

Set systems related to a house allocation problem

Dániel Gerbner¹, Balázs Keszegh^{1,2}, Abhishek Methuku⁴, Dániel T. Nagy¹,
Balázs Patkós¹, Casey Tompkins^{3,4}, and Chuanqi Xiao⁵

¹Alfréd Rényi Institute of Mathematics, Hungary

²MTA-ELTE Lendület Combinatorial Geometry Research Group, Hungary

³Karlsruhe Institute of Technology, Germany

⁴Discrete Mathematics Group, Institute for Basic Science (IBS), Daejeon, Republic of Korea

⁵Central European University, Hungary

Abstract

We are given a set A of buyers, a set B of houses, and for each buyer a preference list, i.e., an ordering of the houses. A house allocation is an injective mapping τ from A to B , and τ is strictly better than another house allocation $\tau' \neq \tau$ if for every buyer i , $\tau'(i)$ does not come before $\tau(i)$ in the preference list of i . A house allocation is Pareto optimal if there is no strictly better house allocation.

Let $s(\tau)$ be the image of τ (i.e., the set of houses sold in the house allocation τ). We are interested in the largest possible cardinality $f(m)$ of the family of sets $s(\tau)$ for Pareto optimal mappings τ taken over all sets of preference lists of m buyers. We improve the earlier upper bound on $f(m)$ given by Asinowski, Keszegh and Miltzow [2], by making a connection between this problem and some problems in extremal set theory.

1 Introduction

In this short note we consider set system properties related to the following house allocation problem: Suppose we are given an m -element set A of buyers, an infinite set B of houses and for each buyer $a \in A$ we are given a preference list of B . Then we say that a matching τ from A to B is *Pareto optimal* (a POM) if there does not exist a nonempty subset A' of A (a *blocking coalition*) and another matching τ' from A to B that differs from τ only on A' such that for every $a \in A'$, $\tau'(a)$ is placed higher in the preference list of a than $\tau(a)$. If for some matching τ , there does not exist a blocking coalition of size at most i , we say that τ is an i -POM.

There is a very natural way to generate POMs: Consider a permutation π of A and let the buyers select their favorite house from those that are not already taken, in order $\pi(1), \pi(2), \dots, \pi(m)$. The corresponding matching is denoted by τ_π . For any matching τ the set $\{\tau(a) : a \in A\}$ of houses which are taken is denoted by $s(\tau)$. The following lemma from [2] summarizes all that we will use in this paper about POMs. We say that a subset E of B is reachable if there exists a POM τ with $s(\tau) = E$.

Lemma 1.1 (Lemma 7 in [2]). *Let $E \subseteq \{1, \dots, n\}$ with $|E| = m$. The following statements are equivalent.*

1. E is reachable, i.e., there exists a POM τ with $s(\tau) = E$.
2. There exists a permutation π such that for the greedy matching τ_π we have $s(\tau_\pi) = E$.
3. There exists a 1-Pareto optimal matching (1-POM) τ with $s(\tau) = E$.

The number of reachable sets depends on the preference lists. Let $f(m)$ denote the maximum number of sets of reachable houses taken over all m -tuples of preference lists. We will elaborate more on the motivation and background of the problem of determining $f(m)$ in the next subsection.

Let us begin by introducing the set system properties that we will be interested in.

Definition 1.2. We say that a family \mathcal{F} of m -element sets has property P if for any integer k and any $F_1, \dots, F_k \in \mathcal{F}$, there exists j , $1 \leq j \leq k$, such that $|F_j \setminus \cup_{i \neq j} F_i| \leq \lfloor m/k \rfloor$. Let $g(m)$ denote the largest cardinality of a family with property P .

Definition 1.3. We say that a family \mathcal{F} of m -element sets has property Q if for any integer k and any $F_1, \dots, F_k \in \mathcal{F}$, we have $|\cup_{i=1}^k F_i| \leq \sum_{i=1}^k \lfloor m/i \rfloor$. Let $h(m)$ denote the largest cardinality of a family with property Q .

