Fair Partitions

Noga Alon

Abstract

A substantial number of results and conjectures deal with the existence of a set of prescribed type which contains a fair share from each member of a finite collection of objects in a space, or the existence of partitions in which this is the case for every part. Examples include the Ham Sandwich Theorem in Measure Theory, the Hobby-Rice Theorem in Approximation Theory, the Necklace Theorem and Ryser's Conjecture in Discrete Mathematics. The techniques in the study of these results combine combinatorial, topological, geometric, probabilistic and algebraic tools. This paper contains a brief description of the topic, focusing on several recent existence results and their algorithmic aspects. This is mainly a survey paper, but it also contains several novel results.

1 Introduction

The problem of the existence of a set with desired properties that has a fair share of each of a family of measures has been studied in several areas. The related notion of fair partitions has also received a considerable amount of attention. Although there have been several earlier results of this type it is common to view the Ham Sandwich Theorem as the initial statement in the area.

Theorem 1.1 (The Ham Sandwich Theorem) *For any collection of d probability measures in R^d, each absolutely continuous with respect to the Lebesgue measure, there is a hyperplane that bisects all measures.*

Thus, each of the two half-spaces determined by the separating hyperplane contains a fair share of each of the measures. This was conjectured by Steinhaus and proved by Banach, using the Borsuk-Ulam Theorem, a fundamental result in Topology which asserts that any continuous function from S^n to R^n maps two antipodal points to the same image.

The Ham Sandwich Theorem is first mentioned in [45], where Steinhaus attributes the proof to Banach (for $d = 3$, but the proof for general d is essentially identical).

There are numerous results and questions dealing with partitions of prescribed types of Euclidean spaces and the ways they can split measures. See [40] for a comprehensive recent survey of the subject. The formulation of most of these results is geometric, dealing with sets or measures in Euclidean spaces. There are, however, also purely combinatorial results and conjectures of the same flavor. Here we focus on questions of this type. The following examples of two results and two conjectures illustrate the diversity of the topic.

Theorem 1.2 (The Cycle and Triangles Theorem) *Let G be a cycle of length $3m$ and let P be an arbitrary partition of its set of vertices into pairwise disjoint sets $P_1, P_2, \ldots, P_m$, each of size 3. Then there is an independent set S of G that contains exactly one vertex of each set P_i. Moreover, all vertices of G can be partitioned into 3 independent sets S_1, S_2, S_3, each containing exactly one point of each P_i.*

This result (for one set S) was conjectured by Du, Hsu and Hwang in [16], the stronger conjecture is due to Erdős [18]. It was proved (in a strong form) by Fleischner and Steibitz in [21], using the algebraic technique of [8]. Additional proofs of the initial conjecture of [16] and of some variants appear in [42], [2] and [1].

Theorem 1.3 (The Necklace Theorem) *Let N be an open necklace with ka_i beads of type i, for $1 \le i \le t$. Then it is possible to cut N in at most $(k-1)t$ points and partition the resulting intervals into k collections, each containing exactly a_i beads of type i, for all $1 \le i \le t$.*

A continuous version of this result for $k = 2$ has been proved in [29], the discrete result for $k = 2$ is proved in [24], and a short derivation of it from the Borsuk-Ulam Theorem appears in [9]. The general result is proved in [3].

Conjecture 1.4 (Rota's Basis Conjecture) *Let $B_1, B_2, \ldots, B_n$ be n bases of a matroid of rank n. Then there is a partition of the elements in the (multi)set $B_1 \cup B_2 \cup \ldots \cup B_n$ into n pairwise disjoint bases $A_1, A_2, \ldots, A_n$ of the matroid, where each A_i contains exactly one element of each B_j.*

This was conjectured in [28]. It has been proved in several special cases. It is also known that there are always at least $n/2$ disjoint bases A_i satisfying the desired property [11] and that there are $(1-o(1))n$ pairwise disjoint independent sets A_i, each of size $(1-o(1))n$, and each containing at most one element from any A_i [35].

A *Latin Square of order* n is an n by n matrix in which each row and each column is a permutation of the n symbols $[n] = \{1, 2, \ldots, n\}$. A *Latin transversal* in such a square is a set of n entries containing one element in each row, one in each column and one copy of each symbol.

