



Some mathematical principles, such as the two in the title of this chapter, are so obvious that you might think they would only produce equally obvious results. To convince you that “It ain’t necessarily so” we illustrate them with examples that were suggested by Paul Erdős to be included in The Book. We will encounter instances of them also in later chapters.

Pigeon-hole principle

If n objects are placed in r boxes, where $r < n$, then at least one of the boxes contains more than one object.

Well, this is indeed obvious, there is nothing to prove. In the language of mappings our principle reads as follows: Let N and R be two finite sets with

$$|N| = n > r = |R|,$$

and let $f : N \rightarrow R$ be a mapping. Then there exists some $a \in R$ with $|f^{-1}(a)| \geq 2$. We may even state a stronger inequality: There exists some $a \in R$ with

$$|f^{-1}(a)| \geq \left\lceil \frac{n}{r} \right\rceil. \quad (1)$$

In fact, otherwise we would have $|f^{-1}(a)| < \frac{n}{r}$ for all a , and hence $n = \sum_{a \in R} |f^{-1}(a)| < r \frac{n}{r} = n$, which cannot be.

1. Numbers

Claim. *Consider the numbers $1, 2, 3, \dots, 2n$, and take any $n + 1$ of them. Then there are two among these $n + 1$ numbers which are relatively prime.*

This is again obvious. There must be two numbers which are only 1 apart, and hence relatively prime.

But let us now turn the condition around.

Claim. *Suppose again $A \subseteq \{1, 2, \dots, 2n\}$ with $|A| = n + 1$. Then there are always two numbers in A such that one divides the other.*



“The pigeon-holes from a bird’s perspective”

Both results are no longer true if one replaces $n+1$ by n : For this consider the sets $\{2, 4, 6, \dots, 2n\}$, respectively $\{n+1, n+2, \dots, 2n\}$.

This is not so clear. As Erdős told us, he put this question to young Lajos Pósa during dinner, and when the meal was over, Lajos had the answer. It has remained one of Erdős' favorite "initiation" questions to mathematics. The (affirmative) solution is provided by the pigeon-hole principle. Write every number $a \in A$ in the form $a = 2^k m$, where m is an odd number between 1 and $2n - 1$. Since there are $n + 1$ numbers in A , but only n different odd parts, there must be two numbers in A with the *same* odd part. Hence one is a multiple of the other. \square

2. Sequences

Here is another one of Erdős' favorites, contained in a paper of Erdős and Szekeres on Ramsey problems.

Claim. *In any sequence $a_1, a_2, \dots, a_{mn+1}$ of $mn+1$ distinct real numbers, there exists an increasing subsequence*

$$a_{i_1} < a_{i_2} < \dots < a_{i_{m+1}} \quad (i_1 < i_2 < \dots < i_{m+1})$$

of length $m + 1$, or a decreasing subsequence

$$a_{j_1} > a_{j_2} > \dots > a_{j_{n+1}} \quad (j_1 < j_2 < \dots < j_{n+1})$$

of length $n + 1$, or both.

This time the application of the pigeon-hole principle is not immediate. Associate to each a_i the number t_i which is the length of a *longest increasing* subsequence starting at a_i . If $t_i \geq m + 1$ for some i , then we have an increasing subsequence of length $m + 1$. Suppose then that $t_i \leq m$ for all i . The function $f : a_i \mapsto t_i$ mapping $\{a_1, \dots, a_{mn+1}\}$ to $\{1, \dots, m\}$ tells us by (1) that there is some $s \in \{1, \dots, m\}$ such that $f(a_i) = s$ for $\frac{mn}{m} + 1 = n + 1$ numbers a_i . Let $a_{j_1}, a_{j_2}, \dots, a_{j_{n+1}}$ ($j_1 < \dots < j_{n+1}$) be these numbers. Now look at two consecutive numbers $a_{j_i}, a_{j_{i+1}}$. If $a_{j_i} < a_{j_{i+1}}$, then we would obtain an increasing subsequence of length s starting at $a_{j_{i+1}}$, and consequently an increasing subsequence of length $s + 1$ starting at a_{j_i} , which cannot be since $f(a_{j_i}) = s$. We thus obtain a decreasing subsequence $a_{j_1} > a_{j_2} > \dots > a_{j_{n+1}}$ of length $n + 1$. \square

The reader may have fun in proving that for mn numbers the statement remains no longer true in general.