Obviously property P implies property Q . Therefore, $g(m) \leq h(m)$. The following lemma is hidden in the proof of the upper bound on the number of reachable houses from [2].

Lemma 1.4 ([2]). *The family of the sets of reachable houses has property P and thus $f(m) \leq g(m)$.*

The upper bound in [2] on the number of houses uses only property Q , showing that if \mathcal{F} has property Q , then $|\cup \mathcal{F}| \leq m(\ln m + 1)$. This obviously implies $|\mathcal{F}| \leq \binom{m(\ln m + 1)}{m}$ and thus $f(m) \leq g(m) \leq h(m) \leq |\mathcal{F}| \leq \binom{m(\ln m + 1)}{m}$. We improve this bound on $|\mathcal{F}|$ as follows.

Theorem 1.5. (i) *Let $\ell = \ell(m)$ be the largest integer with $\ell^\ell \leq m$. Then $f(m) \leq \binom{2m - \ell + 4}{m}$ holds if m is sufficiently large. In other words, the family of the sets of reachable houses taken over all m -tuples of preference lists contains at most $\binom{2m - \ell + 4}{m}$ sets if m is sufficiently large.*

(ii) *We have $g(m) \leq \binom{2m - 1}{m}$. In other words, if a family \mathcal{F} of m -element sets has property P , then $|\mathcal{F}| \leq \binom{2m - 1}{m}$.*

Note that $\ell \geq \frac{\log m}{\log \log m}$, therefore the bounds of (i) and (ii) differ by a factor of $\frac{\log m}{m}$.

1.1 Applications to pattern matching problems

In the topic of matchings the problem of counting the number of sets of reachable houses has still not received considerable attention. The problem was first raised in [8] (a precursor of [3]) where it was shown that it can be applied to upper bound the complexity of a certain subdivision of the plane which arises from a pattern matching problem. Moreover, this was used to derive efficient algorithms for this pattern matching problem.

Before stating the connection and the implication of our result, we repeat the most important definitions from [3].

Let $A = \{a_1, \dots, a_n\}$ and $B = \{b_1, \dots, b_m\}$ be two sets of points in the plane. We assume that $m \leq n$, and we seek a minimum-weight maximum-cardinality matching of B into A . This is a subset M of edges of the complete bipartite graph with edge set $B \times A$, so that each $b \in B$ appears in exactly one edge of M , and each $a \in A$ appears in at most one edge. The weight of an edge (b, a) is $\|b - a\|^2$, the squared Euclidean distance between b and a , and the weight of a matching is the sum of the weights of its edges.

Allowing the pattern B to be translated, we obtain the problem of computing the minimum partial-matching RMS (Root Mean Square) distance under translation. That is, we want to find a translation vector t and a matching of B into A such that the weight of the matching edges determined by the Euclidean distances between $B(t)$ (B translated by t) and A is minimal.

This problem induces a subdivision of the plane, where two points t_1, t_2 are in the same region if when translating B by t_1 or t_2 , the above defined minimum-weight matching of B to A is the same. This is called the *partial-matching subdivision* and is denoted by $\mathcal{D}_{B,A}$.

In [8, 3] it was shown that the combinatorial complexity of $\mathcal{D}_{B,A}$ can be upper bounded by $O(n^2 m^4 f(m))$ and that computing a global minimum of the partial-patching RMS distance can be done in time $O(n^3 m^8 f(m))$.

Using the result of [2] this implied for them the upper bound $O(n^2 m^{3.5} (e \ln m + e)^m)$ for the complexity of $\mathcal{D}_{B,A}$. Our Theorem 1.5 improves this to $O(n^2 m^{3.5} 4^m)$. Similarly, their upper bound on computing a global minimum was $O(n^3 m^{7.5} (e \ln m + e)^m)$, which Theorem 1.5 improves to $O(n^3 m^{7.5} 4^m)$.