Conjecture 1.5 (Ryser's Conjecture, [41], [13]) *Every Latin Square of odd order contains a Latin transversal.*

An equivalent formulation of this conjecture is that for every proper edge coloring of the complete bipartite graph $K_{n,n}$ by n colors, where n is odd, there is a *rainbow* perfect matching, that is, a perfect matching in which no two edges have the same color. It is known that there is a rainbow matching of size at least $n - O(\log n / \log\log n)$, as proved in [31], improving an estimate of [30]. It is also known that for every n besides 3 there are examples of Latin Squares of order n that cannot be partitioned into Latin Transversals. Therefore while the existence of one fair matching is conjectured to always hold (for odd n) the corresponding partition result here fails. This was proved by Euler for all even n, by Mann for all $n \equiv 1 \bmod 4$ [33], and independently by Wanless and Webb and by Evans [46], [19] for the remaining cases.

In the rest of this paper we describe several recent variants and extensions of the examples above. The next section deals with the Necklace Theorem focusing on the investigation of random necklaces and on the algorithmic aspects of the problem. In Section 3 we consider problems dealing with subgraphs of prescribed type in edge colored graphs that contain a fair or nearly fair share of each color. The results in Subsection 3.2 here are new. The final Section 4 contains a discussion of open problems.

2 Necklaces

The bound $(k-1)t$ in the Necklace Theorem (Theorem 1.3) is tight for all admissible values of the parameters. One example demonstrating this is a necklace in which all the beads of each type appear contiguously. In this case at least $k-1$ cuts are needed somewhere in the interval of beads of type i for every i just in order to ensure that each of the collections contains a positive number of beads of each type. A possible interpretation of the Theorem is the following. Suppose that k mathematically oriented thieves want to distribute the necklace fairly among them. The statement ensures that if the number of beads of each of the t types is divisible by k then they can always do it by opening the necklace at the clasp and making at most $(k-1)t$ cuts between beads. This raises two natural questions. The first is if the bound $(k-1)t$ can typically be improved. The second is the algorithmic problem of finding the cuts and the partition into k fair collections efficiently. In this section we briefly describe recent results about both problems.

2.1 Random Necklaces

As mentioned above the bound $(k-1)t$ in the Necklace Theorem is tight. Is the typical minimum number of required cuts smaller? This is studied in a recent joint work in progress with Dor Elboim, Janós Pach and Gábor Tardos, [6]. The random model considered is a necklace of total length $n = ktm$ consisting of exactly km beads of type i for each $1 \le i \le t$, chosen uniformly among all intervals of n beads as above. Call a set of cuts of such a necklace *fair*, if it is possible to split the resulting intervals into k collections, each containing exactly m beads of each type. For a necklace N, let $X = X(N)$ be the minimum number of cuts in a fair collection. When N is chosen randomly as above, X is a random variable which we denote by $X(k,t,m)$. By Theorem 1.3 we have $X(k,t,m) \le (k-1)t$ with probability 1. In [6] we study the typical behavior of the random variable $X = X(k,t,m)$. The results are asymptotic, where at least one of the three variables k,t,m tends to infinity. As usual, we say that a result holds *with high probability* (*whp*, for short), if the probability that it holds tends to 1 when the relevant parameter(s) tend to infinity.

The problem of determining the asymptotic behavior of $X(k,t,m)$ turns out to be connected to several seemingly unrelated topics, including matchings in nearly regular hypergraphs with small codegrees and random walks in Euclidean spaces.

The first observation in [6] is the following.

Proposition 2.1 *For every fixed k and t, as m tends to infinity, $X = X(k,t,m) \ge \left\lceil \frac{(k-1)(t+1)}{2} \right\rceil$ whp.*

The proof is a simple first moment argument, whose details are omitted.

The main result describes the asymptotic behavior of $X = X(k,t,m)$ for two thieves $(k=2)$ and any fixed number of types t, as m tends to infinity.

Theorem 2.2 *Let t be a fixed positive integer and $m \to \infty$.*

1. *For all $1 \le s < \frac{t+1}{2}$,*

$$\mathbb{P}\big(X(2,t,m) = s\big) = \Theta\big(m^{s-\frac{t+1}{2}}\big). \tag{2.1}$$

2. *When t is odd and $s = \frac{t+1}{2}$,*

$$\mathbb{P}\big(X(2,t,m) = s\big) = \Theta\Big(\frac{1}{\log m}\Big). \tag{2.2}$$

3. *For all $\frac{t+1}{2} < s \leq t$,*

$$\mathbb{P}\big(X(2,t,m) = s\big) = \Theta(1). \tag{2.3}$$

Two additional results deal with the case $m = 1$, in which every thief should get a single bead of each type.

