Jaroslav Nešetřil

A few remarks on the history of MST-problem

Archivum Mathematicum, Vol. 33 (1997), No. 1-2, 15--22

Persistent URL: http://dml.cz/dmlcz/107593

ARCHIVUM MATHEMATICUM (BRNO)
Tomus 33 (1997), 15 – 22

A FEW REMARKS ON THE HISTORY OF MST–PROBLEM

Jaroslav Nešetřil

Dedicated to the memory of Professor Otakar Borůvka

Abstract. On the background of Borůvka's pioneering work we present a survey of the development related to the Minimum Spanning Tree Problem. We also complement the historical paper Graham-Hell [GH] by a few remarks and provide an update of the extensive literature devoted to this problem.

In the contemporary terminology the Minimum Spanning Tree problem can be formulated as follows:

Given a finite set V and a real weight function w on pairs of elements of V find a tree (V, T) of minimal weight $w(t) = \sum \big(w(x, y) : \{x, y\} \in T\big)$.

For example when V is a subset of a metric space and the weight function is defined as the distance then a solution T presents the shortest network connecting all the points V.

Another frequent formulation which also explains its name is

MST PROBLEM:

Given a connected (undirected) graph $G = (V, E)$ with real weights attached to its edges find a spanning tree (V, T) of G (i.e. $T \subseteq E$) such that the total weight $w(T)$ is minimal.

This is a cornerstone problem of Combinatorial Optimization and in a sense its cradle. The problem is important both in its practical and theoretical applications. We want to demonstrate this interest seems not to be dying until now.

The problem was isolated and attacked in the fifties with the vigor and confidence of then newly developing fields theory of algorithms and computer science. The contributions were numerous and illustrious: K. Čulík, G. Dantzig, E. W. Dijkstra, A. Kotzig, J. B. Kruskal, H. W. Kuhn, H. Loberman, A. Weinberger, R. Kalaba, R. C. Prim, E. W. Solomon (see the references : it is only fitting and fortunate that this Borůvka's memorial volume contains a reminiscence of these early days written by J. B. Kruskal). These pioneering works made the MST problem popular and the further development only contributed to it. The paper of R. L. Graham and P. Hell [GH] described accurately the development until 1985. Here are some of the main features that indicate the role and importance of this problem in contemporary discrete mathematics:

i. MST problem may be efficiently solved for large sets by several algorithms. These algorithms were studied even before the right complexity measures and problems were isolated. Very early attempts were made to classify the various algorithms according to their basic underlying idea (see e.g. [ČDF] and [Ta1]). Basically, all known algorithms make use of various combinations of the following two (dual) properties of trees :
(CUT RULE) The optimal solution T to MST problem contains an edge with minimal weight in every cut.
(CIRCUIT RULE) The edge of a circuit C whose weight is larger than the weights of the remaining edges of C cannot belong to the optimal solution T.

ii. There is a variety of algorithms to solve MST problem efficiently. Among those the prominent role is played by Kruskal's Greedy Algoritm [Kr]. Greedy algorithm is perhaps the most thoroughly studied and used heuristic in Combinatorial Optimization. Greedy Algorithm is easy to state : one simply sorts the edges of our graph by increasing weights and then the desired set T is defined recursively as follows : the next edge is added to T iff together with T it does not form a circuit.

iii. MST problem has a polynomial solution regardless of the weight function w (e.g. for negative weights).

iv. Problems analogous to the MST problem were also solved efficiently, particularly the directed version of the problem (i.e. minimal branching from a given root, see [E]).

v. MST problems appears as a subroutine to heuristic and approximate algorithms to other combinatorial optimization problems (such as Traveling Salesman Problem).

vi. The class of problems solvable by Greedy Algorithm were identified with the class of matroids (no such a similar characterization seems to be known for other MST algorithms), greedoids [KLS], and more recently with "jump systems".

