CHAPTER 4

The Pigeonhole Principle

The *pigeonhole principle* (also known as *Dirichlet's principle*) states the "obvious" fact that $n+1$ pigeons cannot sit in n holes so that every pigeon is alone in its hole. More generally, the pigeonhole principle states the following:

> *If a set consisting of at least $rs+1$ objects is partitioned into r classes, then some class receives at least $s+1$ objects.*

Its truth is easy to verify: if every class receives at most s objects, then a total of at most rs objects have been distributed. To see that the result is best possible, observe that a set with at most rs points can be divided into r groups with at most s points in each group; hence none of the groups contains $s+1$ points.

This is one of the oldest "non-constructive" principles: it states only the *existence* of a pigeonhole with more than k items and says nothing about how to *find* such a pigeonhole. Today we have powerful and far reaching generalizations of this principle (Ramsey-like theorems, the probabilistic method, etc.). We will talk about them later.

As trivial as the pigeonhole principle itself may sound, it has numerous nontrivial applications. The hard part in applying this principle is to decide what to take as pigeons and what as pigeonholes. Let us illustrate this by several examples.

1. Some quickies

To "warm-up," let us start with the simplest applications. The *degree* of a vertex x in a graph G is the number $d(x)$ of edges of G adjacent to x.

Proposition 4.1. *In any graph there exist two vertices of the same degree.*

Proof. Given a graph G on n vertices, make n pigeonholes labeled from 0 up to $n-1$ and put a vertex x into the k-th pigeonhole iff $d(x)=k$. If some pigeonhole contains more than one vertex, we are done. So, assume that no pigeonhole has more than one vertex. There are n vertices going into the n pigeonholes; hence each pigeonhole has *exactly one* vertex. Let x and y be the vertices lying in the pigeonholes labeled 0 and $n-1$, respectively. The vertex x has degree 0 and so has no connection with other vertices, including y. But y has degree $n-1$ and hence, is connected with all the remaining vertices, including x, a contradiction. □

If G is a finite graph, the *independence number* $\alpha(G)$ is the maximum number of pairwise nonadjacent vertices of G. The *chromatic number* $\chi(G)$ of G is the minimum number of colors in a coloring of the vertices of G with the property that no two adjacent vertices have the same color.

Proposition 4.2. *In any graph G with n vertices, $n \leq \alpha(G)\cdot\chi(G)$.*

Proof. Consider the vertices of G partitioned into $\chi(G)$ color classes (sets of vertices with the same color). By the pigeonhole principle, one of the classes must contain at least $n/\chi(G)$ vertices, and these vertices are pairwise nonadjacent. Thus $\alpha(G)\geq n/\chi(G)$, as desired. □

A graph is *connected* if there is a path between any two of its vertices.

Proposition 4.3. *Let G be an n-vertex graph. If every vertex has a degree of at least $(n-1)/2$ then G is connected.*

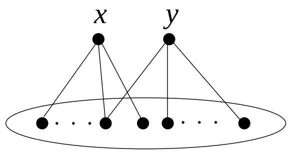


FIGURE 1. There are only $n-2$ vertices and at least $n-1$ edges going to them.

PROOF. Take any two vertices x and y. If these vertices are not adjacent, then at least $n-1$ edges join them to the remaining vertices, because both x and y have a degree of at least $(n-1)/2$. Since there are only $n-2$ other vertices, the pigeonhole principle implies that one of them must be adjacent to both x and y (see Fig. **??**). We have proved that every pair of vertices is adjacent or has a common neighbor, so G is connected. □

REMARK 4.4. A result is *best possible* if the conclusion no longer holds when we weaken one of the conditions. Such is, for example, the result above: let n be even and G be a union of two vertex disjoint complete graphs on $n/2$ vertices; then every vertex has degree $(n-2)/2$, but the graph is disconnected.

Note that, in fact, we have proved more: if every vertex of an n-vertex graph has degree at least $(n-1)/2$ then the graph has diameter at most two. The *diameter* of a graph is the smallest number k such that every two vertices are connected by a path with at most k edges.