Theorem 2.3 *For t and $k/\log t$ tending to infinity, the random variable $X = X(k,t,1)$ is $o(kt)$ whp.*

Theorem 2.4 *The random variable $X = X(2,t,1)$ is at least $2H^{-1}(1/2)t - o(t) = 0.220...t - o(t)$ whp, where $H^{-1}(x)$ is the inverse of the binary entropy function $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ taking values in the interval $[0,1/2]$.*

On the other hand, $X \leq 0.4t + o(t)$ holds whp.

The upper bound above was obtained jointly with Alweiss, Defant and Kravitz, c.f. [6].

The proof of Theorem 2.2 applies the first and second moment and is rather lengthy and technical. A brief outline of the argument for the special case $t = 3$ follows. For $t = 3$ the probability that for the random necklace N, $X(N) = 1$ is easily seen to be $\Theta(1/m)$. By Theorem 1.3 for every N, we have $X(N) \leq 3$. Thus, it remains to show that the probability that $X(N) \leq 2$ is $\Theta(1/\log m)$. In order to estimate this probability, note that two cuts suffice if and only if there is a balanced partition of N into two cyclic intervals that is fair. There are exactly $3m$ balanced partitions into two cyclic intervals. For $0 \leq i < 3m$, we denote by P_i the balanced partition into an interval starting at position $i+1$ and ending at position $i+3m$, and its complement.

Let $Y = Y(N)$ denote the random variable counting the number of fair partitions into cyclic intervals. Clearly, $X(N) \leq 2$ if and only if Y is positive. This probability is lower bounded by the second moment method. It is not too difficult to check that the expectation of Y is $\Theta(1)$ and the expectation of Y^2 is $\Theta(\log m)$. Therefore, by the Paley-Zygmund Inequality [38], [39] the probability that Y is positive is at least $\Omega(1/\log m)$.

The proof of the upper bound for the probability that Y is positive is more interesting. It is done by defining another random variable $Z = Z(N)$. It is then shown that Z is positive with probability $O(1/\log m)$, and that the probability that Y is positive but Z is not, is even lower. The crucial step in bounding the probability that Z is positive, is the analysis of the probability that an appropriate two-dimensional random walk does not return to the origin in a certain number of steps. For this one can apply a slightly modified version of a classical argument of Dvoretzky and Erdős [17]. The details will appear in [6].

The proof of Theorem 2.3 applies a hypergraph edge-coloring result of Pippenger and Spencer [36]. This result asserts that the edges of any hypergraph of constant uniformity and large maximum degree k in which every pair of vertices lie in at most

$o(k)$ common edges, can be partitioned into $(1 + o(1))k$ matchings. By cutting the necklace into intervals of large constant size it is possible to define an appropriate hypergraph and show that whp it satisfies the conditions of the theorem of [36], which provides the required result.

2.2 The Algorithmic Aspects

The proof of Theorem 1.3 is topological. It starts by converting the problem into a continuous one dealing with interval coloring. Let $I = [0, 1]$ be the unit interval. An *interval t-coloring* is a coloring of the points of I by t colors, such that for each $i, 1 \leq i \leq t$, the set of points colored i is (Lebesgue) measurable. Given such a coloring, a *k-splitting of size r* is a sequence of numbers $0 = y_0 \leq y_1 \leq \ldots \leq y_r \leq y_{r+1} = 1$ and a partition of the family of $r + 1$ intervals $F = \{[y_i, y_{i+1}) : 0 \leq i \leq r\}$ into k pairwise disjoint subfamilies $F_1, \ldots, F_k$ whose union is F, such that for each $1 \leq j \leq k$ the union of the intervals in F_j captures precisely $1/k$ of the total measure of each of the t colors. The continuous version of the theorem is then the following.

