise to the possibility of obtaining the extremal properties over the number of triangles. Now we introduce two theorems: Goodman's theorem and Moon and Moser's theorem.

Goodman's Theorem. [7][2] A graph G of order n and its complement contain at least $\frac{1}{24}n(n-1)(n-5)$ triangles.

Proof. We learned formerly that the number of triangles in G and $\bar{G}$ is

$$\sum_{i=1}^{n}\binom{d_i}{2}-(n-2)e+\binom{n}{3}.$$

Notice $\sum_{i=1}^{n}\binom{d_i}{2}$ is at least $n\binom{2e/n}{2}$ when G is a $(2e/n)$-regular graph, according to the AM-GM inequality. Thus, we know the minimum number of triangles should be

$$n\binom{2e/n}{2}-(n-2)e+\binom{n}{3}$$
$$=\binom{n}{3}+e(2e/n-1)-(n-2)e$$

$$
\begin{aligned}
&= \binom{n}{3} - \frac{2e(n^2 - n - 2e)}{2n} \\
&\geq \frac{1}{6}n(n-1)(n-2) - \frac{1}{8}n(n-1)(n-1) \\
&= \frac{1}{24}n(n-1)(n-5)
\end{aligned}
$$

The bound is obtained only when G is a $(2e/n)$-regular graph and $n^2 - n = 4e$. $n^2 - n = 4e$ means $n - 1 = 2 \times \frac{2e}{n}$. Since $\bar{G}$ is a regular graph as well, then both G and $\bar{G}$ are $(2e/n)$-regular graphs. It is clear to see that n is odd. Considering $(n-1)n = 4e$, where $(n-1, n) = 1$, the bound is best possible only when $n = 4k + 1 (k \in \mathbb{Z}^+)$ □

There exists a sharper bound for the number of triangles, as a matter of fact. It was first introduced and proved in A. W. Goodman's 1959 paper [7], and was given a simpler proof (using weights) by L. Sauvé in his 1961 paper [13].

Goodman's Theorem (refined). A graph G of order n and its complement contains

$$
\geq \begin{cases} \frac{1}{24}n(n-2)(n-4), & \text{if } n \equiv 0 (\mathrm{mod}\ 2) \\ \frac{1}{24}n(n-1)(n-5), & \text{if } n \equiv 1 (\mathrm{mod}\ 4). \\ \frac{1}{24}(n+1)(n-3)(n-4), & \text{if } n \equiv 1 (\mathrm{mod}\ 4) \end{cases}
$$

number of triangles. (The inequalities are sharp for all three situations.)

Moon and Moser's Theorem. [8][2] A graph G of order n and size e contains at least $\frac{e}{3n}(4e - n^2)$ triangles.

Proof. We have

$$
3k_3(\bar{G}) \leq \sum_{i=1}^{n} \binom{\bar{d}_i}{2}
$$

because every triangle contains three adjacent pairs of edges, while each adjacent pair of edges may not belong to a triangle.

Remember that $\bar{e} = \binom{n}{2} - e$ and the AM-GM inequality, which will help us get the result

$$
\begin{aligned}
k_3(G) &\geq \sum_{i=1}^{n} \binom{\bar{d}_i}{2} - (n-2)\bar{e} + \binom{n}{3} - \frac{1}{3}\sum_{i=1}^{n} \binom{\bar{d}_i}{2} \\
&\geq \binom{n}{3} - (n-2)\left[\binom{n}{2} - e\right] + \frac{2}{3}n \binom{\frac{2\left[\binom{n}{2} - e\right]}{n}}{2}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{6}n(n-1)(n-2) - \frac{1}{2}n(n-1)(n-2) + e(n-2) + \frac{1}{3}n\left(n-1-\frac{2e}{n}\right)\left(n-2-\frac{2e}{n}\right) \\
&= e(n-2) - \frac{1}{3}n(2n-3)\frac{2e}{n} + \frac{1}{3}n\frac{4e^2}{n^2} \\
&= en - \frac{4}{3}en + \frac{4e^2}{3n} \\
&= \frac{e}{3n}(4e-n^2)
\end{aligned}
$$

This proves that the minimum number of triangles contained in G is $\frac{e}{3n}(4e-n^2)$. The equality is satisfied if and only if G is a regular graph. □

The two theorems above offer us the lower bound of the number of triangles in graphs.