These bounds are still not polynomial in m , but unfortunately we cannot even hope to achieve a polynomial bound on these problems only by improving bounds on $f(m)$ as it was shown in [3] that $f(m) = \Omega(2^m / \sqrt{m})$.

2 Proofs

We start with a lemma about reachable sets. We think of the m buyers' preference lists as a *preference matrix* M of m infinite rows. The (i, j) -entry of M is the index of the house that buyer i has in the j -th place in his preference list.

Lemma 2.1. *Let $\ell = \ell(m)$ be the largest integer with $\ell^\ell \leq m$. Then for any list of preferences of m buyers, we either have $|\bigcap_\tau s(\tau)| \geq \ell - 4$ or there is a set X of houses with $|X| \geq \ell^2$ such that for any POM τ we have $|s(\tau) \cap X| \geq |X| - \ell$.*

Proof. If we can find ℓ rows such that the first $\ell - 4$ columns of the preference matrix are constant restricted to these rows, then obviously the elements in these rows belong to every $s(\tau)$. We are going to work in rounds, and we want to find $A_1 \supseteq A_2 \supseteq \dots \supseteq A_{\ell-4}$ nested sets of rows such that $|A_i| \geq \frac{m}{\ell^i}$ and the first i columns of the preference matrix are constant on the rows in A_i . If we succeed, then we have $|A_{\ell-4}| \geq \ell^4$. It is easy to see that every element from the first $\ell - 4$ columns in the rows of $A_{\ell-4}$ must be in every $s(\tau)$, thus $|\bigcap_\tau s(\tau)| \geq \ell - 4$.

Clearly, every element in the first column belongs to all sets $s(\tau)$, therefore either there are at least $\ell - 4$ elements in the first column and then we are done, or there is an element that appears at least $\frac{m}{\ell-4} \geq \frac{m}{\ell}$ times. These rows form A_1 .

Suppose we have defined A_i such that the first i columns are constant on the rows belonging to A_i . If we are not able to form A_{i+1} , then every element appears less than $\frac{m}{\ell^{i+1}}$ times in the $(i + 1)$ st column on the rows of A_i . Also, any element x that appears at least $i + 1$ times in column $i + 1$ on the rows of A_i must belong to every $s(\tau)$. Indeed, the $i + 1$ buyers corresponding to these rows share their first i preferences, so one of them will have to pick its $(i + 1)$ -st preferred house x . Either that buyer or a buyer from a different row must take x .

Let X be the set of elements appearing in the $(i + 1)$ -st column in the at least $\frac{m}{\ell^i}$ rows belonging to A_i . By the argument above, if $|\bigcap_\tau s(\tau)| < \ell - 4$ and A_{i+1} does not exist with the required properties, then each element of X appears less than $\frac{m}{\ell^{i+1}}$ times, and all but at most $\ell - 5$ elements appear at most i times in the $(i + 1)$ -st column in those rows. Hence

$$(\ell - 5) \frac{m}{\ell^{i+1}} + (|X| - \ell + 5)i \geq \frac{m}{\ell^i}.$$

Rearranging and using that $i \leq \ell - 4$ yields $|X| \geq \frac{m}{\ell^{i+2}} \geq \ell^2$. Observe that for any τ the set $s(\tau)$ contains at least $|X| - i$ elements of X , those that belong to rows of A_i of which the corresponding buyer has not picked their house from its first i preferred houses. \square

We remark that it is easy to obtain some negligible improvement by a more precise analysis of the above arguments. For example, we used that an element that appears at least $i + 1$ times in column $i + 1$ on the rows of A_i must belong to every $s(\tau)$. However, in addition to these elements, the i elements appearing in the first i columns in those rows

also belong to every $s(\tau)$. For simplicity, we did not take this (and similar arguments) into account in our calculation.

Let us now introduce the last and most important notion and result that we need for the proof of Theorem 1.5. We say that a family of k sets A_1, \dots, A_k is disjointly representable, if there are elements x_1, \dots, x_k such that $x_i \in A_j$ if and only if $i = j$ (for all $1 \leq i, j \leq k$). In other words, none of the A_i 's are contained in the union of the others. We will use the following theorem of Frankl and Pach [6].