Theorem 2.5 *Every interval t-coloring has a k-splitting of size $(k - 1) \cdot t$.*

It is not difficult to show that this implies the Necklace Theorem. Indeed, the necklace can be converted to an interval coloring by replacing each bead by a small interval of the corresponding color. If the splitting ensured by the last theorem contains cuts that lie inside intervals corresponding to beads, it can be shown that these can be shifted to produce a splitting of the discrete necklace. The proof of Theorem 2.5 proceeds by first showing, by a simple combinatorial argument, that its validity for (t, k_1) and for (t, k_2) implies its validity for (t, k_1k_2). The main step is a proof that the assertion of the theorem holds for any prime k. This is done by applying a fixed point theorem of Bárány, Shlosman and Szücs [14], which can be viewed as an extension of the Borsuk-Ulam Theorem. Indeed, the case $k = 2$ of Theorem 2.5 admits a short proof using the Borsuk-Ulam Theorem, as shown in [9]. It can also be derived quickly from the Ham Sandwich Theorem, applying it to the measures obtained by placing the interval along the moments curve in R^t. The assertion of the theorem actually holds for general continuous probability measures, and not only for ones corresponding to interval colorings. Indeed, for $k = 2$ this extension is the Hobby-Rice Theorem [29], and the general case is proved in [3]. It is worth noting that a classical result of Liapounoff [32] implies that for any collection of t continuous probability measures μ_i on $[0, 1]$ and any $0 \leq \alpha \leq 1$ there is a subset A of $[0, 1]$ with μ_i measure α for each $1 \leq i \leq t$. The assertion of Theorem 2.5 for general continuous measures shows that for $\alpha = 1/k$ the interval can be partitioned into k such sets A_i, each being a union of a relatively small number of intervals.

The topological proof of the main step in the derivation of Theorem 1.3 is non-constructive, and does not supply any efficient algorithm for finding the required $(k - 1)t$ cuts that provide a fair partition for a given input necklace. For $k = 2$ this algorithmic problem, raised in [4], is called *the Necklace Halving Problem*. A recent result of Filos-Ratsikad and Goldberg [20] shows that this is a hard problem.

PPA and PPAD are two complexity classes introduced by Papadimitriou, [34]. Although this is not our focus here, we include a very brief paragraph about the relevance of these classes to some of the problems discussed here. Both PPA and

PPAD are contained in the class TFNP, which is the complexity class of total search problems, consisting of all problems in NP where a solution exists for every instance. A problem is PPA-complete if and only if it is polynomially equivalent to the canonical problem LEAF, described in [34]. Similarly, a problem is PPAD-complete if and only if it is polynomially equivalent to the problem END-OF-THE-LINE. A problem is PPA-hard or PPAD-hard if the respective canonical problem is polynomially reducible to it. A number of important problems, such as several versions of Nash Equilibrium, have been proved to be PPAD-complete. It is known that PPAD $\subseteq$ PPA. Hence, PPA-hardness implies PPAD-hardness, and if a PPA-hard problem admits an efficient algorithm, so do all problems in PPA (and hence also in PPAD). Filos-Ratsikas and Goldberg [20] showed that the Necklace Halving problem, which is the problem of finding a collection of t cuts that provide a fair partition of a given input necklace with beads of t types and an even number of beads of each type, is PPA-hard [20]. This suggests the problem of finding an efficient algorithm for obtaining a fair partition using a somewhat larger number of cuts. An early result in this direction appears in [12], but it only provides a partition in which the number of beads of each type in the two collections are close to each other, and the number of cuts is exponential in the number of types. A recent improved algorithm is given in [7]. Its performance is described in the next result.

Theorem 2.6 *There is a polynomial time algorithm that given an input necklace with beads of t types, in which the number of beads of each type is an even number that does not exceed m, produces a collection of at most $t(\log m + O(1))$ cuts and a partition of the resulting intervals into two collections, each containing exactly half of the beads of each type.*

The algorithm proceeds by first converting the problem to the continuous interval coloring problem described above. The continuous problem is tackled using a linear algebra procedure based on Carathéodory's Theorem for cones. Its solution can then be rounded to produce a solution of the discrete problem. The details are sketched below.

Proof of Theorem 2.6 (sketch):

Given a necklace with m_i beads of color i for $1 \leq i \leq t$, where $m = \max m_i$, replace each bead of color i by an interval of μ_i-measure $1/m_i$ and μ_j-measure 0 for all $j \neq i$. These intervals are placed next to each other according to the order in the necklace, and their lengths are chosen so that altogether they cover $[0,1]$. We first describe a procedure that splits the interval into two collections so that for every i the difference between the μ_i-measures of the two collections is at most ε, where $\varepsilon = \frac{1}{2m}$. The number of cuts used here is at most $t(\log m + O(1))$. It is then not too difficult to round the cuts and get a solution of the discrete problem without increasing the number of cuts.