This simple-sounding result on monotone subsequences has a highly nonobvious consequence on the *dimension of graphs*. We don't need here the notion of dimension for general graphs, but only for complete graphs K_n . It can be phrased in the following way. Let $N = \{1, \dots, n\}$, $n \geq 3$, and consider m permutations π_1, \dots, π_m of N . We say that the permutations π_i *represent* K_n if to every three distinct numbers i, j, k there exists a permutation π in which k comes *after* both i and j . The dimension of K_n is then the smallest m for which a representation π_1, \dots, π_m exists.

As an example we have $\dim(K_3) = 3$ since any one of the three numbers must come last, as in $\pi_1 = (1, 2, 3)$, $\pi_2 = (2, 3, 1)$, $\pi_3 = (3, 1, 2)$. What

about K_4 ? Note first $\dim(K_n) \leq \dim(K_{n+1})$: just delete $n + 1$ in a representation of K_{n+1} . So, $\dim(K_4) \geq 3$, and, in fact, $\dim(K_4) = 3$, by taking

$$\pi_1 = (1, 2, 3, 4), \quad \pi_2 = (2, 4, 3, 1), \quad \pi_3 = (1, 4, 3, 2).$$

It is not quite so easy to prove $\dim(K_5) = 4$, but then, surprisingly, the dimension stays at 4 up to $n = 12$, while $\dim(K_{13}) = 5$. So $\dim(K_n)$ seems to be a pretty wild function. Well, it is not! With n going to infinity, $\dim(K_n)$ is, in fact, a very well-behaved function — and the key for finding a lower bound is the pigeon-hole principle. We claim

$$\dim(K_n) \geq \log_2 \log_2 n. \quad (2)$$

Since, as we have seen, $\dim(K_n)$ is a monotone function in n , it suffices to verify (2) for $n = 2^{2^p} + 1$, that is, we have to show that

$$\dim(K_n) \geq p + 1 \quad \text{for } n = 2^{2^p} + 1.$$

Suppose, on the contrary, $\dim(K_n) \leq p$, and let π_1, \dots, π_p be representing permutations of $N = \{1, 2, \dots, 2^{2^p} + 1\}$. Now we use our result on monotone subsequences p times. In π_1 there exists a monotone subsequence A_1 of length $2^{2^{p-1}} + 1$ (it does not matter whether increasing or decreasing). Look at this set A_1 in π_2 . Using our result again, we find a monotone subsequence A_2 of A_1 in π_2 of length $2^{2^{p-2}} + 1$, and A_2 is, of course, also monotone in π_1 . Continuing, we eventually find a subsequence A_p of size $2^{2^0} + 1 = 3$ which is monotone in *all* permutations π_i . Let $A_p = (a, b, c)$, then either $a < b < c$ or $a > b > c$ in *all* π_i . But this cannot be, since there must be a permutation where b comes after a and c . \square

The right asymptotic growth was provided by Joel Spencer (upper bound) and by Füredi, Hajnal, Rödl and Trotter (lower bound):

$$\dim(K_n) = \log_2 \log_2 n + \left(\frac{1}{2} + o(1)\right) \log_2 \log_2 \log_2 n.$$

But this is not the whole story: In 1999, Morris and Hoşten found a method which, in principle, establishes the *precise* value of $\dim(K_n)$. Using their result and a computer one can obtain the values given in the margin. This is truly astounding! Just consider how many permutations of size 1422564 there are. How does one decide whether 7 or 8 of them are required to represent $K_{1422564}$?

$$\begin{aligned} \pi_1: & 1 \ 2 \ 3 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 4 \\ \pi_2: & 2 \ 3 \ 4 \ 8 \ 7 \ 6 \ 5 \ 12 \ 11 \ 10 \ 9 \ 1 \\ \pi_3: & 3 \ 4 \ 1 \ 11 \ 12 \ 9 \ 10 \ 6 \ 5 \ 8 \ 7 \ 2 \\ \pi_4: & 4 \ 1 \ 2 \ 10 \ 9 \ 12 \ 11 \ 7 \ 8 \ 5 \ 6 \ 3 \end{aligned}$$