2. The Erdős–Szekeres theorem

Let $A = (a_1, a_2, \ldots, a_n)$ be a sequence of n different numbers. A *subsequence of k terms* of A is a sequence B of k distinct terms of A appearing in the same order in which they appear in A. In symbols, we have $B = (a_{i_1}, a_{i_2}, \ldots, a_{i_k})$, where $i_1 < i_2 < \cdots < i_k$. A subsequence B is said to be *increasing* if $a_{i_1} < a_{i_2} < \cdots < a_{i_k}$, and *decreasing* if $a_{i_1} > a_{i_2} > \cdots > a_{i_k}$.

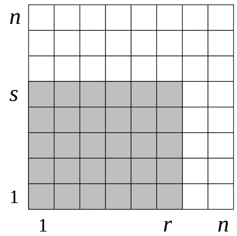
We will be interested in the length of the *longest* increasing and decreasing subsequences of A. It is intuitively plausible that there should be some kind of tradeoff between these lengths. If the longest increasing subsequence is short, say has length s, then *any* subsequence of A of length $s+1$ must contain a pair of decreasing elements, so there are lots of pairs of decreasing elements. Hence, we would expect the longest decreasing sequence to be large. An extreme case occurs when $s = 1$. Then the whole sequence A is decreasing.

How can we quantify the feeling that the length of *both*, longest increasing and longest decreasing subsequences, cannot be small? A famous result of Erdős and Szekeres (1935) gives an answer to this question and was one of the first results in extremal combinatorics.

THEOREM 4.5 (Erdős–Szekeres 1935). *Let $A = (a_1, \ldots, a_n)$ be a sequence of n different real numbers. If $n \geq sr + 1$ then either A has an increasing subsequence of $s+1$ terms or a decreasing subsequence of $r+1$ terms (or both).*

Proof (due to Seidenberg 1959). Associate to each term a_i of A a pair of "scores" (x_i, y_i) where x_i is the number of terms in the longest *increasing* subsequence *ending* at a_i, and y_i is the number of terms in the longest *decreasing* subsequence *starting* at a_i. Observe that no two terms have the same score, i.e., that $(x_i, y_i) \neq (x_j, y_j)$ whenever $i \neq j$. Indeed, if we have $\cdots a_i \cdots a_j \cdots$, then either $a_i < a_j$ and the longest increasing subsequence ending at a_i can be extended by adding on a_j (so that $x_i < x_j$), or $a_i > a_j$ and the longest decreasing subsequence starting at a_j can be preceded by a_i (so that $y_i > y_j$).

Now make a grid of n^2 pigeonholes:



Place each term a_i in the pigeonhole with coordinates (x_i, y_i). Each term of A can be placed in some pigeonhole, since $1 \leq x_i, y_i \leq n$ for all $i = 1, \ldots, n$. Moreover, no pigeonhole can have more than one term because $(x_i, y_i) \neq (x_j, y_j)$ whenever $i \neq j$. Since $|A| = n \geq sr + 1$, we have more items than the pigeonholes shaded in the above picture. So some term a_i will lie outside this shaded region. But this means that either $x_i \geq s + 1$ or $y_i \geq r + 1$ (or both), exactly what we need. □

The set of real numbers is *totally ordered*. That is, for any two distinct numbers x and y, either $x < y$ or $y < x$. The following lemma, due to Dilworth, generalizes the Erdős–Szekeres theorem to sets in which two elements may or may not be comparable.

A *partial order* on a set P is a binary relation $<$ between its elements which is transitive and irreflexive: if $x < y$ and $y < z$ then $x < z$, but $x < y$ and $y < x$ cannot both hold. We write $x \leq y$ if $x < y$ or $x = y$. Elements x and y are *comparable* if either $x \leq y$ or $y \leq x$ (or both) hold. A *chain* in a poset P is a subset $C \subseteq P$ such that any two of its points are comparable. Dually, an *antichain* is a subset $A \subseteq P$ such that no two of its points are comparable.

LEMMA 4.6 (Dilworth 1950). *In any partial order on a set P of $n \geq sr + 1$ elements, there exists a chain of length $s + 1$ or an antichain of size $r + 1$.*

PROOF. A chain is *maximal* if it cannot be prolonged by adding a new element. Let $C_1, \ldots, C_m$ be all maximal chains in P, and suppose there is no chain of length $s + 1$. Since the chains C_i must cover all n points of P, the pigeonhole principle implies that we must have $m \geq r + 1$ such chains. Let $x_i \in C_i$ be the greatest element of C_i. Then no two elements x_i and x_j with $i \neq j$ can be comparable: if $x_i \leq x_j$ then $C_i \cup \{x_j\}$ would also be a chain, a contradiction with the maximality of C_i. Thus, the elements $x_1, \ldots, x_m$ form an antichain of size $m \geq r + 1$. □

This lemma implies the Erdős–Szekeres theorem (we address this question in Exercise **??**).