Now turn back to Mantel's theorem. We can actually generalize Mantel's theorem by replacing K_3 with K_n. The generalization is the famous ***Turán's theorem***, which will be the center of our investigation for the following pages.

Definitions and notation that will be useful in the explanation and proof of Turán's theorem:

Given a graph G and a subset S of the vertex set, the *subgraph of G induced by S*, denoted $\langle S\rangle$, is an induced subgraph of G with vertex set S and with edge set $\{uv \mid u, v \in S \text{ and } uv \in E(G)\}$.

For two vertex sets X and Y, $E(X,Y)$ is defined as the number of edges that have one vertex in X and another vertex in Y.

If two graphs G and H are isomorphic, the relationship is written in the form $G \cong H$, or simply $G = H$.

Definition of Turán Graph: the simple complete k-partite graph on n vertices in which all parts are as equal in size as possible ($\left\lfloor\frac{n}{k}\right\rfloor, \left\lceil\frac{n}{k}\right\rceil$) is called a *Turán graph* and denoted $T_k(n)$ or $T_{k,n}$. Denote the number of edges in Turán graph by $t_k(n) = e(T_{k,n})$.

The origin of Turán's theorem [4]:

Ramsey theory, generally speaking, investigates the smallest order of graph G that ensures every 2-coloring of $E(G)$ (using red and blue to color $E(G)$) let G contain either a red K_i or a blue K_j. Alternatively, Ramsey theory can be interpreted in the following sense: assume we know that graph G of order n contains no k independent vertices. What is the greatest p value at which we can ensure the existence of a K_{p+1} in G? (This is because we may weaken the original Ramsey theory by eliminating all the blue edges. Then we regard the existing edges as the red edges and the deleted edges as the blue edges in the original formulation of Ramsey theory.) In 1941 [9], Turán replaced the condition that G contains no k independent vertices with a simpler condition of the number of edges that G has.

Turán's theorem. [9] Let G be a simple graph of order n that does not contain any K_k $(k \geq 2)$. Then $e(G) \leq e(T_{k-1,n})$, with equality if and only if $G \cong T_{k-1,n}$.

The impact of Turán's theorem is because it is the first theorem to use the following form:

If G_n is a graph containing no S , then G has no more than $f(n)$ edges.

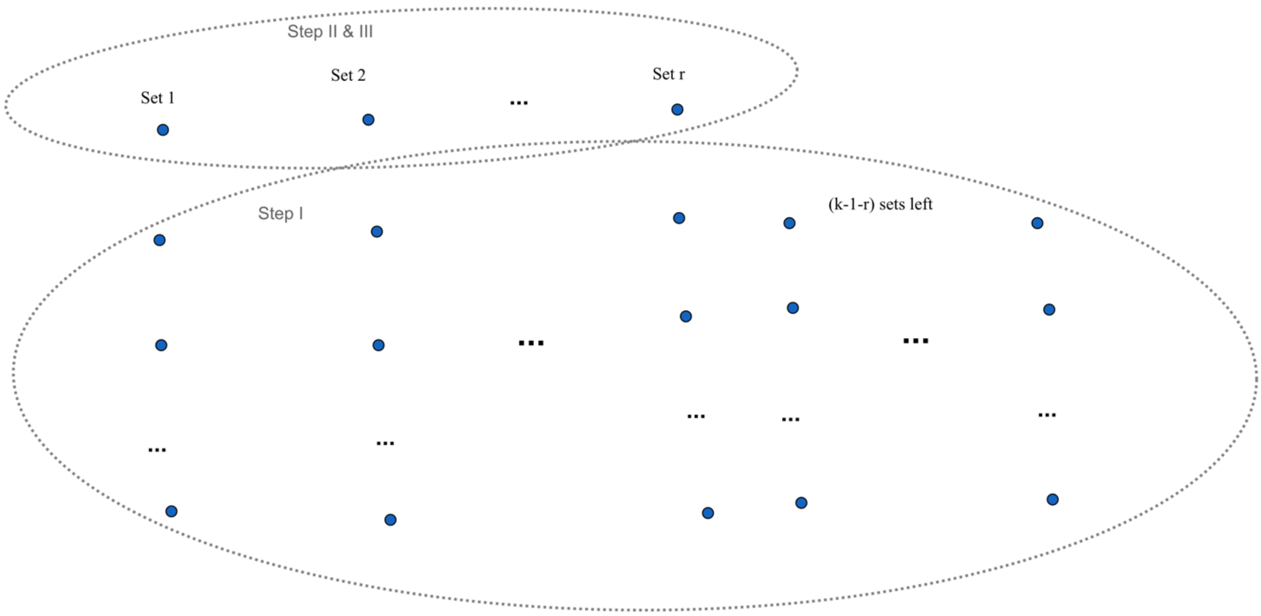
This new form of problems, later called problems of Turán type in graph theory by Paul Erdös and Tibor Gallai [10][6], grabs the attention of mathematicians in this field, leading them to develop a series of theories involving $f(n)$, which is denoted as $ex(n, S)$ in extremal graph theory, where S might be a single *forbidden graph* such as K_k in Turán's theorem, or a family of forbidden graphs. Some famous problems of this type include *Zarankiewicz problem* (initially introduced by K. Zarankiewicz [14]) and *even circuit theorem* (initially introduced by P. Erdös [15]). The former problem is concerned with the bipartite graph while the latter problem is concerned with the cycle.

