On the history of combinatorial optimization (till 1960)

Alexander Schrijver[1]

1. Introduction

As a coherent mathematical discipline, combinatorial optimization is relatively young. When studying the history of the field, one observes a number of independent lines of research, separately considering problems like optimum assignment, shortest spanning tree, transportation, and the traveling salesman problem. Only in the 1950's, when the unifying tool of linear and integer programming became available and the area of operations research got intensive attention, these problems were put into one framework, and relations between them were laid.

Indeed, linear programming forms the hinge in the history of combinatorial optimization. Its initial conception by Kantorovich and Koopmans was motivated by combinatorial applications, in particular in transportation and transshipment. After the formulation of linear programming as generic problem, and the development in 1947 by Dantzig of the simplex method as a tool, one has tried to attack about all combinatorial optimization problems with linear programming techniques, quite often very successfully.

A cause of the diversity of roots of combinatorial optimization is that several of its problems descend directly from practice, and instances of them were, and still are, attacked daily. One can imagine that even in very primitive (even animal) societies, finding short paths and searching (for instance, for food) is essential. A traveling salesman problem crops up when you plan shopping or sightseeing, or when a doctor or mailman plans his tour. Similarly, assigning jobs to men, transporting goods, and making connections, form elementary problems not just considered by the mathematician.

It makes that these problems probably can be traced back far in history. In this survey however we restrict ourselves to the mathematical study of these problems. At the other end of the time scale, we do not pass 1960, to keep size in hand. As a consequence, later important developments, like Edmonds' work on matchings and matroids and Cook and Karp's theory of complexity (NP-completeness) fall out of the scope of this survey.

We focus on six problem areas, in this order: assignment, transportation, maximum flow, shortest tree, shortest path, and the traveling salesman problem.

2. The assignment problem

In mathematical terms, the assignment problem is: given an $n \times n$ 'cost' matrix $C = (c_{i,j})$, find a permutation π of $1, \ldots, n$ for which

[1]CWI and University of Amsterdam. *Mailing address:* CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands.

$$(1) \qquad \sum_{i=1}^{n} c_{i,\pi(i)}$$

is as small as possible.

Monge 1784

The assignment problem is one of the first studied combinatorial optimization problems. It was investigated by G. Monge [1784], albeit camouflaged as a continuous problem, and often called a transportation problem.

Monge was motivated by transporting earth, which he considered as the discontinuous, combinatorial problem of transporting molecules. There are two areas of equal acreage, one filled with earth, the other empty. The question is to move the earth from the first area to the second, in such a way that the total transportation distance is as small as possible. The total transportation distance is the distance over which a molecule is moved, summed over all molecules. Hence it is an instance of the assignment problem, obviously with an enormous cost matrix. Monge described the problem as follows:

> Lorsqu'on doit transporter des terres d'un lieu dans un autre, on a coutime de donner le nom de *Déblai* au volume des terres que l'on doit transporter, & le nom de *Remblai* à l'espace qu'elles doivent occuper après le transport.
>
> Le prix du transport d'une molécule étant, toutes choses d'ailleurs égales, proportionnel à son poids & à l'espace qu'on lui fait parcourir, & par conséquent le prix du transport total devant être proportionnel à la somme des produits des molécules multipliées chacune par l'espace parcouru, il s'ensuit que le déblai & le remblai étant donnés de figure & de position, il n'est pas indifférent que telle molécule du déblai soit transportée dans tel ou tel autre endroit du remblai, mais qu'il y a une certaine distribution à faire des molécules du premier dans le second, d'après laquelle la somme de ces produits sera la moindre possible, & le prix du transport total sera un *minimum.*[2]

Monge gave an interesting geometric method to solve this problem. Consider a line that is tangent to both areas, and move the molecule m touched in the first area to the position x touched in the second area, and repeat, till all earth has been transported. Monge's argument that this would be optimum is simple: if molecule m would be moved to another position, then another molecule should be moved to position x, implying that the two routes traversed by these molecules cross, and that therefore a shorter assignment exists:

> Étant données sur un même plan deux aires égales $ABCD$, & $abcd$, terminées par des contours quelconques, continus ou discontinus, trouver la route que doit suivre chaque molécule M

[2] When one must transport earth from one place to another, one usually gives the name of *Déblai* to the volume of earth that one must transport, & the name of *Remblai* to the space that they should occupy after the transport.

The price of the transport of one molecule being, if all the rest is equal, proportional to its weight & to the distance that one makes it covering, & hence the price of the total transport having to be proportional to the sum of the products of the molecules each multiplied by the distance covered, it follows that, the déblai & the remblai being given by figure and position, it makes difference if a certain molecule of the déblai is transported to one or to another place of the remblai, but that there is a certain distribution to make of the molecules from the first to the second, after which the sum of these products will be as little as possible, & the price of the total transport will be a *minimum.*

de la premiere, & le point m où elle doit arriver dans la seconde, pour que tous les points étant semblablement transportés, ils replissent exactement la seconde aire, & que la somme des produits de chaque molécule multipliée par l'espace parcouru soit un *minimum.*

Si par un point M quelconque de la première aire, on mène une droite Bd, telle que le segment BAD soit égal au segment bad, je dis que pour satisfaire à la question, il faut que toutes les molécules du segment BAD, soient portées sur le segment bad, & que par conséquent les molécules du segment BCD soient portées sur le segment égal bcd; car si un point K quelconque du segment BAD, étoit porté sur un point k de bcd, il faudroit nécessairement qu'un point égal L, pris quelque part dans BCD, fût transporté dans un certain point l de bad, ce qui ne pourroit pas se faire sans que les routes Kk, Ll, ne se coupassent entre leurs extrémités, & la somme des produits des molécules par les espaces parcourus ne seroit pas un *minimum.* Pareillement, si par un point M' infiniment proche du point M, on mène la droite $B'd'$, telle qu'on ait encore le segment $B'A'D'$, égal au segment $b'a'd'$, il faut pour que la question soit satisfaite, que les molécules du segment $B'A'D'$ soient transportées sur $b'a'd'$. Donc toutes les molécules de l'élément $BB'D'D$ doivent être transportées sur l'élément égal $bb'd'd$. Ainsi en divisant le déblai & le remblai en une infinité d'élémens par des droites qui coupent dans l'un & dans l'autre des segmens égaux entr'eux, chaque élément du déblai doit être porté sur l'élément correspondant du remblai.

Les droites Bd & $B'd'$ étant infiniment proches, il est indifférent dans quel ordre les molécules de l'élément $BB'D'D$ se distribuent sur l'élément $bb'd'd$; de quelque manière en effet que se fasse cette distribution, la somme des produits des molécules par les espaces parcourus, est toujours la même, mais si l'on remarque que dans la pratique il convient de débleyer premièrement les parties qui se trouvent sur le passage des autres, & de n'occuper que les dernières les parties du remblai qui sont dans le même cas; la molécule MM' ne devra se transporter que lorsque toute la partie $MM'D'D$ qui la précêde, aura été transportée en $mm'd'd$; donc dans cette hypothèse, si l'on fait $mm'd'd = MM'D'D$, le point m sera celui sur lequel le point M sera transporté.[3]

Although geometrically intuitive, the method is however not fully correct, as was noted by Appell [1928]:

[3]Being given, in the same plane, two equal areas $ABCD$ & $abcd$, bounded by arbitrary contours, continuous or discontinuous, find the route that every molecule M of the first should follow & the point m where it should arrive in the second, so that, all points being transported likewise, they fill precisely the second area & so that the sum of the products of each molecule multiplied by the distance covered, is *minimum.*

If one draws a straight line Bd through an arbitrary point M of the first area, such that the segment BAD is equal to the segment bad, I assert that, in order to satisfy the question, all molecules of the segment BAD should be carried on the segment bad, & hence the molecules of the segment BCD should be carried on the equal segment bcd; for, if an arbitrary point K of segment BAD, is carried to a point k of bcd, then necessarily some point L somewhere in BCD is transported to a certain point l in bad, which cannot be done without that the routes Kk, Ll cross each other between their end points, & the sum of the products of the molecules by the distances covered would not be a *minimum.* Likewise, if one draws a straight line $B'd'$ through a point M' infinitely close to point M, in such a way that one still has that segment $B'A'D'$ is equal to segment $b'a'd'$, then in order to satisfy the question, the molecules of segment $B'A'D'$ should be transported to $b'a'd'$. So all molecules of the element $BB'D'D$ must be transported to the equal element $bb'd'd$. Dividing the déblai & the remblai in this way into an infinity of elements by straight lines that cut in the one & in the other segments that are equal to each other, every element of the déblai must be carried to the corresponding element of the remblai.

The straight lines Bd & $B'd'$ being infinitely close, it does not matter in which order the molecules of element $BB'D'D$ are distributed on the element $bb'd'd$; indeed, in whatever manner this distribution is being made, the sum of the products of the molecules by the distances covered is always the same; but if one observes that in practice it is convenient first to dig off the parts that are in the way of others, & only at last to cover similar parts of the remblai; the molecule MM' must be transported only when the whole part $MM'D'D$ that precedes it will have been transported to $mm'd'd$; hence with this hypothesis, if one has $mm'd'd = MM'D'D$, point m will be the one to which point M will be transported.

> Il est bien facile de faire la figure de manière que les chemins suivis par les deux parcelles dont parle Monge ne se croisent pas.[4]

(cf. Taton [1951]).

Bipartite matching: Frobenius 1912-1917, Kőnig 1915-1931

Finding a largest matching in a bipartite graph can be considered as a special case of the assignment problem. The fundaments of matching theory in bipartite graphs were laid by Frobenius (in terms of matrices and determinants) and Kőnig. We briefly review their work.

In his article *Über Matrizen aus nicht negativen Elementen*, Frobenius [1912] investigated the decomposition of matrices, which led him to the following 'curious determinant theorem':

> *Die Elemente einer Determinante nten Grades seien n^2 unabhängige Veränderliche. Man setze einige derselben Null, doch so, daß die Determinante nicht identisch verschwindet. Dann bleibt sie eine irreduzible Funktion, außer wenn für einen Wert $m < n$ alle Elemente verschwinden, die m Zeilen mit $n - m$ Spalten gemeinsam haben.*[5]

Frobenius gave a combinatorial and an algebraic proof.

In a reaction to this, Dénes Kőnig [1915] realized that Frobenius' theorem can be equivalently formulated in terms of bipartite graphs, by introducing a now quite standard construction of associating a bipartite graph with a matrix $(a_{i,j})$: for each row index i there is a vertex v_i and for each column index j there is a vertex u_j, while vertices v_i and u_j are adjacent if and only if $a_{i,j} \neq 0$. With the help of this, Kőnig gave a proof of Frobenius' result.

According to Gallai [1978], Kőnig was interested in graphs, particularly bipartite graphs, because of his interest in set theory, especially cardinal numbers. In proving Schröder-Bernstein type results on the equicardinality of sets, graph-theoretic arguments (in particular: matchings) can be illustrative. This led Kőnig to studying graphs and its applications in other areas of mathematics.

On 7 April 1914, Kőnig had presented at the *Congrès de Philosophie mathématique* in Paris (cf. Kőnig [1916,1923]) the theorem that each regular bipartite graph has a perfect matching. As a corollary, Kőnig derived that the edge set of any regular bipartite graph can be decomposed into perfect matchings. That is, each k-regular bipartite graph is k-edge-colourable. Kőnig observed that these results follow from the theorem that the edge-colouring number of a bipartite graph is equal to its maximum degree. He gave an algorithmic proof of this.

In order to give an elementary proof of his result described above, Frobenius [1917] proves the following 'Hilfssatz', which now is a fundamental theorem in graph theory:

> II. *Wenn in einer Determinante nten Grades alle Elemente verschwinden, welche p $(\leq n)$ Zeilen mit $n - p + 1$ Spalten gemeinsam haben, so verschwinden alle Glieder der entwickelten Determinante.*

---

[4] It is very easy to make the figure in such a way that the routes followed by the two particles of which Monge speaks, do not cross each other.

[5] *Let the elements of a determinant of degree n be n^2 independent variables. One sets some of them equal to zero, but such that the determinant does not vanish identically. Then it remains an irreducible function, except when for some value $m < n$ all elements vanish that have m rows in common with $n - m$ columns.*

> *Wenn alle Glieder einer Determinante nten Grades verschwinden, so verschwinden alle Elemente, welche p Zeilen mit $n - p + 1$ Spalten gemeinsam haben für $p = 1$ oder $2, \cdots$ oder n.*[6]

That is, if $A = (a_{i,j})$ is an $n \times n$ matrix, and for each permutation π of $\{1, \ldots, n\}$ one has $\prod_{i=1}^{n} a_{i,j} = 0$, then for some p there exist p rows and $n-p+1$ columns of A such that their intersection is all-zero.

In other words, a bipartite graph $G = (V, E)$ with colour classes V_1 and V_2 satisfying $|V_1| = |V_2| = n$ has a perfect matching, if and only if one cannot select p vertices in V_1 and $n - p + 1$ vertices in V_2 such that no edge is connecting two of these vertices.

Frobenius gave a short combinatorial proof (albeit in terms of determinants), and he stated that Kőnig's results follow easily from it. Frobenius also offered his opinion on Kőnig's proof method of his 1912 theorem:

> Die Theorie der Graphen, mittels deren Hr. KŐNIG den obigen Satz abgeleitet hat, ist nach meiner Ansicht ein wenig geeignetes Hilfsmittel für die Entwicklung der Determinantentheorie. In diesem Falle führt sie zu einem ganz speziellen Satze von geringem Werte. Was von seinem Inhalt Wert hat, ist in dem Satze II ausgesprochen.[7]

While Frobenius' result characterizes which bipartite graphs have a perfect matching, a more general theorem characterizing the maximum size of a matching in a bipartite graph was found by Kőnig [1931]:

> Páros körüljárású graphban az éleket kimerítő szögpontok minimális száma megegyezik a páronként közös végpontot nem tartalmazó élek maximális számával.[8]

In other words, the maximum size of a matching in a bipartite graph is equal to the minimum number of vertices needed to cover all edges.

This result can be derived from that of Frobenius [1917], and also from the theorem of Menger [1927] — but, as Kőnig detected, Menger's proof contains an essential hole in the induction basis — see Section 4. This induction basis is precisely the theorem proved by Kőnig.

Egerváry 1931

After the presentation by Kőnig of his theorem at the Budapest Mathematical and Physical Society on 26 March 1931, E. Egerváry [1931] found a weighted version of Kőnig's theorem. It characterizes the maximum *weight* of a matching in a bipartite graph, and thus applies to the assignment problem:

---

[6]II. *If in a determinant of the nth degree all elements vanish that $p(\leq n)$ rows have in common with $n - p + 1$ columns, then all members of the expanded determinant vanish.*

If all members of a determinant of degree n vanish, then all elements vanish that p rows have in common with $n - p + 1$ columns for $p = 1$ or $2, \cdots$ or n.

[7]The theory of graphs, by which Mr KŐNIG has derived the theorem above, is to my opinion of little appropriate help for the development of determinant theory. In this case it leads to a very special theorem of little value. What from its contents has value, is enunciated in Theorem II.

[8]In an even circuit graph, the minimal number of vertices that exhaust the edges agrees with the maximal number of edges that pairwise do not contain any common end point.

Ha az $\|a_{ij}\|$ n-edrendű matrix elemei adott nem negatív egész számok, úgy a

$$\lambda_i + \mu_j \geq a_{ij}, \qquad (i, j = 1, 2, ...n),$$

$(\lambda_i, \mu_j$ nem negatív egész számok)

feltételek mellett

$$\min . \sum_{k=1}^{n} (\lambda_k + \mu_k) = \max .(a_{1\nu_1} + a_{2\nu_2} + \cdots + a_{n\nu_n}).$$

hol $\nu_1, \nu_2, ...\nu_n$ az $1, 2, ...n$ számok összes permutációit befutják.[9]

The proof method of Egerváry is essentially algorithmic. Assume that the $a_{i,j}$ are integer. Let λ_i^*, μ_j^* attain the minimum. If there is a permutation ν of $\{1, \ldots, n\}$ such that $\lambda_i^* + \mu_{\nu_i}^* = a_{i,\nu_i}$ for all i, then this permutation attains the maximum, and we have the required equality. If no such permutation exists, by Frobenius' theorem there are subsets I, J of $\{1, \ldots, n\}$ such that

(2) $\qquad \lambda_i^* + \mu_j^* > a_{i,j}$ for all $i \in I$, $j \in J$

and such that $|I| + |J| = n + 1$. Resetting $\lambda_i^* := \lambda_i^* - 1$ if $i \in I$ and $\mu_j^* := \mu_j^* + 1$ if $j \notin J$, would give again feasible values for the λ_i and μ_j, however with their total sum being decreased. This is a contradiction.

Egerváry's theorem and proof method formed, in the 1950's, the impulse for Kuhn to develop a new, fast method for the assignment problem, which he therefore baptized the *Hungarian method*. But first there were some other developments on the assignment problem.

Easterfield 1946

The first algorithm for the assignment problem might have been published by Easterfield [1946], who described his motivation as follows:

> In the course of a piece of organisational research into the problems of demobilisation in the R.A.F., it seemed that it might be possible to arrange the posting of men from disbanded units into other units in such a way that they would not need to be posted again before they were demobilised; and that a study of the numbers of men in the various release groups in each unit might enable this process to be carried out with a minimum number of postings. Unfortunately the unexpected ending of the Japanese war prevented the implications of this approach from being worked out in time for effective use. The algorithm of this paper arose directly in the course of the investigation.

[9] *If the elements of the matrix $\|a_{ij}\|$ of order n are given nonnegative integers, then under the assumption*

$$\lambda_i + \mu_j \geq a_{ij}, \qquad (i, j = 1, 2, ...n),$$

$(\lambda_i, \mu_j$ nonnegative integers)

we have

$$\min . \sum_{k=1}^{n} (\lambda_k + \mu_k) = \max .(a_{1\nu_1} + a_{2\nu_2} + \cdots + a_{n\nu_n}).$$

where $\nu_1, \nu_2, ...\nu_n$ run over all possible permutations of the numbers $1, 2, ...n$.

Easterfield seems to have worked without knowledge of the existing literature. He formulated and proved a theorem equivalent to Kőnig's theorem and he described a primal-dual type method for the assignment problem from which Egerváry's result given above can be derived. Easterfield's algorithm has running time $O(2^n n^2)$. This is better than scanning all permutations, which takes time $\Omega(n!)$.

Robinson 1949

Cycle reduction is an important tool in combinatorial optimization. In a RAND Report dated 5 December 1949, Robinson [1949] reports that an 'unsuccessful attempt' to solve the traveling salesman problem, led her to the following cycle reduction method for the optimum assignment problem.

Let matrix $(a_{i,j})$ be given, and consider any permutation π. Define for all i, j a 'length' $l_{i,j}$ by: $l_{i,j} := a_{j,\pi(i)} - a_{i,\pi(i)}$ if $j \neq \pi(i)$ and $l_{i,\pi(i)} = \infty$. If there exists a negative-length directed circuit, there is a straightforward way to improve π. If there is no such circuit, then π is an optimal permutation. This clearly is a finite method, and Robinson remarked:

> I believe it would be feasible to apply it to as many as 50 points provided suitable calculating equipment is available.

The simplex method

A breakthrough in solving the assignment problem came when Dantzig [1951a] showed that the assignment problem can be formulated as a linear programming problem that automatically has an integer optimum solution. The reason is a theorem of Birkhoff [1946] stating that the convex hull of the permutation matrices is equal to the set of *doubly stochastic* matrices — nonnegative matrices in which each row and column sum is equal to 1. Therefore, minimizing a linear functional over the set of doubly stochastic matrices (which is a linear programming problem) gives a permutation matrix, being the optimum assignment. So the assignment problem can be solved with the simplex method.

Votaw [1952] reported that solving a 10×10 assignment problem with the simplex method on the SEAC took 20 minutes. On the other hand, in his reminiscences, Kuhn [1991] mentioned the following:

> The story begins in the summer of 1953 when the National Bureau of Standards and other US government agencies had gathered an outstanding group of combinatorialists and algebraists at the Institute for Numerical Analysis (INA) located on the campus of the University of California at Los Angeles. Since space was tight, I shared an office with Ted Motzkin, whose pioneering work on linear inequalities and related systems predates linear programming by more than ten years. A rather unique feature of the INA was the presence of the Standards Western Automatic Computer (SWAC), the entire memory of which consisted of 256 Williamson cathode ray tubes. The SWAC was faster but smaller than its sibling machine, the Standards Eastern Automatic Computer (SEAC), which boasted a liquid mercury memory and which had been coded to solve linear programs.

According to Kuhn:

> the 10 by 10 assignment problem is a linear program with 100 nonnegative variables and 20 equation constraints (of which only 19 are needed). In 1953, there was no machine in the world that had been programmed to solve a linear program this large!

If 'the world' includes the Eastern Coast of the U.S.A., there seems to be some discrepancy with the remarks of Votaw [1952] mentioned above.

The complexity issue

The assignment problem has helped in gaining the insight that a finite algorithm need not be practical, and that there is a gap between exponential time and polynomial time.