Theorem 2.2 (Frankl, Pach [6]). *Let $f(r, k)$ be the maximum size of an r -uniform family that does not contain k disjointly representable sets. Then we have*

$$f(r, k) \leq \binom{r + k - 1}{k - 1}.$$

Proof of Theorem 1.5. A bound of $f(m) \leq g(m) \leq \binom{2m}{m}$ is immediate. Indeed, using only $k = m + 1$ in the definition of property P , we obtain that any $m + 1$ sets in \mathcal{F} are not disjointly representable. Applying Theorem 2.2 with $k = m + 1$, $r = m$ yields $g(m) \leq \binom{2m}{m}$.

To obtain the improvement of (i), observe that if $s(\tau_1), s(\tau_2), \dots$ are the sets of possibly selected houses, then the family $s(\tau_1) \setminus (\bigcap_{\tau} s(\tau)), s(\tau_2) \setminus (\bigcap_{\tau} s(\tau)), \dots$ still does not contain $m + 1$ disjointly representable sets. If Lemma 2.1 yields $|\bigcap_{\tau} (s(\tau))| \geq \ell - 4$, then we can apply Theorem 2.2 with $k = m + 1$, $r = m - \ell + 4$ to get $f(m) \leq \binom{2m - \ell + 4}{m}$. Otherwise Lemma 2.1 yields a set X of at least ℓ^2 elements such that for every POM τ we have $|s(\tau) \cap X| \geq |X| - \ell$. Let $\mathcal{F} = \{s(\tau) : \tau \text{ is a POM}\}$ and for any set Y let $\mathcal{F}_Y = \{F \setminus Y : Y \subset F \in \mathcal{F}\}$. By the pigeonhole principle, there exists a subset $Y \subset X$ of size $|X| - \ell$ such that $|\mathcal{F}| \leq \frac{|\mathcal{F}_Y|}{\binom{|X|}{\ell}}$.

Observe that \mathcal{F}_Y also has property P , thus \mathcal{F}_Y cannot contain $m + 1$ disjointly representable sets. Therefore, by Theorem 2.2 we obtain $|\mathcal{F}_Y| \leq \binom{2m - |X| + \ell}{m}$ and thus $|\mathcal{F}| \leq \binom{|X|}{\ell} \binom{2m - |X| + \ell}{m}$. As $|X| \geq \ell^2$, the latter expression is smaller than $\binom{2m - \ell + 4}{m}$ if m is large enough.

To see the improvement of (ii), we will use only that property P holds for $k = 2$ and $k = m + 1$. As we have seen, $k = m + 1$ means there are no $m + 1$ disjointly representable sets in \mathcal{F} . On the other hand, $k = 2$ means that for every pair of sets, one has at most $\lfloor m/2 \rfloor$ elements not in the other, which means their intersection has size at least $\lceil m/2 \rceil$. We will only use that the intersection is non-empty. As observed by Frankl and Pach [6], if F_1, F_2, \dots, F_t do not contain $m + 1$ disjointly representable sets, then for any $1 \leq i \leq t$ a minimal set $E_i \subset (\cup_j F_j) \setminus F_i$ that meets every F_j ($j \neq i$) has size at most m . Then $F_i \cap E_{i'} \neq \emptyset$ if and only if $i \neq i'$. For pairs with this property, Bollobás's famous inequality states that $\sum_{j=1}^t \frac{1}{\binom{|F_j| + |E_j|}{|E_j|}} \leq 1$, in particular $t \leq \binom{2m}{m}$ as all F_j 's and E_j 's have size at most m .