These four permutations represent K_{12}

$$\begin{aligned} \dim(K_n) \leq 4 & \iff n \leq 12 \\ \dim(K_n) \leq 5 & \iff n \leq 81 \\ \dim(K_n) \leq 6 & \iff n \leq 2646 \\ \dim(K_n) \leq 7 & \iff n \leq 1422564 \end{aligned}$$

3. Sums

Paul Erdős attributes the following nice application of the pigeon-hole principle to Andrew Vázsonyi and Marta Sved:

Claim. *Suppose we are given n integers a_1, \dots, a_n , which need not be distinct. Then there is always a set of consecutive numbers $a_{k+1}, a_{k+2}, \dots, a_\ell$ whose sum $\sum_{i=k+1}^\ell a_i$ is a multiple of n .*

For the proof we set $N = \{0, 1, \dots, n\}$ and $R = \{0, 1, \dots, n-1\}$. Consider the map $f : N \rightarrow R$, where $f(m)$ is the remainder of $a_1 + \dots + a_m$ upon division by n . Since $|N| = n+1 > n = |R|$, it follows that there are two sums $a_1 + \dots + a_k$, $a_1 + \dots + a_\ell$ ($k < \ell$) with the *same* remainder, where the first sum may be the empty sum denoted by 0. It follows that

$$\sum_{i=k+1}^{\ell} a_i = \sum_{i=1}^{\ell} a_i - \sum_{i=1}^k a_i$$

has remainder 0 — end of proof. \square

Let us turn to the second principle: counting in two ways. By this we mean the following.

Double counting

Suppose that we are given two finite sets R and C and a subset $S \subseteq R \times C$. Whenever $(p, q) \in S$, then we say p and q are incident. If r_p denotes the number of elements that are incident to $p \in R$, and c_q denotes the number of elements that are incident to $q \in C$, then

$$\sum_{p \in R} r_p = |S| = \sum_{q \in C} c_q. \quad (3)$$

Again, there is nothing to prove. The first sum classifies the pairs in S according to the first entry, while the second sum classifies the same pairs according to the second entry.

There is a useful way to picture the set S . Consider the matrix $A = (a_{pq})$, the *incidence matrix* of S , where the rows and columns of A are indexed by the elements of R and C , respectively, with

$$a_{pq} = \begin{cases} 1 & \text{if } (p, q) \in S \\ 0 & \text{if } (p, q) \notin S. \end{cases}$$

With this set-up, r_p is the sum of the p -th row of A and c_q is the sum of the q -th column. Hence the first sum in (3) adds the entries of A (that is, counts the elements in S) by rows, and the second sum by columns.

The following example should make this correspondence clear. Let $R = C = \{1, 2, \dots, 8\}$, and set $S = \{(i, j) : i \text{ divides } j\}$. We then obtain the matrix in the margin, which only displays the 1's.

$R \backslash C$	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1	1
2		1		1		1		1
3			1			1		
4				1				1
5					1			
6						1		
7							1	
8								1

4. Numbers again

Look at the table on the left. The number of 1's in column j is precisely the number of divisors of j ; let us denote this number by $t(j)$. Let us ask how

large this number $t(j)$ is on the *average* when j ranges from 1 to n . Thus, we ask for the quantity

$$\bar{t}(n) = \frac{1}{n} \sum_{j=1}^n t(j).$$

n	1	2	3	4	5	6	7	8
$\bar{t}(n)$	1	$\frac{3}{2}$	$\frac{5}{3}$	2	2	$\frac{7}{3}$	$\frac{16}{7}$	$\frac{5}{2}$

The first few values of $\bar{t}(n)$

How large is $\bar{t}(n)$ for arbitrary n ? At first glance, this seems hopeless. For prime numbers p we have $t(p) = 2$, while for 2^k we obtain a large number $t(2^k) = k + 1$. So, $t(n)$ is a wildly jumping function, and we surmise that the same is true for $\bar{t}(n)$. Wrong guess, the opposite is true! Counting in two ways provides an unexpected and simple answer.