While the Greedy Algorithm is esthetically pleasing and perhaps easiest to formulate it is NOT the fastest known algorithm (if only for the fact that we need to sort the edges according to their weights that leads to a nonlinear $n \log n$ lower bound). These complexity considerations revived the interest in alternative procedures and in other algorithms for solving MST problem. It seems that this also revived the interest in the history of MST problem. And it appeared that the pre-computer age history of the problem is as illustrious as the modern development. This is covered carefully in a great detail [GH]. Particularly it appeared that the standard procedure known as Prim's Algorithm [P] was discovered and formulated very clearly and concisely by the prominent number theoretician Vojtěch Jarník in 1930 [Ja]. (Jarník and Kössler [KJ] were also the first to formulate the Euclidean Steiner Tree Problem, see [KN] for the history of Jarník's contribution to Combinatorial Optimization.) Consequently also the work of Otakar Borůvka was reexamined.

Let us recall that Borůvka formulated in [Bo1] and [Bo2] the first efficient solution of MST problem as early as 1926. His contribution was not entirely unrecognized (as opposed to Jarník's work) and both standard early references [Kr] and [Pr] mention Borůvka's paper. However this reference was later dismissed as the Borůvka Algorithm was regarded as "unnecessarily complicated". Well, perhaps a few words of explanation are in order here.

While perhaps not so easy to formulate as the Greedy Algorithm the Borůvka Algorithm is easy to formulate using the present terminology as well :

BORŮVKA ALGORITHM

1. For each vertex v of the given graph G select the edge of minimal weight which is incident with v. (Comment : It is best to formulate the Borůvka Algorithm for graphs with distinct weights of edges. This is either a realistic assumption or it can be solved by a convenient tie breaking procedure. For example we can enumerate the edges and n the case of a tie of edges we select the edge with lower number.)

2. We contract all the selected edges replacing by a single vertex each connected component of the graph defined by the selected edges. In this procedure we eliminate loops (i.e. edges with both ends in the same component) and all the parallel edges (i.e. edges between the same pairs of components) with the exception of the lowest weight edge.

3. We apply the algorithm recursively to find the minimal spanning tree T' of the contracted graph. The minimal spanning tree T is formed by the contracted edges together with the edges of T'.

One should stress that such a concise description was not available in twenties (not only in the pre-computer age but also in the "pre-graph theory" age). One has to see that the operation "contraction" became appreciated much later (in the context of planar graphs and theory of matroids) but even the term "tree" is not mentioned in Borůvka's paper. The later seems to be the main difficulty in [Bo1]. Instead of saying that the selected edges (in Step 1. of the algorithm) form connected components which are (obviously) trees, Borůvka elaborately constructs this tree : first he finds a maximal path P containing a given point then starts with a new vertex and finds a maximal path P' which either is disjoint with P or terminates in a vertex of P and so on. As a result of this the Step 2. has to be tediously described and thus the description of the algorithm takes full 5 pages of [Bo1] !

However all this one should regard as technical difficulties only. Moreover there is an evidence that Borůvka had a simple description in mind as he published a follow up article in an electrotechnical journal [Bo2] where he illustrated his method by an example (of points in the plane together with their distance as weights).

Although each of the iterations of 1. and 2. is more involved than the simpler rule in Greedy Algorithm, we need only $\log n$ of these iterations : in each step we select at least $n/2$ edges and thus the number of vertices of contracted graph is at most half of the size of the original graph.

The following is another view : although we start with many (i.e. n) components (as many as there are blueberries in a forest) the number of components is halved each time and thus we are quickly done. ("Blueberry" is "borůvka" in czech.)

So it appeared that the "simplicity" and effectiveness of Borůvka Algorithm was recognized much later and basically during the last 10 years.

One never knows. Contradicting to the earlier evidence, presently it seems that Borůvka Algorithm is the best algorithm available. This is based on experimental evidence as well as its "parallel" character and its theoretical analysis. Let us be more specific here and let us outline the recent development. It is a spectacular development as it is related to some of the key problems and advances of the modern theory of algorithms.