Let μ_i be the t measures defined above. Our objective is to describe an efficient algorithm that cuts the interval in at most $t(2 + \lceil \log_2 \frac{1}{\varepsilon} \rceil)$ places and splits the resulting intervals into two collections C_0, C_1 so that $\mu_i(C_j) \in [\frac{1}{2} - \frac{\varepsilon}{2}, \frac{1}{2} + \frac{\varepsilon}{2}]$ for all $i \in [t] = \{1, 2, \ldots, t\}, 0 \leq j \leq 1$.

For each interval $I \subset [0,1]$ denote $\mu(I) = \mu_1(I) + \ldots + \mu_t(I)$. Thus $\mu([0,1]) = t$. Using $2t - 1$ cuts split $[0,1]$ into $2t$ intervals $I_1, I_2, \ldots, I_{2t}$ so that $\mu(I_r) = 1/2$ for all

r. Note that it is easy to find these cuts efficiently, since each measure μ_i is uniform on its support. For each interval I_r let v_r denote the t-dimensional vector

$$(\mu_1(I_r), \mu_2(I_r), \dots, \mu_t(I_r)).$$

By a simple linear algebra argument, which is a standard fact about the properties of basic solutions for Linear Programming problems, one can write the vector $(1/2, 1/2, \dots, 1/2)$ as a linear combination of the vectors v_r with coefficients in $[0,1]$, where at most t of them are not in $\{0,1\}$. This follows from Carathédory's Theorem for cones. Here is the simple proof, which also shows that one can find coefficients as above efficiently. Start with all coefficients being $1/2$. Call a coefficient which is not in $\{0,1\}$ *floating* and one in $\{0,1\}$ *fixed*. Thus at the beginning all $2t$ coefficients are floating. As long as there are more than t floating coefficients, find a nontrivial linear dependence among the corresponding vectors and subtract a scalar multiple of it which keeps all floating coefficients in the closed interval $[0,1]$ shifting at least one of them to the boundary $\{0,1\}$, thus fixing it.

This process clearly ends with at most t floating coefficients. The intervals with fixed coefficients with value 1 are now assigned to the collection C_1 and those with coefficient 0 to C_0. The rest of the intervals remain. Split each of the remaining intervals into two intervals, each with μ-value $1/4$. We get a collection $J_1, J_2, \dots, J_m$ of $m \le 2t$ intervals, each of them has the coefficient it inherits from its original interval. Each such interval defines a t-vector as before, and the sum of these vectors with the corresponding coefficients (in $(0,1)$) is exactly what the collection C_1 should still get to have its total vector of measures being $(1/2, \dots, 1/2)$.

As before, we can shift the coefficients until at most t of them are floating, assign the intervals with $\{0,1\}$ coefficients to the collections C_0, C_1 and keep at most t intervals with floating coefficients. Split each of those into two intervals of μ-value $1/8$ each and proceed as before, until we get at most t intervals with floating coefficients, where the μ-value of each of them is at most $\varepsilon/2$. This happens after at most $\lceil \log_2(1/\varepsilon) \rceil$ rounds. In the first one, we have made $2t-1$ cuts and in each additional round at most t cuts. Thus the total number of cuts is at most $t(2 + \lceil \log_2(1/\varepsilon) \rceil) - 1$.

From now on we do not increase the number of cuts, and show how to shift them and allocate the remaining intervals to C_0, C_1. Let $\mathcal{I}$ denote the collection of intervals with floating coefficients. Then $|\mathcal{I}| \le t$ and $\mu(I) \le \varepsilon/2$ for each $I \in \mathcal{I}$. This means that

$$\sum_{i=1}^{t} \sum_{I \in \mathcal{I}} \mu_i(I) \le t\varepsilon/2.$$

It follows that there is at least one measure μ_i so that

$$\sum_{I \in \mathcal{I}} \mu_i(I) \le \varepsilon/2.$$

Observe that for any assignment of the intervals $I \in \mathcal{I}$ to the two collections C_0, C_1, the total μ_i-measure of C_1 (and hence also of C_0) lies in $[1/2-\varepsilon/2, 1/2+\varepsilon/2]$, as this measure with the floating coefficients is exactly $1/2$ and any allocation of the intervals with the floating coefficients changes this value by at most $\varepsilon/2$. We can thus ignore this measure, for ease of notation assume it is measure number t, and replace each measure vector of the members in $\mathcal{I}$ by a vector of length $t-1$ corresponding to the other $t-1$ measures. If $|\mathcal{I}| > t-1$ (that is, if $|\mathcal{I}| = t$), then it is possible to shift the floating coefficients as before until at least one of them reaches the boundary, fix it assigning its interval to C_1 or C_0 as needed, and omit the corresponding interval from $\mathcal{I}$ ensuring its size is at most $t-1$. This means that for the modified $\mathcal{I}$ the sum

