



HAL
open science

Joseph B. Kruskal, Jr., 1956, On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. Version bilingue et commentée

Joseph B. Kruskal, Laurent Beauguitte, Marion Maisonobe

► **To cite this version:**

Joseph B. Kruskal, Laurent Beauguitte, Marion Maisonobe. Joseph B. Kruskal, Jr., 1956, On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. Version bilingue et commentée. 2021. hal-03163666

HAL Id: hal-03163666

<https://hal.science/hal-03163666>

Submitted on 9 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Joseph B. Kruskal, Jr., 1956, On the Shortest
Spanning Subtree of a Graph and the Traveling
Salesman Problem.

Proceedings of the American Mathematical Society

vol. 7, n°1, p. 48-50

Version bilingue et commentée

groupe f.m.r.

A diagram consisting of three colored dots: a blue dot on the left, a red dot in the middle, and a yellow dot on the right. Two curved arrows originate from the red dot; one points to the left towards the blue dot, and the other points to the right towards the yellow dot.

Février 2021

Version traduite et commentée par Laurent Beauquitte, CNRS, UMR Géographie-cités. Relecture : Marion Maisonobe (CNRS, UMR Géographie-cités).

Présentation

Cet article de Joseph B. Kruskal Jr (1928-2010) est issu de son travail de thèse de mathématiques. Soutenue en 1954 à Princeton, *The Theory of Well-Partially-Ordered Sets* est dirigée officiellement par Roger Lyndon et Albert W. Tucker mais Kruskal expliquera que les problèmes traités dans sa thèse sont issus d'échanges avec Paul Erdős. Tout comme nombre de chercheurs des années 1950 travaillant sur la théorie des graphes et les questions d'optimisation des flux¹, il bénéficie de financements de l'armée américaine (*US Office of Naval Research*) pour ses recherches (voir sa biographie sur [web.archives](#)). Après quelques années à l'université, il travaille à partir de 1959 pour les laboratoires Bell, tout en continuant à publier régulièrement.

Cet article présente un algorithme - le terme n'est pas utilisé dans l'article - permettant de détecter l'arbre couvrant minimum sur un graphe valué ; problème issu du jeu mathématique du voyageur de commerce et abondamment traité dans les années 1950 (cf les ouvrages et articles de Ford, Fulkerson, Dijkstra, etc.).

La pagination de la version originale est indiquée entre crochets. Toutes les notes de bas de page en bleu de la version française sont du traducteur.

1. On peut citer notamment Claude Berge qui, à la fin de son article de 1957 "Two Theorems on Graph Theory précise" "This study was prepared at the Economics Research Project, Princeton University, under contract with the *Office of Naval Research*" ; David Gale, 1957, "Flows in networks", consultant pour la *RAND Corporation* puis financé par l'*ONR* ; H. W. Kuhn, 1955, "The Hungarian Method for the Assignment Problem", financé par l'*ONR Logistic Project* - liste non exhaustive.

On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem.

Received by the editors April 11, 1955.

[48] 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 spanning¹ 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 given 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 [49] which solves Problem 1 may be used to prove this theorem.

First I would like to point out that there is no loss of generality in assuming that the given connected graph G is complete, that is, that every pair of vertices is connected by an edge. For if any edge of G is "missing," an edge of great length may be inserted, and this does not alter the graph in any way which is relevant to the present purposes. Also, it is possible and intuitively appealing to think of missing edges as edges of infinite length.

CONSTRUCTION A. Perform the following step as many times as possible: Among the edges of G not yet chosen, choose the shortest edge which does not form any loops with those edges already chosen. Clearly the set of edges eventually chosen must form a spanning tree of G , and in fact it forms a shortest spanning tree.

CONSTRUCTION B. Let V be an arbitrary but fixed (nonempty) subset of the vertices of G . Then perform the following step as many times as possible: Among the edges of G which are not yet chosen but which are connected

1. A subgraph spans a graph if it contains all the vertices of the graph.

either to a vertex of V or to an edge already chosen, pick the shortest edge which does not form any loops with the edges already chosen. Clearly the set of edges eventually chosen forms a spanning tree of G , and in fact it forms a shortest spanning tree. In case V is the set of all vertices of G , then Construction B reduces to Construction A.