Consider the matrix A (as above) for the integers 1 up to n . Counting by columns we get $\sum_{j=1}^n t(j)$. How many 1's are in row i ? Easy enough, the 1's correspond to the multiples of i : $1i, 2i, \dots$, and the last multiple not exceeding n is $\lfloor \frac{n}{i} \rfloor i$. Hence we obtain

$$\bar{t}(n) = \frac{1}{n} \sum_{j=1}^n t(j) = \frac{1}{n} \sum_{i=1}^n \left\lfloor \frac{n}{i} \right\rfloor \leq \frac{1}{n} \sum_{i=1}^n \frac{n}{i} = \sum_{i=1}^n \frac{1}{i},$$

where the error in each summand, when passing from $\lfloor \frac{n}{i} \rfloor$ to $\frac{n}{i}$, is less than 1. Now the last sum is the n -th harmonic number H_n , so we obtain $H_n - 1 < \bar{t}(n) \leq H_n$, and together with the estimates of H_n on page 13 this gives

$$\log n - 1 < H_n - 1 - \frac{1}{n} < \bar{t}(n) \leq H_n < \log n + 1.$$

Thus we have proved the remarkable result that, while $t(n)$ is totally erratic, the average $\bar{t}(n)$ behaves beautifully: It differs from $\log n$ by less than 1.

5. Graphs

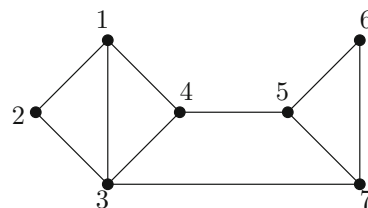
Let G be a finite simple graph with vertex set V and edge set E . We have defined in Chapter 13 the *degree* $d(v)$ of a vertex v as the number of edges which have v as an end-vertex. In the example of the figure, the vertices $1, 2, \dots, 7$ have degrees 3, 2, 4, 3, 3, 2, 3, respectively.

Almost every book in graph theory starts with the following result (that we have already encountered in Chapters 13 and 20):


$$\sum_{v \in V} d(v) = 2|E|. \quad (4)$$

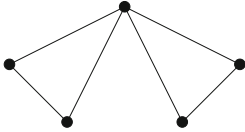
For the proof consider $S \subseteq V \times E$, where S is the set of pairs (v, e) such that $v \in V$ is an end-vertex of $e \in E$. Counting S in two ways gives on the one hand $\sum_{v \in V} d(v)$, since every vertex contributes $d(v)$ to the count, and on the other hand $2|E|$, since every edge has two ends. \square

As simple as the result (4) appears, it has many important consequences, some of which will be discussed as we go along. We want to single out in



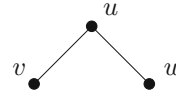
this section the following beautiful application to an *extremal problem* on graphs. Here is the problem:

Suppose $G = (V, E)$ has n vertices and contains no cycle of length 4 (denoted by C_4), that is, no subgraph . How many edges can G have at most?



As an example, the graph in the margin on 5 vertices contains no 4-cycle and has 6 edges. The reader may easily show that on 5 vertices the maximal number of edges is 6, and that this graph is indeed the only graph on 5 vertices with 6 edges that has no 4-cycle.

Let us tackle the general problem. Let G be a graph on n vertices without a 4-cycle. As above we denote by $d(u)$ the degree of u . Now we count the following set S in two ways: S is the set of pairs $(u, \{v, w\})$ where u is adjacent to v and to w , with $v \neq w$. In other words, we count all occurrences of



Summing over u , we find $|S| = \sum_{u \in V} \binom{d(u)}{2}$. On the other hand, every pair $\{v, w\}$ has at most one common neighbor (by the C_4 -condition). Hence $|S| \leq \binom{n}{2}$, and we conclude

$$\sum_{u \in V} \binom{d(u)}{2} \leq \binom{n}{2}$$

or

$$\sum_{u \in V} d(u)^2 \leq n(n-1) + \sum_{u \in V} d(u). \quad (5)$$

Next (and this is quite typical for this sort of extremal problems) we apply the Cauchy–Schwarz inequality to the vectors $(d(u_1), \dots, d(u_n))$ and $(1, 1, \dots, 1)$, obtaining

$$\left(\sum_{u \in V} d(u) \right)^2 \leq n \sum_{u \in V} d(u)^2,$$

and hence by (5)

$$\left(\sum_{u \in V} d(u) \right)^2 \leq n^2(n-1) + n \sum_{u \in V} d(u).$$

Invoking (4) we find

$$4|E|^2 \leq n^2(n-1) + 2n|E|$$

or

$$|E|^2 - \frac{n}{2}|E| - \frac{n^2(n-1)}{4} \leq 0.$$

Solving the corresponding quadratic equation we thus obtain the following result of Istvan Reiman.