$$\sum_{i=1}^{t-1} \sum_{I \in \mathcal{I}} \mu_i(I) \leq (t-1)\varepsilon/2.$$

Hence there is again a measure μ_i, $1 \leq i \leq t-1$ so that

$$\sum_{I \in \mathcal{I}} \mu_i(I) \leq \varepsilon/2.$$

Again, we may assume that $i = t-1$, observe that measure number $t-1$ will stay in its desired range for any future allocation of the remaining intervals, and replace the measure vectors by ones of length $t-2$. This process ends with an allocation of all intervals to C_1 and C_0, ensuring that at the end $\mu_i(C_j) \in [1/2 - \varepsilon/2, 1/2 + \varepsilon/2]$ for all $1 \leq i \leq t$, $0 \leq j \leq 1$. These are the desired collections. It is clear that the procedure for generating them is efficient, requiring only basic linear algebra operations.

This completes the (sketch of the) proof. The full details can be found in [7]. □

3 Graphs

3.1 Fair Representation

Theorem 1.2 and Conjecture 1.5 mentioned in Section 1 are two examples of fair representation problems dealing with graphs. There are quite a few additional results and conjectures of this type. We start this section by discussing several examples.

An *optimal proper edge coloring* of the complete graph K_{2n} on an even number of vertices is a coloring of the edges by $2n-1$ colors, each forming a perfect matching. Given such an edge coloring, the fair share of a spanning tree in each color class is exactly 1. Brualdi and Hollingsworth [10] conjectured that for each such edge coloring of $K = K_{2n}$ where $n > 4$ one can partition all edges of K into n pairwise edge disjoint *rainbow* spanning trees, that is, each tree containing exactly one edge of each color. Constantinos [15] conjectured that it is even possible to find such a partition in which all trees are isomorphic. This is proved for all sufficiently large n in a recent paper of Glock, Kühn, Montgomery and Osthus [23]. The proof is probabilistic, and is based on hypergraph matching results and the so-called absorption technique. This technique starts by removing an appropriate small part of the graph, finding an approximate partition of the rest, and then using the small part to complete it to a precise partition. The details, which require quite some work, can be found in [23]. Similar ideas are useful in the study of several related problems, as described in [23] and its references.

The Cycle and Triangles Theorem (Theorem 1.2) has been proved in [21] using the algebraic approach of [8]. This approach enables one to bound the chromatic number of a graph, and in fact even its so-called list chromatic number, by showing that a certain coefficient of an appropriate polynomial is nonzero. Subsequent proofs of the theorem (at least of the statement about the existence of a single independent set of the required form) apply topological ideas. The shortest proof is the one in [1] where the result is derived from Schrijver's Theorem on vertex critical subgraphs of the Kneser graph. This Theorem, which strengthens the result of Lovász about the chromatic number of the Kneser graph, is proved in [43] using the Borsuk-Ulam Theorem.

Theorem 3.1 (Schrijver [43]) *For $n > 2k$ the family of independent sets of size k in the cycle C_n cannot be partitioned into fewer than $n-2k+2$ intersecting families.*

Now let G be a cycle of length $3m$ and let P be a partition of its set of vertices into pairwise disjoint sets $P_1, P_2, \ldots, P_m$, each of size 3. Assuming the first assertion of Theorem 1.2 fails, there is no independent set of G that contains exactly one vertex of each set P_i. In this case each independent set of size m in the cycle G contains at least two vertices in some set P_i, and we can partition all these independent sets into m families, where a set S belongs to family number i iff i is the smallest index so that $|S \cap P_i| \geq 2$. Note that each such family is intersecting, as each member of it contains at least two vertices among the three vertices of P_i. But since $m < 3m - 2m + 2$ this contradicts Theorem 3.1 with $n = 3m$ and $k = m$, proving the existence of an independent set containing one vertex in each P_i.