CONSTRUCTION A'. This method is in some sense dual to A. Perform the following step as many times as possible : Among the edges not yet chosen, choose the longest edge whose removal will not disconnect them. Clearly the set of edges *not* eventually chosen forms a spanning tree of G , and in fact it forms a shortest spanning tree. It is not clear to me whether Construction B in general has a dual analogous to this.

Before showing how Construction A may be used to prove the theorem of [1], I find it convenient to combine into a theorem a number of elementary facts of graph theory. The reader should have no trouble convincing himself that these are true. For aesthetic reasons, I state considerably more than I need.

PRELIMINARY THEOREM. *If G is a connected graph with n vertices, and T is a subgraph of G , then the following conditions are all equivalent :*

- (a) *T is a spanning tree of G ;*
- (b) *T is a maximal² forest³ in G ; [50]*
- (c) *T is a minimal connected spanning graph of G ;*
- (d) *T is a forest with $n-1$ edges ;*
- (e) *T is a connected spanning graph with $n-1$ edges.*

The theorem to be proved states that if the edges of G all have distinct lengths, then T is unique, where T is any shortest spanning tree of G . Clearly T may be redefined as any shortest forest with $n - 1$ edges.

In Construction A, let the edges chosen be called a_1, \dots, a_{n-1} in the order chosen. Let A_i be the forest consisting of edges a_1 through a_i . It will be proved that $T = A_{n-1}$. From the hypothesis that the edges of G have distinct lengths, it is easily seen that Construction A proceeds in a unique manner. Thus the A_i are unique, and hence also T .

It remains to prove that $T = A_{n-1}$. If $T \neq A_{n-1}$, let a_1 be the first edge of A_{n-1} which is not in T . Then a_1, \dots, a_{i-1} are in T . $T \cup a_i$ must have exactly one loop, which must contain a_i . This loop must also contain some edge e which is not in A_{n-1} . Then $T \cup a_i - e$ is a forest with $n - 1$ edges.

As $A_{i-1} \cup e$ is contained in the last named forest, it is a forest, so from Construction A,

$$\text{length}(e) > \text{length}(a_i).$$

2. A graph is "maximal" if it is not contained in any larger graph of the same sort ; it is "minimal" if it does not contain any smaller graph of the same sort.

3. A "forest" is a graph which does not have any loops.

But then $T \cup a_i - e$ is shorter than T . This contradicts the definition of T , and hence proves indirectly that $T = A_{n-1}$.

BIBLIOGRAPHY

1. Otakar Borůvka, *On a minimal problem*, Práce Moravské Přírodovědecké Společnosti, vol. 3, 1926.

PRINCETON UNIVERSITY

Sur le sous-arbre couvrant minimum d'un graphe et le problème du voyageur de commerce

Reçu par les éditeurs le 11 avril 1955.

Il y a quelques années, une traduction tapuscrite (d'origine obscure) [1] a déclenché un certain intérêt. Ce papier s'intéresse au théorème suivant : si un graphe connexe (fini¹) a un entier positif attribué à chacun de ses liens (la *longueur* de ce lien), et si chacune de ces longueurs est distincte, alors, parmi les arbres (en allemand : Gerüst²) couvrants³ de ce graphe, il en existe un seul dont la somme des longueurs des liens est minimale ; le sous-arbre minimum couvrant de ce graphe est unique. (Dans [1], ce théorème est affirmé et prouvé avec la « matrice des longueurs » du graphe, soit la matrice $\|a_{ij}\|$ où a_{ij} est la longueur du lien connectant les sommets i et j . Bien entendu on suppose que $a_{ij} = a_{ji}$ et que $a_{ii} = 0$ pour tout i et j ⁴.)

La preuve dans [1] est basée sur une méthode non déraisonnable de construction d'un sous-arbre couvrant de longueur minimale. C'est dans cette méthode que réside principalement son intérêt dans la mesure où c'est une solution à un problème (Problème 1 ci-dessous) qui, de manière superficielle, est très proche d'une version (Problème 2 ci-dessous) du fameux problème du voyageur de commerce⁵.

PROBLÈME 1. Donner une méthode pratique pour construire un sous-arbre couvrant de longueur minimale.