Also in other disciplines it was recognized that while the assignment problem is a finite problem, there is a complexity issue. In an address delivered on 9 September 1949 at a meeting of the American Psychological Association at Denver, Colorado, Thorndike [1950] studied the problem of the 'classification' of personnel (being job assignment):

> The past decade, and particularly the war years, have witnessed a great concern about the classification of personnel and a vast expenditure of effort presumably directed towards this end.

He exhibited little trust in mathematicians:

> There are, as has been indicated, a finite number of permutations in the assignment of men to jobs. When the classification problem as formulated above was presented to a mathematician, he pointed to this fact and said that from the point of view of the mathematician there was no problem. Since the number of permutations was finite, one had only to try them all and choose the best. He dismissed the problem at that point. This is rather cold comfort to the psychologist, however, when one considers that only ten men and ten jobs mean over three and a half million permutations. Trying out all the permutations may be a mathematical solution to the problem, it is not a practical solution.

Thorndike presented three heuristics for the assignment problem, the *Method of Divine Intuition*, the *Method of Daily Quotas*, and the *Method of Predicted Yield*.

(Other heuristic and geometric methods for the assignment problem were proposed by Lord [1952], Votaw and Orden [1952], Törnqvist [1953], and Dwyer [1954] (the 'method of optimal regions').)

Von Neumann considered the complexity of the assignment problem. In a talk in the Princeton University Game Seminar on October 26, 1951, he showed that the assignment problem can be reduced to finding an optimum column strategy in a certain zero-sum two-person game, and that it can be found by a method given by Brown and von Neumann [1950]. We give first the mathematical background.

A zero-sum two-person game is given by a matrix A, the 'pay-off matrix'. The interpretation as a game is that a 'row player' chooses a row index i and a 'column player' chooses simultaneously a column index j. After that, the column player pays the row player $A_{i,j}$. The game is played repeatedly, and the question is what is the best strategy.

Let A have order $m \times n$. A *row strategy* is a vector $x \in \mathbb{R}^m_+$ satisfying $\mathbf{1}^\mathsf{T} x = 1$. Similarly, a *column strategy* is a vector $y \in \mathbb{R}^n_+$ satisfying $\mathbf{1}^\mathsf{T} y = 1$. Then

$$(3) \qquad \max_x \min_j (x^\mathsf{T} A)_j = \min_y \max_i (Ay)_i,$$

where x ranges over row strategies, y over column strategies, i over row indices, and j over column indices. Equality (3) follows from LP duality.

It can be derived that the best strategy for the row player is to choose rows with distribution an optimum x in (3). Similarly, the best strategy for the column player is to choose columns with distribution an optimum y in (3). The average pay-off then is the value of (3).

The method of Brown [1951] to determine the optimum strategies is that each player chooses in turn the line that is best with respect to the distribution of the lines chosen by the opponent so far. It was proved by Robinson [1951] that this converges to optimum strategies. The method of Brown and von Neumann [1950] is a continuous version of this, and amounts to solving a system of linear differential equations.

Now von Neumann noted that the following reduces the assignment problem to the problem of finding an optimum column strategy. Let $C = (c_{i,j})$ be an $n \times n$ cost matrix, as input for the assignment problem. We may assume that C is positive. Consider the following pay-off matrix A, of order $2n \times n^2$, with columns indexed by ordered pairs (i, j) with $i, j = 1, \ldots, n$. The entries of A are given by: $A_{i,(i,j)} := 1/c_{i,j}$ and $A_{n+j,(i,j)} := 1/c_{i,j}$ for $i, j = 1, \ldots, n$, and $A_{k,(i,j)} := 0$ for all i, j, k with $k \neq i$ and $k \neq n + j$. Then any minimum-cost assignment, of cost γ say, yields an optimum column strategy y by: $y_{(i,j)} := c_{i,j}/\gamma$ if i is assigned to j, and $y_{(i,j)} := 0$ otherwise. Any optimum column strategy is a convex combination of strategies obtained this way from optimum assignments. So an optimum assignment can in principle be found by finding an optimum column strategy.

According to a transcript of the talk (cf. von Neumann [1951,1953]), von Neumann noted the following on the number of steps:

```
It turns out that this number is a moderate power of n, i.e., considerably smaller
than the "obvious" estimate n!  mentioned earlier.
```

However, no further argumentation is given.

In a Cowles Commission Discussion Paper of 2 April 1953, Beckmann and Koopmans [1953] noted:

```
It should be added that in all the assignment problems discussed, there is, of course,
the obvious brute force method of enumerating all assignments, evaluating the maximand
at each of these, and selecting the assignment giving the highest value.  This is
too costly in most cases of practical importance, and by a method of solution we have
meant a procedure that reduces the computational work to manageable proportions in
a wider class of cases.
```

The Hungarian method: Kuhn 1955-1956, Munkres 1957

The basic combinatorial (nonsimplex) method for the assignment problem is the *Hungarian method*. The method was developed by Kuhn [1955b,1956], based on the work of Egerváry [1931], whence Kuhn introduced the name Hungarian method for it.

In an article "On the origin of the Hungarian method"' Kuhn [1991] gave the following reminiscences from the time starting Summer 1953:

> During this period, I was reading Kőnig's classical book on the theory of graphs and realized that the matching problem for a bipartite graph on two sets of n vertices was exactly the same as an n by n assignment problem with all $a_{ij} = 0$ or 1. More significantly, Kőnig had given a combinatorial algorithm (based on augmenting paths) that produces optimal solutions to the matching problem and its combinatorial (or linear programming) dual. In one of the several

> formulations given by Kőnig (p. 240, Theorem D), given an n by n matrix $A = (a_{ij})$ with all $a_{ij} = 0$ or 1, the maximum number of 1's that can be chosen with no two in the same line (horizontal row or vertical column) is equal to the minimum number of lines that contain all of the 1's. Moreover, the algorithm seemed to be 'good' in a sense that will be made precise later. The problem then was: how could the general assignment problem be reduced to the 0-1 special case?
>
> Reading Kőnig's book more carefully, I was struck by the following footnote (p. 238, footnote 2): "... Eine Verallgemeinerung dieser Sätze gab Egerváry, Matrixok kombinatorius tulajdonságairól (Über kombinatorische Eigenschaften von Matrizen), Matematikai és Fizikai Lapok, *38*, 1931, S. 16-28 (ungarisch mit einem deutschen Auszug) ..." This indicated that the key to the problem might be in Egerváry's paper. When I returned to Bryn Mawr College in the fall, I obtained a copy of the paper together with a large Hungarian dictionary and grammar from the Haverford College library. I then spent two weeks learning Hungarian and translated the paper [1]. As I had suspected, the paper contained a method by which a general assignment problem could be reduced to a finite number of 0-1 assignment problems.
>
> Using Egerváry's reduction and Kőnig's maximum matching algorithm, in the fall of 1953 I solved several 12 by 12 assignment problems (with 3-digit integers as data) by hand. Each of these examples took under two hours to solve and I was convinced that the combined algorithm was 'good'. This must have been one of the last times when pencil and paper could beat the largest and fastest electronic computer in the world.

(Reference [1] is the English translation of the paper of Egerváry [1931].)

The method described by Kuhn is a sharpening of the method of Egerváry sketched above, in two respects: (i) it gives an (augmenting path) method to find either a perfect matching or sets I and J as required, and (ii) it improves the λ_i and μ_j not by 1, but by the largest value possible.

Kuhn [1955b] contented himself with stating that the number of iterations is finite, but Munkres [1957] observed that the method in fact runs in strongly polynomial time ($O(n^4)$).

Ford and Fulkerson [1956b] reported the following computational experience with the Hungarian method:

> The largest example tried was a 20×20 optimal assignment problem. For this example, the simplex method required well over an hour, the present method about thirty minutes of hand computation.

3. The transportation problem

The transportation problem is: given an $m \times n$ 'cost' matrix $C = (a_{i,j})$, a 'supply vector' $b \in \mathbb{R}^m_+$ and a 'demand' vector $d \in \mathbb{R}^n_+$, find a nonnegative $m \times n$ matrix $X = (x_{i,j})$ such that

$$(4) \qquad \text{(i)} \sum_{j=1}^{n} x_{i,j} = b_i \text{ for } i = 1, \ldots, m,$$

$$\text{(ii)} \sum_{i=1}^{m} x_{i,j} = d_j \text{ for } j = 1, \ldots, n,$$

$$\text{(iii)} \sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j} x_{i,j} \text{ is as small as possible.}$$

So the transportation problem is a special case of a linear programming problem.

Tolstoĭ 1930

An early study of the transportation problem was made by A.N. Tolstoĭ [1930]. He published, in a book on transportation planning issued by the National Commissariat of Transportation of the Soviet Union, an article called *Methods of finding the minimal total kilometrage in cargo-transportation planning in space*, in which he formulated and studied the transportation problem, and described a number of solution approaches, including the, now well-known, idea that an optimum solution does not have any negative-cost cycle in its residual graph[10]. He might have been the first to observe that the cycle condition is necessary for optimality. Moreover, he assumed, but did not explicitly state or prove, the fact that checking the cycle condition is also sufficient for optimality.

Tolstoĭ illuminated his approach by applications to the transportation of salt, cement, and other cargo between sources and destinations along the railway network of the Soviet Union. In particular, a, for that time large-scale, instance of the transportation problem was solved to optimality.

We briefly review the article. Tolstoĭ first considered the transportation problem for the case where there are only two sources. He observed that in that case one can order the destinations by the difference between the distances to the two sources. Then one source can provide the destinations starting from the beginning of the list, until the supply of that source has been used up. The other source supplies the remaining demands. Tolstoĭ observed that the list is independent of the supplies and demands, and hence it

> is applicable for the whole life-time of factories, or sources of production. Using this table, one can immediately compose an optimal transportation plan every year, given quantities of output produced by these two factories and demands of the destinations.

Next, Tolstoĭ studied the transportation problem in the case when all sources and destinations are along one circular railway line (cf. Figure 1), in which case the optimum solution is readily obtained by considering the difference of two sums of costs. He called this phenomenon *circle dependency*.

Finally, Tolstoĭ combined the two ideas into a heuristic to solve a concrete transportation problem coming from cargo transportation along the Soviet railway network. The problem has 10 sources and 68 destinations, and 155 links between sources and destinations (all other distances are taken to be infinite).

Tolstoĭ's heuristic also makes use of insight into the geography of the Soviet Union. He goes along all sources (starting with the most remote sources), where, for each source X, he lists those destinations for which X is the closest source or the second closest source. Based on the difference of the distances to the closest and second closest sources, he assigns cargo from X to the destinations, until the supply of X has been used up. (This obviously is equivalent to considering cycles of length 4.) In case Tolstoĭ foresees a negative-cost cycle in the residual graph, he deviates from this rule to avoid such a cycle. No backtracking occurs.

[10] The *residual graph* has arcs from each source to each destination, and moreover an arc from a destination to a source if the transport on that connection is positive; the cost of the 'backward' arc is the negative of the cost of the 'forward' arc.

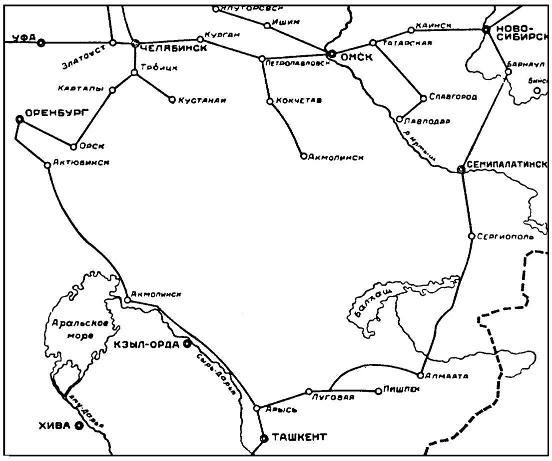


Figure 1

Figure from Tolstoĭ [1930] to illustrate a negative cycle.

After 10 steps, when the transports from all 10 factories have been set, Tolstoĭ 'verifies' the solution by considering a number of cycles in the network, and he concludes that his solution is optimum:

> Thus, by use of successive applications of the method of differences, followed by a verification of the results by the circle dependency, we managed to compose the transportation plan which results in the minimum total kilometrage.

The objective value of Tolstoĭ's solution is 395,052 kiloton-kilometers. Solving the problem with modern linear programming tools (CPLEX) shows that Tolstoĭ's solution indeed is optimum. But it is unclear how sure Tolstoĭ could have been about his claim that his solution is optimum. Geographical insight probably has helped him in growing convinced of the optimality of his solution. On the other hand, it can be checked that there exist feasible solutions that have none of the negative-cost cycles considered by Tolstoĭ in their residual graph, but that are yet not optimum.

Later, Tolstoĭ [1939] described similar results in an article entitled *Methods of removing irrational transportations in planning* in the September 1939 issue of *Sotsialisticheskiĭ Transport*. The methods were also explained in the book *Planning Goods Transportation* by Pariĭskaya, Tolstoĭ, and Mots [1947].

According to Kantorovich [1987], there were some attempts to introduce Tolstoĭ's work by the appropriate department of the People's Commissariat of Transport.

Kantorovich 1939

Apparently unaware (by that time) of the work of Tolstoĭ, L.V. Kantorovich studied a general class of problems, that includes the transportation problem. The transportation problem formed the big motivation for studying linear programming. In his memoirs, Kantorovich [1987] wrote how questions from practice motivated him to formulate these problems:

> Once some engineers from the veneer trust laboratory came to me for consultation with a quite skilful presentation of their problems. Different productivity is obtained for veneer-cutting machines for different types of materials; linked to this the output of production of this group of machines depended, it would seem, on the chance factor of which group of raw materials to which machine was assigned. How could this fact be used rationally?
>
> This question interested me, but nevertheless appeared to be quite particular and elementary, so I did not begin to study it by giving up everything else. I put this question for discussion at a meeting of the mathematics department, where there were such great specialists as Gyunter, Smirnov himself, Kuz'min, and Tartakovskii. Everyone listened but no one proposed a solution; they had already turned to someone earlier in individual order, apparently to Kuz'min. However, this question nevertheless kept me in suspense. This was the year of my marriage, so I was also distracted by this. In the summer or after the vacation concrete, to some extent similar, economic, engineering, and managerial situations started to come into my head, that also required the solving of a maximization problem in the presence of a series of linear constraints.
>
> In the simplest case of one or two variables such problems are easily solved—by going through all the possible extreme points and choosing the best. But, let us say in the veneer trust problem for five machines and eight types of materials such a search would already have required solving about a billion systems of linear equations and it was evident that this was not a realistic method. I constructed particular devices and was probably the first to report on this problem in 1938 at the October scientific session of the Herzen Institute, where in the main a number of problems were posed with some ideas for their solution.
>
> The universality of this class of problems, in conjunction with their difficulty, made me study them seriously and bring in my mathematical knowledge, in particular, some ideas from functional analysis.
>
> What became clear was both the solubility of these problems and the fact that they were widespread, so representatives of industry were invited to a discussion of my report at the university.

This meeting took place on 13 May 1939 at the Mathematical Section of the Institute of Mathematics and Mechanics of the Leningrad State University. A second meeting, which was devoted specifically to problems connected with construction, was held on 26 May 1939 at the Leningrad Institute for Engineers of Industrial Construction. These meetings provided the basis of the monograph *Mathematical Methods in the Organization and Planning of Production* (Kantorovich [1939]).

According to the Foreword by A.R. Marchenko to this monograph, Kantorovich's work was highly praised by mathematicians, and, in addition, at the special meeting industrial workers unanimously evinced great interest in the work.

In the monograph, the relevance of the work for the Soviet system was stressed:

> I want to emphasize again that the greater part of the problems of which I shall speak, relating to the organization and planning of production, are connected specifically with the Soviet system of economy and in the majority of cases do not arise in the economy of a capitalist society. There the choice of output is determined not by the plan but by the interests and profits of individual capitalists. The owner of the enterprise chooses for production those goods which at a given moment have the highest price, can most easily be sold, and therefore give the largest profit. The raw material used is not that of which there are huge supplies in the country, but that which the entrepreneur can buy most cheaply. The question of the maximum utilization of equipment is not raised; in any case, the majority of enterprises work at half capacity.
>
> In the USSR the situation is different. Everything is subordinated not to the interests and advantage of the individual enterprise, but to the task of fulfilling the state plan. The basic

task of an enterprise is the fulfillment and overfulfillment of its plan, which is a part of the general state plan. Moreover, this not only means fulfillment of the plan in aggregate terms (i.e. total value of output, total tonnage, and so on), but the certain fulfillment of the plan for all kinds of output; that is, the fulfillment of the assortment plan (the fulfillment of the plan for each kind of output, the completeness of individual items of output, and so on).

One of the problems studied was a rudimentary form of a transportation problem:

(5)
$$\begin{array}{rll} \text{given:} & \text{an } m \times n \text{ matrix } (c_{i,j}); & \\ \text{find:} & \text{an } m \times n \text{ matrix } (x_{i,j}) \text{ such that:} & \\ & \text{(i)} \quad x_{i,j} \geq 0 & \text{for all } i, j; \\ & \text{(ii)} \quad \displaystyle\sum_{i=1}^{m} x_{i,j} = 1 & \text{for each } j = 1, \ldots, n; \\ & \text{(iii)} \quad \displaystyle\sum_{j=1}^{n} c_{i,j} x_{i,j} \text{ is independent of } i \text{ and is maximized.} & \end{array}$$

Another problem studied by Kantorovich was 'Problem C' which can be stated as follows:

(6)
$$\begin{array}{lll} \text{maximize} & \lambda & \\ \text{subject to} & \displaystyle\sum_{i=1}^{m} x_{i,j} = 1 & (j = 1, \ldots, n) \\ & \displaystyle\sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j,k} x_{i,j} = \lambda & (k = 1, \ldots, t) \\ & x_{i,j} \geq 0 & (i = 1, \ldots, m; j = 1, \ldots, n). \end{array}$$

The interpretation is: let there be n machines, which can do m jobs. Let there be one final product consisting of t parts. When machine i does job j, $c_{i,j,k}$ units of part k are produced ($k = 1, \ldots, t$). Now $x_{i,j}$ is the fraction of time machine i does job j. The number λ is the amount of the final product produced. 'Problem C' was later shown (by H.E. Scarf, upon a suggestion by Kantorovich — see Koopmans [1959]) to be equivalent to the general linear programming problem.

Kantorovich outlined a new method to maximize a linear function under given linear inequality constraints. The method consists of determining dual variables ('resolving multipliers') and finding the corresponding primal solution. If the primal solution is not feasible, the dual solution is modified following prescribed rules. Kantorovich indicated the role of the dual variables in sensitivity analysis, and he showed that a feasible solution for Problem C can be shown to be optimal by specifying optimal dual variables.

The method resembles the simplex method, and a footnote in Kantorovich [1987] by his son V.L. Kantorovich suggests that Kantorovich had found the simplex method in 1938:

> In L.V. Kantorovich's archives a manuscript from 1938 is preserved on "Some mathematical problems of the economics of industry, agriculture, and transport" that in content, apparently, corresponds to this report and where, in essence, the simplex method for the machine problem is described.

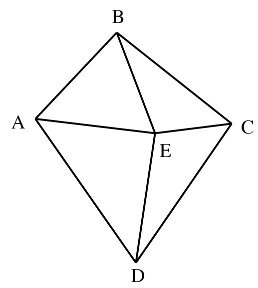
Kantorovich gave a wealth of practical applications of his methods, which he based mainly in the Soviet plan economy:

Here are included, for instance, such questions as the distribution of work among individual machines of the enterprise or among mechanisms, the correct distribution of orders among enterprises, the correct distribution of different kinds of raw materials, fuel, and other factors. Both are clearly mentioned in the resolutions of the 18th Party Congress.

He gave the following applications to transportation problems:

Let us first examine the following question. A number of freights (oil, grain, machines and so on) can be transported from one point to another by various methods; by railroads, by steamship; there can be mixed methods, in part by railroad, in part by automobile transportation, and so on. Moreover, depending on the kind of freight, the method of loading, the suitability of the transportation, and the efficiency of the different kinds of transportation is different. For example, it is particularly advantageous to carry oil by water transportation if oil tankers are available, and so on. The solution of the problem of the distribution of a given freight flow over kinds of transportation, in order to complete the haulage plan in the shortest time, or within a given period with the least expenditure of fuel, is possible by our methods and leads to Problems A or C.

Let us mention still another problem of different character which, although it does not lead directly to questions A, B, and C, can still be solved by our methods. That is the choice of transportation routes.



Let there be several points A, B, C, D, E (Fig. 1) which are connected to one another by a railroad network. It is possible to make the shipments from B to D by the shortest route BED, but it is also possible to use other routes as well: namely, BCD, BAD. Let there also be given a schedule of freight shipments; that is, it is necessary to ship from A to B a certain number of carloads, from D to C a certain number, and so on. The problem consists of the following. There is given a maximum capacity for each route under the given conditions (it can of course change under new methods of operation in transportation). It is necessary to distribute the freight flows among the different routes in such a way as to complete the necessary shipments with a minimum expenditure of fuel, under the condition of minimizing the empty runs of freight cars and taking account of the maximum capacity of the routes. As was already shown, this problem can also be solved by our methods.

As to the reception of his work, Kantorovich [1987] wrote in his memoirs:

The university immediately published my pamphlet, and it was sent to fifty People's Commissariats. It was distributed only in the Soviet Union, since in the days just before the start of the World War it came out in an edition of one thousand copies in all.