Given a connected undirected graph $G = (V, E)$ we denote as usual $n = |V|$ the number of its vertices and $m = |E|$ the number of its edges. As G is connected it is $n - 1 \leq m$ and we can identify m with the size of the input of the graph G. To concentrate on the combinatorial structure of the algorithms we consider the computational model unit-cost RAM with the additional restriction that the only operation allowed on edge weights are binary comparisons of weights. Thus m can be thought as the size of the weighted graph, too. This seems to be the most natural model for solving MST problem. However, one should bear in mind that the detailed complexity analysis is model-dependent as was also shown for MST e.g. in [FW]. The above mentioned algorithms are very efficient, for example the naive implementation of Greedy Algorithm is of order mn and it is easy to turn Borůvka Algorithm into an $m\log n$ deterministic algorithm. However, this also indicated that for MST problem we can hope for very fast algorithms. Here is a summary of the results in this direction mostly related to R. Tarjan:

A. Yao [Ya] was the first to implement Borůvka Algorithm and obtained bound $m \log\log n$. This was further improved by Fredman and Tarjan [FT] and finally by Gabow, Galil, Spencer and Tarjan [GGS] and [GGST] to the bound $m\log\beta(m,n)$ where $\beta(m,n)$ is a very slowly growing function defined as follows:

$$\beta(m,n) = \min\{i; \underbrace{\log\log \ldots \log(n)}_{i} \leq m/n\}$$

Currently this is the best known deterministic algorithm for MST problem. This algorithm also involved an important new data structure Fibonacci Heaps that found its way to standard textbooks of Computer Science.

But one can hope for even more. For example Tarjan [T2] showed that one can implement the Greedy Algorithm for graphs with presorted edge-weights so that its complexity is $m\alpha(m,n)$ where $\alpha(m,n)$ is the functional inverse to the Ackerman function. This function grows much slower than (already very slow) function β. However for general weighted graphs [GGST] still presents the best deterministic algorithm for MST problem and the following seems to be the most important problem in this area:

PROBLEM :

Does there exist a linear deterministic algorithm which solves MST Problem ?

More precisely, does there exist a deterministic algorithm and a constant K such that for a given weighted connected graph G with m edges the algorithm finds a minimum spanning tree of G in at most Km steps ?

One should note that many combinatorial optimization problems can be solved by a linear deterministic algorithm (e.g. shortest path problem or finding of a planar drawing of a graph; see [Ta1]). A bit surprisingly for MST problem this is still an open problem. However the problem has been intensively studied. The key role in the recent development has been played by the following sub problem of MST :

MST VERIFICATION PROBLEM :

Given a weighted graph $G = (V, E)$ and its spanning tree T decide whether T is a minimal.

Building on an earlier work of Tarjan [Ta2] and an algorithm of Komlos [Ko] it has been showed by Dixon, Rauch and Tarjan [DRT] that the MST Verification problem can be solved by a linear deterministic algorithm. Recently a simpler procedure has been found by King [K]. Valerie King observed that the Komlos algorithm is simple and linear for balanced (full branching) trees. In order to apply this she transformed every tree to a full branching tree of at most double size with "preservation" of weights. This transformation is achieved by applying the Borůvka Algorithm to a tree itself, indeed King calls the tree produced in this way Borůvka Tree. (Borůvka tree of a tree (V, T) has all the vertices as leaves and internal vertices correspond to components which appear during Borůvka Algorithm, the edges represent which components produce in the next step a new component.)

This is not the end of the story, perhaps rather beginning of the new interesting period. The combination of the previously obtained methods yields unexpected results. So recently Borůvka Algorithm has been combined with the linear verification algorithm to obtain the first randomized linear algorithm for MST problem, see Klein, Tarjan [KT] and Karger, Klein and Tarjan [KKT]. Also an optimal randomized parallel algorithm has been recently found by Cole, Klein and Tarjan [CKT].

In all these results the Borůvka Algorithm plays a key role. Indeed in order to simplify their complicated parallel algorithm and its analysis Cole, Klein and Tarjan [CKT] call each iteration of Borůvka Algorithm (i.e. each iteration of edge selection and subsequent contraction) "Borůvka Step".

