Referee on Martin Mares Doctoral Thesis: Graph Algorithms

J. Díaz

June 2, 2008

. from static and dynamic models. The last chapter deals with ranking of some combinatorial objects. From my point of view, the thesis deals with an exposition of the theory and applications of efficient RAM data structures for graphs.

The manuscript is a very well written monograph, which besides its new technical contributions, it has a great value in making a serious (and useful) survey on some historical and technical details on the existing algorithms, for some important problems, like the Minimum Spanning Tree (from now on it will be denoted by the acronym MST), which are difficult to find by reading the original sources. Let us comment each chapter.

Chapter 1. The main contribution of this chapter is to give a global presentation of the different algorithms proposed for the MST.

On 1.1.2, the comment on faster algorithms, it should be added that besides some new algorithmic ideas a lot was due to better representation of the data (new data structures).

In section 1.3 the author gives a nice explanation of Tarjan Red-Blue high level algorithm for MST, which will be used through the whole thesis.

1.4.2, In a manuscript like the present one, you should complete the whole list of authors for the algorithm.

Proof of Theorem 1.4.20, through the thesis as proof the author just refer to some other work, it looks terrible. In this case do not need to formally write *Proof* , a brief comment before or after the statement of the Theorem, will be lots better.

1.5.12 The example of the family of graphs for which the *contractive Boruvka* achieve its lower bound, is a nice contribution.

It is nice the Section 1.6, commenting the behavior of Boruvka algorithms for other types of graphs multi-graphs.

Chapter 2. This chapter is a very didactical exposition of things which are well known in CS. As previously said, the author shows a fluent knowledge of the fundaments of data structures. Moreover, it is true that part of the concepts introduced in the Chapter are needed in other sections of the Thesis, but I do not think it was necessary to include sketches of proofs. It looks a bit as a textbook.

2.3 May be when talking about sorting of integers, the author should point that there are RADIX type algorithms, which can sort reasonable data in linear time (and the implementation constants are better that the ones by Thorup and Han. Again in Lemma 2.5.24 skip the *Proof*

Chapter 3. Beginning of 3.1, what does it mean for the author the word *efficient* Does it mean linear?. In general an algorithms running in average time $O(n \ln n)$ with **reasonable small** constant is considered efficient, for instance quick sort, while a linear algorithm with a huge constant, nobody refers to them as efficient (for instance the deterministic version of Quickselect). The other point of view is efficient= poly-time. I would change efficient for linear time, in the sentence at the beginning of 3.1

As it is presented, the "*Proofs*" of Theorems 3.1.7 and 3.1.10 should just be plain comments.

3.1.16 nice the low degree algorithm adaptation from classical Boruvka, due to the author.

Section 3.3 deal with the verification problem: Given G and a ST verify if it is minimum. It is nice, but it is needed the full Komlos' result? why not to go directly into Valerie King linear time result?

3.3.17 Are you sure the last reference (for max-flow algorithm) does not include Karzanov (Dinic-Karzanov) ?

Lemma 3.5.1 I do not think the first algorithm in the proof of the lemma gives any insight. By the way, the reason for the last sentence in your proof is that it follows a geometric distribution.

Chapter 4. The chapter starts by giving a detailed explanation of the data structure soft heap due to Chacelle. At the end the author gives a nice and clear exposition of an optimal algorithm for MST, due to Pettie and Ramachandran. The main value of this chapter is the fact that the author of the thesis does a very good job in explaining clearly difficult concepts and algorithms.

For instance, the presentation of the DS *Soft heaps* is quite nice, (more clear than other versions that I have seeing!).

4.2 deals with Chazelle concept of robust contraction and the algorithm to partition a graph into contractible clusters. Section 4.3 presents the well known technique of decision trees to find lower bounds for comparison based algorithms (see for instance Cormen, Leisserson, Rivest book) and apply the technique to the clustering algorithm. For my test, this section could have been included in 4.2.

Finally in 4.4, using the results in 4.1, 4.2 and 4.3, the author gives a clear exposition of the optimal MST algorithm by Pettie and Ramachandran.

Chapter 5. In this Section, the author makes an incursion in to the area of discrete (picture to picture) dynamic graph motion and the data structure used to implement them. In particular in section 5.2 the author presents the ET-trees from Henzinger and king, which by the way. Again the author does a good effort to present in a clear way difficult constructions.

In section 5.3 the author deals with keeping the best algorithm to keep connectivity of dynamic graphs under insertion and elimination of edges (algorithm of Holm, Lichtenberg, Thorup), and section present the best know algorithm and DS to keep MSF in a dynamic graph (under insertion and deletions of edges) The algorithm is also due to Holm et all. It is a pity that the author does not give a detailed proof of theorem 5.4.6, which in the original paper by Holm et all. is quite messy.

In section 5.5 the author presents his own algorithm to find the 2-lightest spanning tree. His algorithm is a variation of previous known algorithms and techniques. At the end of the section the author generalizes the results to finding the k lightest ST, where $k = \Theta(1)$.

One thing I am missing in he chapter is any reference to the dynamic graphs where vertices are moving, either in a discrete fashion, (time t, t+1) or in a continuous manner (like Brownian motion). This is an exciting area where there are plenty of open problems.

Chapter 6. This chapter, which for my test should have been included as the last section of chapter 5, presents some further applications to other problems of the previous techniques. I think the author should also have included a small sentence about the case of d-regular graphs. Also I beleive, on the topic of Euclidian MST, there are two excellent monographs (talking about more combinatorial aspects of Steiner trees, MST, Hamiltonian circuits etc..) which are worth to mention: M. Steel Probability theory and combinatorial optimization (SIAM 1997) and J.: Yukich: Probability theory of classical Euclidian optimization problems (Springer 1998).

Chapter 7. From my point of view, the last chapter of the manuscript, presents the best new results of M. Mares. The chapter investigate ranking and unranking of combinatorial objects, and uses the word-RAM data structures developed previously in the dissertation. In sections 7.2 and 7.3, Mares presents linear algorithms for the lexicographic ranking of permutations and ranking of k-permutations.

Section 7.4 and 7.5 deals with ranking of restricted permutations, in particular derangement, which the author also denotes as *hatcheck permutations* I guess as an small homage to P. Erdos. However I am not sure the name is the most appropriate, as in the setting of Erdos it was a probabilistic example of an application of the power of indicator random variables, while here it has nothing to do with probability. On the other hand it is true that the example of the hats and clients can be modeled as a derangement.

Contrary to trend in chapters 1 to 6, chapter 7 seems to be written in a hurry, with plenty of details missing. The reason could be that the author was in a hurry to finish the work, as everybody that has written a Doctoral Dissertation knows very well.

As it has been said, the Dissertation contains new results which have been published in very selective conferences (ESA) and in journals. Moreover, the Dissertation contains very clear explanations of cumbersome constructions of data structures an algorithms. Therefore, **I believe this manuscript has the quality to be defended in order that Martin Mares fulfill the requirements to obtain the Doctoral degree.**