3. Mantel's theorem

Here we discuss one typical *extremal* property of graphs. How many edges are possible in a triangle-free graph G on n vertices? A triangle is a set of three vertices, each two of which are connected by an edge. Certainly, G can have $n^2/4$ edges without containing a triangle: just let G be the bipartite complete graph consisting of two sets of $n/2$ vertices each and all the edges between the two sets. Indeed, $n^2/4$ turns out to be the maximum possible number of edges: if we take one more edge then the graph will have a triangle.

We give four proofs of this beautiful result: the first (original) proof is based on double counting, the second uses the inequality $\sqrt{ab} \leq (a + b)/2$ of the arithmetic and geometric mean, the third uses the pigeonhole principle, and the fourth employs the so-called "shifting argument" (we will give this last proof in the Sect. **??** devoted to this argument).

THEOREM 4.7 (Mantel 1907). *If a graph G on n vertices contains more than $n^2/4$ edges, then G contains a triangle.*

First proof. Let G be a graph on a set V of n vertices containing $m > n^2/4$ edges. Assume that G has no triangles. Then adjacent vertices have no common neighbors, so $d(x) + d(y) \leq n$ for each edge $\{x, y\} \in E$. Summing over all edges of G, we have (cf. Equation (**??**))

$$\sum_{x \in V} d(x)^2 = \sum_{\{x,y\} \in E} (d(x) + d(y)) \leq mn\,.$$

On the other hand, using Cauchy–Schwarz inequality (see Notation or Proposition **??**) and Euler's equality $\sum_{x\in V} d(x) = 2m$ (see Theorem **??**), we obtain

$$\sum_{x\in V} d(x)^2 \geq \frac{\left(\sum_{x\in V} d(x)\right)^2}{|V|} = \frac{4m^2}{n}\,.$$

These two inequalities imply that $m \leq n^2/4$, contradicting the hypothesis. □

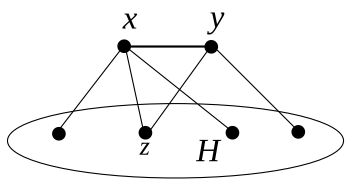
Second proof. Let $G = (V, E)$ be a graph on a set V of n vertices and assume that G has no triangles. Let $A \subseteq V$ be the largest independent set, i.e., a maximal set of vertices, no two of which are adjacent in G. Since G is triangle-free, the neighbors of a vertex $x \in V$ form an independent set, and we infer $d(x) \leq |A|$ for all x.

The set $B = V \setminus A$ meets every edge of G. Counting the edges of G according to their end-vertices in B, we obtain $|E| \leq \sum_{x\in B} d(x)$. The inequality of the arithmetic and geometric mean (**??**) yields

$$|E| \leq \sum_{x\in B} d(x) \leq |A|\cdot|B| \leq \left(\frac{|A|+|B|}{2}\right)^2 = \frac{n^2}{4}\,.$$

□

Third proof. To avoid ceilings and floorings, we will prove the theorem for graphs on an *even* number $2n$ of vertices. We want to prove that every such graph with at least n^2+1 edges must contain a triangle. We argue by induction on n. If $n = 1$, then G cannot have n^2+1 edges; hence the statement is true. Assuming the result for n, we now consider a graph G on $2(n+1)$ vertices with $(n+1)^2+1$ edges. Let x and y be adjacent vertices in G, and let H be the induced subgraph on the remaining $2n$ vertices. If H contains at least n^2+1 edges then we are done by the induction hypothesis. Suppose that H has at most n^2 edges, and therefore at least $2n+1$ edges of G emanate from x and y to vertices in H:



By the pigeonhole principle, among these $2n+1$ edges there must be an edge from x and an edge from y to the same vertex z in H. Hence G contains the triangle $\{x, y, z\}$. □

4. Turán's theorem

A k-*clique* is a graph on k vertices, every two of which are connected by an edge. For example, triangles are 3-cliques. Mantel's theorem says that, if a graph on n vertices has no 3-clique then it has at most $n^2/4$ edges. What about $k > 3$?