Theorem. *If the graph G on n vertices contains no 4-cycles, then*

$$|E| \leq \left\lfloor \frac{n}{4} (1 + \sqrt{4n-3}) \right\rfloor. \quad (6)$$

For $n = 5$ this gives $|E| \leq 6$, and the graph above shows that equality can hold.

Counting in two ways has thus produced in an easy way an upper bound on the number of edges. But how good is the bound (6) in general? The following beautiful example [2] [3] [6] shows that it is almost sharp. As is often the case in such problems, finite geometry leads the way.

In presenting the example we assume that the reader is familiar with the finite field \mathbb{Z}_p of integers modulo a prime p (see page 20). Consider the 3-dimensional vector space X over \mathbb{Z}_p . We construct from X the following graph G_p . The vertices of G_p are the one-dimensional subspaces $[v] := \text{span}_{\mathbb{Z}_p}\{v\}$, $0 \neq v \in X$, and we connect two such subspaces $[v] \neq [w]$ by an edge if

$$\langle v, w \rangle = v_1w_1 + v_2w_2 + v_3w_3 = 0.$$

Note that it does not matter which vector $\neq 0$ we take from the subspace. In the language of geometry, the vertices are the *points* of the projective plane over \mathbb{Z}_p , and $[v]$ is adjacent to $[w]$ if w lies on the *polar line* of v .

As an example, the graph G_2 has no 4-cycle and contains 9 edges, which almost reaches the bound 10 given by (6). We want to show that this is true for any prime p .

Let us first prove that G_p satisfies the C_4 -condition. If $[u]$ is a common neighbor of $[v]$ and $[w]$, then u is a solution of the linear equations

$$\begin{aligned} v_1x + v_2y + v_3z &= 0 \\ w_1x + w_2y + w_3z &= 0. \end{aligned}$$

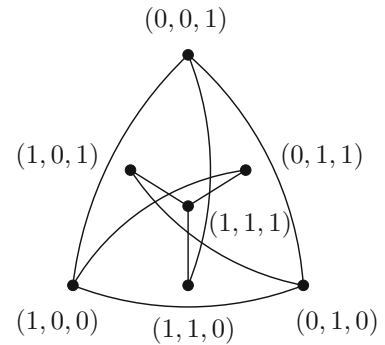
Since v and w are linearly independent, we infer that the solution space has dimension 1, and hence that the common neighbor $[u]$ is unique.

Next, we ask how many vertices G_p has. It's double counting again. The space X contains $p^3 - 1$ vectors $\neq 0$. Since every one-dimensional subspace contains $p - 1$ vectors $\neq 0$, we infer that X has $\frac{p^3-1}{p-1} = p^2 + p + 1$ one-dimensional subspaces, that is, G_p has $n = p^2 + p + 1$ vertices. Similarly, any two-dimensional subspace contains $p^2 - 1$ vectors $\neq 0$, and hence $\frac{p^2-1}{p-1} = p + 1$ one-dimensional subspaces.

It remains to determine the number of edges in G_p , or, what is the same by (4), the degrees. By the construction of G_p , the vertices adjacent to $[u]$ are the solutions of the equation

$$u_1x + u_2y + u_3z = 0. \quad (7)$$

The solution space of (7) is a two-dimensional subspace, and hence there are $p + 1$ vertices adjacent to $[u]$. But beware, it may happen that u itself is a solution of (7). In this case there are only p vertices adjacent to $[u]$.



The graph G_2 : its vertices are all seven nonzero triples (x, y, z) .

In summary, we obtain the following result: If \mathbf{u} lies on the conic given by $x^2 + y^2 + z^2 = 0$, then $d([\mathbf{u}]) = p$, and, if not, then $d([\mathbf{u}]) = p + 1$. So it remains to find the number of one-dimensional subspaces on the conic

$$x^2 + y^2 + z^2 = 0.$$

Let us anticipate the result which we shall prove in a moment.