PROBLÈME 2. Donner une méthode pratique pour construire un sous-arbre couvrant sans ramifications⁶ de longueur minimale.

1. La précision peut sembler étrange dans la mesure où la définition contemporaine d'un graphe suppose un ensemble fini non vide de sommets et un ensemble fini éventuellement vide de liens entre ces sommets. Mais en 1956, le manuel de référence - cf note suivante - est *Théorie des graphes finis et infinis*.

2. Pourquoi donner le terme allemand alors que le papier d'origine est en hongrois? Lorsque Kruskal écrit, le seul manuel de théorie des graphes disponible est en allemand, il s'agit de *Theorie der endlichen und unendlichen Graphen* de Denés König, édité en 1936 à Leipzig et réédité à New York - mais non traduit - en 1950.

3. Un sous-graphe est dit couvrant d'un graphe s'il contient tous les sommets de ce graphe.

4. Si l'on reformule, on a donc un graphe non orienté et simple - pas de boucle -, valué, où les intensités des liens sont des entiers positifs tous différents. Lorsque Kruskal par la suite utilise le terme de boucle (*loop*), il évoque ce qu'on nommerait aujourd'hui un cycle.

5. Sur les nombreuses solutions algorithmiques proposés depuis presque un siècle à ce problème, voir Graham, Ronald L. et Pavol Hell. « On the history of the minimum spanning tree problem. » *Annals of the History of Computing* 7.1 (1985) : 43-57. et Mareš, Martin. « The saga of minimum spanning trees. » *Computer Science Review* 2.3 (2008) : 165-221.

6. Le terme *unbranched* est peu utilisé en théorie des graphes, il est par exemple absent de la version anglaise des manuels de Claude Berge, *The Theory of Graphs and Its Applications* (paru en 1958, traduit en 1962) et *Graphs and Hypergraphs*. (paru en 1970, traduit en 1973).

La proposition donnée dans [1] est inutilement compliquée. Dans cet article, je fournis plusieurs propositions plus simples qui résolvent le problème 1, et je montre comment une de ces propositions peut être utilisée pour prouver le théorème présent dans [1]. Il est probable que toute proposition résolvant le problème 1 puisse être utilisée pour prouver ce théorème.

Premièrement, je voudrais souligner qu'il n'y a aucune perte de généralité à assumer que le graphe connexe G soit complet, c'est-à-dire que toute paire de sommets soit connectée par un lien. Si un lien quelconque de G est « manquant », un lien de plus grande taille peut être inséré et ceci n'altère en rien le graphe vu les objectifs présents. Il est également possible et intuitivement stimulant de considérer les liens manquants comme des liens de longueur infinie⁷.

PROPOSITION A. Effectuer l'étape suivante autant de fois que possible : parmi les liens de G non encore choisis, choisir le lien le plus court ne formant aucune boucle⁸ avec les liens déjà choisis. Il est clair que l'ensemble de liens sélectionnés doit former un arbre couvrant de G , et de fait il forme l'arbre couvrant minimal.

PROPOSITION B. Soit V un sous-ensemble arbitraire mais fini (non vide) de sommets de G . Puis effectuer l'étape suivante autant de fois que possible : parmi les liens de G qui n'ont pas encore été sélectionnés mais qui sont connectés soit à un sommet de V soit à un lien déjà choisi, prendre le lien le plus court ne formant aucune boucle avec les liens déjà choisis. Il est clair que l'ensemble des liens finalement choisis forme un arbre couvrant de G , et en fait il forme l'arbre couvrant minimal. Si V est l'ensemble des sommets de G , alors la Proposition B se réduit à la Proposition A.

PROPOSITION A'. Cette méthode est en quelque sorte duale à A. Effectuer l'étape suivante autant de fois que possible : parmi les liens non encore choisis, choisir le lien le plus long dont la suppression ne déconnectera pas le graphe. Il est clair que l'ensemble des liens *non* choisis forme un arbre couvrant de G , et en fait il forme l'arbre couvrant minimal. Savoir dans quelle mesure la Proposition B a un dual analogue n'est pas clair pour moi.

Avant de montrer comment la Proposition A peut être utilisée pour prouver le théorème de [1], j'ai trouvé pratique de combiner dans un théorème certains faits élémentaires de la théorie des graphes. Le lecteur ne devrait pas avoir de problème pour se convaincre de leur véracité. Pour des raisons esthétiques, je développe beaucoup plus que je n'en ai besoin.