The answer is given by a fundamental result of Paul Turán, which initiated extremal graph theory.

THEOREM 4.8 (Turán 1941). *If a graph $G = (V, E)$ on n vertices has no $(k+1)$-clique, $k \geq 2$, then*

$$|E| \leq \left(1 - \frac{1}{k}\right)\frac{n^2}{2}. \tag{36}$$

Like Mantel's theorem, this result was rediscovered many times with various different proofs. Here we present the original one due to Turán. The proof based on so-called "weight shifting" argument is addressed in Exercise **??**. In Sect. **??** we will give a proof which employs ideas of a totally different nature – the probabilistic argument.

PROOF. We use induction on n. Inequality (**??**) is trivially true for $n = 1$. The case $k = 2$ is Mantel's theorem. Suppose now that the inequality is true for all graphs on at most $n-1$ vertices, and let $G = (V, E)$ be a graph on n vertices without $(k+1)$-cliques and with a maximal number of edges. This graph certainly contains k-cliques, since otherwise we could add edges. Let A be a k-clique, and set $B = V \setminus A$.

Since each two vertices of A are joined by an edge, A contains $e_A = \binom{k}{2}$ edges. Let e_B be the number of edges joining the vertices of B and $e_{A,B}$ the number of edges between A and B. By induction, we have

$$e_B \leq \left(1 - \frac{1}{k}\right) \frac{(n-k)^2}{2}.$$

Since G has no $(k+1)$-clique, every $x \in B$ is adjacent to at most $k-1$ vertices in A, and we obtain

$$e_{A,B} \leq (k-1)(n-k).$$

Summing up and using the identity

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

we conclude that

$$\begin{aligned} |E| &\leq e_A + e_B + e_{A,B} \leq \binom{k}{2} + \binom{k}{2} \left(\frac{n-k}{k}\right)^2 + (k-1)(n-k) \\ &= \binom{k}{2} \left(1 + \frac{n-k}{k}\right)^2 = \left(1 - \frac{1}{k}\right) \frac{n^2}{2}. \quad \square \end{aligned}$$

An n-vertex graph $T(n,k)$ that does not contain any $(k+1)$-clique may be formed by partitioning the set of vertices into k parts of equal or nearly-equal size, and connecting two vertices by an edge whenever they belong to two different parts. Thus, Turán's theorem states that the graph $T(n,k)$ has the largest number of edges among all n-vertex graphs without $(k+1)$-cliques.

CHAPTER 5

Systems of Distinct Representatives

A *system of distinct representatives* for a sequence of (not necessarily distinct) sets $S_1, S_2, \ldots, S_m$ is a sequence of distinct elements $x_1, x_2, \ldots, x_m$ such that $x_i \in S_i$ for all $i = 1, 2, \ldots, m$.

When does such a system exist? This problem is called the "marriage problem" because an easy reformulation of it asks whether we can marry each of m girls to a boy she knows; boys are the elements and S_i is the set of boys known to the i-th girl.

Clearly, if the sets $S_1, S_2, \ldots, S_m$ have a system of distinct representatives then the following *Hall's Condition* is fulfilled:

$(*)$ for every $k = 1, 2, \ldots, m$ the union of any k sets has at least k elements:

$$\left| \bigcup_{i \in I} S_i \right| \geq |I| \quad \text{for all } I \subseteq \{1, \ldots, m\}.$$

Surprisingly, this obvious necessary condition is also sufficient.

1. The marriage theorem

The following fundamental result is known as *Hall's marriage theorem* (Hall 1935), though an equivalent form of it was discovered earlier by König (1931) and Egerváry (1931), and the result is also a special case of Menger's theorem (1927). The case when we have the same number of girls as boys was proved by Frobenius (1917).

THEOREM 5.1 (**Hall's Theorem**). *The sets $S_1, S_2, \ldots, S_m$ have a system of distinct representatives if and only if $(*)$ holds.*

PROOF. We prove the sufficiency of Hall's condition $(*)$ by induction on m. The case $m = 1$ is clear. Assume that the claim holds for any collection with less than m sets.