We need one more definition (related to the MST verification algorithm) : Given a weighted graph $G = (V, E)$ and a spanning forest F (i.e. (V, F) contains no circuits but is need not be connected), we say that an edge e not in the forest F is F-heavy if the endpoints of e are connected by a path in F and the weight of every edge on that path is less than the weight of e.

It follows from the Circuit Rule that a tree T is a minimal spanning tree iff every edge outside T is T-heavy. All the MST verification algorithms determine all heavy edges (with respect to a particular tree or forest).

Let us end this paper by a description of the first linear randomized algorithm [KT], [KKT]:

LINEAR RANDOMIZED ALGORITHM FOR MST

1. (Borůvka Step):

For each vertex v of the weighted graph G select the edge with minimal weight which is incident with v. Contract all the selected edges replacing by a single vertex each connected component of the graph defined by the selected edges. Eliminate all the loops and all the parallel edges with the exception of the lowest weight edge (between a given pair of vertices).

2. If the density of the contracted graph is less than 6 then continue with 3. (The density is the number of edges divided by the number of vertices, i.e. m/n.) Otherwise we choose a random subgraph H of the contracted graph by including each edge with probability 1/2. Apply then the algorithm recursively to the graph H. Then let F be a minimum spanning forests of H. Using linear MST verification algorithm find all F-heavy edges in the whole contracted graph and delete them. (According to the CIRCUIT rule these edges cannot be contained in the minimum spanning tree of the whole graph.) Proceed with 3.

3. Apply the algorithm recursively to the remaining graph (i.e. either the contracted graph or the contracted graph without F-heavy edges) to obtain a minimum spanning tree T'. The minimum spanning forest of G consists from those edges contracted in the Borůvka Step 1. together with the edges of T'.

It follows from the correctness of the Borůvka Algorithm and from from the CIRCUIT rule that the algorithm correctly computes a minimum spanning tree of the given graph. It has been shown in [KT] that the expected number of the F-heavy edges of the graph H is large (as the expected number of F-light edges is bounded by $2n$). One can the prove that the expected length of the algorithm is linear and even that the algorithm runs in linear time except with the exponential small probability. The analysis of the algorithm and related algorithms given in [KT], [KKT] and [CKT] is quite involved.

The Combinatorial Optimization has gone a long way in its relatively short history. But it is a bit surprising how persistent are the classical motivations and algorithms. But possibly for a (positive) solution of some of the key problems (such as the linearity of MST problem) some new combinatorial trics are needed.

Acknowledgement: I thank to P. Hell, M. Karpinski and R. Tarjan for their help with the recent literature about MST problem.
This work was supported by GAUK 194 and GAČR 0194.

References

[Bo1] O. Borůvka, *O jistém problému minimálním (About a certain minimal problem)*, Práce mor. přírodověd. spol. v Brně III, 3 (1926), 37–58.

[Bo2] O. Borůvka, *Příspěvek k řešení otázky ekonomické stavby elektrovodných sítí (Contribution to the solution of a problem of economical construction of electrical networks)*, Elektrotechnický obzor 15(1926), 153–154.

[C] K. Čulík, *K jednomu minimálnímu problému O. Borůvky*, Čas. pro pěst. mat. 85(1960), 93–94.

[CDF] K. Čulík, V. Doležal, M. Fiedler, *Kombinatorická analýza v praxi*, SNTL, Prague, 1967.

[CKT] R. Cole, P. N. Klein, R. E. Tarjan, *A linear-work parallel algorithm for finding minimum spanning trees*, Proc. of SPAA, 1994.

[Da] G. Dantzig, *Discrete variable extremum problems*, Oper. Research 5(1957).

[Di] E. W. Dijkstra, *Some theorems on spanning subtrees of a graph*, Indag. Math. XXII, 2(1960), 196–199.

[DRT] B. Dixon, M. Rauch, R. Tarjan, *Verification and sensitivity analysis of minimum spanning trees in linear time*, SIAM J. of Computing 21, 6(1992), 1184–1192.

[E] J. Edmonds, *Optimum branchings*, J. Res. Nat. Bur. Standards 71B(1967), 233–240.

