On the history of the transportation and maximum flow problems

Alexander Schrijver[1]

Abstract. We review two papers that are of historical interest for combinatorial optimization: an article of A.N. Tolstoĭ from 1930, in which the transportation problem is studied, and a negative cycle criterion is developed and applied to solve a (for that time) large-scale (10×68) transportation problem to optimality; and an, until recently secret, RAND report of T.E. Harris and F.S. Ross from 1955, that Ford and Fulkerson mention as motivation to study the maximum flow problem. The papers have in common that they both apply their methods to the Soviet railway network.

1. Transportation

The transportation problem and cycle cancelling methods are classical in optimization. The usual attributions are to the 1940's and later[2]. However, as early as 1930, A.N. Tolstoĭ [1930][3] 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 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[4]. 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.

[1] CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands, and Department of Mathematics, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands.

[2] The transportation problem was formulated by Hitchcock [1941], and a cycle criterion for optimality was considered by Kantorovich [1942] (Kantorovich and Gavurin [1949]), Koopmans [1948] (Koopmans and Reiter [1951]), Robinson [1949,1950], Gallai [1957,1958], Lur'e [1959], Fulkerson [1961], and Klein [1967].

[3] Later, Tolstoĭ described similar results in an article entitled *Methods of removing irrational transportations in planning* [1939], in the September 1939 issue of *Sotsialisticheskiĭ Transport*.

[4] 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.

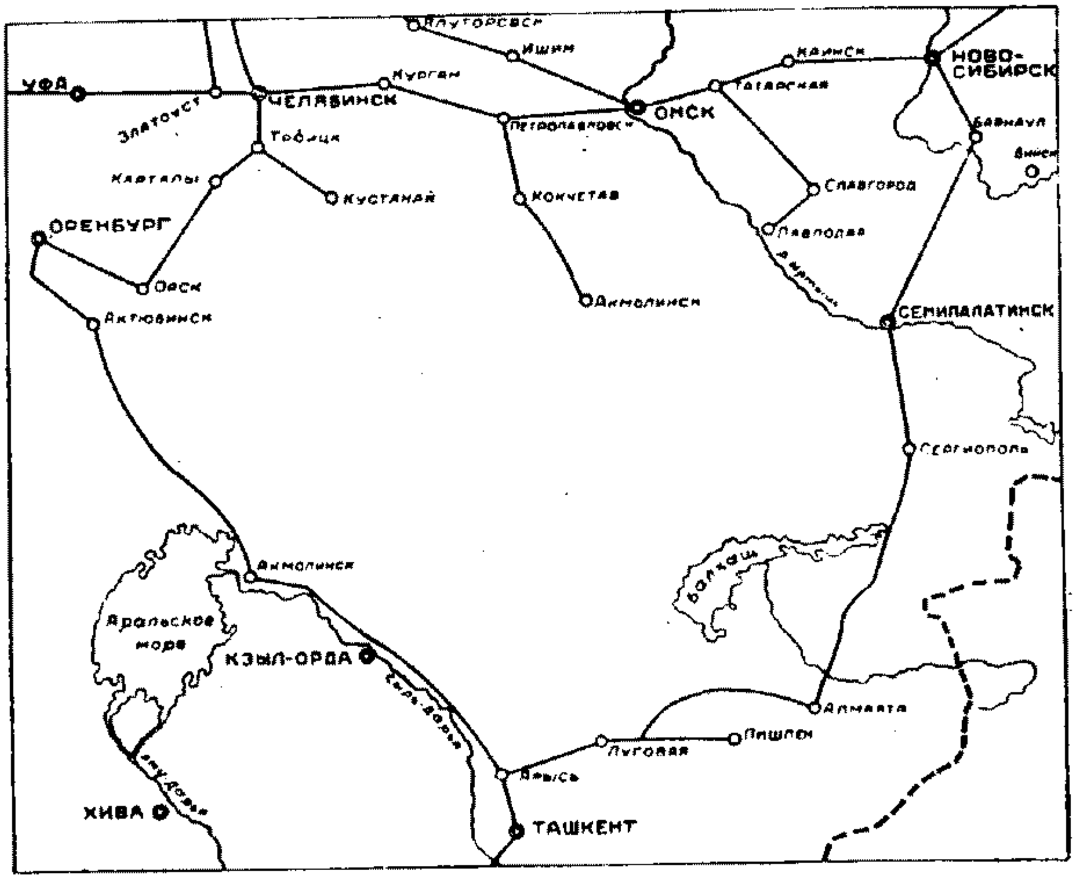

Figure 1
Figure from Tolstoĭ [1930] to illustrate a negative cycle.