Case 1: For each k, $1 \leq k < m$, the union of any k sets contains more than k elements.

Take any of the sets, and choose any of its elements x as its representative, and remove x from all the other sets. The union of any $s \leq m - 1$ of the remaining $m - 1$ sets has at least s elements, and therefore the remaining sets have a system of distinct representatives, which together with x give a system of distinct representatives for the original family.

Case 2: The union of some k, $1 \leq k < m$, sets contains exactly k elements.

By the induction hypothesis, these k sets have a system of distinct representatives. Remove these k elements from the remaining $m - k$ sets. Take any s of these sets. Their union contains at least s elements, since otherwise the union of these s sets and the k sets would have less than $s + k$ elements. Consequently, the remaining $m - k$ sets also have a system of distinct representatives by the induction hypothesis. Together these two systems of distinct representatives give a system of distinct representatives for the original family. $\square$

In general, Hall's condition $(*)$ is hard to verify: we must check if the union of *any* k, $1 \leq k \leq m$, of the sets $S_1, \ldots, S_m$ contains at least k elements. But if we know more about these sets, then (sometimes) the situation is much better. Here is an example.

COROLLARY 5.2. *Let $S_1, \ldots, S_m$ be r-element subsets of an n-element set such that each element belongs to the same number d of these sets. If $m \leq n$, then the sets $S_1, \ldots, S_m$ have a system of distinct representatives.*

1	5	2	4	?
				3

FIGURE 1. A partial 2×5 Latin square that cannot be completed

PROOF. By the double counting argument (**??**), $mr = nd$, and hence, $m \leq n$ implies that $d \leq r$. Now suppose that $S_1, \dots, S_m$ does not have a system of distinct representatives. By Hall's theorem, the union $Y = S_{i_1} \cup \cdots \cup S_{i_k}$ of some k ($1 \leq k \leq m$) sets contains strictly less than k elements. For $x \in Y$, let d_x be the number of these sets containing x. Then, again, using (**??**), we obtain

$$rk = \sum_{j=1}^{k} |S_{i_j}| = \sum_{x \in Y} d_x \leq d|Y| < dk,$$

a contradiction with $d \leq r$. □

Hall's theorem was generalized in different ways. Suppose, for example, that each of the elements of the underlying set is colored either in red or in blue. Interpret red points as "bad" points. Given a system of subsets of this (colored) set, we would like to come up with a system of distinct representatives which has as few bad elements as possible.

THEOREM 5.3 (Chvátal–Szemerédi 1988). *The sets $S_1, \dots, S_m$ have a system of distinct representatives with at most t red elements if and only if they have a system of distinct representatives and for every $k = 1, 2, \dots, m$ the union of any k sets has at least $k - t$ blue elements.*

PROOF. The "only if" part is obvious. To prove the "if" part, let R be the set of red elements. We may assume that $|R| > t$ (otherwise the conclusion is trivial). Now enlarge $S_1, \dots, S_m$ to $S_1, \dots, S_m, S_{m+1}, \dots, S_{m+r}$ by adding $r = |R| - t$ copies of the set R. Observe that the sequence $S_1, \dots, S_m$ has a system of distinct representatives with at most t red elements if and only if the extended sequence has a system of distinct representatives (without any restriction). Hence, Hall's theorem reduces our task to proving that the extended sequence fulfills Hall's condition $(*)$, i.e., that for any set of indices $I \subseteq \{1, \dots, m+r\}$, the union $Y = \bigcup_{i \in I} S_i$ contains at least $|I|$ elements. Let $J = I \cap \{1, \dots, m\}$. If $J = I$ then, by the first assumption, the sets S_i ($i \in I$) have a system of distinct representatives, and hence, $|Y| \geq |I|$. Otherwise, by the second assumption,

$$\begin{aligned} |Y| &= \left| \bigcup_{i \in J} (S_i \setminus R) \right| + |R| \geq (|J| - t) + |R| \\ &= |J| + (|R| - t) \geq |J| + |I \setminus J| = |I|; \end{aligned}$$

hence $(*)$ holds again. □

2. Two applications

In this section we present two applications of Hall's theorem to prove results whose statement does not seem to be related at all to set systems and their representatives.