The number of responses was not very large. There was quite an interesting reference from the People's Commissariat of Transportation in which some optimization problems directed at decreasing the mileage of wagons was considered, and a good review of the pamphlet appeared in the journal "The Timber Industry."

At the beginning of 1940 I published a purely mathematical version of this work in Doklady Akad. Nauk [76], expressed in terms of functional analysis and algebra. However, I did not even put in it a reference to my published pamphlet—taking into account the circumstances I did not want my practical work to be used outside the country.

> In the spring of 1939 I gave some more reports—at the Polytechnic Institute and the House of Scientists, but several times met with the objection that the work used mathematical methods, and in the West the mathematical school in economics was an anti-Marxist school and mathematics in economics was a means for apologists of capitalism. This forced me when writing a pamphlet to avoid the term "economic" as much as possible and talk about the organization and planning of production; the role and meaning of the Lagrange multipliers had to be given somewhere in the outskirts of the second appendix and in the semi Aesopian language.

(Here reference [76] is Kantorovich [1940].)

Kantorovich mentions that the new area opened by his work played a definite role in forming the Leningrad Branch of the Mathematical Institute (LOMI), where he worked with M.K. Gavurin on this area. The problem they studied occurred to them by itself, but they soon found out that railway workers were already studying the problem of planning haulage on railways, applied to questions of driving empty cars and transport of heavy cargoes.

Kantorovich and Gavurin developed a method (the method of 'potentials'), which they wrote down in a paper "Application of mathematical methods in questions of analysis of freight traffic". This paper was presented in January 1941 to the mathematics section of the Leningrad House of Scientists, but according to Kantorovich [1987] there were political problems in publishing it:

> The publication of this paper met with many difficulties. It had already been submitted to the journal "Railway Transport" in 1940, but because of the dread of mathematics already mentioned it was not printed then either in this or in any other journal, despite the support of Academicians A.N. Kolmogorov and V.N. Obraztsov, a well-known transport specialist and first-rank railway General.

(The paper was finally published as Kantorovich and Gavurin [1949].) Kantorovich [1987] said that he fortunately made an abstract version of the problem, which was published as Kantorovich [1942]. In this, he considered the following generalization of the transportation problem.

Let R be a compact metric space, with two measures μ and μ'. Let $\mathcal{B}$ be the collection of measurable sets in R. A *translocation (of masses)* is a function $\Psi : \mathcal{B} \times \mathcal{B} \to \mathbb{R}_+$ such that for each $X \in \mathcal{B}$ the functions $\Psi(X, .)$ and $\Psi(., X)$ are measures and such that

$$(7) \qquad \Psi(X, R) = \mu(X) \text{ and } \Psi(R, X) = \mu'(X)$$

for each $X \in \mathcal{B}$.

Let a continuous function $r : R \times R \to \mathbb{R}_+$ be given. The value $r(x, y)$ represents the work necessary to transfer a unit mass from x to y. The *work* of a *translocation* Ψ is defined by:

$$(8) \qquad \int_R \int_R r(x, y)\Psi(d\mu, d\mu').$$

Kantorovich argued that, if there exists a translocation, then there exists a *minimal* translocation, that is, a translocation Ψ minimizing (8).

He called a translocation Ψ *potential* if there exists a function $p : R \to \mathbb{R}$ such that for all $x, y \in R$:

(9) (i) $|p(x) - p(y)| \leq r(x, y)$;
(ii) $p(y) - p(x) = r(x, y)$ if $\Psi(U_x, U_y) > 0$ for any neighbourhoods U_x and U_y of x and y.

Kantorovich showed that a translocation Ψ is minimal if and only if it is potential. This framework applies to the transportation problem (when $m = n$), by taking for R the space $\{1, \ldots, n\}$, with the discrete topology. Kantorovich seems to assume that r satisfies the triangle inequality.

Kantorovich remarked that his method in fact is algorithmic:

> The theorem just demonstrated makes it easy for one to prove that a given mass translocation is or is not minimal. He has only to try and construct the potential in the way outlined above. If this construction turns out to be impossible, i.e. the given translocation is not minimal, he at least will find himself in the possession of the method how to lower the translocation work and eventually come to the minimal translocation.

Kantorovich gave the transportation problem as application:

> Problem 1. *Location of consumption stations with respect to production stations.* Stations $A_1, A_2, \cdots, A_m$, attached to a network of railways deliver goods to an extent of $a_1, a_2, \cdots, a_m$ carriages per day respectively. These goods are consumed at stations $B_1, B_2, \cdots, B_n$ of the same network at a rate of $b_1, b_2, \cdots, b_n$ carriages per day respectively ($\sum a_i = \sum b_k$). Given the costs $r_{i,k}$ involved in moving one carriage from station A_i to station B_k, assign the consumption stations such places with respect to the production stations as would reduce the total transport expenses to a minimum.

Kantorovich [1942] also gave a cycle reduction method for finding a minimum-cost transshipment (which is a *uncapacitated* minimum-cost flow problem). He restricted himself to symmetric distance functions.

Kantorovich's work remained unnoticed for some time by Western researchers. In a note introducing a reprint of the article of Kantorovich [1942], in *Management Science* in 1958, the following reassuring remark was made:

> It is to be noted, however, that the problem of determining an effective method of actually acquiring the solution to a specific problem is *not* solved in this paper. In the category of development of such methods we seem to be, currently, ahead of the Russians.

Hitchcock 1941

Independently of Kantorovich, the transportation problem was studied by Hitchcock and Koopmans.

Hitchcock [1941] might be the first giving a precise mathematical description of the problem. The interpretation of the problem is, in Hitchcock's words:

> When several factories supply a product to a number of cities we desire the least costly manner of distribution. Due to freight rates and other matters the cost of a ton of product to a particular city will vary according to which factory supplies it, and will also vary from city to city.

Hitchcock showed that the minimum is attained at a vertex of the feasible region, and he outlined a scheme for solving the transportation problem which has much in common with the simplex method for linear programming. It includes pivoting (eliminating and introducing basic variables) and the fact that nonnegativity of certain dual variables implies optimality. He showed that the *complementary slackness* condition characterizes optimality.

Hitchcock gave a method to find an initial basic solution of (4), now known as the *north-west rule*: set $x_{1,1} := \min\{a_1, b_1\}$; if the minimum is attained by a_1, reset $b_1 := b_1 - a_1$ and recursively find a basic solution $x_{i,j}$ satisfying $\sum_{j=1}^{n} x_{i,j} = a_i$ for each $i = 2, \ldots, m$ and $\sum_{i=2}^{m} x_{i,j} = b_j$ for each $j = 1, \ldots, n$; if the minimum is attained by b_1, proceed symmetrically. (The north-west rule was also described by Salvemini [1939] and Fréchet [1951] in a statistical context, namely in order to complete correlation tables given the marginal distributions.)

Hitchcock however seems to have overlooked the possibility of cycling of his method, although he pointed at an example in which some dual variables are negative while yet the primal solution is optimum.

Koopmans 1942-1948

Koopmans was appointed, in March 1942, as a statistician on the staff of the British Merchant Shipping Mission, and later the Combined Shipping Adjustment Board (CSAB), a British-American agency dealing with merchant shipping problems during the Second World War. Influenced by his teacher J. Tinbergen (cf. Tinbergen [1934]) he was interested in tanker freights and capacities (cf. Koopmans [1939]). Koopmans' wrote in August 1942 in his diary that, while the Board was being organized, there was not much work for the statisticians,

> and I had a fairly good time working out exchange ratio's between cargoes for various routes, figuring how much could be carried monthly from one route if monthly shipments on another route were reduced by one unit.

At the Board he studied the assignment of ships to convoys so as to accomplish prescribed deliveries, while minimizing empty voyages. According to the memoirs of his wife (Wanningen Koopmans [1995]), when Koopmans was with the Board,

> he had been appalled by the way the ships were routed. There was a lot of redundancy, no intensive planning. Often a ship returned home in ballast, when with a little effort it could have been rerouted to pick up a load elsewhere.

In his autobiography (published posthumously), Koopmans [1992] wrote:

> My direct assignment was to help fit information about losses, deliveries from new construction, and employment of British-controlled and U.S-controlled ships into a unified statement. Even in this humble role I learned a great deal about the difficulties of organizing a large-scale effort under dual control—or rather in this case four-way control, military and civilian cutting across U.S. and U.K. controls. I did my study of optimal routing and the associated shadow costs of transportation on the various routes, expressed in ship days, in August 1942 when an impending redrawing of the lines of administrative control left me temporarily without urgent duties. My memorandum, cited below, was well received in a meeting of the Combined Shipping Adjustment Board (that I did not attend) as an explanation of the "paradoxes of shipping" which were always difficult to explain to higher authority. However, I have no knowledge of any systematic use of my ideas in the combined U.K.-U.S. shipping problems thereafter.

In the memorandum for the Board, Koopmans [1942] analyzed the sensitivity of the optimum shipments for small changes in the demands. In this memorandum (first published in Koopmans' Collected Works), Koopmans did not yet give a method to find an optimum shipment.

Further study led him to a 'local search' method for the transportation problem, stating that it leads to an optimum solution. Koopmans found these results in 1943, but, due to wartime restrictions, published them only after the war (Koopmans [1948], Koopmans and Reiter [1949a,1949b,1951]). Wanningen Koopmans [1995] writes that

> Tjalling said that it had been well received by the CSAB, but that he doubted that it was ever applied.

As Koopmans [1948] wrote:

> Let us now for the purpose of argument (since no figures of war experience are available) assume that one particular organization is charged with carrying out a world dry-cargo transportation program corresponding to the actual cargo flows of 1925. How would that organization solve the problem of moving the empty ships economically from where they become available to where they are needed? It seems appropriate to apply a procedure of trial and error whereby one draws tentative lines on the map that link up the surplus areas with the deficit areas, trying to lay out flows of empty ships along these lines in such a way that a minimum of shipping is at any time tied up in empty movements.

He gave an optimum solution for the following supplies and demands:

Net receipt of dry cargo in overseas trade, 1925

Unit: Millions of metric tons per annum

Harbour	Received	Dispatched	Net receipts
New York	23.5	32.7	−9.2
San Francisco	7.2	9.7	−2.5
St. Thomas	10.3	11.5	−1.2
Buenos Aires	7.0	9.6	−2.6
Antofagasta	1.4	4.6	−3.2
Rotterdam	126.4	130.5	− 4.1
Lisbon	37.5	17.0	20.5
Athens	28.3	14.4	13.9
Odessa	0.5	4.7	−4.2
Lagos	2.0	2.4	−0.4
Durban	2.1	4.3	−2.2
Bombay	5.0	8.9	−3.9
Singapore	3.6	6.8	−3.2
Yokohama	9.2	3.0	6.2
Sydney	2.8	6.7	−3.9
Total	266.8	266.8	0.0

So Koopmans solved a 3×12 transportation problem.

Koopmans stated that if no improvement on a solution can be obtained by a cyclic rerouting of ships, then the solution is optimum. It was observed by Robinson [1950] that this gives a finite algorithm.

Koopmans moreover claimed that there exist *potentials* $p_1, \ldots, p_n$ and $q_1, \ldots, q_m$ such that $c_{i,j} \geq p_i - q_j$ for all i, j and such that $c_{i,j} = p_i - q_j$ for each i, j for which any optimum solution x has $x_{i,j} > 0$.

Koopmans and Reiter [1951] investigated the economic implications of the model and the method:

> For the sake of definiteness we shall speak in terms of the transportation of cargoes on ocean-going ships. In considering only shipping we do not lose generality of application since ships may be "translated" into trucks, aircraft, or, in first approximation, trains, and ports into the various sorts of terminals. Such translation is possible because all the above examples involve particular types of movable transportation equipment.

In a footnote they contemplate the application of graphs in economic theory:

> The cultural lag of economic thought in the application of mathematical methods is strikingly illustrated by the fact that linear graphs are making their entrance into transportation theory just about a century after they were first studied in relation to electrical networks, although organized transportation systems are much older than the study of electricity.

Linear programming and the simplex method 1949-1950

The transportation problem was pivotal in the development of the more general problem of linear programming. The simplex method, found in 1947 by G.B. Dantzig, extends the methods of Kantorovich, Hitchcock, and Koopmans. It was published in Dantzig [1951b]. In another paper, Dantzig [1951a] described a direct implementation of the simplex method as applied to the transportation problem.

Votaw and Orden [1952] reported on early computational results (on the SEAC), and claimed (without proof) that the simplex method is polynomial-time for the transportation problem (a statement refuted by Zadeh [1973]):

> As to computation time, it should be noted that for moderate size problems, say $m \times n$ up to 500, the time of computation is of the same order of magnitude as the time required to type the initial data. The computation time on a sample computation in which m and n were both 10 was 3 minutes. The time of computation can be shown by study of the computing method and the code to be proportional to $(m+n)^3$.

The new ideas of applying linear programming to the transportation problem were quickly disseminated, although in some cases applicability to practice was met by scepticism. At a Conference on Linear Programming in May 1954 in London, Land [1954] presented a study of applying linear programming to the problem of transporting coal for the British Coke Industry:

> The real crux of this piece of research is whether the saving in transport cost exceeds the cost of using linear programming.

In the discussion which followed, T. Whitwell of Powers Samas Accounting Machines Ltd remarked

> that in practice one could have one's ideas of a solution confirmed or, much more frequently, completely upset by taking a couple of managers out to lunch.

Alternative methods for the transportation problem were designed by Gleyzal [1955] (a primal-dual method), and by Ford and Fulkerson [1955,1956a,1956b], Munkres [1957], and Egerváry [1958] (extensions of the Hungarian method for the assignment problem). It was also observed that the problem is a special case of the minimum-cost flow problem, for which several new algorithms were developed — see Section 4.

4. Menger's theorem and maximum flow

Menger's theorem 1927

Menger's theorem forms an important precursor of the max-flow min-cut theorem found in the 1950's by Ford and Fulkerson.

The topologist Karl Menger published his theorem in an article called *Zur allgemeinen Kurventheorie* (On the general theory of curves) (Menger [1927]) in the following form:

> Satz β. *Ist K ein kompakter regulär eindimensionaler Raum, welcher zwischen den beiden endlichen Mengen P und Q n-punktig zusammenhängend ist, dann enthält K n paarweise fremde Bögen, von denen jeder einen Punkt von P und einen Punkt von Q verbindet.*[11]

The result can be formulated in terms of graphs as: Let $G = (V, E)$ be an undirected graph and let $P, Q \subseteq V$. Then the maximum number of disjoint $P - Q$ paths is equal to the minimum cardinality of a set W of vertices such that each $P - Q$ path intersects W.

Menger's interest in this question arose from his research on what he called 'curves': a *curve* is a connected, compact topological space X with the property that for each $x \in X$, each neighbourhood of x contains a neighbourhood of x with totally disconnected boundary.

It was however noticed by Kőnig [1932] that Menger's proof of 'Satz β' is incomplete. Menger applied induction on $|E|$, where E is the edge set of the graph G. The basis of the induction is when P and Q contain all vertices. Menger overlooked that this constitutes a nontrivial case. It amounts to the theorem of Kőnig [1931] that in a bipartite graph $G = (V, E)$, the maximum size of a matching is equal to the minimum number of vertices needed to cover all edges. (According to Kőnig [1932], Menger informed him that he was aware of the hole in his proof.)

In his reminiscences on the origin of the 'n-arc theorem', Menger [1981] wrote:

> In the spring of 1930, I came through Budapest and met there a galaxy of Hungarian mathematicians. In particular, I enjoyed making the acquaintance of Dénes Kőnig, for I greatly admired the work on set theory of his father, the late Julius Kőnig — to this day one of the most significant contributions to the continuum problem—and I had read with interest some of Dénes' papers. Kőnig told me that he was about to finish a book that would include all that was known about graphs. I assured him that such a book would fill a great need; and I brought up my n-Arc Theorem which, having been published as a lemma in a curve-theoretical paper, had not yet come to his attention. Kőnig was greatly interested, but did not believe that the theorem was correct. "This evening," he said to me in parting, "I won't go to sleep before having constructed a counterexample." When we met again the next day he greeted me with the words, "A sleepless night!" and asked me to sketch my proof for him. He then said that he would add to his book a final section devoted to my theorem. This he did; and it is largely thanks to Kőnig's valuable book that the n-Arc Theorem has become widely known among graph theorists.

Variants of Menger's theorem 1927-1938

In a paper presented 7 May 1927 to the American Mathematical Society, Rutt [1927,1929] gave the following variant of Menger's theorem, suggested by Kline. Let $G = (V, E)$ be a

[11] Theorem β. *If K is a compact regular one-dimensional space which is n-point connected between the two finite sets P and Q, then K contains n disjoint curves, each of which connects a point in P and a point in Q.*

planar graph and let $s, t \in V$. Then the maximum number of internally disjoint $s-t$ paths is equal to the minimum number of vertices in $V \setminus \{s, t\}$ intersecting each $s-t$ path.

In fact, the theorem follows quite easily from Menger's theorem by deleting s and t and taking for P and Q the sets of neighbours of s and t respectively. (Rutt referred to Menger and gave an independent proof of the theorem.)

This construction was also observed by Knaster [1930] who showed that, conversely, Menger's theorem would follow from Rutt's theorem for general (not necessarily planar) graphs. A similar theorem was published by Nöbeling [1932], using Menger's result.

A result implied by Menger's theorem was presented by Whitney [1932] on 28 February 1931 to the American Mathematical Society: a graph is n-connected if and only if any two vertices are connected by n internally disjoint paths. While referring to the papers of Menger and Rutt, Whitney gave a direct proof.

Other proofs of Menger's theorem were given by Hajós [1934] and Grünwald [1938] (= T. Gallai) — the latter gave an algorithmic proof similar to the flow-augmenting path method for finding a maximum flow of Ford and Fulkerson [1955].

Gallai observed, in a footnote, that the theorem also holds for directed graphs:

> Die ganze Betrachtung lässt sich auch bei orientierten Graphen durchführen und liefert dann eine Verallgemeinerung des Mengerschen Satzes.[12]

Maximum flow 1954

The maximum flow problem is: given a graph, with a 'source' vertex s and a 'terminal' vertex t specified, and given a capacity function c defined on its edges, find a flow from s to t subject to c, of maximum value.

In their basic paper *Maximal Flow through a Network* (published first as a RAND Report of 19 November 1954), Ford and Fulkerson [1954] mentioned that the maximum flow problem was formulated by T.E. Harris as follows:

> Consider a rail network connecting two cities by way of a number of intermediate cities, where each link of the network has a number assigned to it representing its capacity. Assuming a steady state condition, find a maximal flow from one given city to the other.

In their 1962 book *Flows in Networks*, Ford and Fulkerson [1962] give a more precise reference to the origin of the problem[13]:

> It was posed to the authors in the spring of 1955 by T.E. Harris, who, in conjunction with General F.S. Ross (Ret.), had formulated a simplified model of railway traffic flow, and pinpointed this particular problem as the central one suggested by the model [11].

Ford-Fulkerson's reference [11] is a secret report by Harris and Ross [1955] entitled *Fundamentals of a Method for Evaluating Rail Net Capacities*, dated 24 October 1955[14] and written for the US Air Force. At our request, the Pentagon downgraded it to 'unclassified' on 21 May 1999.

[12] The whole consideration lets itself carry out also for oriented graphs and then yields a generalization of Menger's theorem.

[13] There seems to be some discrepancy between the date of the RAND Report of Ford and Fulkerson (19 November 1954) and the date mentioned in the quotation (spring of 1955).

[14] In their book, Ford and Fulkerson incorrectly date the Harris-Ross report 24 October 1956.

In fact, the Harris-Ross report solves a relatively large-scale maximum flow problem coming from the railway network in the Western Soviet Union and Eastern Europe ('satellite countries'). Unlike what Ford and Fulkerson said, the interest of Harris and Ross was not to find a maximum flow, but rather a minimum cut ('interdiction') of the Soviet railway system. We quote:

> Air power is an effective means of interdicting an enemy's rail system, and such usage is a logical and important mission for this Arm.
>
> As in many military operations, however, the success of interdiction depends largely on how complete, accurate, and timely is the commander's information, particularly concerning the effect of his interdiction-program efforts on the enemy's capability to move men and supplies. This information should be available at the time the results are being achieved.
>
> The present paper describes the fundamentals of a method intended to help the specialist who is engaged in estimating railway capabilities, so that he might more readily accomplish this purpose and thus assist the commander and his staff with greater efficiency than is possible at present.

First, much attention is given in the report to modeling a railway network: taking each railway junction as a vertex would give a too refined network (for their purposes). Therefore, Harris and Ross proposed to take 'railway divisions' (organizational units based on geographical areas) as vertices, and to estimate the capacity of the connections between any two adjacent railway divisions. In 1996, Ted Harris remembered (Alexander [1996]):

> We were studying rail transportation in consultation with a retired army general, Frank Ross, who had been chief of the Army's Transportation Corps in Europe. We thought of modeling a rail system as a network. At first it didn't make sense, because there's no reason why the crossing point of two lines should be a special sort of node. But Ross realized that, in the region we were studying, the "divisions" (little administrative districts) should be the nodes. The link between two adjacent nodes represents the total transportation capacity between them. This made a reasonable and manageable model for our rail system. Problems about the effect of cutting links turned out to be linear programming, so we asked for help from George Dantzig and other LP specialists at Rand.