Note that Turán's theorem can be reformulated by replacing $e(T_{k-1,n})$ with

$$\frac{1}{2}\left(1-\frac{1}{k-1}\right)(n^2-r^2)+\binom{r}{2},$$

as Turán marks in his original paper, where r is the smallest positive integer that satisfies $n \equiv r(\text{mod } (k-1))$.

We may use the figure here to comprehend why this reformulation is correct.



We can easily determine, according to the definition of r, that there are $(k-1-r)$ parts of $\frac{n-r}{k-1}$ vertices and r parts of $(\frac{n-r}{k-1}+1)$ vertices. ($(k-1-r)$ parts are the $(k-1-r)$ *sets left* on the right side and r parts are the *set* 1 *to* r on the left side in the figure.) We can examine three cases to get the result.

First, we may see all parts to be of $\frac{n-r}{k-1}$ vertices. This means that the $T_{k-1,n}$ contains $\binom{k-1}{2}\left(\frac{n-r}{k-1}\right)^2$ edges, since every time we may pick two sets of vertices and form $\left(\frac{n-r}{k-1}\right)^2$ edges, and we need to repeat the procedure $\binom{k-1}{2}$ times.

Second, we look at how the additional r vertices form edges with the existing $(n-r)$ vertices. Every vertex in the additional set of r vertices forms $(k-2)\frac{n-r}{k-1}$ edges with the existing $(n-r)$ vertices since all such vertices have the other $(k-2)$ sets of vertices as their neighbors. Thus, there are an additional $r(k-2)\frac{n-r}{k-1}$ edges.

Third, we know that the additional r vertices themselves will form $\binom{r}{2}$ edges.

In total, we get $\frac{1}{2}\left(1-\frac{1}{k-1}\right)(n^2-r^2)+\binom{r}{2}$ number of edges.

Note that the number of edges reaches its maximum $\frac{(k-2)n^2}{2(k-1)}$ when $(k-1) \mid n$ $(r=0)$. We may also replace the $e(T_{k-1,n})$ with this expression; however, this is weaker than the original Turán's theorem.

The "complementary" form of Turán's theorem in his original paper [9][4]:

If $\overline{G_n}$ has no k independent vertices, then $e(\overline{G_n}) \geq e(\overline{T_{k-1,n}})$, and the equality is satisfied only when $G_n \cong T_{k-1,n}$.

The first proof of Turán's theorem is the most widely used one. It is concerned with the vertex with the maximum degree in G (as maximal properties often bring interesting results in combinatorics). It uses induction that is simpler than the one used in proof II, which will occur later.

Proof I (Modified upon Alexander Zykov's proof [11][3], 1949).

Lemma. The complete k-partite graph on n vertices with the largest number of edges is the Turán graph $T_{k,n}$.

Proof. For k parts of the graph, let the numbers of vertices they have be $p_1, p_2, \dots, p_k$. Then the number of edges in the graph will be

$$\sum_{1\leq p_i<p_j\leq k} p_i p_j,$$

which is equal to

$$\frac{1}{2}\left[\left(\sum_{i=1}^{k} p_i\right)^2-\sum_{i=1}^{k} p_i^2\right]. \tag{1}$$

From the Cauchy-Schwarz Inequality, we get the following inequality

$$n \sum_{i=1}^{k} p_i^2 \geq \left(\sum_{i=1}^{k} p_i\right)^2 \tag{2}$$

with equality when $p_1 = p_2 = \cdots = p_k$.