2.1. Latin rectangles. An $r \times n$ *Latin rectangle* is an $r \times n$ matrix with entries in $\{1, \dots, n\}$ such that each of the numbers $1, 2, \dots, n$ occurs once in each row and at most once in each column. A *Latin square* is a Latin $r \times n$-rectangle with $r = n$. This is one of the oldest combinatorial objects, whose study goes back to ancient times.

Suppose somebody gives us an $n \times n$ matrix, *some* of whose entries are filled with the numbers from $\{1, \dots, n\}$ so that no number occurs more than once in a row or column. Our goal is to fill the remaining entries so that to get a Latin square. When is this possible? Of course, the fewer entries are filled, the more chances we have to complete the matrix. Fig. **??** shows that, in general, it is possible to fill n entries so that the resulting partial matrix cannot be completed.

In 1960, Trevor Evans raised the following question: if fewer than n entries in an $n \times n$ matrix are filled, can one then always complete it to obtain a Latin square? The assertion that a completion is always possible became known as the Evans conjecture, and was proved by Smetaniuk (1981) using a quite subtle induction argument.

On the other hand, it was long known that if a partial Latin square has no partially filled rows (that is, each row is either completely filled or completely free) then it can always be completed. That is, we can build Latin squares by adding rows one-by-one. And this can be easily derived from Hall's theorem.

THEOREM 5.4 (Ryser 1951). *If $r < n$, then any given $r \times n$ Latin rectangle can be extended to an $(r+1) \times n$ Latin rectangle.*

PROOF. Let R be an $r \times n$ Latin rectangle. For $j = 1, \ldots, n$, define S_j to be the set of those integers $1, 2, \ldots, n$ which do *not* occur in the j-th column of R. It is sufficient to prove that the sets $S_1, \ldots, S_n$ have a system of distinct representatives. But this follows immediately from Corollary **??**, because: every set S_j has precisely $n - r$ elements, and each element belongs to precisely $n - r$ sets S_j (since it appears in precisely r columns of the rectangle R). □

2.2. Decomposition of doubly stochastic matrices. Using Hall's theorem we can obtain a basic result of polyhedral combinatorics, due to Birkhoff (1949) and von Neumann (1953).

An $n \times n$ matrix $A = \{a_{ij}\}$ with real non-negative entries $a_{ij} \geq 0$ is *doubly stochastic* if the sum of entries along any row and any column equals 1. A *permutation matrix* is a doubly stochastic matrix with entries 0 and 1; such a matrix has exactly one 1 in each row and in each column. Doubly stochastic matrices arise in the theory of Markov chains: a_{ij} is the transition probability from the state i to the state j. A matrix A is a convex combination of matrices $A_1, \ldots, A_s$ if there exist non-negative reals $\lambda_1, \ldots, \lambda_s$ such that $A = \sum_{i=1}^{s} \lambda_i A_i$ and $\sum_{i=1}^{s} \lambda_i = 1$.

BIRKHOFF–VON NEUMANN THEOREM. *Every doubly stochastic matrix is a convex combination of permutation matrices.*

PROOF. We will prove a more general result that every $n \times n$ non-negative matrix $A = (a_{ij})$ having all row and column sums equal to some positive value $\gamma > 0$ can be expressed as a linear combination $A = \sum_{i=1}^{s} \lambda_i P_i$ of permutation matrices $P_1, \ldots, P_s$, where $\lambda_1, \ldots, \lambda_s$ are non-negative reals such that $\sum_{i=1}^{s} \lambda_i = \gamma$.

To prove this, we apply induction on the number of non-zero entries in A. Since $\gamma > 0$, we have at least n such entries. If there are exactly n non-zero entries then $A = \gamma P$ for some permutation matrix P, and we are done. Now suppose that A has more than n non-zero entries and that the result holds for matrices with a smaller number of such entries. Define

$$S_i = \{j : a_{ij} > 0\}, \; i = 1, 2, \ldots, n,$$

and observe that the sets $S_1, \ldots, S_n$ fulfill Hall's condition. Indeed, if the union of some k $(1 \leq k \leq n)$ of these sets contained less than k elements, then all the non-zero entries of the corresponding k rows of A would occupy no more than $k-1$ columns; hence, the sum of these entries by columns would be at most $(k-1)\gamma$, whereas the sum by rows is $k\gamma$, a contradiction.

