Khaled Elbassioni
Khalifa University of Science and Technology
Electrical Engineering and Computer Science Department
SAN Campus
P.O. Box 127788, Abu Dhabi, UAE
Email: khaled.elbassioni@ku.ac.ae

August 20, 2020

Professor Tomasz Jurdziński
Dean, Faculty of Mathematics and Computer Sience
ul. F. Joliot-Curie 15
50-383 Wrocław | Poland

Dear Professor Jurdziński,

Please find below my report on the doctoral dissertation "Efficient algorithms for combinatorial optimization problems related to rank-maximal matchings and rectangle tiling" by Pratik Ghosal.

Summary of the results. The thesis studies two combinatorial optimization problems on bipartite graphs: (i) computing a rank-maximal matching in a dynamic setting, and (ii) computing a partition of the edge-set subject to some restriction on the parts.

Matchings in bipartite graphs are among the most studied subjects in Theoretical Computer Science, due to their wide applicability; for instance, it is a situation one encounters when assigning posts to applicants. Typically, applicants rank posts according to their preferences, and a rank-maximal matching is an assignment of posts to applicants such that the largest possible number of applicants are granted their first-choice, then, subject to this, the largest possible number of applicants are granted their second-choice, and so on. In a dynamic setting, the graph can be changed in an incremental way, such as adding/deleting an applicant/post, or modifying the rank of one edge. Assuming that a rank-maximal matching was already available before the change, and that it is required to compute a new one for the updated graph, the thesis provides an efficient algorithm for doing this without building a new rank-maximal matching for the updated graph from scratch. The thesis also considers the (naturally arising) situation when the applicants can lie about their preferences, and presents some strategies by which an applicant can fool a rank-maximal matching algorithm to give her/him a better post than what (s)he would get by being truthful.

The second problem considered in the thesis can be best described as a tiling problem on a binary matrix, where the objective is to partition the matrix into a given number of rectangular parts (called tiles), such that the maximum number of ones in each part is minimized. The problem arises in a number of applications, such as load balancing, data compression, etc., and is generally NP-hard, so the thesis considers efficient heuristics (approximation algorithms) which are guaranteed only to produce near-optimal approximate solutions. The main result here is an approximation algorithm that improves on previous results for the case when the number of tiles is very small relative to the total number of ones in the matrix.

The thesis is organized in 5 chapters. The first chapter introduces the problems considered and outlines the results and the organization of the thesis. The second chapter fixes the terminology and introduces some preliminariers needed for the third and forth chapters. In particular, the most important ingredients are the Gallai-Edmonds (GE) decomposition of a bipartite graph, and its use in [21] for computing a rank-maximal matching.

The new results are presented starting from Chapter 3. The main result of that chapter is an efficient algorithm to update an existing rank-maximal matching after an update operation has been applied to the graph. The update operations are of 4 types: add/delete a vertex/edge, but it is shown in the chapter that they all can be reduced to a constant number of updates of the first type, namely, adding a vertex and its incident edges. Thus, the chapter focuses on this last type of update. If one would apply the algorithm in [21] for computing a rank-maximal matching from scratch then the running time would be $O(\min\{n, c\sqrt{n}\}m)$, where n is the number of vertices, m is the number of edges and c is the maximum rank of an edge in an optimal solution. The first idea to improve on this is to update the intermediate graphs constructed by the algorithm of [21]; this was the approach taken in an independent work by other authors [32] and yields a running time of $O(r(n+m))$, where r is the maximum rank of an edge. The main result in this chapter is an algorithm to update the existing rank-maximal matching in time $O(\min(cn, n^2) + m)$, which a significant improvement over the previous two approaches for a graph with large number of edges (the leading coefficient of m in the last approach is 1 as compared to $\min\{n, c\sqrt{n}\}$ and r in the previous two approaches, respectively). The thesis presents an elegant method to achieve this result. First, it studies what happens to the GE decomposition when a single edge is added to the graph. Then based on this an algorithm is given to check in $O(m)$ time if the existing rank-maximal matching remains optimal after the update. At a very high level, one can try now to apply this check to the intermediate matchings computed in the algorithm in [21] and obtain augmenting paths needed for improving these matchings. This idea is established further to show how it can be done more efficiently by first constructing a carefully chosen subgraph containing all needed augmenting paths, whose application to the exiting matching yields a rank-maximal for the new graph. The details of the algorithm and the proof of correctness are a bit technical and require establishing a number of structural results. Examples are given to illustrate the workings of the algorithm and the proof. Finally, a reduction is given to show that essentially the same algorithm can be used for another commonly studied type of matching, called ”popular” matching.