Since $\left(\sum_{i=1}^{k} p_i\right)^2$ is a constant n^2, we may use inequality (2) in inequality (1) and get the maximum number of edges, when all parts of the k-partite graph are of the equal size. As n might not be divisible by k, we determine that only when the size of each part is either $\lfloor n/k \rfloor$ or $\lceil n/k \rceil$ (as equal as possible) will the number of edges be maximized.

Therefore, the Turán graph $T_{k,n}$ is the graph that will maximize the number of edges. Q.E.D.

Based on the lemma, we may now prove Turán's theorem.

Obviously, the theorem holds for $k = 2$. Assume it holds for all integers less than k. Choose the vertex x of the maximum degree Δ in G, and set $X \coloneqq N(x)$ and $Y \coloneqq V - X$. We can get the following formula

$$e(G) = e(X) + e(X, Y) + e(Y).$$

Since G does not contain any K_k, $\langle X \rangle$ does not contain any K_{k-1} (If not, the additional vertex x together with the K_{k-1} in $\langle X \rangle$ will form a K_k in G). Thus, by the inductive hypothesis we obtain

$$e(X) \leq e(T_{k-2,\Delta}),$$

with equality if and only if $\langle X \rangle \cong T_{k-2,\Delta}$. Notice that each edge that is incident with vertices in Y belongs to either $E(X, Y)$ or $E(Y)$, meaning that

$$e(X, Y) + e(Y) \leq \Delta(n - \Delta)$$

with equality if and only if Y is an independent set and all vertices of Y is of degree Δ. Therefore, $e(G) \leq e(H)$, where H is defined as the combination of $T_{k-2,\Delta}$ and an independent set with $(n - \Delta)$ vertices, each of which is connected to all the Δ vertices in X. We see that H is a complete $(k-1)$-partite graph on n vertices. From the lemma, we learn that $e(H) \leq e(T_{k-1,n})$, with equality if and only if $H \cong T_{k-1,n}$. Thus, we get $e(G) \leq e(T_{k-1,n})$, with equality if and only if $G \cong T_{k-1,n}$. $\square$

The following proof is from Turán's original paper. It uses a completely different method from Alexander Zykov's proof. It uses induction, which results from a meticulous observation that every additional $(k-1)$ vertex will be individually distributed into the existing $(k-1)$ parts of vertices, according to the definition of Turán's graph.

Proof II (Modified upon Paul Turán's proof [9][12], 1941).

Obviously, the theorem holds for $n = k - 1$.

Lemma. We can prove that if G, the graph that doesn't contain K_k, has the maximum number of edges, it must have at least a K_{k-1}.

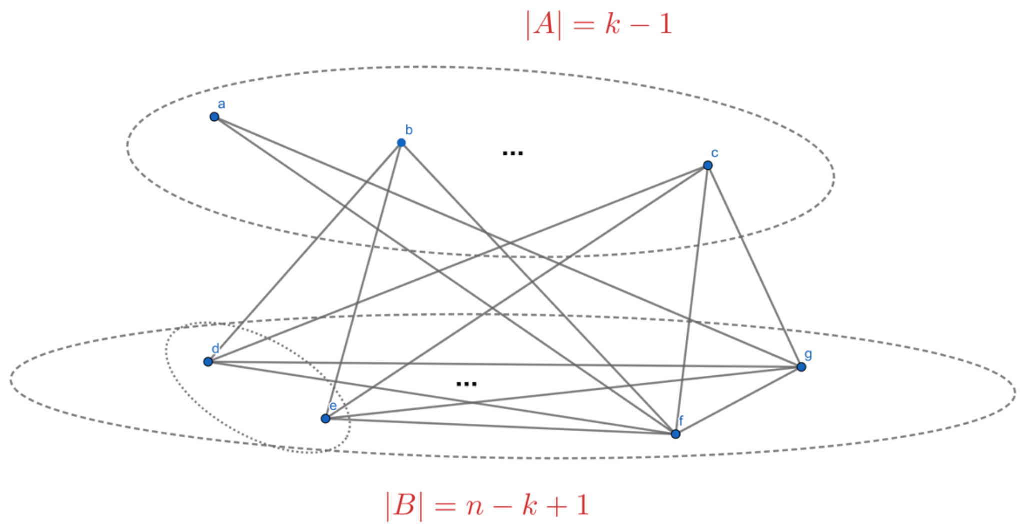
Proof. Pick out all the induced subgraphs of order $(k-1)$ from G and choose the one with the maximum number of edges p. For any such subgraph of order $(k-1)$ to be a K_{k-1}, it needs to add at least p edges. This means that it needs at least $(p+1)$ edges to be a K_k since every K_k with one edge deleted is always a K_{k-1}. As $(p+1)$ is greater than p, the G with p edges added can never form a K_k accidentally. Note that we may always increase the number of edges in G by adding these p edges; thus, there must at least be one K_{k-1} in the graph G with the maximum number of edges. Q.E.D.