By Hall's theorem, there is a system of distinct representatives

$$j_1 \in S_1, \ldots, j_n \in S_n.$$

Take the permutation matrix $P_1 = \{p_{ij}\}$ with entries $p_{ij} = 1$ if and only if $j = j_i$. Let $\lambda_1 = \min\{a_{1j_1}, \ldots, a_{nj_n}\}$, and consider the matrix $A_1 = A - \lambda_1 P_1$. By the definition of the sets S_i, $\lambda_1 > 0$. So, this new matrix A_1 has less non-zero entries than A. Moreover, the matrix A_1 satisfies the condition of the theorem with $\gamma_1 = \gamma - \lambda_1$. We can therefore apply the induction hypothesis to A_1, which yields a decomposition $A_1 = \lambda_2 P_2 + \cdots + \lambda_s P_s$, and hence, $A = \lambda_1 P_1 + A_1 = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_s P_s$, as desired. □

3. Min–max theorems

The early results of Frobenius and König have given rise to a large number of *min-max theorems* in combinatorics, in which the minimum of one quantity equals the maximum of another. Celebrated among these are:

- *Menger's theorem* (Menger 1927): the minimum number of vertices separating two given vertices in a graph is equal to the maximum number of vertex-disjoint paths between them;
- *König–Egerváry's min-max theorem* (König 1931, Egerváry 1931): the size of a largest matching in a bipartite graph is equal to the smallest set of vertices which together touch every edge;
- *Dilworth's theorem* for partially ordered sets (Dilworth 1950): the minimum number of *chains* (totally ordered sets) which cover a partially ordered set is equal to the maximum size of an antichain (set of incomparable elements).

Here we present the proof of König–Egerváry's theorem (stated not for bipartite graphs but for their adjacency matrices); the proof of Dilworth's theorem is given in Sect. **??**.

By Hall's theorem, we know whether each of the girls can be married to a boy she knows. If so, all are happy (except for the boys not chosen ...). But what if not? In this sad situation it would be nice to make as many happy marriages as possible. So, given a sequence of sets $S_1, S_2, \dots, S_m$, we try to find a system of distinct representatives for as many of these sets as possible. In terms of 0-1 matrices this problem is solved by the following result.

Let A be an $m \times n$ matrix, all whose entries have value 0 or 1. Two 1s are *dependent* if they are on the same row or on the same column; otherwise, they are *independent*. The size of the largest set of independent 1s is also known as the *term rank* of A.

Theorem 5.5 (König 1931, Egerváry 1931). *Let A be an $m \times n$ 0-1 matrix. The maximum number of independent 1s is equal to the minimum number of rows and columns required to cover all the 1s in A.*

Proof. Let r denote the maximum number of independent 1s and R the minimum number of rows and columns required to cover all the 1s. Clearly, $R \geq r$, because we can find r independent 1s in A, and any row or column covers at most one of them.

We need to prove that $r \geq R$. Assume that some a rows and b columns cover all the 1s and $a + b = R$. Because permuting the rows and columns changes neither r nor R, we may assume that the first a rows and the first b columns cover the 1s. Write A in the form

$$A = \begin{pmatrix} B_{a\times b} & C_{a\times(n-b)} \\ D_{(m-a)\times b} & E_{(m-a)\times(n-b)} \end{pmatrix}.$$

We know that there are no 1s in E. We will show that there are a independent 1s in C. The same argument shows – by symmetry – that there are b independent 1s in D. Since altogether these $a + b$ 1s are independent, this shows that $r \geq a + b = R$, as desired.

We use Hall's theorem. Define

$$S_i = \{j : c_{ij} = 1\} \subseteq \{1, 2, \dots, n - b\},$$

as the set of locations of the 1s in the i-th row of $C = (c_{ij})$. We claim that the sequence $S_1, S_2, \dots, S_a$ has a system of distinct representatives, i.e., we can choose a 1 from each row, no two in the same column. Otherwise, Hall's theorem tells us that the 1s in some k $(1 \leq k \leq a)$ of these rows can all be covered by less than k columns. But then we obtain a covering of all the 1s in A with fewer than $a + b$ rows and columns, a contradiction. $\square$