Chapter 4 considers manipulation strategies an applicant can employ to improve her/his outcome obtained by a rank-maximal matching algorithm. As shown at the beginning of the chapter, it is easy to construct simple examples in which one applicant can benefit from untruthfully reporting her/his preferences. A few simple strategies are given for such manipulation, such as providing a short list of preferences or a list with gaps. Thus, it is assumed next that each applicant will provide a full list of preferences with no gaps. Under this assumption, three manipulation strategies are provided that can benefit the applicant. The first strategy ”best nonfirst” guarantees that the applicant is always matched to a post that is not worse than the most preferred post of rank greater than one, obtainable with a truthful preference list. An example is given to show that such a strategy is not always optimal. Next, a second manipulation strategy ”min-max” is given, which optimizes the worst post the applicant can be matched to in a rank-maximal matching. A key notion introduced there is that of a ”critical rank” of an edge of the manipulator, which is defined in terms of the subgraphs constructed by the rank-maximal matching algorithm of [21]. Based on this, a strategy is developed that guarantees that the post the manipulator is matched to in any rank-maximal matching is as small as possible (the smaller the better for the manipulator). However this latter strategy does not guarantee that that manipulator is matched to her/his most preferred post, and it is proved that, in general, no manipulation strategy (with a full preference list) exists that guarantees her/him this in every rank-maximal matching. Finally, a strategy ”Improve best” is presented that matches the manipulator to her/his most preferred post, but only in some rank-maximal matching.

Chapter 5 considers the rectangle tiling problem. In the basic version, called RTILE, it is required to partition a binary matrix A into a specified number p of rectangular subarrays, called tiles, such that the maximum number of ones over all tiles is minimized. An efficient approximation algorithm is given that guarantees a solution within a factor $1.5 + \beta$ from the optimum, where β is a parameter that depends on how large the sought number of tiles p is, compared to the square root of the total number of ones in the matrix. When β is sufficiently small, this improves on the currently best approximation ratio of 2.125 in [33]. The idea, which seems to be an extension of the one in [33], is to use a simple lower bound L on the optimum solution (namely the ratio between the total number of ones in the matrix and the number of tiles) and a guessed value α of the sought approximation ratio (in this case $\alpha = 1.5 + \beta$), to define a set of tiles in terms of carefully chosen boundaries that slice the given matrix horizontally and vertically, such that each tile contains at most αL ones. It remains to argue that the number of tiles obtained by this tiling procedure is not more than the sought number p. This is done elegantly by noting that the way the boundaries are defined imposes a set of linear constraints on the distribution of the ones in the matrix. This leads to a linear program whose dual can be easily (explicitly) solved to yield a lower bound on the number of ones in the matrix, implying that the number of tiles obtained by the procedure is no more than p. An example is given to show that the approximation ratio cannot be improved upon unless another lower bound on the optimum is used (which is not known so far). A simple reduction is given next to show how the $(1.5+\beta)$-approximation algorithm for RTILE can be used to give an approximation algorithm for the version where it is required to tile the matrix into the minimum number of tiles such that each tile contains no more than a given number W of ones. When W is very large, the approximation ratio obtained is roughly 1.5, improving (only in the case of large W) on the previously known ratio of 2 in [24]. Finally, an extension of the approximation algorithm for the d-dimensional version of RTILE is given.

Recommendation. The thesis contains a number of interesting results, most impressively, the algorithm to dynamically update a rank-maximal matching in $O(\min(cn, n^2) + m)$ time, which is a significant improvement over the previous results for dense graphs. This result requires a deep understanding of structural graph theory results related to rank-maximal matchings, and introduces some techniques that may prove useful for obtaining similar results on other dynamic matching problems.

The thesis is a bit hard to read, which, in my opinion, is due mainly to the technical nature of the contents, even though I believe that the writing of chapter 5 can be improved a little bit, with the current description being enough to capture the main ideas. A list of minor comments, meant mainly to improve readability, has been sent separately.

The results of Chapter 4 were published in a good Theoretical Computer Science conference (24th International Computing and Combinatorics Conference (COCOON 2018)), while the results in chapter 3 and 5 are under review in SIAM Journal on Discrete Math and Information Processing Journal, respectively.

In summary, the candidate demonstrated a deep knowledge of the literature and techniques of this area, and obtained several impressive new results improving on earlier ones, and standing as the currently best results for the problems studied.

I strongly recommend that the Ph.D. degree is granted to the candidate for this thesis.

Yours Sincerely,

Khaled Elbassioni

Khaled Elbassioni
Professor of Computer Science