Based on the lemma, we may now prove Turán's theorem.

Suppose the theorem holds for all "n"s less than a certain value bigger than $(k-1)$. Let A be the vertices of a $(k-1)$-clique in G and let B be $V-A$. Thus, $|B| = n-k+1$.

Since G has no k-clique, every vertex in B has at most $(k-2)$ vertices in A as its neighbors. Because the number of edges in $\langle B\rangle$ has already been maximized and $e(\langle A\rangle) = \binom{k-1}{2}$, we must let every vertex in B have $(k-2)$ neighbors in A. Also note that all the vertices of K_{k-1} in B must not have the same neighbor in A, or there will be a K_k in G.

We can only realize the condition above with the following arrangement. (The figure here may serve as an example.)



Let all vertices in every $(k-1)$ partitioned set of vertices in B (note that some of these sets might be empty sets for lower cases), which is a $T_{k-1,n-k+1}$, from the inductive hypothesis, have the same $(k-2)$ neighbors in A, while making every set of vertices in B have different sets of neighbors. Thus, we will get a single but different unconnected vertex in A for every partitioned set of vertices in B. (An example of this is in the figure above. Vertices d and e belong to a partitioned set called P_1 in B. The two vertices are connected to all the vertices in A except vertex a; thus, vertex a belongs to P_1 as well. It is true from the figure also that vertex f and g have different unconnected vertex in A.) Combine the unconnected vertices in A with the corresponding sets of vertices in B, we still get $(k-1)$ evenly distributed sets of vertices that forms a $T_{k-1,n}$.

Through induction, we know that if and only if $G \cong T_{k-1,n}$ will the number of edges in G be maximized. $\square$

From Turán's theorem, we get an immediate result on the bound of the minimum degree in G, denoted by $\delta(G)$.

Deduction on the lower bound of $\boldsymbol{\delta(G)}$. [2]

If K_k is not a subgraph of G_n, then $\delta(G_n) \leq \frac{k-2}{k-1}n$.

Proof. From the weaker version of Turán's theorem, we know that $e(G_n) \leq \frac{(k-2)n^2}{2(k-1)}$. In turn we get

$$\sum_{v \in V} d(v) = 2e(G_n) \leq \frac{(k-2)n^2}{(k-1)}.$$

Since $n\delta(G_n) = \sum_{v \in V} d(v)$, we then reach our result

$$\delta(G_n) \leq \frac{k-2}{k-1}n.$$

$\square$

Turán's graph not only determines the bound to the number of edges, but also yields the bounds to the clique number, the chromatic number, and the independence number, which are defined as follows (for a graph G):

Clique number, denoted by $\omega(G)$, is the order of the largest complete graph that is a subgraph of G.

Chromatic number, denoted by $\chi(G)$, is the minimum integer such that G is $\chi(G)$-colorable.

Independence number, denoted by $\alpha(G)$, is the largest size of an independent set of vertices in G.

Theorem on the lower bounds of $\boldsymbol{\omega(G)}$, $\boldsymbol{\chi(G)}$, and $\boldsymbol{\alpha(G)}$. [5]

For a graph G_n with e edges, it follows that

a) $\omega(G) \geq \left\lceil \frac{n^2}{n^2-2e} \right\rceil$, and $\chi(G) \geq \left\lceil \frac{n^2}{n^2-2e} \right\rceil$

b) $\alpha(G) \geq \left\lceil \frac{n^2}{n+2e} \right\rceil$.

Proof.