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), as given in the following table.

	Arkhangelsk	Yaroslavl’	Murom	Balakhonikha	Dzerzhinsk	Kishert’	Sverdlovsk	Artemovsk	Iledzhk	Dekonskaya	*demand:*
Agryz				709	1064	693					2
Aleksandrov					397			1180			4
Almaznaya								81		65	1.5
Alchevskaya								106		114	4
Baku								1554		1563	10
Barybino								985		968	2
Berendeevo		135			430						10
Bilimbai						200	59				1
Bobrinskaya								655		663	10
Bologoe		389						1398			1
Verkhov’e								678		661	1
Volovo								757		740	3
Vologda	634					1236					2
Voskresensk				427				1022		1005	1
V.Volochek		434						1353		1343	5
Galich	815	224				1056					0.5
Goroblagodatskaya						434	196				0.5
Zhlobin								882		890	8
Zverevo								227		235	5
Ivanovo					259						6
Inza				380	735					1272	2
Kagan								2445	2379		0.5
Kasimov			0								1
Kinel’				752		1208			454	1447	2
Kovylkino				355						1213	2
Kyshtym						421	159				3
Leningrad	1237	709						1667		1675	55
Likino			223		328						15
Liski								443		426	1
Lyuberdzhy			268		411					1074	1
Magnitogorskaya						932	678		818		1
Mauk						398	136				5
Moskva			288	378	405			1030		1022	141
Navashino			12	78							2
Nizhegol’								333		316	1
Nerekhta		50			349						5
Nechaevskaya			92								0.5
N.-Novgorod					32						25
Omsk						1159	904		1746		5
Orenburg									76		1.5
Penza				411				1040	883	1023	7
Perm’	1749					121					1
Petrozavodsk	1394										1
Poltoradzhk								1739	3085	1748	4
Pskov								1497		1505	10
Rostov/Don								287		296	20
Rostov/Yarosl		56			454						2
Rtishchevo								880		863	1
Savelovo		325						1206		1196	5
Samara				711					495	1406	7
San-Donato						416	157				1
Saratov								1072		1055	15
Sasovo				504				1096		1079	1
Slavyanoserbsk								119		115	1.1
Sonkovo		193						1337			0.5
Stalingrad								624		607	15.4
St.Russa		558						1507		1515	5
Tambov								783		766	4
Tashkent								3051	1775		3
Tula								840		848	8
Tyumen’						584	329				6
Khar’kov								251		259	60
Chelyabinsk						511	257		949		2
Chishmy				1123		773			889		0.5
Shchigry								566		549	4
Yudino				403	757	999					0.5
Yama								44		52	5
Yasinovataya								85		93	6
supply:	5	11.5	8.5	12	100	12	15	314	10	55	543

Table of distances (in kilometers) between sources and destinations, and of supplies and demands (in kilotons). (Tolstoĭ did not give any distance for Kasimov. We have inserted a distance 0 to Murom, since from Tolstoĭ’s solution it appears that Kasimov is connected only to Murom, by a waterway.)

Tolstoĭ's heuristic also makes use of insight in 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.

In the following quotation, Tolstoĭ considers the cycles Dzerzhinsk-Rostov-Yaroslavl'-Leningrad-Artemovsk-Moscow-Dzerzhinsk and Dzerzhinsk-Nerekhta-Yaroslavl'-Leningrad-Artemovsk-Moscow-Dzerzhinsk. It is the sixth step in his method, after the transports from the factories in Iletsk, Sverdlovsk, Kishert, Balakhonikha, and Murom have been set:

> 6. The Dzerzhinsk factory produces 100,000 tons. It can forward its production only in the Northeastern direction, where it sets its boundaries in interdependency with the Yaroslavl' and Artemovsk (or Dekonskaya) factories.
>
> | | From Dzerzhinsk | From Yaroslavl' | Difference to Dzerzhinsk |
> |---|---|---|---|
> | Berendeevo | 430 km | 135 km | −295 km |
> | Nerekhta | 349 ,, | 50 ,, | −299 ,, |
> | Rostov | 454 ,, | 56 ,, | −398 ,, |
>
> | | From Dzerzhinsk | From Artemovsk | Difference to Dzerzhinsk |
> |---|---|---|---|
> | Aleksandrov | 397 km | 1,180 km | +783 km |
> | Moscow | 405 ,, | 1,030 ,, | +625 ,, |
>
> The method of differences does not help to determine the boundary between the Dzerzhinsk and Yaroslavl' factories. Only the circle dependency, specified to be an interdependency between the Dzerzhinsk, Yaroslavsl' and Artemovsk factories, enables us to exactly determine how far the production of the Dzerzhinsk factory should be advanced in the Yaroslavl' direction.
>
> Suppose we attach point Rostov to the Dzerzhinsk factory; then, by the circle dependency, we get:
>
> | | | | | | |
> |---|---|---|---|---|---|
> | Dzerzhinsk-Rostov | 454 km | −398 km | Nerekhta | 349 km | −299 km |
> | Yaroslavl'- ,, | 56 ,, | | ,, | 50 ,, | |
> | Yaroslavl'-Leningrad | 709 ,, | +958 ,, | These points remain | | |
> | Artemovsk- ,, | 1,667 ,, | | unchanged because only the | | |
> | Artemovsk-Moscow | 1,030 ,, | −625 ,, | quantity of production sent | | |
> | Dzerzhinsk- ,, | 405 ,, | | by each factory changes | | |
> | Total | | −65 km | | | +34 km |
>
> Therefore, the attachment of Rostov to the Dzerzhinsk factory causes over-run in 65 km, and only Nerekhta gives a positive sum of differences and hence it is the last point supplied by the Dzerzhinsk factory in this direction.
>
> As a result, the following points are attached to the Dzerzhinsk factory:

N. Novgorod	25,000	tons	
Ivanova	6,000	,,	
Nerekhta	5,000	,,	
Aleksandrov	4,000	,,	
Berendeevo	10,000	,,	
Likino	15,000	,,	
Moscow	35,000	,,	(remainder of factory's production)
Total	100,000	tons	

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[5].

2. Max-Flow Min-Cut

The Soviet rail system also roused the interest of the Americans, and again it inspired fundamental research in optimization.

In their basic paper *Maximal Flow through a Network* (published first as a RAND Report of November 19, 1954), Ford and Fulkerson [1954] mention 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[6]:

> 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 October 24, 1955[7] and

[5] The maximum objective value of a feasible solution, whose residual graph does not contain any nonnegative-cost cycle of length 4, and not any of the seven longer nonnegative-length cycles considered by Tolstoĭ (of lengths 6 and 8), is equal to 397,226.

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

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

written for the US Air Force. At our request, the Pentagon downgraded it to "unclassified" on May 21, 1999.

As is known (Billera and Lucas [1976]), the motivation for the maximum flow problem came from the Soviet railway system. 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 say, 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 propose 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.

The Harris-Ross report stresses that specialists remain needed to make up the model (which is always a good tactics to get a new method 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 these lines would imply 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 August 5, 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 (December 29, 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. So the flow value and the cut capacity are equal, hence optimum.

In the report, the minimum cut is indicated as 'the bottleneck' (Figure 2). While Tolstoĭ and Harris-Ross had the same railway network as object, their objectives were dual.