The Harris-Ross report stresses that specialists remain needed to make up the model (which is always a good strategy to get new methods accepted):

> The ability to estimate with relative accuracy the capacity of single railway lines is largely an art. Specialists in this field have no authoritative text (insofar as the authors are informed) to guide their efforts, and very few individuals have either the experience or talent for this type of work. The authors assume that this job will continue to be done by the specialist.

The authors next dispute the naive belief that a railway network is just a set of disjoint through lines, and that cutting them implies cutting the network:

> It is even more difficult and time-consuming to evaluate the capacity of a railway network comprising a multitude of rail lines which have widely varying characteristics. Practices among individuals engaged in this field vary considerably, but all consume a great deal of time. Most, if not all, specialists attack the problem by viewing the railway network as an aggregate of through lines.
>
> The authors contend that the foregoing practice does not portray the full flexibility of a large network. In particular it tends to gloss over the fact that even if every one of a set of independent through lines is made inoperative, there may exist alternative routings which can still move the traffic.

> This paper proposes a method that departs from present practices in that it views the network as an aggregate of railway operating divisions. All trackage capacities within the divisions are appraised, and these appraisals form the basis for estimating the capability of railway operating divisions to receive trains from and concurrently pass trains to each neighboring division in 24-hour periods.

Whereas experts are needed to set up the model, to solve it is routine (when having the 'work sheets'):

> The foregoing appraisal (accomplished by the expert) is then used in the preparation of comparatively simple work sheets that will enable relatively inexperienced assistants to compute the results and thus help the expert to provide specific answers to the problems, based on many assumptions, which may be propounded to him.

For solving the problem, the authors suggested applying the 'flooding technique', a heuristic described in a RAND Report of 5 August 1955 by A.W. Boldyreff [1955a]. It amounts to pushing as much flow as possible greedily through the network. If at some vertex a 'bottleneck' arises (that is, more trains arrive than can be pushed further through the network), the excess trains are returned to the origin. The technique does not guarantee optimality, but Boldyreff speculates:

> In dealing with the usual railway networks a single flooding, followed by removal of bottlenecks, should lead to a maximal flow.

Presenting his method at an ORSA meeting in June 1955, Boldyreff [1955b] claimed simplicity:

> The mechanics of the solutions is formulated as a simple game which can be taught to a ten-year-old boy in a few minutes.

The well-known flow-augmenting path algorithm of Ford and Fulkerson [1955], that does guarantee optimality, was published in a RAND Report dated only later that year (29 December 1955). As for the simplex method (suggested for the maximum flow problem by Ford and Fulkerson [1954]), Harris and Ross remarked:

> The calculation would be cumbersome; and, even if it could be performed, sufficiently accurate data could not be obtained to justify such detail.

The Harris-Ross report applied the flooding technique to a network model of the Soviet and Eastern European railways. For the data it refers to several secret reports of the Central Intelligence Agency (C.I.A.) on sections of the Soviet and Eastern European railway networks. After the aggregation of railway divisions to vertices, the network has 44 vertices and 105 (undirected) edges.

The application of the flooding technique to the problem is displayed step by step in an appendix of the report, supported by several diagrams of the railway network. (Also work sheets are provided, to allow for future changes in capacities.) It yields a flow of value 163,000 tons from sources in the Soviet Union to destinations in Eastern European 'satellite' countries (Poland, Czechoslovakia, Austria, Eastern Germany), together with a cut with a capacity of, again, 163,000 tons. (This cut is indicated as 'The bottleneck' in Figure 2 from the Harris-Ross report.) So the flow value and the cut capacity are equal, hence optimum.

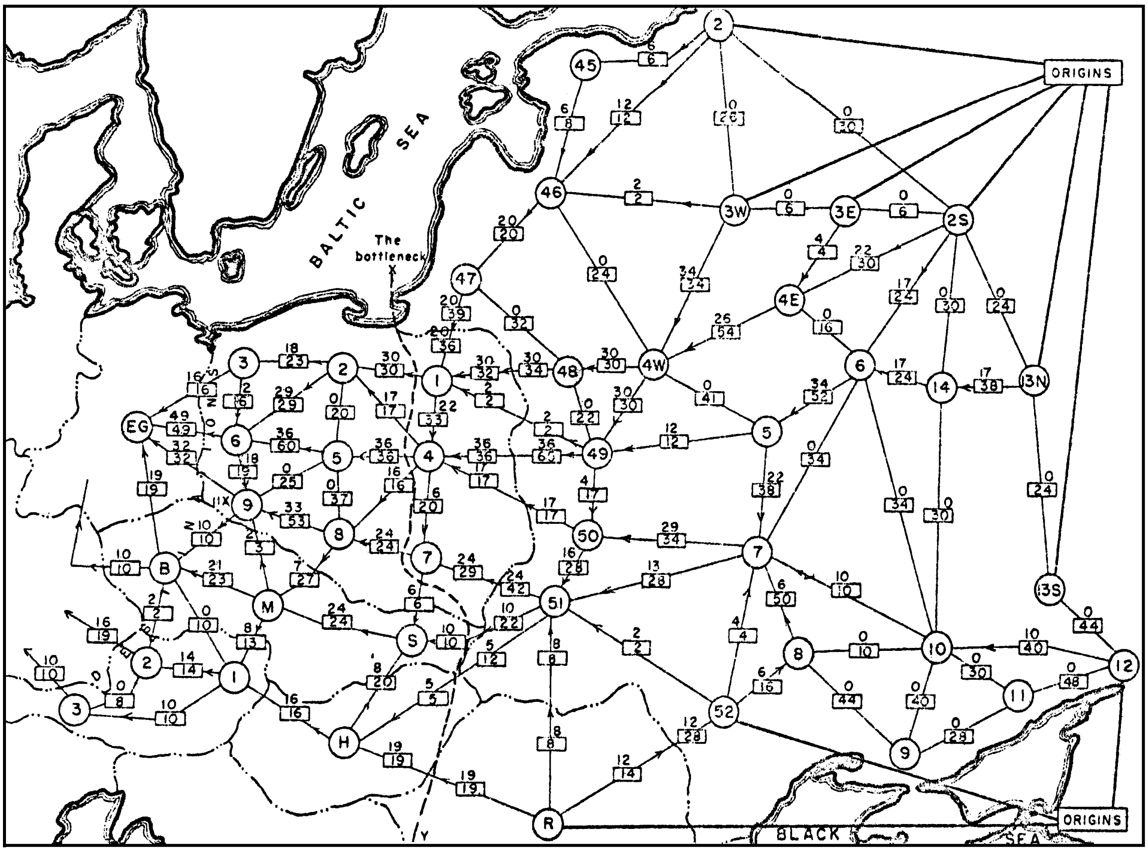


Figure 2

From Harris and Ross [1955]: Schematic diagram of the railway network of the Western Soviet Union and Eastern European countries, with a maximum flow of value 163,000 tons from Russia to Eastern Europe, and a cut of capacity 163,000 tons indicated as 'The bottleneck'.

The max-flow min-cut theorem

In the RAND Report of 19 November 1954, Ford and Fulkerson [1954] gave (next to defining the maximum flow problem and suggesting the simplex method for it) the max-flow min-cut theorem for undirected graphs, saying that the maximum flow value is equal to the minimum capacity of a cut separating source and terminal. Their proof is not constructive, but for planar graphs, with source and sink on the outer boundary, they give a polynomial-time, constructive method. In a report of 26 May 1955, Robacker [1955a] showed that the max-flow min-cut theorem can be derived also from the vertex-disjoint version of Menger's theorem.

As for the directed case, Ford and Fulkerson [1955] observed that the max-flow min-cut theorem holds also for directed graphs. Dantzig and Fulkerson [1955] showed, by extending the results of Dantzig [1951a] on integer solutions for the transportation problem to the maximum-flow problem, that if the capacities are integer, there is an integer maximum flow (the 'integrity theorem'). Hence, the arc-disjoint version of Menger's theorem for directed graphs follows as a consequence.

Also Kotzig gave the edge-disjoint version of Menger's theorem, but restricted to undirected graphs. In his dissertation for the degree of Academical Doctor, Kotzig [1956] defined, for any undirected graph G and any pair u, v of vertices of G, $\sigma_G(u, v)$ to be the minimum size of a $u - v$ cut. He stated:

> Veta 35. Nech G je l'ubovol'ný graf obsahujúci uzly $u \neq v$, o ktorých platí $\sigma_G(u, v) = k > 0$, potom existuje systém ciest $\{C_1, C_2, \ldots, C_k\}$ taký že každa cesta spojuje uzly u, v a žiadne dve rôzne cesty systému nemajú spoločnej hrany. Takýto systém ciest v G existuje len vtedy, keď je $\sigma_G(u, v) \geq k$.[15]

The proof method is to consider a minimal graph satisfying the cut condition, and next to orient it so as to make a directed graph in which each vertex (except u and v) has indegree equal to outdegree, while u has outdegree k and indegree 0. This then gives the paths.

Although the dissertation has several references to Kőnig's book, which book contains the vertex-disjoint version of Menger's theorem, Kotzig did not link his result to that of Menger.

An alternative proof of the max-flow min-cut theorem was given by Elias, Feinstein, and Shannon [1956] ('manuscript received by the PGIT, July 11,1956'), who claimed that the result was known by workers in communication theory:

> This theorem may appear almost obvious on physical grounds and appears to have been accepted without proof for some time by workers in communication theory. However, while the fact that this flow cannot be exceeded is indeed almost trivial, the fact that it can actually be achieved is by no means obvious. We understand that proofs of the theorem have been given by Ford and Fulkerson and Fulkerson and Dantzig. The following proof is relatively simple, and we believe different in principle.

The proof of Elias, Feinstein, and Shannon is based on a reduction technique similar to that used by Menger [1927] in proving his theorem.

Minimum-cost flows

The *minimum-cost* flow problem was studied, in rudimentary form, by Dantzig and Fulkerson [1954], in order to determine the minimum number of tankers to meet a fixed schedule. Similarly, Bartlett [1957] and Bartlett and Charnes [1957] gave methods to determine the minimum railway stock to run a given schedule.

It was noted by Orden [1955] and Prager [1957] that the minimum-cost flow problem is equivalent to the capacitated transportation problem.

A basic combinatorial minimum-cost flow algorithm was given (in disguised form) by Ford and Fulkerson [1957]. It consists of repeatedly finding a zero-length $s - t$ path in the residual graph, making lengths nonnegative by translating the cost with the help of a potential. If no zero-length path exists, the potential is updated. The complexity of this method was studied in a report by Fulkerson [1958].

[15] Theorem 35. Let G be an arbitrary graph containing vertices $u \neq v$ for which $\sigma_G(u, v) = k > 0$, then there exists a system of paths $\{C_1, C_2, ..., C_k\}$ such that each path connects vertices u, v and no two distinct paths have an edge in common. Such a system of paths in G exists only if $\sigma_G(u, v) \geq k$.

5. Shortest spanning tree

The problem of finding a shortest spanning tree came up in several applied areas, like in the construction of road, energy, and communication networks and in the clustering of data in anthropology and taxonomy.

We refer to Graham and Hell [1985] for an extensive historical survey of shortest tree algorithms, with several quotes (with translations) from old papers. Our notes below have profited from their investigations.

Borůvka 1926

Borůvka [1926a] seems to be the first to consider the shortest spanning tree problem. His interest came from a question of the Electric Power Company of Western Moravia in Brno, at the beginning of the 1920's, asking for the most economical construction of an electric power network (see Borůvka [1977]).

Borůvka formulated the problem as follows:

> In dieser Arbeit löse ich folgendes Problem:
> Es möge eine Matrix der bis auf die Bedingungen $r_{\alpha\alpha} = 0$, $r_{\alpha\beta} = r_{\beta\alpha}$ positiven und von einander verschiedenen Zahlen $r_{\alpha\beta}$ $(\alpha, \beta = 1, 2, \ldots n; n \geq 2)$ gegeben sein.
>
> Aus dieser ist eine Gruppe von einander und von Null verschiedener Zahlen auszuwählen, so dass
>
> 1° in ihr zu zwei willkürlich gewählten natürlichen Zahlen p_1, p_2 $(\leq n)$ eine Teilgruppe von der Gestalt
>
> $$r_{p_1c_2}, r_{c_2c_3}, r_{c_3c_4}, \ldots r_{c_{q-2}c_{q-1}}, r_{c_{q-1}p_2}$$
>
> existiere,
>
> 2° die Summe ihrer Glieder kleiner sei als die Summe der Glieder irgendeiner anderen, der Bedingung 1° genügenden Gruppe von einander und von Null verschiedenen Zahlen.[16]

So Borůvka stated that the spanning tree found is the unique shortest. He assumed that all edge lengths are different.

As a method, Borůvka proposed *parallel merging*: connect each component to its nearest neighbouring component, and iterate. His description is somewhat complicated, but in a follow-up paper, Borůvka [1926b] gave an easier description of his method.

Jarník 1929

In a reaction to Borůvka's work, Jarník wrote on 12 February 1929 a letter to Borůvka in which he described a 'new solution of a minimal problem discussed by Mr Borůvka.'

[16]In this work, I solve the following problem:
A matrix may be given of positive distinct numbers $r_{\alpha\beta}$ $(\alpha, \beta = 1, 2 \ldots n;\ n \geq 2)$, besides the conditions $r_{\alpha\alpha} = 0$, $r_{\alpha\beta} = r_{\beta\alpha}$.
From this, a group of numbers, different from each other and from zero, should be selected such that
1° for arbitrarily chosen natural numbers p_1, p_2 $(\leq n)$ a subgroup of it exist of the form

$$r_{p_1c_2}, r_{c_2c_3}, r_{c_3c_4}, \ldots r_{c_{q-2}c_{q-1}}, r_{c_{q-1}p_2},$$

2° the sum of its members be smaller than the sum of the members of any other group of numbers different from each other and from zero, satisfying condition 1°.

The 'new solution' amounts to *tree growing*: keep a tree on a subset of the vertices, and iteratively extend it by adding a shortest edge joining the tree with a vertex outside of the tree.

An extract of the letter was published as Jarník [1930]. We quote from the German summary:

> a_1 ist eine beliebige unter den Zahlen $1, 2, \ldots, n$.
> a_2 ist durch
> $$r_{a_1,a_2} = \min_{\left(\begin{array}{c} l = 1, 2, \ldots, n \\ l \neq a_1 \end{array}\right)} r_{a_1,l}$$
> definiert.
> Wenn $2 \leq k < n$ und wenn $[a_1, a_2], \ldots, [a_{2k-3}, a_{2k-2}]$ bereits bestimmt sind, so wird $[a_{2k-1}, a_{2k}]$ durch
> $$r_{a_{2k-1},a_{2k}} = \min r_{i,j},$$
> definiert, wo i alle Zahlen $a_1, a_2, \ldots, a_{2k-2}$, j aber alle übrigen von den Zahlen $1, 2, \ldots, n$ durchläuft.[17]

(For a detailed discussion and a translation of the article of Jarník [1930] (and of Jarník and Kössler [1934] on the Steiner tree problem), see Korte and Nešetřil [2001].)

Parallel merging was also described by Choquet [1938] (without proof) and Florek, Łukaszewicz, Perkal, Steinhaus, and Zubrzycki [1951a,1951b]. Choquet gave as a motivation the construction of road systems:

> Étant donné n villes du plan, il s'agit de trouver un réseau de routes permettant d'aller d'une quelconque de ces villes à une autre et tel que:
> 1° la longueur globale du réseau soit minimum;
> 2° exception faite des villes, on ne peut partir d'aucun point dans plus de deux directions, afin d'assurer la sûreté de la circulation; ceci entraîne, par exemple, que lorsque deux routes semblent se croiser en un point qui n'est pas une ville, elles passent en fait l'une au-dessus de l'autre et ne communiquent pas entre elles en ce point, qu'on appellera faux-croisement.[18]

Choquet might be the first concerned with the complexity of the method:

---

[17] a_1 is an arbitrary one among the numbers $1, 2, \ldots, n$.
a_2 is defined by
$$r_{a_1,a_2} = \min_{\left(\begin{array}{c} l = 1, 2, \ldots, n \\ l \neq a_1 \end{array}\right)} r_{a_1,l}.$$
If $2 \leq k < n$ and if $[a_1, a_2], \ldots, [a_{2k-3}, a_{2k-2}]$ are determined already, then $[a_{2k-1}, a_{2k}]$ is determined by
$$r_{a_{2k-1},a_{2k}} = \min r_{i,j},$$
where i runs through all numbers $a_1, a_2, \ldots, a_{2k-2}$, j however through all remaining of the numbers $1, 2, \ldots, n$.

[18] Being given n cities of the plane, the point is to find a network of routes allowing to go from an arbitrary of these cities to another and such that:
1° the global length of the network be minimum;
2° except for the cities, one cannot depart from any point in more than two directions, in order to assure the certainty of the circulation; this entails, for instance, that when two routes seem to cross each other in a point which is not a city, they pass in fact one above the other and do not communicate among them in this point, which we shall call a false crossing.

> Le réseau cherché sera tracé après $2n$ opérations élémentaires au plus, en appelant opération élémentaire la recherche du continu le plus voisin d'un continu donné.[19]

Florek *et al.* were motivated by clustering in anthropology, taxonomy, etc. They applied the method to:

> 1° the capitals of Poland's provinces, 2° two collections of excavated skulls, 3° 42 archeological finds, 4° the liverworts of Silesian Beskid mountains with forests as their background, and to the forests of Silesian Beskid mountains with the liverworts appearing in them as their background.

Shortest spanning trees 1956-1959

In the years 1956-1959 a number of papers appeared that again presented methods for the shortest spanning tree problem. Several of the results overlap, also with the earlier papers of Borůvka and Jarník, but also a few new and more general methods were given.

Kruskal [1956] was motivated by Borůvka's first paper and by the application to the traveling salesman problem, described as follows (where [1] is reference Borůvka [1926a]):

> Several years ago a typewritten translation (of obscure origin) of [1] raised some interest. This paper is devoted to the following theorem: If a (finite) connected graph has a positive real number attached to each edge (the *length* of the edge), and if these lengths are all distinct, then among the spanning trees (German: Gerüst) of the graph there is only one, the sum of whose edges is a minimum; that is, the shortest spanning tree of the graph is unique. (Actually in [1] this theorem is stated and proved in terms of the "matrix of lengths" of the graph, that is, the matrix $\|a_{ij}\|$ where a_{ij} is the length of the edge connecting vertices i and j. Of course, it is assumed that $a_{ij} = a_{ji}$ and that $a_{ii} = 0$ for all i and j.)
>
> The proof in [1] is based on a not unreasonable method of constructing a spanning subtree of minimum length. It is in this construction that the interest largely lies, for it is a solution to a problem (Problem 1 below) which on the surface is closely related to one version (Problem 2 below) of the well-known traveling salesman problem.
>
> Problem 1. Give a practical method for constructing a spanning subtree of minimum length.
>
> Problem 2. Give a practical method for constructing an unbranched spanning subtree of minimum length.
>
> The construction in [1] is unnecessarily elaborate. In the present paper I give several simpler constructions which solve Problem 1, and I show how one of these constructions may be used to prove the theorem of [1]. Probably it is true that any construction which solves Problem 1 may be used to prove this theorem.

Kruskal next described three algorithms: Construction A: choose iteratively the shortest edge that can be added so as not to create a circuit; Construction B: fix a nonempty set U of vertices, and choose iteratively the shortest edge leaving some component intersecting U; Construction A$'$: remove iteratively the longest edge that can be removed without making the graph disconnected.

In his reminiscences, Kruskal [1997] wrote about Borůvka's method:

> In one way, the method of construction was very elegant. In another way, however, it was unnecessarily complicated. A goal which has always been important to me is to find simpler ways

[19]The network looked for will be traced after at most $2n$ elementary operations, calling the search for the continuum closest to a given continuum an elementary operation.

> to describe complicated ideas, and that is all I tried to do here. I simplified the construction down to its essence, but it seems to me that the idea of Professor Borůvka's method is still present in my version.

Another paper on the minimum spanning tree problem was published by Prim [1957], who was at Bell Laboratories, and who was motivated by the problem of finding a shortest telecommunication network:

> A problem of inherent interest in the planning of large-scale communication, distribution and transportation networks also arises in connection with the current rate structure for Bell System leased-line services.

He described the following algorithm: choose a component of the current forest, and connect it to the nearest other component. He observed that Kruskal's constructions A and B are special cases of this.

Prim noticed that in fact only the order of the lengths determines if a spanning tree is shortest:

> The *shortest spanning subtree* of a connected labelled graph also minimizes all increasing symmetric functions, and maximizes all decreasing symmetric functions, of the edge "lengths."

Prim preferred the tree growing method for computational reasons:

> This computational procedure is easily programmed for an automatic computer so as to handle quite large-scale problems. One of its advantages is its avoidance of checks for closed cycles and connectedness. Another is that it never requires access to more than two rows of distance data at a time — no matter how large the problem.

The implementation described by Prim has $O(n^2)$ running time.

