Discrete Mathematics 339 (2016) 2919–2932

Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc

Counting houses of Pareto optimal matchings in the house allocation problem

Andrei Asinowski [a,1], Balázs Keszegh [b], Tillmann Miltzow [a,*]

[a] Institut für Informatik, Freie Universität Berlin, Takustraße 9, 14195 Berlin, Germany

[b] Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Reáltanoda u. 13-15, 1053 Budapest, Hungary

ARTICLE INFO

Article history:
Received 7 October 2015
Received in revised form 17 March 2016
Accepted 19 May 2016
Available online 2 July 2016

Keywords:
Pareto optimal matching
Allocation problem
Complexity
Algorithm
Combinatorics
Counting problem

ABSTRACT

In an instance of the house allocation problem, two sets A and B are given. The set A is referred to as *applicants* and the set B is referred to as *houses*. We denote by m and n the size of A and B, respectively. In the house allocation problem, we assume that every *applicant* $a \in A$ has a preference list over the set of houses B. We call an injective mapping τ from A to B a matching. A *blocking coalition* of τ is a non-empty subset A' of A such that there exists a matching τ' that differs from τ only on elements of A', and every element of A' improves in τ', compared to τ, according to its preference list. If there exists no blocking coalition, we call the matching τ a *Pareto optimal matching* (POM).

A house $b \in B$ is *reachable* if there exists a Pareto optimal matching using b. The set of all reachable houses is denoted by E^*. We show

$$|E^*| \le \sum_{i=1,\dots,m} \left\lfloor \frac{m}{i} \right\rfloor = \Theta(m \log m).$$

This is asymptotically tight. A set $E \subseteq B$ is *reachable* (respectively *exactly reachable*) if there exists a Pareto optimal matching τ whose image contains E as a subset (respectively equals E). We give bounds for the number of exactly reachable sets. We find that our results hold in the more general setting of multi-matchings, when each applicant a of A is matched with ℓ_a elements of B instead of just one. Furthermore, we give complexity results and algorithms for corresponding algorithmic questions. Finally, we characterize *unavoidable* houses, i.e., houses that are used by all POMs. We obtain efficient algorithms to determine all unavoidable elements.

1. Introduction

1.1. Definitions

The house allocation problem is motivated by the following setup: a set of people wish to be allocated to a certain set of houses. Each person ranks the set of houses and wants to be assigned to the house with her highest preference. As soon as

* Correspondence to: Hungarian Academy of Sciences, Institute for Computer Science and Control, (MTA SZTAKI), Kende utca 13-17, 1111 Budapest, Hungary.

E-mail addresses: asinowski@fu-berlin.de, andrei.asinowski@tuwien.ac.at (A. Asinowski), keszegh@renyi.hu (B. Keszegh), t.m@fu-berlin.de, t.miltzow@gmail.com (T. Miltzow).

1 Current address: Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien, Wiedner Hauptstraße 8–10/104, 1040 Vienna, Austria.

http://dx.doi.org/10.1016/j.disc.2016.05.027

two people have the same favorite house this is not possible. Motivated by this picture, we generalize the setup. We start with some definitions.

In an instance of the house allocation problem, two sets A and B are given. The set A represents applicants and the set B represents houses. We denote by m and n the size of A and B, respectively. In the house allocation problem, we assume that every $a \in A$ has a preference list over the set B. A preference list can be formally defined as a total order on B. We call an injective mapping τ from A to B a *matching*. A *blocking coalition* of τ is a non-empty subset A' of A such that there exists a matching τ' that differs from τ only on elements of A', and every element of A' improves in τ', compared to τ according to its preference list. If there exists no blocking coalition, τ is called a *Pareto optimal matching* (POM).

We represent the preference lists by an $m \times n$ matrix. Every row represents the preference list of one of the applicants in A, i.e., in a given row r corresponding to some applicant $a \in A$, the leftmost house is the one that a prefers most, etc., that is house b_1 is left of b_2 in r if and only if a prefers b_1 over b_2. Note that no row contains an element from B twice. We usually denote this matrix by M and following this interpretation we usually denote the members of A by $r_1, r_2, \dots, r_m$ and the members of B by $1, 2, \dots, n$. Because of this matrix representation, we usually refer to members of A only as rows and to members of B as elements (of the matrix).