Since the F_j 's are intersecting, one can define another collection $\{(A_i, B_i)\}_{i=1}^{2t}$ with $A_i = F_i$, $B_i = E_i$ for $1 \leq i \leq t$ and $B_i = F_{2t-i+1}$, $A_i = E_{2t-i+1}$ for $t + 1 \leq i \leq 2t$. Then this collection is *skew cross-intersecting*, i.e., for any $i < j$ we have $A_i \cap B_j \neq \emptyset$ and $A_i \cap B_i = \emptyset$

for all $1 \leq i \leq 2t$. As proved by Frankl [5] and Kalai [9], if in such collections all sets have size m , then the number of pairs is at most $\binom{2m}{m}$. Thus $|\mathcal{F}| = t \leq \frac{1}{2} \binom{2m}{m} = \binom{2m-1}{m}$, as claimed. \square

Note that Theorem 2.2 is very close to being sharp (also shown in [6]). We believe that both bounds in Theorem 1.5 are far from being sharp. The best construction we are aware of is summarized in the following simple proposition.

Proposition 2.3. $h(m) \geq \binom{\lfloor 3m/2 \rfloor}{m}$.

Proof. Consider the family $\binom{\lfloor 3m/2 \rfloor}{m}$. Property Q holds for this family for $k = 1$ because of the uniformity, and for larger k because $\sum_{i=1}^k |F_i|$ cannot be larger than the underlying set. \square

Many variants, generalizations and strengthenings [5, 7, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19] of Bollobás's inequality have been established. By the remarks above, to improve our bounds, we would be interested in versions of Bollobás's inequality when the A_i 's satisfy some extra intersection property. Let $j(m)$ denote the largest cardinality of an $(\lceil m/2 \rceil)$ -intersecting family without $m + 1$ disjointly representable sets. Then by the arguments above we have $g(m) \leq j(m) \leq \binom{2m-1}{m}$.

We finish our note with two conjectures starting with the 1-intersecting case.

Conjecture 2.4. Let $\{(A_i, B_i)\}_{i=1}^m$ be a collection of pairs of sets such that $A_i \cap B_j = \emptyset$ if and only if $i = j$ and $A_i \cap A_j \neq \emptyset$ for all i and j . Let $|A_i| = a$ and $|B_i| = b$ and $a \leq b$, then

$$m \leq \binom{a+b-1}{a-1}$$

By the reasoning in the proof of Theorem 1.5 (ii) (the idea of which goes back to Poljak and Tuza [14]), Conjecture 2.4 is true if $a = b$. Moreover, as Scott and Wilmer [15] have recently established a Bollobás type inequality for skew intersecting set pairs $\{A_i, B_i\}_{i=1}^m$ with $|A_1| \leq |A_2| \leq \dots \leq |A_m|$ and $|B_1| \geq |B_2| \geq \dots \geq |B_m|$, their result implies that in Conjecture 2.4 the bound $m \leq \frac{1}{2} \binom{a+b}{a}$ holds.

We propose a straightforward generalization for the t -intersecting case. Let $AK(n, k, t)$ denote the maximum size of a k -uniform t -intersecting family $\mathcal{F} \subseteq \binom{[n]}{k}$. The value of $AK(n, k, t)$ for any n, k, t was determined by Ahlswede and Khatchatrian [1].

Conjecture 2.5. Let $\{(A_i, B_i)\}_{i=1}^m$ be a collection of pairs of sets such that $A_i \cap B_j = \emptyset$ if and only if $i = j$ and $|A_i \cap A_j| \geq t$ for all i and j . Let $|A_i| = a$ and $|B_i| = b$ and $a \leq b$, then

$$m \leq AK(a+b, a, t).$$

Conjecture 2.5, if true, would yield an upper bound $f(m) \leq j(m) \leq \max\{|\mathcal{F}_i| : i = 0, 1, \dots, m/2\}$, where $\mathcal{F}_i = \{F \in \binom{[2m]}{m} : |F \cap [\lceil \frac{m}{2} \rceil + 2i]| \geq \lceil \frac{m}{2} \rceil + i\}$. Note that the maximum is taken approximately at $i = \frac{m}{8}$.