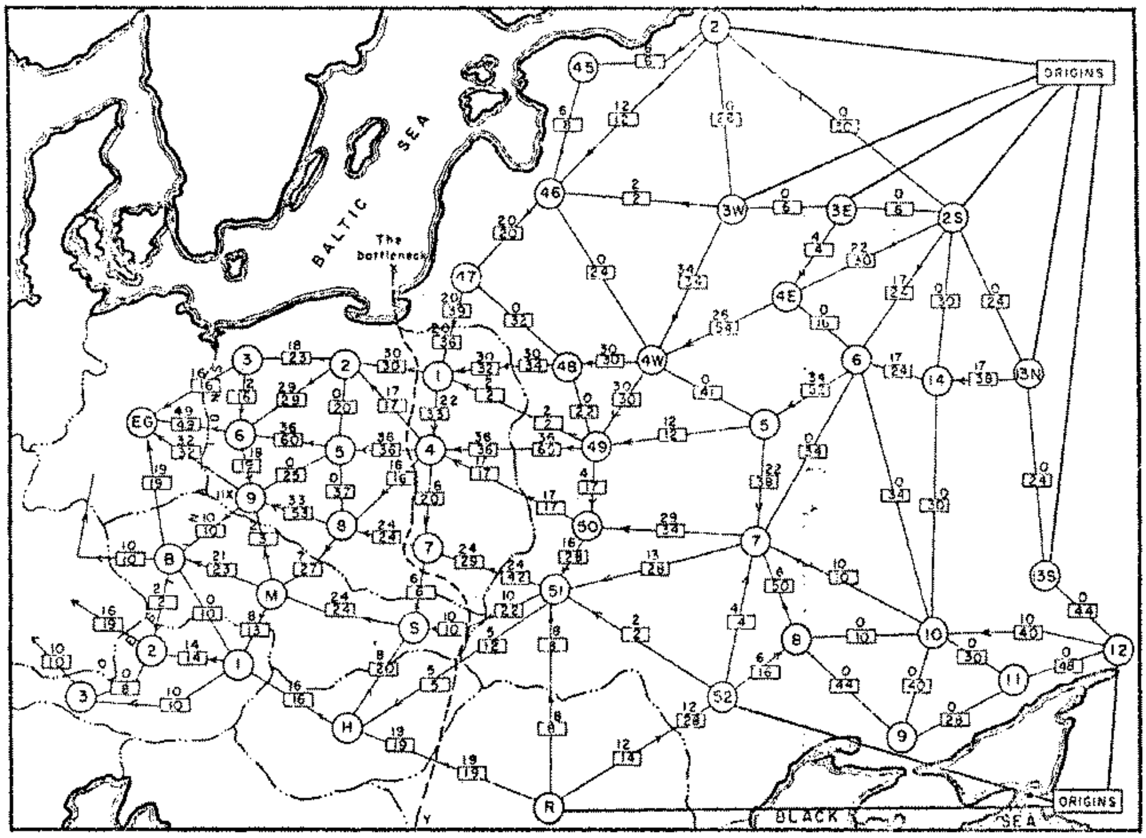

Figure 2

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".

Acknowledgements. I thank Sasha Karzanov for his efficient help in finding Tolstoĭ's paper in the (former) Lenin Library in Moscow, Irina V. Karzanova for accurately providing me with an English translation of it, Kim H. Campbell and Joanne McLean at Air Force Pentagon for declassifying the Harris-Ross report, and Richard Bancroft and Gustave Shubert at RAND Corporation for their mediation in this.

References

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

[1976] L.J. Billera, W.F. Lucas, Delbert Ray Fulkerson August 14, 1924 - January 10, 1976, *Mathematics of Operations Research* 1 (1976) 299–310.

[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.

[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].

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

[1961] D.R. Fulkerson, An out-of-kilter method for minimal-cost flow problems, *Journal of the Society for Industrial and Applied Mathematics* 9 (1961) 18–27.

[1957] T. Gallai, Gráfokkal kapcsolatos maximum-minimum tételek (I. rész) [Hungarian; Maximum-minimum theorems for networks (part I)], *A Magyar Tudományos Akadémia Matematikai és Fizikai Tudományok Osztályának Közleményei* 7 (1957) 305–338.

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

[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, 1955.

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

[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]].

[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.

[1967] M. Klein, A primal method for minimal cost flows with applications to the assignment and transportation problems, *Management Science* 14 (1967) 205–220.

[1948] Tj.C. Koopmans, Optimum utilization of the transportation system, in: *The Econometric Society Meeting* (Washington, D.C., September 6–18, 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].

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

[1959] A.L. Lur'e, Methods of establishing the shortest running distances for freights on setting up transportation systems [in Russian], 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. 349–382 [English translation in: *The Use of Mathematics in Economics* (V.S. Nemchinov, ed.), Oliver and Boyd, Edinburgh, 1964, pp. 323–355].

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

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

[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].