Hint for (a). Remember Turán's theorem indirectly encompasses the clique number of a graph.

a) We know formerly from the weaker version of Turán's theorem that $e \leq \left(1 - \frac{1}{\omega(G)}\right)\frac{n^2}{2}$, which in turn gives us the lower bound to the clique number:

$$\omega(G) \geq \frac{1}{1 - \frac{2e}{n^2}} = \frac{n^2}{n^2 - 2e}.$$

This is equivalent to $\omega(G) \geq \left\lceil \frac{n^2}{n^2-2e} \right\rceil$. Notice that this inequality is sharp if and only if $T_{p,n}$ has at least e edges and $e \geq \binom{p}{2}$, which ensures the use of the largest number of edges in Turán's graph. We may delete edges in $T_{p,n}$ until there are e edges left, making $\chi(G) = \omega(G) \geq \left\lceil \frac{n^2}{n^2-2e} \right\rceil$ sharp. □

Remarks. In fact, Turán's theorem is related to graph coloring (vertex coloring). Turán's graph is the solution to finding the graph with the maximum number of edges that *avoid* needing k colors for vertices, since every part in the Turán's graph may be colored with one single color but is different from the other existing colors. This is opposite to finding the chromatic number, which means we need to find the smallest number of colors that give a proper coloring to the graph.

Hint for (b). Note that every clique in G corresponds to every independent set in $\bar{G}$. This idea is conducive to our proof of b).

b) Let $H = \bar{G}$, and suppose $e' = e(H) = \binom{n}{2} - e$. $\alpha(H) = \alpha(\bar{G}) = \omega(G) \geq \frac{n^2}{n^2-2e} = \frac{n^2}{n+2e'}$. This means for every graph G with n vertices, $\alpha(G) \geq \left\lceil \frac{n^2}{n+2e} \right\rceil$. □

The simplicity of Turán's theorem probably indicates that there are quite a few extensions that can be developed upon this theorem. In 1970, P. Erdös published the paper *On the Graph Theory of Turán*, where he takes the degree sequence of k-partite graphs into the picture.

Erdös theorem (1970). [9][2][4] If G is a graph of order n with vertex set $V = \{x_1, x_2, x_3, \dots, x_n\}$ ($d(x_1) \leq d(x_2) \leq d(x_3) \leq \cdots \leq d(x_n)$) that contains no K_p, then there exists a $(p-1)$-partite graph (We can replace this with a $(p-1)$-chromatic graph as well since we can paint every partitioned set of vertices with the same color.) G' with vertex set $V' = \{x_1', x_2', x_3', \dots, x_n'\}$ ($d(x_1') \leq d(x_2') \leq d(x_3') \leq \cdots \leq d(x_n')$) such that

$$d(x_i) \leq d(x_i'), \qquad i = 1,2,\dots,n.$$

Proof. We apply induction on p. When $p = 2$, the case is obviously true. Suppose the case holds for $2 < i \leq k$. Now consider the case for $i = k+1$. At this stage, we pick the vertex v with the maximum degree and use the letter N to denote the set of the vertices that are adjacent to v. Here we consider the induced subgraph $\langle N \rangle$, which contains no K_{p-1} since all the elements in N combined with v will form a K_p. Since $\langle N \rangle$ contains no K_{p-1}, by the inductive hypothesis, every vertex in $\langle N \rangle$ will achieve their maximum degrees when $\langle N \rangle$ is rearranged into a $(p-2)$-partite graph. For the vertex set $V-N$, let every element in this set be adjacent to every vertex in N. Thus, the vertex set $V-N$ itself will be another partitioned set in G apart from the existing $(p-2)$ partitioned set of vertices in $\langle N \rangle$, while every vertex x_i' in $V-N$ will have degree $d(v)$, the largest degree in G. The newly formed G' is a $(p-1)$-partite graph, and we clearly know that the degree of every vertex reaches its maximum, meaning that

$$d(x_i) \leq d(x_i'), \qquad i = 1,2,\dots,n.$$

□

It's notable that this theorem directly implies Turán's theorem, as

$$2e(G) = \sum_{i=1}^{n} d(x_i) \leq \sum_{i=1}^{n} d(x_i') = 2e(G') \leq 2e(T_{p-1,n}).$$

Hence, Erdös Theorem actually gives the third proof of Turán's theorem.

Some other properties were introduced in G. Dirac's 1963 paper *Extension of Turán's Theorem on Graphs* [6]. The several theorems in the following pages are due to him. The foundation of these theorems lies on the

$$t_{k-1}(n) = \frac{1}{2}\left(1 - \frac{1}{k-1}\right)(n^2 - r^2) + \binom{r}{2},$$

which has been explained formerly.

Theorem I. Let k and n be integers, where $n \geq k+1 \geq 4$. Any graph with n vertices and at least $t_{k-1}(n) + \alpha$ edges, where α is an integer less than or equal to 1, contains as least one subgraph with n' vertices and at least $t_{k-1}(n') + \alpha$ edges for every $n' = k, k+1, \dots, n-1$.