Claim. *There are precisely p^2 solutions (x, y, z) of the equation $x^2 + y^2 + z^2 = 0$, and hence (excepting the zero solution) precisely $\frac{p^2-1}{p-1} = p + 1$ vertices in G_p of degree p .*

With this, we complete our analysis of G_p . There are $p + 1$ vertices of degree p , hence $(p^2 + p + 1) - (p + 1) = p^2$ vertices of degree $p + 1$. Using (4), we obtain

$$\begin{aligned} |E| &= \frac{(p+1)p}{2} + \frac{p^2(p+1)}{2} = \frac{(p+1)^2 p}{2} \\ &= \frac{(p+1)p}{4} (1 + (2p+1)) = \frac{p^2+p}{4} (1 + \sqrt{4p^2 + 4p + 1}). \end{aligned}$$

Setting $n = p^2 + p + 1$, the last equation reads

$$|E| = \frac{n-1}{4} (1 + \sqrt{4n-3}),$$

and we see that this almost agrees with (6).

Now to the proof of the claim. The following argument is a beautiful application of linear algebra involving symmetric matrices and their eigenvalues. We will encounter the same method in Chapter 44, which is no coincidence: both proofs are from the same paper by Erdős, Rényi and Sós.

We represent the one-dimensional subspaces of X as before by vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{p^2+p+1}$, any two of which are linearly independent. Similarly, we may represent the two-dimensional subspaces by the *same* set of vectors, where the subspace corresponding to $\mathbf{u} = (u_1, u_2, u_3)$ is the set of solutions of the equation $u_1x + u_2y + u_3z = 0$ as in (7). (Of course, this is just the duality principle of linear algebra.) Hence, by (7), a one-dimensional subspace, represented by \mathbf{v}_i , is contained in the two-dimensional subspace, represented by \mathbf{v}_j , if and only if $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$.

Consider now the matrix $A = (a_{ij})$ of size $(p^2+p+1) \times (p^2+p+1)$, defined as follows: The rows and columns of A correspond to $\mathbf{v}_1, \dots, \mathbf{v}_{p^2+p+1}$ (we use the same numbering for rows and columns) with

$$a_{ij} := \begin{cases} 1 & \text{if } \langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0, \\ 0 & \text{otherwise.} \end{cases}$$

A is thus a real symmetric matrix, and we have $a_{ii} = 1$ if $\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 0$, that is, precisely when \mathbf{v}_i lies on the conic $x^2 + y^2 + z^2 = 0$. Thus, all that remains to show is that

$$\text{trace } A = p + 1.$$

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

The matrix for G_2

From linear algebra we know that the trace equals the sum of the eigenvalues. And here comes the trick: While A looks complicated, the matrix A^2 is easy to analyze. We note two facts:

- Any row of A contains precisely $p+1$ 1's. This implies that $p+1$ is an eigenvalue of A , since $A\mathbf{1} = (p+1)\mathbf{1}$, where $\mathbf{1}$ is the vector consisting of 1's.
- For any two distinct rows v_i, v_j there is exactly one column with a 1 in both rows (the column corresponding to the unique subspace spanned by v_i, v_j).

Using these facts we find

$$A^2 = \begin{pmatrix} p+1 & 1 & \cdots & 1 \\ 1 & p+1 & & \vdots \\ \vdots & & \ddots & \\ 1 & \cdots & & p+1 \end{pmatrix} = pI + J,$$

where I is the identity matrix and J is the all-ones-matrix. Now, J has the eigenvalue $p^2 + p + 1$ (of multiplicity 1) and 0 (of multiplicity $p^2 + p$). Hence A^2 has the eigenvalues $p^2 + 2p + 1 = (p+1)^2$ of multiplicity 1 and p of multiplicity $p^2 + p$. Since A is real and symmetric, hence diagonalizable, we find that A has the eigenvalue $p+1$ or $-(p+1)$ and $p^2 + p$ eigenvalues $\pm\sqrt{p}$. From Fact 1 above, the first eigenvalue must be $p+1$. Suppose that \sqrt{p} has multiplicity r , and $-\sqrt{p}$ multiplicity s , then