Acknowledgements. The results presented in this note were obtained during the 9th Emléktábla Workshop, in Gárdony, June 2019. The research was supported by the National Research, Development and Innovation Office NKFIH under the grants K 116769, KH130371 and SNN 129364. The research of Gerbner was supported by the János Bolyai Research Fellowship. The research of Keszegh was supported by the Lendület program of the Hungarian Academy of Sciences (MTA), under the grant LP2017-19/2017. The research of Methuku and Tompkins was supported by IBS-R029-C1.

References

- [1] R. Ahlswede, L. Khachatrian. The complete intersection theorem for systems of finite sets. *European Journal of Combinatorics*, 18(2) (1997), 125–136.
- [2] A. Asinowski, B. Keszegh, T. Miltzow. Counting houses of Pareto optimal matchings in the house allocation problem. *Discrete Mathematics*, 339(12) (2016), 2919–2932.
- [3] R. Ben-Avraham, M. Henze, R. Jaume, B. Keszegh, O. Raz, M. Sharir, I. Tubis. Partial-Matching RMS Distance Under Translation. *Combinatorics and Algorithms, Algorithmica*, 80(8) (2018), 2400–2421; 22nd European Symposium on Algorithms - ESA 2014, *Lecture Notes in Computer Science*, 8737 (2014), 100–111.
- [4] B. Bollobás. On generalized graphs. *Acta Math. Hung.*, 16 (1965), 447–452.
- [5] P. Frankl. An extremal problem for two families of sets. *European Journal of Combinatorics*, 3(2) (1982), 125–127.
- [6] P. Frankl, J. Pach. On disjointly representable sets. *Combinatorica*, 4(1) (1984), 39–45.
- [7] Z. Füredi. Geometrical solution of an intersection problem for two hypergraphs. *Eur. J. Combin.*, 5 (1984), 133–136.
- [8] M. Henze, R. Jaume, B. Keszegh. On the complexity of the partial least-squares matching Voronoi diagram. In *Proceedings of the 29th European Workshop on Computational Geometry, EuroCG*, (2013) 193–196.
- [9] G. Kalai. Intersection patterns of convex sets. *Israel Journal of Mathematics*, 48(2-3) (1984), 161–174.

- [10] D. Kang, J. Kim, Y. Kim. On the Erdős-Ko-Rado theorem and the Bollobás theorem for t -intersecting families. *European J. Combin.*, 47 (2015), 68–74.
- [11] Z. Király, Z. L. Nagy, D. Pálvölgyi, M. Visontai. On families of weakly cross-intersecting set-pairs. *Fundamenta Informaticae*, 117(1-4) (2012), 189–198.
- [12] L. Lovász. Flats in matroids and geometric graphs. In *Combinatorial surveys (Proc. Sixth British Combinatorial Conf., Royal Holloway Coll., Egham, 1977)*, pages 4586. Academic Press London, 1977.
- [13] B. Patkós, On the general position problem on Kneser graphs, arXiv:1903.08056
- [14] S. Poljak, Zs. Tuza, On the maximum number of qualitatively independent partitions. *Journal of Combinatorial Theory, Series A*, 51(1) (1989), 111–116.
- [15] A. Scott, E. Wilmer, Combinatorics in the Exterior Algebra and the Bollobás Two Families Theorem, arXiv:1907.06019
- [16] J. Talbot, A new Bollobás-type inequality and applications to t -intersecting families of sets, *Discrete Mathematics* 285 (2004) 349–353.
- [17] Zs. Tuza, Inequalities for two-set systems with prescribed intersections, *Graphs Combin.* 3 (1987), no. 1, 7580.
- [18] Zs. Tuza, Applications of the set-pair method in extremal hypergraph theory, *Extremal problems for finite sets (Visegrád, 1991)*, *Bolyai Soc. Math. Stud.*, vol. 3, János Bolyai Math. Soc., Budapest, 1994, 479–514.
- [19] Zs. Tuza, Applications of the set-pair method in extremal problems. II, *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, *Bolyai Soc. Math. Stud.*, vol. 2, János Bolyai Math. Soc., Budapest, 1996, 459–490.