Lines in the plane and decompositions of graphs

Chapter 11

Perhaps the best-known problem on configurations of lines was raised by Sylvester in 1893 in a column of mathematical problems.

QUESTIONS FOR SOLUTION.

11851. (Professor SYLVESTER.)—Prove that it is not possible to arrange any finite number of real points so that a right line through every two of them shall pass through a third, unless they all lie in the same right line.

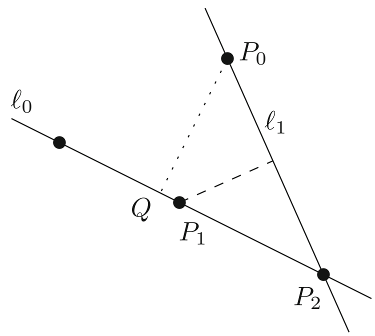
Whether Sylvester himself had a proof is in doubt, but a correct proof was given by Tibor Gallai [Grünwald] some 40 years later. Therefore the following theorem is commonly attributed to Sylvester and Gallai. Subsequent to Gallai's proof several others appeared, but the following argument due to L. M. Kelly may be "simply the best."

J. J. Sylvester

Theorem 1. *In any configuration of n points in the plane, not all on a line, there is a line which contains exactly two of the points.*

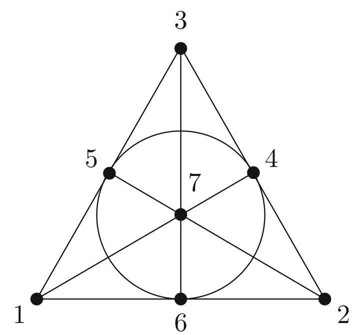
■ **Proof.** Let $\mathcal{P}$ be the given set of points and consider the set $\mathcal{L}$ of all lines which pass through at least two points of $\mathcal{P}$. Among all pairs (P, ℓ) with P not on ℓ, choose a pair (P_0, ℓ_0) such that P_0 has the smallest distance to ℓ_0, with Q being the point on ℓ_0 closest to P_0 (that is, on the line through P_0 vertical to ℓ_0).

Claim. *This line ℓ_0 does it!*



If not, then ℓ_0 contains at least three points of $\mathcal{P}$, and thus two of them, say P_1 and P_2, lie on the same side of Q. Let us assume that P_1 lies between Q and P_2, where P_1 possibly coincides with Q. The figure on the right shows the configuration. It follows that the distance of P_1 to the line ℓ_1 determined by P_0 and P_2 is smaller than the distance of P_0 to ℓ_0, and this contradicts our choice for ℓ_0 and P_0. □

In the proof we have used metric axioms (shortest distance) and order axioms (P_1 lies between Q and P_2) of the real plane. Do we really need these properties beyond the usual incidence axioms of points and lines? Well, some additional condition is required, as the famous Fano plane depicted in the margin demonstrates. Here $\mathcal{P} = \{1, 2, \ldots, 7\}$ and $\mathcal{L}$ consists of the 7 three-point lines as indicated in the figure, including the "line" $\{4, 5, 6\}$. Any two points determine a unique line, so the incidence axioms are satisfied, but there is no 2-point line. The Sylvester–Gallai theorem therefore shows that the Fano configuration cannot be embedded into the real plane such that the seven collinear triples lie on real lines: there must always be a "crooked" line.



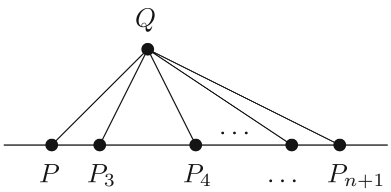
M. Aigner, G. M. Ziegler, *Proofs from THE BOOK*, https://doi.org/10.1007/978-3-662-57265-8_11

However, it was shown by Coxeter that the order axioms will suffice for a proof of the Sylvester–Gallai theorem. Thus one can devise a proof that does not use any metric properties — see also the proof that we will give in Chapter 13, using Euler's formula.

Armed with Theorem 1, we may ask how many such two-point lines every n-point configuration in the plane must contain. After many partial results, the definitive answer was given very recently by Ben Green and Terence Tao: There is a constant n_0 such that every configuration of $n \geq n_0$ points, not all on a line, contains at least $n/2$ two-point lines, and this is best possible — if n is even. In the case when n is odd, they prove that there are even at least $3\lfloor n/4 \rfloor$ such lines, and again this is best possible.