The short proof above can be extended in several ways. In particular the following holds.

Proposition 3.2 ([1]) *If $V = V_1 \cup V_2 \cup \cdots \cup V_m$ is a partition of the vertex set of a cycle C into m pairwise disjoint sets, and $|V_i|$ is odd for all i, then for any vertex v of C there is an independent set S of C so that $v \notin S$ and $|S \cap V_i| = (|V_i| - 1)/2$ for all i.*

In an attempt to strengthen Ryser's Conjectures (Conjecture 1.5), Stein [44] suggested the stronger conjecture that for any partition P of the edges of the complete bipartite graph $K_{n,n}$ into n pairwise disjoint color classes, each containing exactly n edges, there exists a rainbow matching of size $n-1$. This turned out to be too strong. A counterexample was found by Pokrovskiy and Sudakov in [37], where the authors describe a coloring as above so that every matching misses at least $\Omega(\log n)$ color classes. This shows that in some natural cases tight fair representations fail and suggests relaxed versions of questions of this type, as discussed in the following subsection.

3.2 Nearly Fair Representation

A special case of the approach described here was initiated in discussions with Eli Berger and Paul Seymour. Let $G = (V, E)$ be a graph and let P be an arbitrary partition of its set of edges into m pairwise disjoint subsets $E_1, E_2, \ldots, E_m$. The sets E_i are called the color classes of the partition. For any subgraph $H' = (V', E')$ of G, let $x(H', P)$ denote the vector $(x_1, x_2, \ldots, x_m)$, where $x_i = |E_i \cap E'|$ is the number of edges of H' that lie in E_i. Thus, in particular, $x(G,P) = (|E_1|, \ldots, |E_m|)$. In a completely fair representation of the sets E_i in H', each entry x_i of the vector $x(H',P)$ should be equal to $|E_i| \cdot \frac{|E'|}{|E|}$. Of course such equality can hold only if all these numbers are integers. But even when this is not the case the equality may hold up to a small additive error.

We are interested in results and conjectures asserting that when G is either the complete graph K_n or the complete bipartite graph $K_{n,n}$, then for certain graphs H and for any partition P of $E(G)$ into color classes $E_1, \ldots, E_m$, there is a subgraph H' of G which is isomorphic to H so that the vector $x(H',P)$ is close (or equal) to the vector $x(G,P)\frac{|E(H')|}{|E(G)|}$. As mentioned in the previous subsection, Stein [44] conjectured that if $G = K_{n,n}$ and P is any partition of the edges of G into n sets, each of size n, then there is always a rainbow matching of size $n-1$ in G. However, this turned out to be false as shown by a clever counter-example of Pokrovskiy and Sudakov [37].

In [1] it is conjectured that when $G = K_{n,n}$, P is arbitrary, and H is a matching of size n, then there is always a copy H' of H (that is, a perfect matching H' in G), so that

$$\|x(H',P) - \frac{1}{n}x(G,P)\|_\infty < 2.$$

This is proved in [1] (in a slightly stronger form) for partitions P with 2 or 3 color classes. Here we first prove the following, showing that when allowing a somewhat larger additive error (which grows with the number of colors m but is independent of n) a similar result holds for partitions with any fixed number of classes.

Theorem 3.3 *For any partition P of the edges of the complete bipartite graph $K_{n,n}$ into m color classes, there is a perfect matching M so that*

$$\|x(M,P) - \frac{1}{n}x(K_{n,n},P)\|_\infty \le \|x(M,P) - \frac{1}{n}x(K_{n,n},P)\|_2 < (m-1)2^{(3m-2)/2}.$$

It is worth noting that a random perfect matching M typically satisfies

$$\|x(M,P) - \frac{1}{n}x(K_{n,n},P)\|_\infty \le O_m(\sqrt{n}).$$

The main challenge addressed in the theorem is to get an upper bound independent of n.

Theorem 3.3 is a special case of a general result which we describe next, starting with the following definition.

Definition 3.4 Let G be a graph and let H be a subgraph of it. Call a family of graphs $\mathcal{H}$ (which may have repeated members) a *uniform cover of width s of the pair* (G,H) if the following four conditions hold.

- Every member H' of $\mathcal{H}$ is a subgraph of G which is isomorphic to H.
- The number of edges of each such H' which are not edges of H is at most s.
- Every edge of H belongs to the same number of members of $\mathcal{H}$.