To illustrate the notion consider the following matrix and observe that the matching indicated by circles is indeed Pareto optimal.

$$\begin{pmatrix} \textcircled{1} & 5 & 3 & 2 & 4 \\ 3 & 1 & \textcircled{4} & 5 & 2 \\ 1 & \textcircled{3} & 5 & 4 & 2 \end{pmatrix}.$$

We denote tuples $p = (a, k)$ as positions of the matrix, if a is some row and k is some natural number. A matching corresponds to a set of positions P_τ if there is exactly one position for each row and no two positions contain the same house. The image set of τ corresponds to the set of houses of B in these positions. Thus, we say that τ *selects* some position p of M (resp. some element b of B), if p is in P_τ (resp. b is in the image set of τ). Similarly, given some matching τ we say that a row a *selects* a position p in row a (resp. element b) if $p = (a, k) \in P_\tau$ for some k (resp. b is in the image of τ). We denote by $s(\tau)$ the image set of τ.

In a POM the positions after the mth column will never be assigned, because at least one of the previous m elements in that row is preferred and not assigned to any other element on A. Therefore, it is sufficient to consider only $m \times m$ square matrices. (In other words, only the first m elements of the preference lists matter.)

If some POM τ selects p (resp. b), then it is a *reachable* position (resp. *reachable* element). More generally, a set $E \subseteq B$ is *(exactly) reachable* if there exists a POM τ with $E \subseteq s(\tau)$ ($E = s(\tau)$). In this case, we also say that τ reaches E. An element b is *unavoidable* if it belongs to the set $s(\tau)$ for every Pareto optimal matching τ of M and b is called *avoidable* if there exists a Pareto optimal matching τ with $b \notin s(\tau)$. A set E is *avoidable* if there exists a POM τ with $s(\tau) \cap E = \varnothing$. Note that for a set E such that $|E| = m$ it is exactly reachable if and only if $B \setminus E$ is avoidable. These notions can be generalized naturally to the case of *multi-matchings*, when each element a has to be paired up with $\ell_a \geq 1$ elements of B. For the precise definitions see Section 5. These *Pareto Optimal Multi-matchings (POMMs)* can also appear naturally in practical applications and we will see that our results about POMs generalize naturally to POMMs. We will also study matrices with fewer than m columns, precise definitions will be given in Section 1.4. In this case, preference lists are incomplete, i.e., it can happen that some elements of A are not assigned. (In this case, it might be interesting to compute a POM of maximum size.)

Example 1. To illustrate the notions and to avoid confusion, we give here a detailed example. The reader can use it to verify that she understood all important notions. The example can be skipped, as we will not refer to it again. Consider the following matrix.

$$M = \begin{pmatrix} 1 & 5 & 3 & 2 \\ 3 & 1 & 5 & 4 \\ 1 & 6 & 5 & 4 \\ 3 & 6 & 2 & 4 \end{pmatrix}.$$

The elements 1 and 3 are unavoidable, as they are both in the first column of M and thus picked by every POM. A quick check reveals that every other element of M can be avoided. Each set E with $1 \in E$ or $3 \in E$ is also unavoidable. A simple argument reveals that every set $E \subseteq \{2, 4, 5, 6\}$ with $|E| \geq 3$ is unavoidable, because every POM picks 4 elements of M. In order to determine all unavoidable sets, there remain only six sets $E \subset \{2, 4, 5, 6\}$ of size 2 to be examined. It turns out that only $\{5, 6\}$ and $\{2, 5\}$ are unavoidable.

It is easy to see that all elements of M are reachable. In order to specify the reachable sets, note that if some set E is reachable then are all its subsets as well. Thus, it is sufficient to specify all reachable sets of size 4. Also note that if $E \subseteq \{1, 2, 3, 4, 5, 6\}$ of size four is reachable then $D = \{1, 2, 3, 4, 5, 6\} \setminus E$ is avoidable. Thus, by the discussion above, the four reachable sets of size four are exactly

$$\{1, 3, 4, 5\}, \{1, 3, 2, 5\}, \{1, 3, 2, 6\}, \{1, 3, 5, 6\}.$$