The Sylvester–Gallai theorem directly implies another famous result on points and lines in the plane, due to Paul Erdős and Nicolaas G. de Bruijn. But in this case the result holds more generally for arbitrary point-line systems, as was observed already by Erdős and de Bruijn. We will discuss the more general result in a moment.

Theorem 2. *Let $\mathcal{P}$ be a set of $n \geq 3$ points in the plane, not all on a line. Then the set $\mathcal{L}$ of lines passing through at least two points contains at least n lines.*



■ **Proof.** For $n = 3$ there is nothing to show. Now we proceed by induction on n. Let $|\mathcal{P}| = n + 1$. By the previous theorem there exists a line $\ell_0 \in \mathcal{L}$ containing exactly two points P and Q of $\mathcal{P}$. Consider the set $\mathcal{P}' = \mathcal{P} \backslash \{Q\}$ and the set $\mathcal{L}'$ of lines determined by $\mathcal{P}'$. If the points of $\mathcal{P}'$ do not all lie on a single line, then by induction $|\mathcal{L}'| \geq n$ and hence $|\mathcal{L}| \geq n + 1$ because of the additional line ℓ_0 in $\mathcal{L}$. If, on the other hand, the points in $\mathcal{P}'$ are all on a single line, then we have the "pencil" which results in precisely $n + 1$ lines. □

Now, as promised, here is the general result, which applies to much more general "incidence geometries."

Theorem 3. *Let X be a set of $n \geq 3$ elements, and let $A_1, \ldots, A_m$ be proper subsets of X, such that every pair of elements of X is contained in precisely one set A_i. Then $m \geq n$ holds.*

■ **Proof.** The following proof, variously attributed to Motzkin or Conway, is almost one-line and truly inspired. For $x \in X$ let r_x be the number of sets A_i containing x. (Note that $2 \leq r_x < m$ by the assumptions.) Now if $x \notin A_i$, then $r_x \geq |A_i|$ because the $|A_i|$ sets containing x and an element of A_i must be distinct. Suppose $m < n$, then $m|A_i| < n\, r_x$ and thus $m(n - |A_i|) > n(m - r_x)$ for $x \notin A_i$, and we find

$$1 = \sum_{x \in X} \frac{1}{n} = \sum_{x \in X} \sum_{A_i : x \notin A_i} \frac{1}{n(m-r_x)} > \sum_{A_i} \sum_{x : x \notin A_i} \frac{1}{m(n-|A_i|)} = \sum_{A_i} \frac{1}{m} = 1,$$

which is absurd. □

There is another very short proof for this theorem that uses linear algebra. Let B be the *incidence matrix* of $(X; A_1, \dots, A_m)$, that is, the rows in B are indexed by the elements of X, the columns by $A_1, \dots, A_m$, where

$$B_{xA} := \begin{cases} 1 & \text{if} \quad x \in A \\ 0 & \text{if} \quad x \notin A. \end{cases}$$

Consider the product BB^T. For $x \neq x'$ we have $(BB^T)_{xx'} = 1$, since x and x' are contained in precisely one set A_i, hence

$$BB^T = \begin{pmatrix} r_{x_1}-1 & 0 & \dots & 0 \\ 0 & r_{x_2}-1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & r_{x_n}-1 \end{pmatrix} + \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & & \vdots \\ \vdots & & \ddots & 1 \\ 1 & \dots & 1 & 1 \end{pmatrix}$$

where r_x is defined as above. Since the first matrix is positive definite (it has only positive eigenvalues) and the second matrix is positive semi-definite (it has the eigenvalues n and 0), we deduce that BB^T is positive definite and thus, in particular, invertible, implying $\text{rank}(BB^T) = n$. It follows that the rank of the $n \times m$ matrix B is at least n, and we conclude that indeed $n \leq m$, since the rank cannot exceed the number of columns.

Let us go a little beyond and turn to graph theory. (We refer to the review of basic graph concepts in the appendix to this chapter.) A moment's thought shows that the following statement is really the same as Theorem 3:

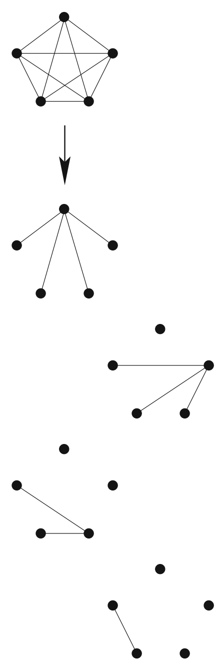
> *If we decompose a complete graph K_n into m cliques different from K_n, such that every edge is in a unique clique, then $m \geq n$.*

Indeed, let X correspond to the vertex set of K_n and the sets A_i to the vertex sets of the cliques, then the statements are identical.