A paper by Loberman and Weinberger [1957] gave minimizing wire connections as motivation:

> In the construction of a digital computer in which high-frequency circuitry is used, it is desirable and often necessary when making connections between terminals to minimize the total wire length in order to reduce the capacitance and delay-line effects of long wire leads.

They described two methods: tree growing and *forest merging*: keep a forest, and iteratively add a shortest edge connecting two components.

Only after they had designed their algorithms, Loberman and Weinberger discovered that their algorithms were given earlier by Kruskal [1956]:

> However, it is felt that the more detailed implementation and general proofs of the procedures justify this paper.

They next described how to implement Kruskal's method, in particular, how to merge forests. And, like Prim, they observed that the minimality of a spanning tree depends only on the order of the lengths, and not on their specific values:

> After the initial sorting into a list where the branches are of monotonically increasing length, the actual value of the length of any branch no longer appears explicitly in the subsequent manipulations. As a result, some other parameter such as the square of the length could have been used. More generally, the same minimum tree will persist for all variations in branch lengths that do not disturb the original relative order.

Dijkstra [1959] gave again the tree growing method, which he prefers (for computational reasons) to the methods given by Kruskal and Loberman and Weinberger (overlooking the fact that these authors also gave the tree growing method):

> The solution given here is to be preferred to the solution given by J.B. KRUSKAL [*1*] and those given by H. LOBERMAN and A. WEINBERGER [*2*]. In their solutions all the — possibly $\frac{1}{2}n(n-1)$ — branches are first of all sorted according to length. Even if the length of the branches is a computable function of the node coordinates, their methods demand that data for all branches are stored simultaneously.

(Dijkstra's references [1] and [2] are Kruskal [1956] and Loberman and Weinberger [1957].) Also Dijkstra described an $O(n^2)$ implementation.

Extension to matroids: Rado 1957

Rado [1957] noticed that the methods of Borůvka and Kruskal can be extended to finding a minimum-weight basis in a matroid. He first showed that if the elements of a matroid are linearly ordered by $<$, there is a unique minimal basis $\{b_1, \ldots, b_r\}$ with $b_1 < b_2 < \cdots < b_r$ such that for each $i = 1, \ldots, r$ all elements $s < b_i$ belong to $\mathrm{span}(\{b_1, \ldots, b_{i-1}\})$. Rado derived that for any independent set $\{a_1, \ldots, a_k\}$ with $a_1 < \cdots < a_k$ one has $b_i \leq a_i$ for $i = 1, \ldots, k$. According to Rado, this 'leads to the result of' Borůvka [1926a] and Kruskal [1956].

6. Shortest path

Compared with other combinatorial optimization problems, like shortest spanning tree, assignment and transportation, mathematical research in the shortest path problem started relatively late. This might be due to the fact that the problem is elementary and relatively easy, which is also illustrated by the fact that at the moment that the problem came into the focus of interest, several researchers independently developed similar methods.

Yet, the problem has offered some substantial difficulties. For some considerable period heuristical, nonoptimal approaches have been investigated (cf. for instance Rosenfeld [1956], who gave a heuristic approach for determining an optimal trucking route through a given traffic congestion pattern).

Path finding, in particular searching in a maze, belongs to the classical graph problems, and the classical references are Wiener [1873], Lucas [1882] (describing a method due to C.P. Trémaux), and Tarry [1895] — see Biggs, Lloyd, and Wilson [1976]. They form the basis for depth-first search techniques.

Path problems were also studied at the beginning of the 1950's in the context of 'alternate routing', that is, finding a second shortest route if the shortest route is blocked. This applies to freeway usage (Trueblood [1952]), but also to telephone call routing. At that time making long-distance calls in the U.S.A. was automatized, and alternate routes for telephone calls over the U.S. telephone network nation-wide should be found automatically. Quoting Jacobitti [1955]:

> When a telephone customer makes a long-distance call, the major problem facing the operator is how to get the call to its destination. In some cases, each toll operator has two main routes

by which the call can be started towards this destination. The first-choice route, of course, is the most direct route. If this is busy, the second choice is made, followed by other available choices at the operator's discretion. When telephone operators are concerned with such a call, they can exercise choice between alternate routes. But when operator or customer toll dialing is considered, the choice of routes has to be left to a machine. Since the "intelligence" of a machine is limited to previously "programmed" operations, the choice of routes has to be decided upon, and incorporated in, an automatic alternate routing arrangement.

Matrix methods for unit-length shortest path 1946-1953

Matrix methods were developed to study relations in networks, like finding the transitive closure of a relation; that is, identifying in a directed graph the pairs of points s, t such that t is reachable from s. Such methods were studied because of their application to communication nets (including neural nets) and to animal sociology (e.g. peck rights).

The matrix methods consist of representing the directed graph by a matrix, and then taking iterative matrix products to calculate the transitive closure. This was studied by Landahl and Runge [1946], Landahl [1947], Luce and Perry [1949], Luce [1950], Lunts [1950, 1952], and by A. Shimbel.

Shimbel's interest in matrix methods was motivated by their applications to neural networks. He analyzed with matrices which sites in a network can communicate to each other, and how much time it takes. To this end, let S be the $0,1$ matrix indicating that if $S_{i,j} = 1$ then there is direct communication from i to j (including $i = j$). Shimbel [1951] observed that the positive entries in S^t correspond to pairs between which there exists communication in t steps. An *adequate* communication system is one for which the matrix S^t is positive for some t. One of the other observations of Shimbel [1951] is that in an adequate communication system, the time it takes that all sites have all information, is equal to the minimum value of t for which S^t is positive. (A related phenomenon was observed by Luce [1950].)

Shimbel [1953] mentioned that the distance from i to j is equal to the number of zeros in the i, j position in the matrices $S^0, S^1, S^2, \ldots, S^t$. So essentially he gave an $O(n^4)$ algorithm to find all distances in a directed graph with *unit lengths*.

Shortest-length paths

If a directed graph $D = (V, A)$ and a length function $l : A \to \mathbb{R}$ are given, one may ask for the distances and shortest-length paths from a given vertex s.

For this, there are two well-known methods: the 'Bellman-Ford method' and 'Dijkstra's method'. The latter one is faster but is restricted to nonnegative length functions. The former method only requires that there is no directed circuit of negative length.

The general framework for both methods is the following scheme, described in this general form by Ford [1956]. Keep a provisional distance function d. Initially, set $d(s) := 0$ and $d(v) := \infty$ for each $v \neq s$. Next, iteratively,

(10) choose an arc (u, v) with $d(v) > d(u) + l(u, v)$ and reset $d(v) := d(u) + l(u, v)$.

If no such arc exists, d is the distance function.

The difference in the methods is the rule by which the arc (u, v) with $d(v) > d(u)+l(u, v)$ is chosen. The Bellman-Ford method consists of considering all arcs consecutively and applying (10) where possible, and repeating this (at most $|V|$ rounds suffice). This is the method described by Shimbel [1955], Bellman [1958], and Moore [1959].

Dijkstra's method prescribes to choose an arc (u, v) with $d(u)$ smallest (then each arc is chosen at most once, if the lengths are nonnegative). This was described by Leyzorek, Gray, Johnson, Ladew, Meaker, Petry, and Seitz [1957] and Dijkstra [1959]. A related method, but slightly slower than Dijkstra's method when implemented, was given by Dantzig [1958], and chooses an arc (u, v) with $d(u) + l(u, v)$ smallest.

Parallel to this, a number of further results were obtained on the shortest path problem, including a linear programming approach and 'good characterizations'. We review the articles in a more or less chronological order.

Shimbel 1955

The paper of Shimbel [1955] was presented in April 1954 at the Symposium on Information Networks in New York. Extending his matrix methods for unit-length shortest paths, he introduced the following 'min-sum algebra':

> <u>Arithmetic</u>
> For any arbitrary real or infinite numbers x and y
>
> $$x + y \equiv \min(x, y) \text{ and}$$
> $$xy \equiv \text{the algebraic sum of } x \text{ and } y.$$

He transferred this arithmetic to the matrix product. Calling the distance matrix associated with a given length matrix S the 'dispersion', he stated:

> It follows trivially that S^k $k \geq 1$ is a matrix giving the <u>shortest</u> paths from site to site in S given that $k-1$ <u>other</u> sites may be traversed in the process. It also follows that for any S there exists an integer k such that $S^k = S^{k+1}$. Clearly, the dispersion of S (let us label it $D(S)$) will be the matrix S^k such that $S^k = S^{k+1}$.

This is equivalent to the Bellman-Ford method.

Although Shimbel did not mention it, one trivially can take $k \leq |V|$, and hence the method yields an $O(n^4)$ algorithm to find the distances between all pairs of points.

Shortest path as linear programming problem 1955-1957

Orden [1955] observed that the shortest path problem is a special case of a transshipment problem (= uncapacitated minimum-cost flow problem), and hence can be solved by linear programming. Dantzig [1957] described the following graphical procedure for the simplex method applied to this problem. Let T be a rooted spanning tree on $\{1, \ldots, n\}$, with root 1. For each $i = 1, \ldots, n$, let u_i be equal to the length of the path from 1 to i in T. Now if $u_j \leq u_i + d_{i,j}$ for all i, j, then for each i, the $1 - i$ path in T is a shortest path. If $u_j > u_i + d_{i,j}$, replace the arc of T entering j by the arc (i, j), and iterate with the new tree.

Trivially, this process terminates (as $\sum_{j=1}^{n} u_j$ decreases at each iteration, and as there are only finitely many rooted trees). Dantzig illustrated his method by an example of sending a package from Los Angeles to Boston. (Edmonds [1970] showed that this method may take exponential time.)

In a reaction to the paper of Dantzig [1957], Minty [1957] proposed an 'analog computer' for the shortest path problem:

> Build a string model of the travel network, where knots represent cities and string lengths represent distances (or costs). Seize the knot 'Los Angeles' in your left hand and the knot 'Boston' in your right and pull them apart. If the model becomes entangled, have an assistant untie and re-tie knots until the entanglement is resolved. Eventually one or more paths will stretch tight — they then are alternative shortest routes.
>
> Dantzig's 'shortest-route tree' can be found in this model by weighting the knots and picking up the model by the knot 'Los Angeles'.
>
> It is well to label the knots since after one or two uses of the model their identities are easily confused.

A similar method was proposed by Bock and Cameron [1958].

Ford 1956

In a RAND report dated 14 August 1956, Ford [1956] described a method to find a shortest path from P_0 to P_N, in a network with vertices $P_0, \ldots, P_N$, where l_{ij} denotes the length of an arc from i to j. We quote:

> Assign initially $x_0 = 0$ and $x_i = \infty$ for $i \neq 0$. Scan the network for a pair P_i and P_j with the property that $x_i - x_j > l_{ji}$. For this pair replace x_i by $x_j + l_{ji}$. Continue this process. Eventually no such pairs can be found, and x_N is now minimal and represents the minimal distance from P_0 to P_N.

So this is the general scheme described above ((10)). No selection rule for the arc (u, v) in (10) is prescribed by Ford.

Ford showed that the method terminates. It was shown however by Johnson [1973a, 1973b,1977] that Ford's liberal rule can take exponential time.

The correctness of Ford's method also follows from a result given in the book *Studies in the Economics of Transportation* by Beckmann, McGuire, and Winsten [1956]: given a length matrix $(l_{i,j})$, the distance matrix is the unique matrix $(d_{i,j})$ satisfying

(11) $d_{i,i} = 0$ for all i;
$d_{i,k} = \min_j(l_{i,j} + d_{j,k})$ for all i, k with $i \neq k$.

Good characterizations for shortest path 1956-1958

It was noticed by Robacker [1956] that shortest paths allow a theorem dual to Menger's theorem: the minimum length of an $P_0 - P_n$ path in a graph N is equal to the maximum number of pairwise disjoint $P_0 - P_n$ cuts. In Robacker's words:

> the maximum number of mutually disjunct cuts of N is equal to the length of the shortest chain of N from P_0 to P_n.

A related 'good characterization' was found by Gallai [1958]: A length function $l : A \to \mathbb{Z}$ on the arcs of a directed graph (V, A) does not give negative-length directed circuits, if and only if there is a function ('potential') $p : V \to \mathbb{Z}$ such that $l(u, v) \geq p(v) - p(u)$ for each arc (u, v).

Case Institute of Technology 1957

The shortest path problem was also investigated by a group of researchers at the Case Institute of Technology in Cleveland, Ohio, in the project *Investigation of Model Techniques*, performed for the Combat Development Department of the Army Electronic Proving Ground. In their *First Annual Report*, Leyzorek, Gray, Johnson, Ladew, Meaker, Petry, and Seitz [1957] presented their results.

First, they noted that Shimbel's method can be speeded up by calculating S^k by iteratively raising the current matrix to the square (in the min-sum matrix algebra). This solves the all-pairs shortest path problem in time $O(n^3 \log n)$.

Next, they gave a rudimentary description of a method equivalent to Dijkstra's method. We quote:

> (1) All the links joined to the origin, a, may be given an outward orientation. ...
>
> (2) Pick out the link or links radiating from a, $a_{a\alpha}$, with the smallest delay. ... Then it is impossible to pass from the origin to any other node in the network by any "shorter" path than $a_{a\alpha}$. Consequently, the minimal path to the general node α is $a_{a\alpha}$.
>
> (3) All of the other links joining α may now be directed outward. Since $a_{a\alpha}$ must necessarily be the minimal path to α, there is no advantage to be gained by directing any other links toward α. ...
>
> (4) Once α has been evaluated, it is possible to evaluate immediately all other nodes in the network whose minimal values do not exceed the value of the second-smallest link radiating from the origin. Since the minimal values of these nodes are less than the values of the second-smallest, third-smallest, and all other links radiating directly from the origin, only the smallest link, $a_{a\alpha}$, can form a part of the minimal path to these nodes. Once a minimal value has been assigned to these nodes, it is possible to orient all other links except the incoming link in an outward direction.
>
> (5) Suppose that all those nodes whose minimal values do not exceed the value of the second-smallest link radiating from the origin have been evaluated. Now it is possible to evaluate the node on which the second-smallest link terminates. At this point, it can be observed that if conflicting directions are assigned to a link, in accordance with the rules which have been given for direction assignment, that link may be ignored. It will not be a part of the minimal path to either of the two nodes it joins. ...
>
> Following these rules, it is now possible to expand from the second-smallest link as well as the smallest link so long as the value of the third-smallest link radiating from the origin is not exceeded. It is possible to proceed in this way until the entire network has been solved.

(In this quotation we have deleted sentences referring to figures.)

Bellman 1958

After having published several papers on dynamic programming (which is, in some sense, a generalization of shortest path methods), Bellman [1958] eventually focused on the shortest path problem by itself, in a paper in the *Quarterly of Applied Mathematics*. He described the following 'functional equation approach' for the shortest path problem, which is the same as that of Shimbel [1955].

There are N cities, numbered $1, \ldots, N$, every two of which are linked by a direct road. A matrix $T = (t_{i,j})$ is given, where $t_{i,j}$ is time required to travel from i to j (not necessarily symmetric). Find a path between 1 and N which consumes minimum time.

Bellman remarked:

> Since there are only a finite number of paths available, the problem reduces to choosing the smallest from a finite set of numbers. This direct, or enumerative, approach is impossible to execute, however, for values of N of the order of magnitude of 20.

He gave a 'functional equation approach"

> The basic method is that of successive approximations. We choose an initial sequence $\{f_i^{(0)}\}$, and then proceed iteratively, setting
>
> $$f_i^{(k+1)} = \underset{j \neq i}{\text{Min}}\ (t_{ij} + f_j^{(k)}], \qquad i = 1, 2, \cdots, N-1,$$
>
> $$f_N^{(k+1)} = 0,$$
>
> for $k = 0, 1, 2 \cdots,$.

As initial function $f_i^{(0)}$ Bellman proposed (upon a suggestion of F. Haight) to take $f_i^{(0)} = t_{i,N}$ for all i. Bellman noticed that, for each fixed i, starting with this choice of $f_i^{(0)}$ gives that $f_i^{(k)}$ is monotonically nonincreasing in k, and stated:

> It is clear from the physical interpretation of this iterative scheme that at most $(N-1)$ iterations are required for the sequence to converge to the solution.

Since each iteration can be done in time $O(N^2)$, the algorithm takes time $O(N^3)$. As for the complexity, Bellman said:

> It is easily seen that the iterative scheme discussed above is a feasible method for either hand or machine computation for values of N of the order of magnitude of 50 or 100.

In a footnote, Bellman mentioned:

> *Added in proof* (*December* 1957): After this paper was written, the author was informed by Max Woodbury and George Dantzig that the particular iterative scheme discussed in Sec. 5 had been obtained by them from first principles.

Dantzig 1958

The paper of Dantzig [1958] gives an $O(n^2 \log n)$ algorithm for the shortest path problem with nonnegative length function. It consists of choosing in (10) an arc with $d(u) + l(u, v)$ as small as possible. Dantzig assumed

> (a) that one can write down without effort for each node the arcs leading to other nodes in increasing order of length and (b) that it is no effort to ignore an arc of the list if it leads to a node that has been reached earlier.

He mentioned that, beside Bellman, Moore, Ford, and himself, also D. Gale and D.R. Fulkerson proposed shortest path methods, 'in informal conversations'.

Dijkstra 1959

Dijkstra [1959] gave a concise and clean description of 'Dijkstra's method', yielding an $O(n^2)$-time implementation. Dijkstra stated:

> The solution given above is to be preferred to the solution by L.R. FORD [*3*] as described by C. BERGE [*4*], for, irrespective of the number of branches, we need not store the data for all branches simultaneously but only those for the branches in sets I and II, and this number is always less than n. Furthermore, the amount of work to be done seems to be considerably less.

(Dijkstra's references [3] and [4] are Ford [1956] and Berge [1958].)

Dijkstra's method is easier to implement (as an $O(n^2)$ algorithm) than Dantzig's, since we do not need to store the information in lists: in order to find a next vertex v minimizing $d(v)$, we can just scan all vertices.

Moore 1959

At the International Symposium on the Theory of Switching at Harvard University in April 1957, Moore [1959] of Bell Laboratories, presented a paper "The shortest path through a maze":

> The methods given in this paper require no foresight or ingenuity, and hence deserve to be called algorithms. They would be especially suited for use in a machine, either a special-purpose or a general-purpose digital computer.

The motivation of Moore was the routing of toll telephone traffic. He gave algorithms A, B, C, and D.

First, Moore considered the case of an undirected graph $G = (V, E)$ with no length function, in which a path from vertex A to vertex B should be found with a minimum number of edges. Algorithm A is: first give A label 0. Next do the following for $k = 0, 1, \ldots$: give label $k + 1$ to all unlabeled vertices that are adjacent to some vertex labeled k. Stop as soon as vertex B is labeled.

> If it were done as a program on a digital computer, the steps given as single steps above would be done serially, with a few operations of the computer for each city of the maze; but, in the case of complicated mazes, the algorithm would still be quite fast compared with trial-and-error methods.

In fact, a direct implementation of the method would yield an algorithm with running time $O(m)$. Algorithms B and C differ from A in a more economical labeling (by fewer bits).

Moore's algorithm D finds a shortest route for the case where each edge of the graph has a nonnegative length. This method is a refinement of Bellman's method described above: (i) it extends to the case that not all pairs of vertices have a direct connection; that is, if there is an underlying graph $G = (V, E)$ with length function; (ii) at each iteration only those $d_{i,j}$ are considered for which u_i has been decreased at the previous iteration.

The method has running time $O(nm)$. Moore observed that the algorithm is suitable for parallel implementation, yielding a decrease in running time bound to $O(n\Delta(G))$, where $\Delta(G)$ is the maximum degree of G. Moore concluded:

> The origin of the present methods provides an interesting illustration of the value of basic research on puzzles and games. Although such research is often frowned upon as being frivolous, it seems plausible that these algorithms might eventually lead to savings of very large sums of money by permitting more efficient use of congested transportation or communication systems. The actual problems in communication and transportation are so much complicated by timetables, safety requirements, signal-to-noise ratios, and economic requirements that in the past those seeking to solve them have not seen the basic simplicity of the problem, and have continued to use trial-and-error procedures which do not always give the true shortest path. However, in the case of a simple geometric maze, the absence of these confusing factors permitted algorithms A, B, and C to be obtained, and from them a large number of extensions, elaborations, and modifications are obvious.
>
> The problem was first solved in connection with Claude Shannon's maze-solving machine. When this machine was used with a maze which had more than one solution, a visitor asked why it had not been built to always find the shortest path. Shannon and I each attempted to find economical methods of doing this by machine. He found several methods suitable for analog computation, and I obtained these algorithms. Months later the applicability of these ideas to practical problems in communication and transportation systems was suggested.

Among the further applications of his method, Moore described the example of finding the fastest connections from one station to another in a given railroad timetable. A similar method was given by Minty [1958].

In May 1958, Hoffman and Pavley [1959] reported, at the Western Joint Computer Conference in Los Angeles, the following computing time for finding the distances between all pairs of vertices by Moore's algorithm (with nonnegative lengths):

> It took approximately three hours to obtain the minimum paths for a network of 265 vertices on an IBM *704*.