Hint. In this question, since the result is satisfied for every n' $(k \leq n' \leq n-1)$, and the starting graph is of order n, we might consider the problem reversely from $(n-1)$ down to k (similar to the method of reverse induction).

Proof. For the graph that has more than $t_{k-1}(n) + \alpha$ edges, we may delete some edges to convert the graph into the subgraph of itself with only $t_{k-1}(n) + \alpha$ edges left. Denote these newly generated subgraph as H.

We may substitute n using $(k-1)t + r$, where $1 \leq r \leq k-1$. Thus,

$$t_{k-1}(n) = \frac{1}{2}(k-1)(k-2)t^2 + (k-2)tr + \frac{1}{2}r(r-1).$$

At the same time, we get

$$t_{k-1}(n) - t_{k-1}(n-1) = (k-2)t + r - 1 = n - t + 1. \tag{1}$$

(At this stage we need to consider proving the existence of a vertex with degree less than or equal to $n-t+1$, since this will ensure that once this vertex is eliminated, the induced subgraph will face the same situation as the former graph, making the reverse induction possible.)

In the following lines we deduce that there must exist at least one vertex in H that is of degree $\leq n-t+1$.

Suppose that all vertices have a degree $\geq n-t$. Hence, we have edge number $e(H)$ at least $\frac{1}{2}n(n-t)=\frac{1}{2}[(k-1)t+r][(k-2)t+r]=\frac{1}{2}(k-1)(k-2)t^2+\frac{1}{2}(2k-3)tr+\frac{1}{2}r^2$, and in turn we obtain

$$e(H)-t_{k-1}(n)=\frac{1}{2}tr+\frac{1}{2}r=\frac{1}{2}r(t+1).$$

It is clear that when $t=1, r\geq 2$ from $n\geq k+1$, and when $t\geq 2, r\geq 1$. This means that $e(H)-t_{k-1}(n)\geq 2$, which contradicts to the definition that $e(H)\leq t_{k-1}(n)+\alpha$.

Therefore, there exists at least one vertex with degree $\leq n-t+1$.Q.E.D.

We may delete this vertex x_1 and attain the induced subgraph $\langle V(H)-x_1\rangle$, which has edge number $\geq t_{k-1}(n)+\alpha-(n-t+1)$. Remember it has been shown in (1) that

$$t_{k-1}(n)-(n-t+1)=t_{k-1}(n-1).$$

As a result, $e(\langle V(H)-x_1\rangle)\geq t_{k-1}(n-1)+\alpha$, meaning that H, a subgraph of the original graph, contains a subgraph of order $(n-1)$ and at least $t_{k-1}(n-1)+\alpha$ edges. We have proven that the original graph contains a $(n-1)$-vertex subgraph with at least $t_{k-1}(n-1)+\alpha$ edges.

At this stage, we notice that for the new subgraph, it is possible to follow the same steps as shown above simply through replacing n by $(n-1)$ (and $(n-1)$ by $(n-2)$ as well). After the procedure, we will get that a $(n-2)$-vertex subgraph with at least $t_{k-1}(n-2)+\alpha$ edges is a subgraph of the original graph as well.

By repetition, we will prove that for every integer n' $(k\leq n'\leq n-1)$, there exists at least one subgraph of the original graph with n' vertices and at least $t_{k-1}(n')+\alpha$ edges. $\square$

Theorem II. Let k, n, q, and α be integers such that $k\geq 3, 1\leq q\leq k-1, n\geq k+q-1$, and $\alpha\leq 1$. Any graph with n vertices and at least $t_{k-1}(n)+\alpha$ edges contains at least one K_{k+q-1} with $(q-\alpha)$ edges missing as a subgraph.

Proof. The theorem holds for the case $n=k$. When $n=k$, $q=1$ and $t_{k-1}(n)=\frac{1}{2}k(k-1)-1$. This indicates that the original graph G_n has at least $t_{k-1}(n)=\frac{1}{2}k(k-1)-(1-\alpha)$ edges, meaning that it does contain a subgraph of K_k with $(1-\alpha)$ edges missing.

For $n\geq k+1$ we may use the result deduced from theorem I. We know that the original graph G_n contains a graph with $(k+q-1)$ vertices and $t_k(k+q-1)+\alpha$ edges. We know from the theorem I that

$$t_{k-1}(n) = \frac{1}{2}(k-1)(k-2)t^2 + (k-2)tr + \frac{1}{2}r(r-1),$$

where $n = (k-1)t + r$ and $1 \leq r \leq k-1$.