Our next task is to decompose K_n into complete bipartite graphs such that again every edge is in exactly one of these graphs. There is an easy way to do this. Number the vertices $\{1, 2, \dots, n\}$. First take the complete bipartite graph joining 1 to all other vertices. Thus we obtain the graph $K_{1,n-1}$ which is called a *star*. Next join 2 to $3, \dots, n$, resulting in a star $K_{1,n-2}$. Going on like this, we decompose K_n into stars $K_{1,n-1}, K_{1,n-2}, \dots, K_{1,1}$. This decomposition uses $n-1$ complete bipartite graphs. Can we do better, that is, use fewer graphs? No, as the following result of Ron Graham and Henry O. Pollak says:

Theorem 4. *If K_n is decomposed into complete bipartite subgraphs $H_1, \dots, H_m$, then $m \geq n - 1$.*

The interesting thing is that, in contrast to the Erdős–de Bruijn theorem, no combinatorial proof for this result is known! All of them use linear algebra in one way or another. Of the various more or less equivalent ideas let us look at the proof due to Tverberg, which may be the most transparent.



A decomposition of K_5 into 4 complete bipartite subgraphs

■ **Proof.** Let the vertex set of K_n be $\{1,\dots,n\}$, and let L_j, R_j be the defining vertex sets of the complete bipartite graph H_j, $j = 1,\dots,m$. To every vertex i we associate a variable x_i. Since $H_1,\dots,H_m$ decompose K_n, we find

$$\sum_{i<j} x_i x_j \;=\; \sum_{k=1}^{m} \Big(\sum_{a\in L_k} x_a \cdot \sum_{b\in R_k} x_b\Big). \tag{1}$$

Now suppose the theorem is false, $m < n-1$. Then the system of linear equations

$$\begin{aligned} x_1 + \dots + x_n &= 0, \\ \sum_{a\in L_k} x_a &= 0 \qquad (k = 1,\dots,m) \end{aligned}$$

has fewer equations than variables, hence there exists a nontrivial solution $c_1,\dots,c_n$. From (1) we infer

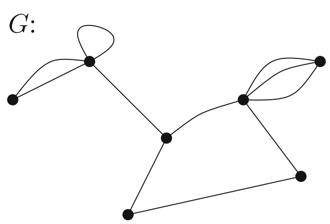
$$\sum_{i<j} c_i c_j \;=\; 0.$$

But this implies

$$0 \;=\; (c_1 + \dots + c_n)^2 \;=\; \sum_{i=1}^{n} c_i^2 + 2\sum_{i<j} c_i c_j \;=\; \sum_{i=1}^{n} c_i^2 \;>\; 0,$$

a contradiction, and the proof is complete. □

Appendix: Basic graph concepts

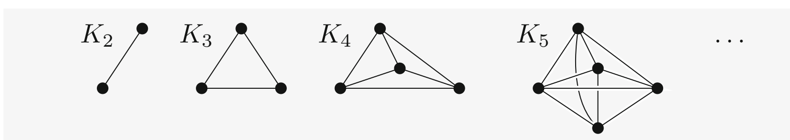


A graph G with 7 vertices and 11 edges. It has one loop, one double edge and one triple edge.

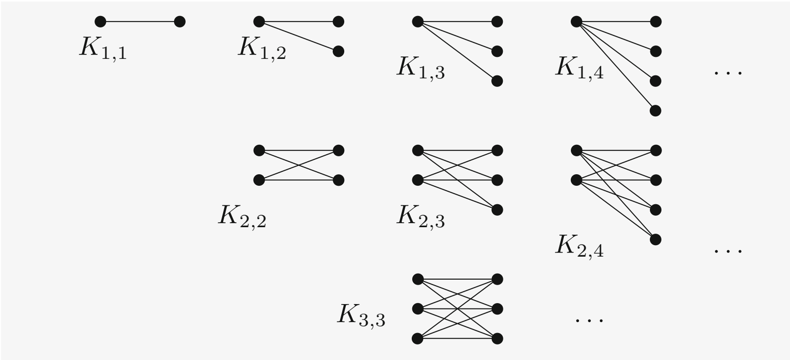
Graphs are among the most basic of all mathematical structures. Correspondingly, they have many different versions, representations, and incarnations. Abstractly, a *graph* is a pair $G = (V, E)$, where V is the set of *vertices*, E is the set of *edges*, and each edge $e \in E$ "connects" two vertices $v, w \in V$. We consider only finite graphs, where V and E are finite.