7. The traveling salesman problem

The *traveling salesman problem* (TSP) is: given n cities and their intermediate distances, find a shortest route traversing each city exactly once. Mathematically, the traveling salesman problem is related to, in fact generalizes, the question for a Hamiltonian circuit in a graph. This question goes back to Kirkman [1856] and Hamilton [1856,1858] and was also studied by Kowalewski [1917b,1917a] — see Biggs, Lloyd, and Wilson [1976]. We restrict our survey to the traveling salesman problem in its general form.

The mathematical roots of the traveling salesman problem are obscure. Dantzig, Fulkerson, and Johnson [1954] say:

> It appears to have been discussed informally among mathematicians at mathematics meetings for many years.

A 1832 manual

The traveling salesman problem has a natural interpretation, and Müller-Merbach [1983] detected that the problem was formulated in a 1832 manual for the successful traveling salesman, *Der Handlungsreisende — wie er sein soll und was er zu thun hat, um Aufträge zu erhalten und eines glücklichen Erfolgs in seinen Geschäften gewiß zu sein — von einem* *alten Commis-Voyageur*[20] [1832]. (Whereas the politically correct nowadays prefer to speak of the traveling sales*person* problem, the manual presumes that the 'Handlungsreisende' is male, and it warns about the risks of women in or out of business.)

The booklet contains no mathematics, and formulates the problem as follows:

> Die Geschäfte führen die Handlungsreisenden bald hier, bald dort hin, und es lassen sich nicht füglich Reisetouren angeben, die für alle vorkommende Fälle passend sind; aber es kann durch eine zweckmäßige Wahl und Eintheilung der Tour, manchmal so viel Zeit gewonnen werden, daß wir es nicht glauben umgehen zu dürfen, auch hierüber einige Vorschriften zu geben. Ein Jeder möge so viel davon benutzen, als er es seinem Zwecke für dienlich hält; so viel glauben wir aber davon versichern zu dürfen, daß es nicht wohl thunlich sein wird, die Touren durch Deutschland in Absicht der Entfernungen und, worauf der Reisende hauptsächlich zu sehen hat, des Hin- und Herreisens, mit mehr Oekonomie einzurichten. Die Hauptsache besteht immer darin: so viele Orte wie möglich mitzunehmen, ohne den nämlichen Ort zweimal berühren zu müssen.[21]

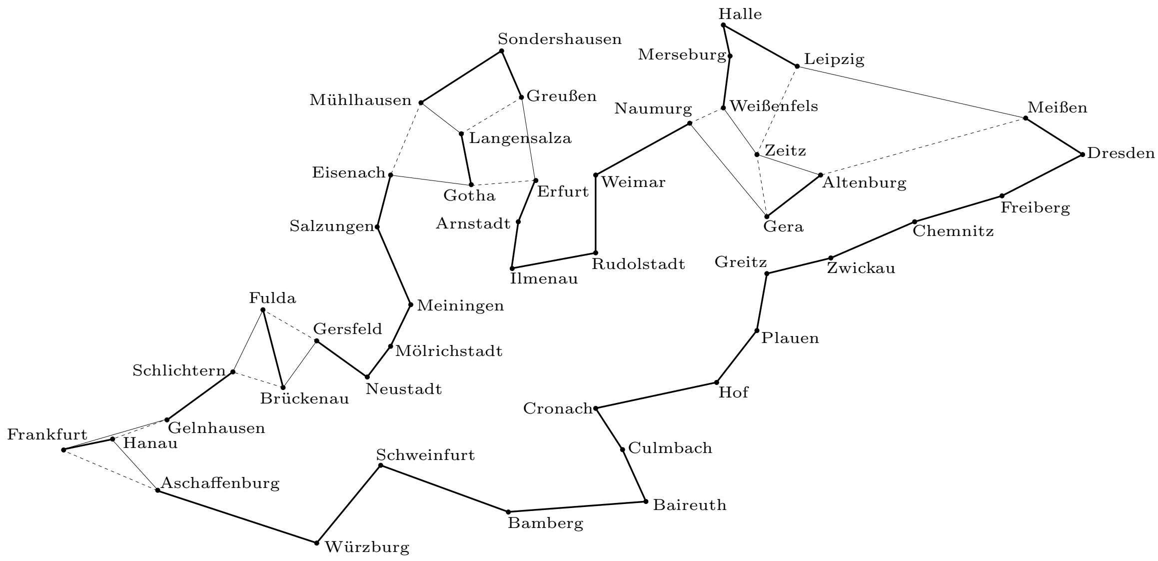


Figure 3

A tour along 45 German cities, as described in the 1832 traveling salesman manual, is given by the unbroken (bold and thin) lines (1285 km). A shortest tour is given by the unbroken bold and by the dashed lines (1248 km). We have taken geodesic distances — taking local conditions into account, the 1832 tour might be optimum.

The manual suggests five tours through Germany (one of them partly through Switzerland).

[20] "The traveling salesman — how he should be and what he has to do, to obtain orders and to be sure of a happy success in his business — by an old traveling salesman"

[21] Business brings the traveling salesman now here, then there, and no travel routes can be properly indicated that are suitable for all cases occurring; but sometimes, by an appropriate choice and arrangement of the tour, so much time can be gained, that we don't think we may avoid giving some rules also on this. Everybody may use that much of it, as he takes it for useful for his goal; so much of it however we think we may assure, that it will not be well feasible to arrange the tours through Germany with more economy in view of the distances and, which the traveler mainly has to consider, of the trip back and forth. The main point always consists of visiting as many places as possible, without having to touch the same place twice.

In Figure 3 we compare one of the tours with a shortest tour, found with ‘modern’ methods. (Most other tours given in the manual do not qualify for ‘die Hauptsache’ as they contain subtours, so that some places are visited twice.)

Menger’s Botenproblem 1930

K. Menger seems to be the first mathematician to have written about the traveling salesman problem. The root of his interest is given in his paper Menger [1928b]. In this, he studies the *length* $l(C)$ of a simple curve C in a metric space S, which is, by definition,

$$(12) \qquad l(C) := \sup \sum_{i=1}^{n-1} \mathrm{dist}(x_i, x_{i+1}),$$

where the supremum ranges over all choices of $x_1, \ldots, x_n$ on C *in the order determined by* C. What Menger showed is that we may relax this to finite subsets X of C and minimize over all possible orderings of X. To this end he defined, for any finite subset X of a metric space, $\lambda(X)$ to be the shortest length of a path through X (in graph terminology: a *Hamitonian path*), and he showed that

$$(13) \qquad l(C) = \sup_X \lambda(X),$$

where the supremum ranges over all finite subsets X of C. It amounts to showing that for each $\varepsilon > 0$ there is a finite subset X of C such that $\lambda(X) \geq l(C) - \varepsilon$.

Menger [1929a] sharpened this to:

$$(14) \qquad l(C) = \sup_X \kappa(X),$$

where again the supremum ranges over all finite subsets X of C, and where $\kappa(X)$ denotes the minimum length of a *spanning tree* on X.

These results were reported also in Menger [1930]. In a number of other papers, Menger [1928a,1929b,1929a] gave related results on these new characterizations of the length function.

The parameter $\lambda(X)$ clearly is close to the practical application of the traveling salesman problem. This relation was mentioned explicitly by Menger in the session of 5 February 1930 of his *mathematisches Kolloquium* in Vienna (organized at the desire of some students). According to the report in Menger [1931a,1932], he first asked if a further relaxation is possible by replacing $\kappa(X)$ by the minimum length of an (in current terminology) *Steiner tree* connecting X — a spanning tree on a superset of X in S. (So Menger toured along some basic combinatorial optimization problems.) This problem was solved for Euclidean spaces by Mimura [1933].

Next Menger posed the traveling salesman problem, as follows:

> Wir bezeichnen als *Botenproblem* (weil diese Frage in der Praxis von jedem Postboten, übrigens auch von vielen Reisenden zu lösen ist) die Aufgabe, für endlichviele Punkte, deren paarweise

> Abstände bekannt sind, den kürzesten die Punkte verbindenden Weg zu finden. Dieses Problem ist natürlich stets durch endlichviele Versuche lösbar. Regeln, welche die Anzahl der Versuche unter die Anzahl der Permutationen der gegebenen Punkte herunterdrücken würden, sind nicht bekannt. Die Regel, man solle vom Ausgangspunkt erst zum nächstgelegenen Punkt, dann zu dem diesem nächstgelegenen Punkt gehen usw., liefert im allgemeinen nicht den kürzesten Weg.[22]

So Menger asked for a shortest Hamiltonian path through the given points. He was aware of the complexity issue in the traveling salesman problem, and he knew that the now well-known nearest neighbour heuristic might not give an optimum solution.

Harvard, Princeton 1930-1934

Menger spent the period September 1930-February 1931 as visiting lecturer at Harvard University. In one of his seminar talks at Harvard, Menger presented his results on lengths of arcs and shortest paths through finite sets of points quoted above. According to Menger [1931b], a suggestion related to this was given by Hassler Whitney, who at that time did his Ph.D. research in graph theory at Harvard. This paper however does not mention if the practical interpretation was given in the seminar talk.

The year after, 1931-1932, Whitney was a National Research Council Fellow at Princeton University, where he gave a number of seminar talks. In a seminar talk, he mentioned the problem of finding the shortest route along the 48 States of America.

There are some uncertainties in this story. It is not sure if Whitney spoke about the 48 States problem during his 1931-1932 seminar talks (which talks he did give), or later, in 1934, as is said by Flood [1956] in his article on the traveling salesman problem:

> This problem was posed, in 1934, by Hassler Whitney in a seminar talk at Princeton University.

That memory can be shaky might be indicated by the following two quotes. Dantzig, Fulkerson, and Johnson [1954] remark:

> Both Flood and A.W. Tucker (Princeton University) recall that they heard about the problem first in a seminar talk by Hassler Whitney at Princeton in 1934 (although Whitney, recently queried, does not seem to recall the problem).

However, when asked by David Shmoys, Tucker replied in a letter of 17 February 1983 (see Hoffman and Wolfe [1985]):

> I cannot confirm or deny the story that I heard of the TSP from Hassler Whitney. If I did (as Flood says), it would have occurred in 1931-32, the first year of the old Fine Hall (now Jones Hall). That year Whitney was a postdoctoral fellow at Fine Hall working on Graph Theory, especially planarity and other offshoots of the 4-color problem. ... I was finishing my thesis with Lefschetz on n-manifolds and Merrill Flood was a first year graduate student. The Fine Hall Common Room was a very lively place — 24 hours a day.

[22] We denote by *messenger problem* (since in practice this question should be solved by each postman, anyway also by many travelers) the task to find, for finitely many points whose pairwise distances are known, the shortest route connecting the points. Of course, this problem is solvable by finitely many trials. Rules which would push the number of trials below the number of permutations of the given points, are not known. The rule that one first should go from the starting point to the closest point, then to the point closest to this, etc., in general does not yield the shortest route.

(Whitney finished his Ph.D. at Harvard University in 1932.)

Another uncertainty is in which form Whitney has posed the problem. That he might have focused on finding a shortest route along the 48 states in the U.S.A., is suggested by the reference by Flood, in an interview on 14 May 1984 with Tucker [1984], to the problem as the "48 States Problem of Hassler Whitney". In this respect Flood also remarked:

> I don't know who coined the peppier name 'Traveling Salesman Problem' for Whitney's problem, but that name certainly has caught on, and the problem has turned out to be of very fundamental importance.

TSP, Hamiltonian paths, and school bus routing

Flood [1956] mentioned a number of connections of the TSP with Hamiltonian games and Hamiltonian paths in graphs, and continues:

> I am indebted to A.W. Tucker for calling these connections to my attention, in 1937, when I was struggling with the problem in connection with a schoolbus routing study in New Jersey.

In the following quote from the interview by Tucker [1984], Flood referred to school bus routing in a different state (West Virginia), and he mentioned the involvement in the TSP of Koopmans, who spent 1940-1941 at the Local Government Surveys Section of Princeton University ("the Princeton Surveys"):

> Koopmans first became interested in the "48 States Problem" of Hassler Whitney when he was with me in the Princeton Surveys, as I tried to solve the problem in connection with the work by Bob Singleton and me on school bus routing for the State of West Virginia.

1940

In 1940, some papers appeared that study the traveling salesman problem, in a different context. They seem to be the first containing mathematical results on the problem.

In the American continuation of Menger's *mathematisches Kolloquium*, Menger [1940] returned to the question of the shortest path through a given set of points in a metric space, followed by investigations of Milgram [1940] on the shortest Jordan curve that covers a given, not necessarily finite, set of points in a metric space. As the set may be infinite, a shortest curve need not exist.

Fejes [1940] investigated the problem of a shortest curve through n points in the unit square. In consequence of this, Verblunsky [1951] showed that its length is less than $2 + \sqrt{2.8n}$. Later work in this direction includes Few [1955] and Beardwood, Halton, and Hammersley [1959].

Lower bounds on the expected value of a shortest path through n random points in the plane were studied by Mahalanobis [1940] in order to estimate the cost of a sample survey of the acreage under jute in Bengal. This survey took place in 1938 and one of the major costs in carrying out the survey was the transportation of men and equipment from one survey point to the next. He estimated (without proof) the minimum length of a tour along n random points in the plane, for Euclidean distance:

> It is also easy to see in a general way how the journey time is likely to behave. Let us suppose that n sampling units are scattered at random within any given area ; and let us assume

> that we may treat each such sample unit as a geometrical point. We may also assume that arrangements will usually be made to move from one sample point to another in such a way as to keep the total distance travelled as small as possible ; that is, we may assume that the path traversed in going from one sample point to another will follow a straight line. In this case it is easy to see that the mathematical expectation of the total length of the path travelled in moving from one sample point to another will be $(\sqrt{n} - 1/\sqrt{n})$. The cost of the journey from sample to sample will therefore be roughly proportional to $(\sqrt{n} - 1/\sqrt{n})$. When n is large, that is, when we consider a sufficiently large area, we may expect that the time required for moving from sample to sample will be roughly proportional to $\sqrt{n}$, where n is the total number of samples in the given area. If we consider the journey time per sq. mile, it will be roughly proportional to $\sqrt{y}$, where y is the density of number of sample units per sq. mile.

This research was continued by Jessen [1942], who estimated empirically a similar result for l_1-distance (Manhattan distance), in a statistical investigation of a sample survey for obtaining farm facts in Iowa:

> If a route connecting y points located at random in a fixed area is minimized, the total distance, D, of that route is[23]
>
> $$D = d\left(\frac{y-1}{\sqrt{y}}\right)$$
>
> where d is a constant.
>
> This relationship is based upon the assumption that points are connected by direct routes. In Iowa the road system is a quite regular network of mile square mesh. There are very few diagonal roads, therefore, routes between points resemble those taken on a checkerboard. A test wherein several sets of different members of points were located at random on an Iowa county road map, and the minimum distance of travel from a given point on the border of the county through all the points and to an end point (the county border nearest the last point on route), revealed that
>
> $$D = d\sqrt{y}$$
>
> works well. Here y is the number of randomized points (border points not included). This is of great aid in setting up a cost function.

Marks [1948] gave a proof of Mahalanobis' bound. In fact he showed that $\sqrt{\frac{1}{2}A}(\sqrt{n}-1/\sqrt{n})$ is a lower bound, where A is the area of the region. Ghosh [1949] showed that asymptotically this bound is close to the expected value, by giving a heuristic for finding a tour, yielding an upper bound of $1.27\sqrt{An}$. He also observed the complexity of the problem:

> After locating the n random points in a map of the region, it is very difficult to find out *actually* the shortest path connecting the points, unless the number n is very small, which is seldom the case for a large-scale survey.

TSP, transportation, and assignment

As is the case for many other combinatorial optimization problems, the RAND Corporation in Santa Monica, California, played an important role in the research on the TSP. Hoffman and Wolfe [1985] write that

> John Williams urged Flood in 1948 to popularize the TSP at the RAND Corporation, at least partly motivated by the purpose of creating intellectual challenges for models outside the theory of games. In fact, a prize was offered for a significant theorem bearing on the TSP. There is no doubt that the reputation and authority of RAND, which quickly became the intellectual center of much of operations research theory, amplified Flood's advertizing.

[23] at this point, Jessen referred in a footnote to Mahalanobis [1940].

At RAND, researchers considered the idea of transferring the successful methods for the transportation problem to the traveling salesman problem. Flood [1956] mentioned that this idea was brought to his attention by Koopmans in 1948. In the interview with Tucker [1984], Flood remembered:

> George Dantzig and Tjallings Koopmans met with me in 1948 in Washington, D.C., at the meeting of the International Statistical Institute, to tell me excitedly of their work on what is now known as the linear programming problem and with Tjallings speculating that there was a significant connection with the Traveling Salesman Problem.

(This meeting was in fact held 6–18 September 1947.)

The issue was taken up in a RAND Report by Julia Robinson [1949], who, in an 'unsuccessful attempt' to solve the traveling salesman problem, considered, as a relaxation, the assignment problem, for which she found a cycle reduction method. The relation is that the assignment problem asks for an optimum permutation, and the TSP for an optimum *cyclic* permutation.

Robinson's RAND report might be the earliest mathematical reference using the term 'traveling salesman problem':

```
The purpose of this note is to give a method for solving a problem related to the
traveling salesman problem.  One formulation is to find the shortest route for a
salesman starting from Washington, visiting all the state capitals and then
returning to Washington.  More generally, to find the shortest closed curve containing
n given points in the plane.
```

Flood wrote (in a letter of 17 May 1983 to E.L. Lawler) that Robinson's report stimulated several discussions on the TSP of him with his research assistant at RAND, D.R. Fulkerson, during 1950-1952[24].

It was noted by Beckmann and Koopmans [1952] that the TSP can be formulated as a quadratic assignment problem, for which however no fast methods are known.

Dantzig, Fulkerson, Johnson 1954

Fundamental progress on the traveling salesman was made in a seminal paper by the RAND researchers Dantzig, Fulkerson, and Johnson [1954] — according to Hoffman and Wolfe [1985] 'one of the principal events in the history of combinatorial optimization'. The paper introduced several new methods for solving the traveling salesman problem that are now basic in combinatorial optimization. In particular, it shows the importance of *cutting planes* for combinatorial optimization.

By a theorem of Birkhoff [1946], the convex hull of the $n \times n$ permutation matrices is precisely the set of doubly stochastic matrices — nonnegative matrices with all row and column sums equal to 1. In other words, the convex hull of the permutation matrices is determined by:

$$(15) \qquad x_{i,j} \geq 0 \text{ for all } i,j; \; \sum_{j=1}^{n} x_{i,j} = 1 \text{ for all } i; \; \sum_{i=1}^{n} x_{i,j} = 1 \text{ for all } j.$$

[24] Fulkerson started at RAND only in March 1951.

This makes it possible to solve the assignment problem as a linear programming problem. It is tempting to try the same approach to the traveling salesman problem. For this, one needs a description in linear inequalities of the *traveling salesman polytope* — the convex hull of the *cyclic* permutation matrices. To this end, one may add to (15) the following *subtour elimination constraints*:

$$(16) \qquad \sum_{i\in I, j\notin I} x_{i,j} \geq 1 \text{ for each } I \subseteq \{1,\ldots,n\} \text{ with } \emptyset \neq I \neq \{1,\ldots,n\}.$$

However, while these inequalities are enough to cut off the noncyclic permutation matrices from the polytope of doubly stochastic matrices, they yet do not yield all facets of the traveling salesman polytope (if $n \geq 5$), as was observed by Heller [1953a]: there exist doubly stochastic matrices, of any order $n \geq 5$, that satisfy (16) but are not a convex combination of cyclic permutation matrices.

The inequalities (16) can nevertheless be useful for the TSP, since we obtain a lower bound for the optimum tour length if we minimize over the constraints (15) and (16). This lower bound can be calculated with the simplex method, taking the (exponentially many) constraints (16) as *cutting planes* that can be added during the process when needed. In this way, Dantzig, Fulkerson, and Johnson were able to find the shortest tour along cities chosen in the 48 U.S. states and Washington, D.C. Incidentally, this is close to the problem mentioned by Julia Robinson in 1949 (and maybe also by Whitney in the 1930's).

The Dantzig-Fulkerson-Johnson paper does not give an algorithm, but rather gives a tour and proves its optimality with the help of the subtour elimination constraints. This work forms the basis for most of the later work on large-scale traveling salesman problems.

Early studies of the traveling salesman polytope were made by Heller [1953a,1953b, 1955a,1956b,1955b,1956a], Kuhn [1955a], Norman [1955], and Robacker [1955b], who also made computational studies of the probability that a random instance of the traveling salesman problem needs the constraints (16) (cf. Kuhn [1991]). This made Flood [1956] remark on the intrinsic complexity of the traveling salesman problem:

> Very recent mathematical work on the traveling-salesman problem by I. Heller, H.W. Kuhn, and others indicates that the problem is fundamentally complex. It seems very likely that quite a different approach from any yet used may be required for succesful treatment of the problem. In fact, there may well be no general method for treating the problem and impossibility results would also be valuable.

Flood mentioned a number of other applications of the traveling salesman problem, in particular in machine scheduling, brought to his attention in a seminar talk at Columbia University in 1954 by George Feeney.

Other work on the traveling salesman problem in the 1950's was done by Morton and Land [1955] (a linear programming approach with a 3-exchange heuristic), Barachet [1957] (a graphic solution method), Bock [1958], Croes [1958] (a heuristic), and Rossman and Twery [1958]. In a reaction to Barachet's paper, Dantzig, Fulkerson, and Johnson [1959] showed that their method yields the optimality of Barachet's (heuristically found) solution.