[FW] M. Fredman, D. E. Willard, *Trans-dichotomous algorithms for minimum spanning trees and shortest paths*, Proc. 31st Annual IEEE Symp. on Found. of Comp. Sci., 1966, 719–725.

[FT] M. Fredman, R. E. Tarjan, *Fibonacci heaps and their uses in network optimization algorithms*, Proc. 25th Annual IEEE Symp. on Found. of Comp. Sci., 1984, 338–346.

[GGS] H. N. Gabov, Z. Galil, T. H. Spencer, *Efficient implementation of graph algorithms using contraction*, Proc. 25th Annual IEEE Symp. on Foundations of Computer Sci., 1984, 347–357.

[GGST] H. N. Gabov, Z. Galil, T. H. Spencer, R. E. Tarjan, *Efficient algorithms for finding minimum spanning trees in undirected and directed graphs*, Combinatorica 6(1986), 109–122.

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

[Ja] V. Jarník, *O jistém problému minimálním*, Práce mor. přírodověd. spol. v Brně VI, 4(1930), 57–63.

[JK] V. Jarník, M. Kössler, *O minimálních grafech obsahujících n daných bodů*, Čas. pro pěst. mat. 63(1934), 223–235.

[K] R. Kalaba, *On some communication network problem*, Proc. Symp. Applied Math. (1960), 261–280.

[Ka] D. R. Karger, *Random sampling in matroids, with applications to graph connectivity and minimumspanning trees*, Proc. 34th Annual IEEE Symp. on Found. of Computer Sci. 1993, 84–93.

[KKT] D. Karger, P. N. Klein, R. E. Tarjan, *A randomized linear-time algorithm to find minimum spanning trees*, J. Assoc. Comp. Mach. 42(1995), 321–328.

[Ki] V. King, *A simpler minimum spanning tree verification algorithm*, manuscript 1993.

[KT] P. N. Klein, R. E. Tarjan, *A randomized linear-time algorithm for finding minimum spanning trees*, Proc. 26th Annual ACM Symp. on theory of Computing, 1994, 9–15.

[Ko] J. Komlos, *Linear verification for spanning trees*, Combinatorica 5(1985), 57–65.

[KLS] B. Korte, L. Lovasz, R. Schrader, *Greedoids*, Springer Verlag (1991).

[KN] B. Korte, J. Nešetřil, *Vojtěch Jarník's work in combinatorial optimization*, KAM Series No. 96–315.

[Ko] A. Kotzig, *Súvislé grafy s minimálnou hodnotou v konečnom súvislom grafe*, Čas pro pěst. mat. (1961), 1–6.

[Kr] J. B. Kruskal, *On the shortest spanning subtree of a graph and the travelling salesman problem*, Proc. Amer. Math. Soc. 7(1956), 48–50.

[LFPSZ] J. Lukasiewicz, K. Florek, J. Perkal. H. Steinhaus, S. Zubrzycky, *Sur la liaison et la division des points d'un ensemble fini*, Colloq. Math. 2(1949-1951), 282–285.

[M] E. Milková, *Prohledávání, třídění a optimalizace stromů*, doctoral dissertation, Prague, 1997.

[Pr] R. C. Prim, *The shortest connecting network and some generalisations*, Bell. Syst. Tech. J. 36(1957), 1389–1401.

[So] E. W. Solomon, *A comprehensive program for network problems*, Computer J. 3(1960), 89–97.

[Ta1] R. E. Tarjan, *Data structures and network algorithms*, CBMS-NSF Regional Conf. Series in Applied Math., SIAM 44(1983).

[Ta2] R. E. Tarjan, *Applications of path compressions on balanced trees*, J. Assoc. Comput. Math. 26(1979), 690–715.

[Y] A. Yao, *An $O(|E| \log\log |V|)$ algorithm for finding minimum spanning trees*, Inform. Process. Lett. 4(1975), 21–23.

Department of Applied Mathematics
Charles University
Malostranské nám. 25
118 00 Praha 1, CZECH REPUBLIC