THÉORÈME PRÉLIMINAIRE. *Si G est un graphe connexe avec n sommets et T un sous-graphe de G , alors les conditions suivantes sont toutes équiva-*

7. Rappelons qu'en théorie des graphes, la distance entre deux sommets appartenant à des composantes connexes disjointes est supposée infinie. Ici Kruskal pousse la logique en imaginant que cette propriété pourrait s'appliquer au sein d'une même composante connexe pour toute paire de sommets ij qui ne soient pas voisins.

8. On écrirait aujourd'hui cycle plutôt que boucle (loop) - cf note supra.

lentes :

- (a) T est un arbre couvrant de G ;
- (b) T est une forêt⁹ maximale¹⁰ de G ; [50]
- (c) T est un graphe couvrant connexe minimal de G ;
- (d) T est une forêt avec $n-1$ liens ;
- (e) T est un graphe couvrant connexe avec $n-1$ liens.

Le théorème à démontrer affirme que si les liens de G ont tous des longueurs distinctes, alors T est unique, où T est n'importe quel arbre couvrant minimal de G . Il apparaît clairement que T doit être redéfini comme toute forêt minimale avec $n-1$ liens.

Dans la Proposition A, appelons les liens choisis et ordonnés a_1, \dots, a_{n-1} . Soit A_i la forêt constituée par les liens de a_1 à a_i . Il sera prouvé que $T = A_{n-1}$. Étant donné l'hypothèse que les liens de G ont des longueurs distinctes, il est facile de voir que la Proposition A admet une seule solution. Donc A_1 est unique, et T également.

Il reste à prouver que $T = A_{n-1}$. Si $T \neq A_{n-1}$, considérons que a_1 est le premier lien de A_{n-1} qui n'est pas dans T . Alors a_1, \dots, a_{i-1} sont dans T . $T \cup a_i$ doit avoir exactement une boucle, qui doit contenir a_i . Cette boucle doit également contenir un lien e qui n'est pas dans A_{n-1} . Alors $T \cup a_i - e$ est une forêt avec $n-1$ liens.

Comme $A_{i-1} \cup e$ est contenu dans la dernière forêt évoquée, c'est une forêt, donc d'après la Proposition A,

$$\text{longueur}(e) > \text{longueur}(a_i).$$

Mais alors $T \cup a_i - e$ est plus court que T . Cela contredit la définition de T et prouve ainsi indirectement que $T = A_{n-1}$.

BIBLIOGRAPHIE

1. Otakar Borůvka, *On a minimal problem*, Práce Moravské Pridovedecké Spolecnosti, vol. 3, 1926. ¹¹

PRINCETON UNIVERSITY

9. Une « forêt » est un graphe n'ayant aucune boucle.

10. Un graphe est « maximal » s'il n'est pas contenu dans un graphe plus grand ayant les mêmes propriétés ; il est « minimal » s'il ne contient aucun graphe plus petit de même type.

11. Cet article paru en hongrois en 1926 a été traduit - dans une version accessible et non dans la version introuvable citée par Kruskal - en anglais en 2001 par Nešetřil, Jaroslav, Eva Milková, et Helena Nešetřilová. "Otkar Borůvka on minimum spanning tree problem translation of both the 1926 papers, comments, history." *Discrete mathematics* 233.1-3 (2001) : 3-36. Il est accessible en ligne, contrairement à l'article original titré O jistém problému minimálním.

groupe f.m.r.



La collection « [textes](#) » du [groupe fmr](#) ([flux](#), [matrices](#), [réseaux](#)) propose des rééditions bilingues d'articles consacrés à l'analyse de réseaux.

Parus

- L. Beauguitte, P. Beauguitte et P. Gourdon, 2021, « [William L. Garrison, 1960, Connectivity of the Interstate Highway System](#) ».
- L. Beauguitte et M. Maisonobe, 2021, « Joseph B. Kruskal Jr., 1956, On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem ».

À paraître

- F. Bahoken et L. Beauguitte, 2021, « John D. Nystuen et F. Dacey, 1961, A graph theory interpretation of nodal regions ».