Acknowledgements. I thank Sasha Karzanov for his efficient help in finding Tolstoĭ's and several other papers in the (former) Lenin Library in Moscow, Irina V. Karzanova for accurately providing me with an English translation of Tolstoĭ's 1930 paper, Alexander Rosa for sending me a copy of Kotzig's thesis and for providing me with translations of excerpts of it, András Frank and Tibor Jordán for translating parts of Hungarian articles, Adri Steenbeek and Bill Cook for finding the shortest traveling salesman tour along the 45 German towns from the 1832 manual, Karin van Gemert and Wouter Mettrop at CWI's Library for providing me with bibliographic information and copies of numerous papers, Alfred B. Lehman for giving me copies of old reports of the Case Institute of Technology, Jan Karel Lenstra for giving me copies of letters of Albert Tucker to David Shmoys and of Merrill M. Flood to Eugene L. Lawler on TSP history, Alan Hoffman and David Williamson for helping me to understand Gleyzal's paper on transportation, Steve Brady (RAND) and Dick Cottle for their help in obtaining classical RAND Reports, Kim H. Campbell and Joanne McLean at Air Force Pentagon for declassifying the Harris-Ross report, Richard Bancroft and Gustave Shubert at RAND Corporation for their mediation in this, Bruno Simeone for sending me Salvemini's paper, and Truus Wanningen Koopmans for imparting to me her "Stories and Memories" and quotations from the diary of Tj.C. Koopmans.

References

[1996] K.S. Alexander, A conversation with Ted Harris, *Statistical Science* 11 (1996) 150–158.

[1928] P. Appell, *Le problème géométrique des déblais et remblais* [Mémorial des Sciences Mathématiques XXVII], Gauthier-Villars, Paris, 1928.

[1957] L.L. Barachet, Graphic solution to the traveling-salesman problem, *Operations Research* 5 (1957) 841–845.

[1957] T.E. Bartlett, An algorithm for the minimum number of transport units to maintain a fixed schedule, *Naval Research Logistics Quarterly* 4 (1957) 139–149.

[1957] T.E. Bartlett, A. Charnes, [Cyclic scheduling and combinatorial topology: assignment and routing of motive power to meet scheduling and maintenance requirements] Part II Generalization and analysis, *Naval Research Logistics Quarterly* 4 (1957) 207–220.

[1959] J. Beardwood, J.H. Halton, J.M. Hammersley, The shortest path through many points, *Proceedings of the Cambridge Philosophical Society* 55 (1959) 299–327.

[1952] M. Beckmann, T.C. Koopmans, *A Note on the Optimal Assignment Problem*, Cowles Commission Discussion Paper: Economics 2053, Cowles Commission for Research in Economics, Chicago, Illinois, [October 30] 1952.

[1953] M. Beckmann, T.C. Koopmans, *On Some Assignment Problems*, Cowles Commission Discussion Paper: Economics No. 2071, Cowles Commission for Research in Economics, Chicago, Illinois, [April 2] 1953.

[1956] M. Beckmann, C.B. McGuire, C.B. Winsten, *Studies in the Economics of Transportation*, Cowles Commission for Research in Economics, Yale University Press, New Haven, Connecticut, 1956.

[1958] R. Bellman, On a routing problem, *Quarterly of Applied Mathematics* 16 (1958) 87–90.

[1958] C. Berge, *Théorie des graphes et ses applications*, Dunod, Paris, 1958.

[1976] N.L. Biggs, E.K. Lloyd, R.J. Wilson, *Graph Theory 1736–1936*, Clarendon Press, Oxford, 1976.

[1946] G. Birkhoff, Tres observaciones sobre el algebra lineal, *Revista Facultad de Ciencias Exactas, Puras y Aplicadas Universidad Nacional de Tucuman, Serie A (Matematicas y Fisica Teorica)* 5 (1946) 147–151.

[1958] F. Bock, An algorithm for solving "travelling-salesman" and related network optimization problems [abstract], *Operations Research* 6 (1958) 897.

[1958] F. Bock, S. Cameron, Allocation of network traffic demand by instant determination of optimum paths [paper presented at the 13th National (6th Annual) Meeting of the Operations Research Society of America, Boston, Massachusetts, 1958], *Operations Research* 6 (1958) 633–634.

[1955a] A.W. Boldyreff, *Determination of the Maximal Steady State Flow of Traffic through a Railroad Network*, Research Memorandum RM-1532, The RAND Corporation, Santa Monica, California, [5 August] 1955 [published in *Journal of the Operations Research Society of America* 3 (1955) 443–465].

[1955b] A.W. Boldyreff, The gaming approach to the problem of flow through a traffic network [abstract of lecture presented at the Third Annual Meeting of the Society, New York, June 3–4, 1955], *Journal of the Operations Research Society of America* 3 (1955) 360.

[1926a] O. Borůvka, O jistém problému minimálním [Czech, with German summary; On a minimal problem], *Práce Moravské Přírodovědecké Společnosti Brno [Acta Societatis Scientiarum Naturalium Moravi[c]ae]* 3 (1926) 37–58.

[1926b] O. Borůvka, Příspěvek k řešení otázky ekonomické stavby elektrovodných sítí [Czech; Contribution to the solution of a problem of economical construction of electrical networks], *Elektrotechnický Obzor* 15:10 (1926) 153–154.

[1977] O. Borůvka, Několik vzpomínek na matematický život v Brně, *Pokroky Matematiky, Fyziky a Astronomie* 22 (1977) 91–99.

[1951] G.W. Brown, Iterative solution of games by fictitious play, in: *Activity Analysis of Production and Allocation — Proceedings of a Conference* (Proceedings Conference on Linear Programming, Chicago, Illinois, 1949; Tj.C. Koopmans, ed.), Wiley, New York, 1951, pp. 374–376.

[1950] G.W. Brown, J. von Neumann, Solutions of games by differential equations, in: *Contributions to the Theory of Games* (H.W. Kuhn, A.W. Tucker, eds.) [Annals of Mathematics Studies 24], Princeton University Press, Princeton, New Jersey, 1950, pp. 73–79.

[1938] G. Choquet, Étude de certains réseaux de routes, *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* 206 (1938) 310–313.

[1832] ['ein alter Commis-Voyageur'], *Der Handlungsreisende — wie er sein soll und was er zu thun hat, um Aufträge zu erhalten und eines glücklichen Erfolgs in seinen Geschäften gewiß zu sein — Von einem alten Commis-Voyageur*, B.Fr. Voigt, Ilmenau, 1832 [reprinted: Verlag Bernd Schramm, Kiel, 1981].

[1958] G.A. Croes, A method for solving traveling-salesman problems, *Operations Research* 6 (1958) 791–812.

[1951a] G.B. Dantzig, Application of the simplex method to a transportation problem, in: *Activity Analysis of Production and Allocation — Proceedings of a Conference* (Proceedings Conference on Linear Programming, Chicago, Illinois, 1949; Tj.C. Koopmans, ed.), Wiley, New York, 1951, pp. 359–373.

[1951b] G.B. Dantzig, Maximization of a linear function of variables subject to linear inequalities, in: *Activity Analysis of Production and Allocation — Proceedings of a Conference* (Proceedings Conference on Linear Programming, Chicago, Illinois, 1949; Tj.C. Koopmans, ed.), Wiley, New York, 1951, pp. 339–347.

[1957] G.B. Dantzig, Discrete-variable extremum problems, *Operations Research* 5 (1957) 266–277.

[1958] G.B. Dantzig, *On the Shortest Route through a Network*, Report P-1345, The RAND Corporation, Santa Monica, California, [April 12] 1958 [Revised April 29, 1959] [published in *Management Science* 6 (1960) 187–190].

[1954] G.B. Dantzig, D.R. Fulkerson, *Notes on Linear Programming: Part XV — Minimizing the Number of Carriers to Meet a Fixed Schedule*, Research Memorandum RM-1328, The RAND Corporation, Santa Monica, California, [24 August] 1954 [published in *Naval Research Logistics Quarterly* 1 (1954) 217–222].

[1955] G.B. Dantzig, D.R. Fulkerson, *On the Max Flow Min Cut Theorem of Networks*, Research Memorandum RM-1418, The RAND Corporation, Santa Monica, California, [1 January] 1955 [revised: Research Memorandum RM-1418-1 (= Paper P-826), The RAND Corporation, Santa Monica, California, [15 April] 1955 [published in in: *Linear Inequalities and Related Systems* (H.W. Kuhn, A.W. Tucker, eds.) [Annals of Mathematics Studies 38], Princeton University Press, Princeton, New Jersey, 1956, pp. 215–221]].

[1954] G. Dantzig, R. Fulkerson, S. Johnson, *Solution of a Large Scale Traveling Salesman Problem*, Paper P-510, The RAND Corporation, Santa Monica, California, [12 April] 1954 [published in *Journal of the Operations Research Society of America* 2 (1954) 393–410].

[1959] G.B. Dantzig, D.R. Fulkerson, S.M. Johnson, *On a Linear-Programming-Combinatorial Approach to the Traveling-Salesman Problem: Notes on Linear Programming and Extensions-Part 49*, Research Memorandum RM-2321, The RAND Corporation, Santa Monica, California, 1959 [published in *Operations Research* 7 (1959) 58–66].

[1959] E.W. Dijkstra, A note on two problems in connexion with graphs, *Numerische Mathematik* 1 (1959) 269–271.

[1954] P.S. Dwyer, Solution of the personnel classification problem with the method of optimal regions, *Psychometrika* 19 (1954) 11–26.

[1946] T.E. Easterfield, A combinatorial algorithm, *The Journal of the London Mathematical Society* 21 (1946) 219–226.

[1970] J. Edmonds, *Exponential growth of the simplex method for shortest path problems*, manuscript [University of Waterloo, Waterloo, Ontario], 1970.

[1931] J. Egerváry, Matrixok kombinatorius tulajdonságairól [Hungarian, with German summary], *Matematikai és Fizikai Lapok* 38 (1931) 16–28 [English translation [by H.W. Kuhn]: On combinatorial properties of matrices, Logistics Papers, George Washington University, issue 11 (1955), paper 4, pp. 1–11].

[1958] E. Egerváry, Bemerkungen zum Transportproblem, *MTW Mitteilungen* 5 (1958) 278–284.

[1956] P. Elias, A. Feinstein, C.E. Shannon, A note on the maximum flow through a network, *IRE Transactions on Information Theory* IT-2 (1956) 117–119.

[1940] L. Fejes, Über einen geometrischen Satz, *Mathematische Zeitschrift* 46 (1940) 83–85.

[1955] L. Few, The shortest path and the shortest road through n points, *Mathematika [London]* 2 (1955) 141–144.

[1956] M.M. Flood, The traveling-salesman problem, *Operations Research* 4 (1956) 61–75 [also in: *Operations Research for Management — Volume IICase Histories, Methods, Information Handling* (J.F. McCloskey, J.M. Coppinger, eds.), Johns Hopkins Press, Baltimore, Maryland, 1956, pp. 340–357].

[1951a] K. Florek, J. Łukaszewicz, J. Perkal, H. Steinhaus, S. Zubrzycki, Sur la liaison et la division des points d'un ensemble fini, *Colloquium Mathematicum* 2 (1951) 282–285.

[1951b] K. Florek, J. Łukaszewicz, J. Perkal, H. Steinhaus, S. Zubrzycki, Taksonomia Wrocławska [Polish, with English and Russian summaries], *Przeglad Antropologiczny* 17 (1951) 193–211.

[1956] L.R. Ford, Jr, *Network Flow Theory*, Paper P-923, The RAND Corporation, Santa Monica, California, [August 14,] 1956.

[1954] L.R. Ford, D.R. Fulkerson, *Maximal Flow through a Network*, Research Memorandum RM-1400, The RAND Corporation, Santa Monica, California, [19 November] 1954 [published in *Canadian Journal of Mathematics* 8 (1956) 399–404].

[1955] L.R. Ford, Jr, D.R. Fulkerson, *A Simple Algorithm for Finding Maximal Network Flows and an Application to the Hitchcock Problem*, Research Memorandum RM-1604, The RAND Corporation, Santa Monica, California, [29 December] 1955 [published in *Canadian Journal of Mathematics* 9 (1957) 210–218].

[1956a] L.R. Ford, Jr, D.R. Fulkerson, *A Primal Dual Algorithm for the Capacitated Hitchcock Problem* [Notes on Linear Programming: Part XXXIV], Research Memorandum RM-1798 [ASTIA Document Number AD 112372], The RAND Corporation, Santa Monica, California, [September 25] 1956 [published in *Naval Research Logistics Quarterly* 4 (1957) 47–54].

[1956b] L.R. Ford, Jr, D.R. Fulkerson, *Solving the Transportation Problem* [Notes on Linear Programming — Part XXXII], Research Memorandum RM-1736, The RAND Corporation, Santa Monica, California, [June 20] 1956 [published in *Management Science* 3 (1956-57) 24–32].

[1957] L.R. Ford, Jr, D.R. Fulkerson, *Construction of Maximal Dynamic Flows in Networks*, Paper P-1079 [= Research Memorandum RM-1981], The RAND Corporation, Santa Monica, California, [May 7,] 1957 [published in *Operations Research* 6 (1958) 419–433].

[1962] L.R. Ford, Jr, D.R. Fulkerson, *Flows in Networks*, Princeton University Press, Princeton, New Jersey, 1962.

[1951] M. Fréchet, Sur les tableaux de corrélation dont les marges sont données, *Annales de l'Université de Lyon, Section A, Sciences Mathématiques et Astronomie* (3) 14 (1951) 53–77.

[1912] F.G. Frobenius, Über Matrizen aus nicht negativen Elementen, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin* (1912) 456–477 [reprinted in: *Ferdinand Georg Frobenius, Gesammelte Abhandlungen*, Band III (J.-P. Serre, ed.), Springer, Berlin, 1968, pp. 546–567].

[1917] G. Frobenius, Über zerlegbare Determinanten, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin* (1917) 274–277 [reprinted in: *Ferdinand Georg Frobenius, Gesammelte Abhandlungen*, Band III (J.-P. Serre, ed.), Springer, Berlin, 1968, pp. 701–704].

[1958] D.R. Fulkerson, *Notes on Linear Programming: Part XLVI — Bounds on the Primal-Dual Computation for Transportation Problems*, Research Memorandum RM-2178, The RAND Corporation, Santa Monica, California, 1958.

[1958] T. Gallai, Maximum-minimum Sätze über Graphen, *Acta Mathematica Academiae Scientiarum Hungaricae* 9 (1958) 395–434.

[1978] T. Gallai, The life and scientific work of Dénes Kőnig (1884–1944), *Linear Algebra and Its Applications* 21 (1978) 189–205.

[1949] M.N. Ghosh, Expected travel among random points in a region, *Calcutta Statistical Association Bulletin* 2 (1949) 83–87.

[1955] A. Gleyzal, An algorithm for solving the transportation problem, *Journal of Research National Bureau of Standards* 54 (1955) 213–216.

[1985] R.L. Graham, P. Hell, On the history of the minimum spanning tree problem, *Annals of the History of Computing* 7 (1985) 43–57.

[1938] T. Grünwald, Ein neuer Beweis eines Mengerschen Satzes, *The Journal of the London Mathematical Society* 13 (1938) 188–192.

[1934] G. Hajós, Zum Mengerschen Graphensatz, *Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae, Sectio Scientiarum Mathematicarum [Szeged]* 7 (1934-35) 44–47.

[1856] W.R. Hamilton, Memorandum respecting a new system of roots of unity (the Icosian calculus), *Philosophical Magazine* 12 (1856) 446 *Proceedings of the Royal Irish Academy* 6 (1858) 415–416 [reprinted in: *The Mathematical Papers of Sir William Rowan Hamilton — Vol. III Algebra* (H. Halberstam, R.E. Ingram, eds.), Cambridge University Press, Cambridge, 1967, p. 610].

[1858] W.R. Hamilton, On a new system of roots of unity, *Proceedings of the Royal Irish Academy* 6 (1858) 415–416 [reprinted in: *The Mathematical Papers of Sir William Rowan Hamilton — Vol. III Algebra* (H. Halberstam, R.E. Ingram, eds.), Cambridge University Press, Cambridge, 1967, p. 609].

[1955] T.E. Harris, F.S. Ross, *Fundamentals of a Method for Evaluating Rail Net Capacities*, Research Memorandum RM-1573, The RAND Corporation, Santa Monica, California, [October 24,] 1955.

[1953a] I. Heller, On the problem of shortest path between points. I [abstract], *Bulletin of the American Mathematical Society* 59 (1953) 551.

[1953b] I. Heller, On the problem of shortest path between points. II [abstract], *Bulletin of the American Mathematical Society* 59 (1953) 551–552.

[1955a] I. Heller, Geometric characterization of cyclic permutations [abstract], *Bulletin of the American Mathematical Society* 61 (1955) 227.

[1955b] I. Heller, Neighbor relations on the convex of cyclic permutations, *Bulletin of the American Mathematical Society* 61 (1955) 440.

[1956a] I. Heller, Neighbor relations on the convex of cyclic permutations, *Pacific Journal of Mathematics* 6 (1956) 467–477.

[1956b] I. Heller, On the travelling salesman's problem, in: *Proceedings of the Second Symposium in Linear Programming* (Washington, D.C., 1955; H.A. Antosiewicz, ed.), Vol. 2, National Bureau of Standards, U.S. Department of Commerce, Washington, D.C., 1956, pp. 643–665.

[1941] F.L. Hitchcock, The distribution of a product from several sources to numerous localities, *Journal of Mathematics and Physics* 20 (1941) 224–230.

[1959] W. Hoffman, R. Pavley, Applications of digital computers to problems in the study of vehicular traffic, in: *Proceedings of the Western Joint Computer Conference* (Los Angeles, California, 1958), American Institute of Electrical Engineers, New York, 1959, pp. 159–161.

[1985] A.J. Hoffman, P. Wolfe, History, in: *The Traveling Salesman Problem — A Guided Tour of Combinatorial Optimization* (E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan, D.B. Shmoys, eds.), Wiley, Chichester, 1985, pp. 1–15.

[1955] E. Jacobitti, Automatic alternate routing in the 4A crossbar system, *Bell Laboratories Record* 33 (1955) 141–145.

[1930] V. Jarník, O jistém problému minimálním (Z dopisu panu O. Borůvkovi) [Czech; On a minimal problem (from a letter to Mr Borůvka)], *Práce Moravské Přírodovědecké Společnosti Brno [Acta Societatis Scientiarum Naturalium Moravicae]* 6 (1930-31) 57–63.

[1934] V. Jarník, M. Kössler, O minimálních grafech, obsahujíciích n daných bodů, *Časopis pro Pěstování Matematiky a Fysiky* 63 (1934) 223–235.

[1942] R.J. Jessen, *Statistical Investigation of a Sample Survey for Obtaining Farm Facts* Research Bulletin 304, Iowa State College of Agriculture and Mechanic Arts, Ames, Iowa, 1942.

[1973a] D.B. Johnson, A note on Dijkstra's shortest path algorithm, *Journal of the Association for Computing Machinery* 20 (1973) 385–388.

[1973b] D.B. Johnson, *Algorithms for Shortest Paths*, Ph.D. Thesis [Technical Report CU-CSD-73-169, Department of Computer Science], Cornell University, Ithaca, New York, 1973.

[1977] D.B. Johnson, Efficient algorithms for shortest paths in sparse networks, *Journal of the Association for Computing Machinery* 24 (1977) 1–13.

[1939] L.V. Kantorovich, *Matematicheskie metody organizatsii i planirovaniia proizvodstva* [Russian], Publication House of the Leningrad State University, Leningrad, 1939 [reprinted (with minor changes) in: *Primenenie matematiki v èkonomicheskikh issledovaniyakh* [Russian; Application of Mathematics in Economical Studies] (V.S. Nemchinov, ed.), Izdatel'stvo Sotsial'no-Èkonomicheskoĭ Literatury, Moscow, 1959, pp. 251–309] [English translation: Mathematical methods of organizing and planning production, *Management Science* 6 (1959-60) 366–422 [also in: *The Use of Mathematics in Economics* (V.S. Nemchinov, ed.), Oliver and Boyd, Edinburgh, 1964, pp. 225–279]].

[1940] L.V. Kantorovich, Ob odnom èffektivnom metode resheniya nekotorykh klassov èkstremal'nykh problem [Russian], *Doklady Akademii Nauk SSSR* 28 (1940) 212–215 [English translation: An effective method for solving some classes of extremal problems, *Comptes Rendus (Doklady) de l'Académie des Sciences de l'U.R.S.S.* 28 (1940) 211–214].

[1942] L.V. Kantorovich, O peremeshchenii mass [Russian], *Doklady Akademii Nauk SSSR* 37:7-8 (1942) 227–230 [English translation: On the translocation of masses, *Comptes Rendus (Doklady) de l'Académie des Sciences de l'U.R.S.S.* 37 (1942) 199–201 [reprinted: *Management Science* 5 (1958) 1–4]].