$$\text{trace } A = (p+1) + r\sqrt{p} - s\sqrt{p}.$$

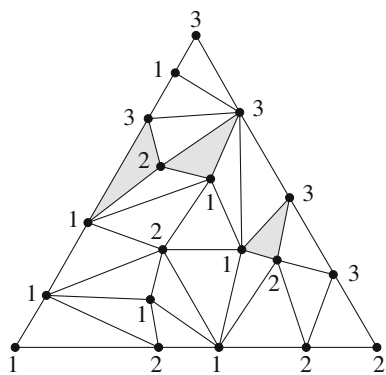
But now we are home: Since the trace is an integer, we must have $r = s$, so $\text{trace } A = p+1$. \square

6. Sperner's Lemma

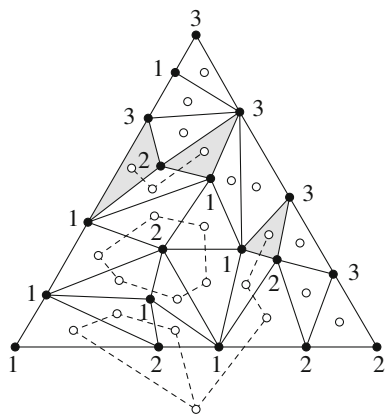
In 1912, Luitzen Brouwer published his famous fixed point theorem:

Every continuous function $f: B^n \rightarrow B^n$ of an n -dimensional ball to itself has a fixed point (a point $\mathbf{x} \in B^n$ with $f(\mathbf{x}) = \mathbf{x}$).

For dimension 1, that is for an interval, this follows easily from the intermediate value theorem, but for higher dimensions Brouwer's proof needed some sophisticated machinery. It was therefore quite a surprise when in 1928 young Emanuel Sperner (he was 23 at the time) produced a simple combinatorial result from which both Brouwer's fixed point theorem and the invariance of the dimension under continuous bijective maps could be deduced. And what's more, Sperner's ingenious lemma is matched by an equally beautiful proof — it is just double counting.



The tricolored triangles are shaded.



We discuss Sperner's lemma, and Brouwer's theorem as a consequence, for the first interesting case, that of dimension $n = 2$. The energetic reader should find it not too difficult to extend the proofs to higher dimensions (by induction on the dimension).

Sperner's Lemma.

Suppose that some "big" triangle with vertices V_1, V_2, V_3 is triangulated (that is, decomposed into a finite number of "small" triangles that fit together edge-by-edge).

Assume that the vertices in the triangulation get "colors" from the set $\{1, 2, 3\}$ such that V_i receives the color i (for each i), and only the colors i and j are used for vertices along the edge from V_i to V_j (for $i \neq j$), while the interior vertices are colored arbitrarily with 1, 2 or 3.

Then in the triangulation there must be a small "tricolored" triangle, which has all three different vertex colors.

■ **Proof.** We will prove a stronger statement: The number of tricolored triangles is not only nonzero, it is always *odd*.

Consider the dual graph to the triangulation, but don't take all its edges — only those which cross an edge that has endvertices with the (different) colors 1 and 2. Thus we get a "partial dual graph" which has degree 1 at all vertices that correspond to tricolored triangles, degree 2 for all triangles in which the two colors 1 and 2 appear, and degree 0 for triangles that do not have both colors 1 and 2. Thus only the tricolored triangles correspond to vertices of odd degree (of degree 1).

However, the vertex of the dual graph which corresponds to the outside of the triangulation has odd degree: in fact, along the big edge from V_1 to V_2 , there is an odd number of changes between 1 and 2. Thus an odd number of edges of the partial dual graph crosses this big edge, while the other big edges cannot have both 1 and 2 occurring as colors.

Now since the number of odd-degree vertices in any finite graph is even (by equation (4)), we find that the number of small triangles with three different colors (corresponding to odd inside vertices of our dual graph) is odd. \square

With this lemma, it is easy to derive Brouwer's theorem.

■ **Proof of Brouwer's fixed point theorem (for $n = 2$).** Let Δ be the triangle in \mathbb{R}^3 with vertices $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, and $e_3 = (0, 0, 1)$. It suffices to prove that every continuous map $f: \Delta \rightarrow \Delta$ has a fixed point, since Δ is homeomorphic to the two-dimensional ball B_2 .