We can substitute r by q and t by 1, replacing the equation by

$$\begin{aligned} &t_{k-1}(k+q-1) \\ &= \frac{1}{2}(k-1)(k-2) + (k-2)q + \frac{1}{2}q(q-1) \\ &= \frac{1}{2}(k+q-1)(k+q-2) - q \end{aligned}$$

The subgraph of G_n, as a result, contains $\frac{1}{2}(k+q-1)(k+q-2) - (q-\alpha)$, which proves that the subgraph is a K_{k+q-1} with $(q-\alpha)$ edges missing. □

We may replace α by 1 and every q individually by integers from 1 to $(k-1)$, which formulates a highly understandable theorem that serves as an extension of Turán's theorem.

Theorem III. Assume that $e(G_n) > t_{k-1}(n)$, where $n \geq k$. Then, for every q that satisfies $1 \leq q \leq k-1$, G_n contains a K_k, a K_{k+1} with 1 edge missing, …, and a K_{k+q-1} with $(q-1)$ edges missing, until n is no longer larger than $k+q-1$. □

This theorem directly implies Turán's theorem for the reason that under the condition $n \geq k \geq 3$, once $e(G_n) > t_{k-1}(n)$, G_n contains a K_k. According to logic, this is equivalent to the proposition

"*If G_n does not contain a K_k, $e(G_n) \leq t_{k-1}(n)$,*"

which is the same as Turán's theorem.

Dirac's theorem III gives the fourth proof of Turán's theorem in this paper.

Conclusion:

In this paper, we have briefly introduced the field of extremal graph theory. We have proven Mantel's theorem and Turán's theorem in detail and related these two theorems to other fields in graph theory. We have combined the theories mentioned in the paper with the historical development of extremal graph theory and also looked at the extensions of Turán's theorem. The paper has clearly presented the importance of extremal graph theory in graph theory and the great potential the investigation of this field deserves, as there exist many unsolved problems within this field.

References:

[1] Asaf Shapira, Guy Rutenberg. Extremal Graph Theory (course notes).

http://www.cs.tau.ac.il/~asafico/ext-graph-theory/notes.pdf

[2] Béla Bollobás. Chapter 6: Complete Subgraphs. *Extremal Graph Theory*. Academic Press Inc. (1978).

[3] J.A. Bondy, U.S.R. Murty. *Graph Theory*, Springer (2008), pp. 301–305.

[4] Miklós Simonovits. Paul Turán's influence in combinatorics, *Number Theory, Analysis, and Combinatorics*. De Gruyter (2013).

[5] Douglas B. West. Chapter 5, *Introduction to Graph Theory* (second edition and solution manual). Prentice Hall (2001, 2005).

[6] G. Dirac. Extensions of Turán's Theorem on Graphs, *Acta Math. Acad. Sci. Hung.* **14** (1963), pp. 417–422.

[7] A. W. Goodman. On sets of acquittances and strangers at any party, *Amer. Math. Monthly* **66** (1959), pp. 778–783.

[8] J. W. Moon and L. Moser. On a problem of Turán, *Publ. Math. Inst. Hungar. Acad. Sci.* **7** (1962), pp. 283–286.

[9] P. Erdös. On the graph theorem of Turán (in Hungarian), *Mat. Lapok* **21** (1970), pp. 249–251.

[10] P. Erdös and T. Gallai. On maximal paths and circuits of graphs, *Acta Math. Acad. Sci. Hung.* **10** (1959), pp. 337–356.

[11] A. A. Zykov. On some properties of linear complexes, *Mat. Sbornik N.S.* **24(66)** (1949), pp. 163–188.

[12] M. Aigner. Turán's graph theorem, *Amer. Math. Monthly* **102** (1995), pp. 808–816.

[13] L. Sauvé. On Chromatic Graphs, *Amer. Math. Monthly* **68** (1961), pp. 107–111.

[14] K. Zarankiewicz. Problem P 101, *Colloq. Math.* **2** (1951), pp. 301.

[15] P. Erdös. Extremal problems in graph theory, in: "Theory of Graphs and its Applications" (Fiedler, M., ed.), Acadmeic Press, New York (1965), pp. 29–36.

Remarks. If there are multiple references at the same time in the paper, the first one is where the original source comes from.