[1987] L.V. Kantorovich, Moĭ put' v nauke (Predpolagavshiĭsya doklad v Moskovskom matematicheskom obshchestve) [Russian; My journey in science (proposed report to the Moscow Mathematical Society)], *Uspekhi Matematicheskikh Nauk* 42:2 (1987) 183–213 [English translation: *Russian Mathematical Surveys* 42:2 (1987) 233–270 [reprinted in: *Functional Analysis, Optimization, and Mathematical Economics, A Collection of Papers Dedicated to the Memory of Leonid Vital'evich Kantorovich* (L.J. Leifman, ed.), Oxford University Press, New York, 1990, pp. 8–45]; also in: *L.V. Kantorovich Selected Works Part I* (S.S. Kutateladze, ed.), Gordon and Breach, Amsterdam, 1996, pp. 17–54].

[1949] L.V. Kantorovich, M.K. Gavurin, Primenenie matematicheskikh metodov v voprosakh analiza gruzopotokov [Russian; The application of mathematical methods to freight flow analysis], in: *Problemy povysheniya effectivnosti raboty transporta* [Russian; Collection of Problems of Raising the Efficiency of Transport Performance], Akademiia Nauk SSSR, Moscow-Leningrad, 1949, pp. 110–138.

[1856] T.P. Kirkman, On the representation of polyhedra, *Philosophical Transactions of the Royal Society of London Series A* 146 (1856) 413–418.

[1930] B. Knaster, Sui punti regolari nelle curve di Jordan, in: *Atti del Congresso Internazionale dei Matematici* [Bologna 3–10 Settembre 1928] Tomo II, Nicola Zanichelli, Bologna, [1930,] pp. 225–227.

[1915] D. Kőnig, Vonalrendszerek és determinánsok [Hungarian; Line systems and determinants], *Mathematikai és Természettudományi Értesitő* 33 (1915) 221–229.

[1916] D. Kőnig, Graphok és alkalmazásuk a determinánsok és a halmazok elméletére [Hungarian], *Mathematikai és Természettudományi Értesitő* 34 (1916) 104–119 [German translation: Über Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre, *Mathematische Annalen* 77 (1916) 453–465].

[1923] D. Kőnig, Sur un problème de la théorie générale des ensembles et la théorie des graphes [Communication faite, le 7 avril 1914, au Congrès de Philosophie mathématique à Paris], *Revue de Métaphysique et de Morale* 30 (1923) 443–449.

[1931] D. Kőnig, Graphok és matrixok [Hungarian; Graphs and matrices], *Matematikai és Fizikai Lapok* 38 (1931) 116–119.

[1932] D. Kőnig, Über trennende Knotenpunkte in Graphen (nebst Anwendungen auf Determinanten und Matrizen), *Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae, Sectio Scientiarum Mathematicarum [Szeged]* 6 (1932-34) 155–179.

[1939] T. Koopmans, *Tanker Freight Rates and Tankship Building — An Analysis of Cyclical Fluctuations*, Publication Nr 27, Netherlands Economic Institute, De Erven Bohn, Haarlem, 1939.

[1942] Tj.C. Koopmans, Exchange ratios between cargoes on various routes (non-refrigerating dry cargoes), Memorandum for the Combined Shipping Adjustment Board, Washington D.C., 1942, 1–12 [first published in: *Scientific Papers of Tjalling C. Koopmans*, Springer, Berlin, 1970, pp. 77–86].

[1948] Tj.C. Koopmans, Optimum utilization of the transportation system, in: *The Econometric Society Meeting* (Washington, D.C., 1947; D.H. Leavens, ed.) [Proceedings of the International Statistical Conferences — Volume V], 1948, pp. 136–146 [reprinted in: *Econometrica* 17 (Supplement) (1949) 136–146] [reprinted in: *Scientific Papers of Tjalling C. Koopmans*, Springer, Berlin, 1970, pp. 184–193].

[1959] Tj.C. Koopmans, A note about Kantorovich's paper, "Mathematical methods of organizing and planning production", *Management Science* 6 (1959-60) 363–365.

[1992] Tj.C. Koopmans, [autobiography] in: *Nobel Lectures including presentation speeches and laureates' biographies — Economic Sciences 1969—1980* (A. Lindbeck, ed.), World Scientific, Singapore, 1992, pp. 233–238.

[1949a] T.C. Koopmans, S. Reiter, *Allocation of Resources in Production, I*, Cowles Commission Discussion Paper. Economics: No. 264, Cowles Commission for Research in Economics, Chicago, Illinois, [May 4] 1949.

[1949b] T.C. Koopmans, S. Reiter, *Allocation of Resources in Production II Application to Transportation*, Cowles Commission Discussion Paper: Economics: No. 264A, Cowles Commission for Research in Economics, Chicago, Illinois, [May 19] 1949.

[1951] Tj.C. Koopmans, S. Reiter, A model of transportation, in: *Activity Analysis of Production and Allocation — Proceedings of a Conference* (Proceedings Conference on Linear Programming, Chicago, Illinois, 1949; Tj.C. Koopmans, ed.), Wiley, New York, 1951, pp. 222–259.

[2001] B. Korte, J. Nešetřil, Vojtěch Jarník's work in combinatorial optimization, *Discrete Mathematics* 235 (2001) 1–17.

[1956] A. Kotzig, *Súvislosť a Pravidelná Súvislosť Konečných Grafov* [Slovak; Connectivity and Regular Connectivity of Finite Graphs], Academical Doctorate Dissertation, Vysoká Škola Ekonomická, Bratislava, [September] 1956.

[1917a] A. Kowalewski, Topologische Deutung von Buntordnungsproblemen, *Sitzungsberichte Kaiserliche Akademie der Wissenschaften in Wien Mathematisch-naturwissenschaftliche Klasse Abteilung IIa* 126 (1917) 963–1007.

[1917b] A. Kowalewski, W.R. Hamilton's Dodekaederaufgabe als Buntordnungsproblem, *Sitzungsberichte Kaiserliche Akademie der Wissenschaften in Wien Mathematisch-naturwissenschaftliche Klasse Abteilung IIa* 126 (1917) 67–90.

[1956] J.B. Kruskal, Jr, On the shortest spanning subtree of a graph and the traveling salesman problem, *Proceedings of the American Mathematical Society* 7 (1956) 48–50.

[1997] J.B. Kruskal, A reminiscence about shortest spanning subtrees, *Archivum Mathematicum (Brno)* 33 (1997) 13–14.

[1955a] H.W. Kuhn, On certain convex polyhedra [abstract], *Bulletin of the American Mathematical Society* 61 (1955) 557–558.

[1955b] H.W. Kuhn, The Hungarian method for the assignment problem, *Naval Research Logistics Quarterly* 2 (1955) 83–97.

[1956] H.W. Kuhn, Variants of the Hungarian method for assignment problems, *Naval Research Logistics Quarterly* 3 (1956) 253–258.

[1991] H.W. Kuhn, On the origin of the Hungarian method, in: *History of Mathematical Programming — A Collection of Personal Reminiscences* (J.K. Lenstra, A.H.G. Rinnooy Kan, A. Schrijver, eds.), CWI, Amsterdam and North-Holland, Amsterdam, 1991, pp. 77–81.

[1954] A.H. Land, A problem in transportation, in: *Conference on Linear Programming May 1954* (London, 1954), Ferranti Ltd., London, 1954, pp. 20–31.

[1947] H.D. Landahl, A matrix calculus for neural nets: II, *Bulletin of Mathematical Biophysics* 9 (1947) 99–108.

[1946] H.D. Landahl, R. Runge, Outline of a matrix algebra for neural nets, *Bulletin of Mathematical Biophysics* 8 (1946) 75–81.

[1957] M. Leyzorek, R.S. Gray, A.A. Johnson, W.C. Ladew, S.R. Meaker, Jr, R.M. Petry, R.N. Seitz, *Investigation of Model Techniques — First Annual Report — 6 June 1956 – 1 July 1957 — A Study of Model Techniques for Communication Systems*, Case Institute of Technology, Cleveland, Ohio, 1957.

[1957] H. Loberman, A. Weinberger, Formal procedures for connecting terminals with a minimum total wire length, *Journal of the Association for Computing Machinery* 4 (1957) 428–437.

[1952] F.M. Lord, Notes on a problem of multiple classification, *Psychometrika* 17 (1952) 297–304.

[1882] É. Lucas, *Récréations mathématiques, deuxième édition*, Gauthier-Villars, Paris, 1882–1883.

[1950] R.D. Luce, Connectivity and generalized cliques in sociometric group structure, *Psychometrika* 15 (1950) 169–190.

[1949] R.D. Luce, A.D. Perry, A method of matrix analysis of group structure, *Psychometrika* 14 (1949) 95–116.

[1950] A.G. Lunts, Prilozhen ie matrichnoĭ bulevskoĭ algebry k analizu i sintezu releĭno-kontaktiykh skhem [Russian; Application of matrix Boolean algebra to the analysis and synthesis of relay-contact schemes], *Doklady Akademii Nauk SSSR (N.S.)* 70 (1950) 421–423.

[1952] A.G. Lunts, Algebraicheskie metody analiza i sinteza kontaktiykh skhem [Russian; Algebraic methods of analysis and synthesis of relay contact networks], *Izvestiya Akademii Nauk SSSR, Seriya Matematicheskaya* 16 (1952) 405–426.

[1940] P.C. Mahalanobis, A sample survey of the acreage under jute in Bengal, *Sankhyā* 4 (1940) 511–530.

[1948] E.S. Marks, A lower bound for the expected travel among m random points, *The Annals of Mathematical Statistics* 19 (1948) 419–422.

[1927] K. Menger, Zur allgemeinen Kurventheorie, *Fundamenta Mathematicae* 10 (1927) 96–115.

[1928a] K. Menger, Die Halbstetigkeit der Bogenlänge, *Anzeiger — Akademie der Wissenschaften in Wien — Mathematisch-naturwissenschaftliche Klasse* 65 (1928) 278–281.

[1928b] K. Menger, Ein Theorem über die Bogenlänge, *Anzeiger — Akademie der Wissenschaften in Wien — Mathematisch-naturwissenschaftliche Klasse* 65 (1928) 264–266.

[1929a] K. Menger, Eine weitere Verallgemeinerung des Längenbegriffes, *Anzeiger — Akademie der Wissenschaften in Wien — Mathematisch-naturwissenschaftliche Klasse* 66 (1929) 24–25.

[1929b] K. Menger, Über die neue Definition der Bogenlänge, *Anzeiger — Akademie der Wissenschaften in Wien — Mathematisch-naturwissenschaftliche Klasse* 66 (1929) 23–24.

[1930] K. Menger, Untersuchungen über allgemeine Metrik. Vierte Untersuchung. Zur Metrik der Kurven, *Mathematische Annalen* 103 (1930) 466–501.

[1931a] K. Menger, Bericht über ein mathematisches Kolloquium, *Monatshefte für Mathematik und Physik* 38 (1931) 17–38.

[1931b] K. Menger, Some applications of point-set methods, *Annals of Mathematics* (2) 32 (1931) 739–760.

[1932] K. Menger, Eine neue Definition der Bogenlänge, *Ergebnisse eines Mathematischen Kolloquiums* 2 (1932) 11–12.

[1940] K. Menger, On shortest polygonal approximations to a curve, *Reports of a Mathematical Colloquium* (2) 2 (1940) 33–38.

[1981] K. Menger, On the origin of the n-arc theorem, *Journal of Graph Theory* 5 (1981) 341–350.

[1940] A.N. Milgram, On shortest paths through a set, *Reports of a Mathematical Colloquium* (2) 2 (1940) 39–44.

[1933] Y. Mimura, Über die Bogenlänge, *Ergebnisse eines Mathematischen Kolloquiums* 4 (1933) 20–22.

[1957] G.J. Minty, A comment on the shortest-route problem, *Operations Research* 5 (1957) 724.

[1958] G.J. Minty, A variant on the shortest-route problem, *Operations Research* 6 (1958) 882–883.

[1784] G. Monge, Mémoire sur la théorie des déblais et des remblais, *Histoire de l'Académie Royale des Sciences* [année 1781. Avec les Mémoires de Mathématique & de Physique, pour la même Année] (2e partie) (1784) [*Histoire:* 34–38, *Mémoire:*] 666–704.

[1959] E.F. Moore, The shortest path through a maze, in: *Proceedings of an International Symposium on the Theory of Switching, 2–5 April 1957, Part II* [The Annals of the Computation Laboratory of Harvard University Volume XXX] (H. Aiken, ed.), Harvard University Press, Cambridge, Massachusetts, 1959, pp. 285–292.

[1955] G. Morton, A. Land, A contribution to the 'travelling-salesman' problem, *Journal of the Royal Statistical Society Series B* 17 (1955) 185–194.

[1983] H. Müller-Merbach, Zweimal travelling Salesman, *DGOR-Bulletin* 25 (1983) 12–13.

[1957] J. Munkres, Algorithms for the assignment and transportation problems, *Journal of the Society for Industrial and Applied Mathematics* 5 (1957) 32–38.

[1951] J. von Neumann, *The Problem of Optimal Assignment and a Certain 2-Person Game*, unpublished manuscript, [October 26] 1951.

[1953] J. von Neumann, A certain zero-sum two-person game equivalent to the optimal assignment problem, in: *Contributions to the Theory of Games Volume II* (H.W. Kuhn, A.W. Tucker, eds.) [Annals of Mathematics Studies 28], Princeton University Press, Princeton, New Jersey, 1953, pp. 5–12 [reprinted in: *John von Neumann, Collected Works, Volume VI* (A.H. Taub, ed.), Pergamon Press, Oxford, 1963, pp. 44–49].

[1932] G. Nöbeling, Eine Verschärfung des n-Beinsatzes, *Fundamenta Mathematicae* 18 (1932) 23–38.

[1955] R.Z. Norman, On the convex polyhedra of the symmetric traveling salesman problem [abstract], *Bulletin of the American Mathematical Society* 61 (1955) 559.

[1955] A. Orden, The transhipment problem, *Management Science* 2 (1955-56) 276–285.

[1947] Z.N. Pariĭskaya, A.N. Tolstoĭ, A.B. Mots, *Planirovanie Tovarnykh Perevozok — Metody Opredeleniya Ratsionaljiykh Puteĭ Tovarodvizheniya* [Russian; Planning Goods Transportation — Methods of Determining Efficient Routes of Goods Traffic], Gostorgizdat, Moscow, 1947.

[1957] W. Prager, A generalization of Hitchcock's transportation problem, *Journal of Mathematics and Physics* 36 (1957) 99–106.

[1957] R.C. Prim, Shortest connection networks and some generalizations, *The Bell System Technical Journal* 36 (1957) 1389–1401.

[1957] R. Rado, Note on independence functions, *Proceedings of the London Mathematical Society* (3) 7 (1957) 300–320.

[1955a] J.T. Robacker, *On Network Theory*, Research Memorandum RM-1498, The RAND Corporation, Santa Monica, California, [May 26,] 1955.

[1955b] J.T. Robacker, *Some Experiments on the Traveling-Salesman Problem*, Research Memorandum RM-1521, The RAND Corporation, Santa Monica, California, [28 July] 1955.

[1956] J.T. Robacker, *Min-Max Theorems on Shortest Chains and Disjoint Cuts of a Network*, Research Memorandum RM-1660, The RAND Corporation, Santa Monica, California, [12 January] 1956.

[1949] J. Robinson, *On the Hamiltonian Game (A Traveling Salesman Problem)*, Research Memorandum RM-303, The RAND Corporation, Santa Monica, California, [5 December] 1949.

[1950] J. Robinson, *A Note on the Hitchcock-Koopmans Problem*, Research Memorandum RM-407, The RAND Corporation, Santa Monica, California, [15 June] 1950.

[1951] J. Robinson, An iterative method of solving a game, *Annals of Mathematics* 54 (1951) 296–301 [reprinted in: *The Collected Works of Julia Robinson* (S. Feferman, ed.), American Mathematical Society, Providence, Rhode Island, 1996, pp. 41–46].

[1956] L. Rosenfeld, Unusual problems and their solutions by digital computer techniques, in: *Proceedings of the Western Joint Computer Conference* (San Francisco, California, 1956), The American Institute of Electrical Engineers, New York, 1956, pp. 79–82.

[1958] M.J. Rossman, R.J. Twery, A solution to the travelling salesman problem by combinatorial programming [abstract], *Operations Research* 6 (1958) 897.

[1927] N.E. Rutt, Concerning the cut points of a continuous curve when the arc curve, ab, contains exactly n independent arcs [abstract], *Bulletin of the American Mathematical Society* 33 (1927) 411.

[1929] N.E. Rutt, Concerning the cut points of a continuous curve when the arc curve, AB, contains exactly N independent arcs, *American Journal of Mathematics* 51 (1929) 217–246.

[1939] T. Salvemini, Sugl'indici di omofilia, *Supplemento Statistico* 5 (Serie II) (1939) [= Atti della Prima Riunione Scientifica della Società Italiana di Statistica, Pisa, 1939] 105–115 [English translation: On the indexes of homophilia, in: *Tommaso Salvemini — Scritti Scelti*, Cooperativa Informazione Stampa Universitaria, Rome, 1981, pp. 525–537].

[1951] A. Shimbel, Applications of matrix algebra to communication nets, *Bulletin of Mathematical Biophysics* 13 (1951) 165–178.

[1953] A. Shimbel, Structural parameters of communication networks, *Bulletin of Mathematical Biophysics* 15 (1953) 501–507.

[1955] A. Shimbel, Structure in communication nets, in: *Proceedings of the Symposium on Information Networks* (New York, 1954), Polytechnic Press of the Polytechnic Institute of Brooklyn, Brooklyn, New York, 1955, pp. 199–203.

[1895] G. Tarry, Le problème des labyrinthes, *Nouvelles Annales de Mathématiques* (3) 14 (1895) 187–190 [English translation in: N.L. Biggs, E.K. Lloyd, R.J. Wilson, *Graph Theory 1736–1936*, Clarendon Press, Oxford, 1976, pp. 18–20].

[1951] R. Taton, *L'Œuvre scientifique de Monge*, Presses universitaires de France, Paris, 1951.

[1950] R.L. Thorndike, The problem of the classification of personnel, *Psychometrika* 15 (1950) 215–235.

[1934] J. Tinbergen, Scheepsruimte en vrachten, *De Nederlandsche Conjunctuur* (1934) maart 23–35.

[1930] A.N. Tolstoĭ, Metody nakhozhdeniya naimen'shego summovogo kilometrazha pri planirovanii perevozok v prostranstve [Russian; Methods of finding the minimal total kilometrage in cargo-transportation planning in space], in: *Planirovanie Perevozok, Sbornik pervyĭ* [Russian; Transportation Planning, Volume I], Transpechat' NKPS [TransPress of the National Commissariat of Transportation], Moscow, 1930, pp. 23–55.

[1939] A. Tolstoĭ, Metody ustraneniya neratsional'nykh perevozok pri planirovanii [Russian; Methods of removing irrational transportation in planning], *Sotsialisticheskiĭ Transport* 9 (1939) 28–51 [also published as 'pamphlet': *Metody ustraneniya neratsional'nykh perevozok pri sostavlenii operativnykh planov* [Russian; Methods of Removing Irrational Transportation in the Construction of Operational Plans], Transzheldorizdat, Moscow, 1941].

[1953] L. Törnqvist, *How to Find Optimal Solutions to Assignment Problems*, Cowles Commission Discussion Paper: Mathematics No. 424, Cowles Commission for Research in Economics, Chicago, Illinois, [August 3] 1953.

[1952] D.L. Trueblood, The effect of travel time and distance on freeway usage, *Public Roads* 26 (1952) 241–250.

[1984] Albert Tucker, Merrill Flood (with Albert Tucker) — This is an interview of Merrill Flood in San Francisco on 14 May 1984, in: *The Princeton Mathematics Community in the 1930s — An Oral-History Project* [located at Princeton University in the Seeley G. Mudd Manuscript Library web at the URL: `http://www.princeton.edu/mudd/math`], Transcript Number 11 (PMC11), 1984.

[1951] S. Verblunsky, On the shortest path through a number of points, *Proceedings of the American Mathematical Society* 2 (1951) 904–913.

[1952] D.F. Votaw, Jr, Methods of solving some personnel-classification problems, *Psychometrika* 17 (1952) 255–266.

[1952] D.F. Votaw, Jr, A. Orden, The personnel assignment problem, in: *Symposium on Linear Inequalities and Programming* [Scientific Computation of Optimum Programs, Project SCOOP, No. 10] (Washington, D.C., 1951; A. Orden, L. Goldstein, eds.), Planning Research Division, Director of Management Analysis Service, Comptroller, Headquarters U.S. Air Force, Washington, D.C., 1952, pp. 155–163.

[1995] T. Wanningen Koopmans, *Stories and Memories*, type set manuscript, [May] 1995.

[1932] H. Whitney, Congruent graphs and the connectivity of graphs, *American Journal of Mathematics* 54 (1932) 150–168 [reprinted in: *Hassler Whitney Collected Works Volume I* (J. Eells, D. Toledo, eds.), Birkhäuser, Boston, Massachusetts, 1992, pp. 61–79].

[1873] Chr. Wiener, Ueber eine Aufgabe aus der Geometria situs, *Mathematische Annalen* 6 (1873) 29–30.

[1973] N. Zadeh, A bad network problem for the simplex method and other minimum cost flow algorithms, *Mathematical Programming* 5 (1973) 255–266.