We use $\delta(\mathcal{T})$ to denote the maximal length of an edge in a triangulation \mathcal{T} . One can easily construct an infinite sequence of triangulations $\mathcal{T}_1, \mathcal{T}_2, \dots$ of Δ such that the sequence of maximal diameters $\delta(\mathcal{T}_k)$ converges to 0. Such a sequence can be obtained by explicit construction, or inductively, for example by taking \mathcal{T}_{k+1} to be the barycentric subdivision of \mathcal{T}_k .

For each of these triangulations, we define a 3-coloring of their vertices v by setting $\lambda(v) := \min\{i : f(v)_i < v_i\}$, that is, $\lambda(v)$ is the smallest index i such that the i -th coordinate of $f(v) - v$ is negative. If this smallest index i does not exist, then we have found a fixed point and are done: To see this,

note that every $v \in \Delta$ lies in the plane $x_1 + x_2 + x_3 = 1$, hence $\sum_i v_i = 1$. So if $f(v) \neq v$, then at least one of the coordinates of $f(v) - v$ must be negative (and at least one must be positive).

Let us check that this coloring satisfies the assumptions of Sperner's lemma. First, the vertex e_i must receive color i , since the only possible negative component of $f(e_i) - e_i$ is the i -th component. Moreover, if v lies on the edge opposite to e_i , then $v_i = 0$, so the i -th component of $f(v) - v$ cannot be negative, and hence v does not get the color i .

Sperner's lemma now tells us that in each triangulation \mathcal{T}_k there is a tri-colored triangle $\{v^{k:1}, v^{k:2}, v^{k:3}\}$ with $\lambda(v^{k:i}) = i$. The sequence of points $(v^{k:1})_{k \geq 1}$ need not converge, but since the simplex Δ is compact some subsequence has a limit point. After replacing the sequence of triangulations \mathcal{T}_k by the corresponding subsequence (which for simplicity we also denote by \mathcal{T}_k) we can assume that $(v^{k:1})_k$ converges to a point $v \in \Delta$. Now the distance of $v^{k:2}$ and $v^{k:3}$ from $v^{k:1}$ is at most the mesh length $\delta(\mathcal{T}_k)$, which converges to 0. Thus the sequences $(v^{k:2})$ and $(v^{k:3})$ converge to the same point v .

But where is $f(v)$? We know that the first coordinate $f(v^{k:1})$ is smaller than that of $v^{k:1}$ for all k . Now since f is continuous, we derive that the first coordinate of $f(v)$ is smaller or equal to that of v . The same reasoning works for the second and third coordinates. Thus none of the coordinates of $f(v) - v$ is positive — and we have already seen that this contradicts the assumption $f(v) \neq v$. \square

References

- [1] L. E. J. BROUWER: *Über Abbildungen von Mannigfaltigkeiten*, Math. Annalen **71** (1912), 97-115.
- [2] W. G. BROWN: *On graphs that do not contain a Thomsen graph*, Canadian Math. Bull. **9** (1966), 281-285.
- [3] P. ERDŐS, A. RÉNYI & V. SÓS: *On a problem of graph theory*, Studia Sci. Math. Hungar. **1** (1966), 215-235.
- [4] P. ERDŐS & G. SZEKERES: *A combinatorial problem in geometry*, Compositio Math. (1935), 463-470.
- [5] S. HOŞTEN & W. D. MORRIS: *The order dimension of the complete graph*, Discrete Math. **201** (1999), 133-139.
- [6] I. REIMAN: *Über ein Problem von K. Zarankiewicz*, Acta Math. Acad. Sci. Hungar. **9** (1958), 269-273.
- [7] J. SPENCER: *Minimal scrambling sets of simple orders*, Acta Math. Acad. Sci. Hungar. **22** (1971), 349-353.
- [8] E. SPERNER: *Neuer Beweis für die Invarianz der Dimensionszahl und des Gebietes*, Abh. Math. Sem. Hamburg **6** (1928), 265-272.
- [9] W. T. TROTTER: *Combinatorics and Partially Ordered Sets: Dimension Theory*, John Hopkins University Press, Baltimore and London 1992.