Usually, we deal with *simple graphs*: Then we do not admit *loops*, i. e., edges for which both ends coincide, and no *multiple edges* that have the same set of endvertices. Vertices of a graph are called *adjacent* or *neighbors* if they are the endvertices of an edge. A vertex and an edge are called *incident* if the edge has the vertex as an endvertex.

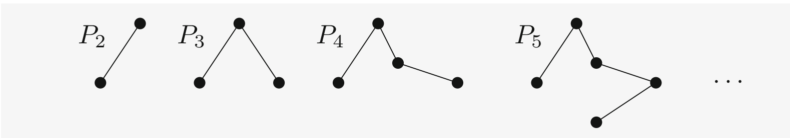
Here is a little picture gallery of important (simple) graphs:



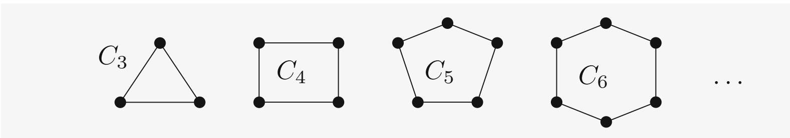
The *complete graphs* K_n on n vertices and $\binom{n}{2}$ edges



The *complete bipartite graphs* $K_{m,n}$ with $m + n$ vertices and mn edges

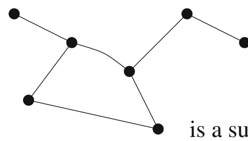


The *paths* P_n with n vertices

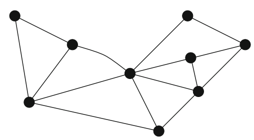


The *cycles* C_n with n vertices

Two graphs $G = (V, E)$ and $G' = (V', E')$ are considered *isomorphic* if there are bijections $V \to V'$ and $E \to E'$ that preserve the incidences between edges and their endvertices. (It is a major unsolved problem whether there is an efficient test to decide whether two given graphs are isomorphic.) This notion of isomorphism allows us to talk about *the* complete graph K_5 on 5 vertices, etc.



is a subgraph of



$G' = (V', E')$ is a *subgraph* of $G = (V, E)$ if $V' \subseteq V, E' \subseteq E$, and every edge $e \in E'$ has the same endvertices in G' as in G. G' is an *induced subgraph* if, additionally, *all* edges of G that connect vertices of G' are also edges of G'.

Many notions about graphs are quite intuitive: for example, a graph G is *connected* if every two distinct vertices are connected by a path in G, or equivalently, if G cannot be split into two nonempty subgraphs whose vertex sets are disjoint. Any graph decomposes into its *connected components*.

We end this survey of basic graph concepts with a few more pieces of terminology: A *clique* in G is a complete subgraph. An *independent set* in G is an induced subgraph without edges, that is, a subset of the vertex set such that no two vertices are connected by an edge of G. A graph is a *forest* if it does not contain any cycles. A *tree* is a connected forest. Finally, a graph $G = (V, E)$ is *bipartite* if it is isomorphic to a subgraph of a complete bipartite graph, that is, if its vertex set can be written as a union $V = V_1 \cup V_2$ of two independent sets.

References

[1] N. G. DE BRUIJN & P. ERDŐS: *On a combinatorial problem,* Proc. Kon. Ned. Akad. Wetensch. **51** (1948), 1277-1279.

[2] H. S. M. COXETER: *A problem of collinear points,* Amer. Math. Monthly **55** (1948), 26-28 (contains Kelly's proof).

[3] P. ERDŐS: *Problem 4065 — Three point collinearity,* Amer. Math. Monthly **51** (1944), 169-171 (contains Gallai's proof).

[4] R. L. GRAHAM & H. O. POLLAK: *On the addressing problem for loop switching,* Bell System Tech. J. **50** (1971), 2495-2519.

[5] B. GREEN & T. TAO: *On sets defining few ordinary lines,* Discrete Comput. Geometry **50** (2013), 409-468.

[6] J. J. SYLVESTER: *Mathematical Question 11851,* The Educational Times **46** (1893), 156.

[7] H. TVERBERG: *On the decomposition of K_n into complete bipartite graphs,* J. Graph Theory **6** (1982), 493-494.