A Note on the Union-Closed Sets Conjecture

Theresa P. Vaughan
Department of Mathematics
University of North Carolina at Greensboro
Greensboro, NC 27412

January 22, 2002

Abstract

A union closed (UC) family $\mathcal{A}$ is a finite family of sets such that the union of any two sets in $\mathcal{A}$ is also in $\mathcal{A}$. Peter Frankl conjectured that for every union closed family $\mathcal{A}$, there exists some x contained in at least half the members of $\mathcal{A}$. This is the union-closed sets conjecture.

An FC family is a UC family $\mathcal{B}$ such that for every UC family $\mathcal{A}$, if $\mathcal{B} \subseteq \mathcal{A}$, then $\mathcal{A}$ satisfies the union-closed sets conjecture. We give a heuristic method for identifying possible FC families, and apply it to families in $\mathcal{P}(5)$ and $\mathcal{P}(6)$.

1 Introduction

A union closed (UC) family $\mathcal{A}$ is a finite family of sets, such that the union of any two sets in $\mathcal{A}$ is also in $\mathcal{A}$.

In 1979, Peter Frankl conjectured that for every UC family $\mathcal{A}$, there exists some x contained in at least half the members of $\mathcal{A}$.

This is one of those fascinating conjectures which has a simple elementary statement, but which has resisted (so far) all attempts to prove it. There are many results for special cases; that is, quite a few sufficient conditions have been established for the property " there exists some x contained in at least half the members of $\mathcal{A}$", but as far as I know, no necessary conditions.

The conjecture has been verified for various small values of $|\mathcal{A}|$ and $\cup\mathcal{A}$ (see e.g. [5], [7], [3]). A UC family can be identified with a lattice [5], and the lattice form of the conjecture has been verified for some special types of lattices, e.g. lower semi-modular lattices and lower quasi-semimodular lattices ([6], [1]).

A result of a somewhat different kind is given in [4]; we showed that a UC family $\mathcal{A}$ has a dual UC family of the same cardinality, and that one or the other of $\mathcal{A}$ or its dual, must satisfy the conjecture.

JCMCC **45** (2003), pp. 95-108

The most general of all the sufficient conditions I have found so far, was given by Bjorn Poonen. Poonen gives a characterization of a special type of UC family, which we call FC families. An FC family $\mathcal{B}$ is a UC family having the property that for every UC family $\mathcal{A}$ containing $\mathcal{B}$, it is true that one of the elements of $\cup\mathcal{B}$ is in at least half the members of $\mathcal{A}$. To state Poonen's result, we define: $N_i(\mathcal{A})$ is the number of members of $\mathcal{A}$ which contain the element i, and if $\mathcal{A}$ and $\mathcal{B}$ are UC families, then $\mathcal{A} \uplus \mathcal{B} = \{X \cup Y : X \in \mathcal{A}, Y \in \mathcal{B}\}$. We use the notation $\mathcal{P}(n)$ to denote the power set of $\{1, 2, \ldots, n\}$.

Theorem 1.1 *(Poonen, [5])*
$\mathcal{B} \in \mathcal{P}(n)$ is an FC family if and only if there exist non-negative real numbers $c_1, c_2, \cdots, c_n$ with sum 1 such that

(P) for every UC family $\mathcal{A} \in \mathcal{P}(n)$ satisfying $\mathcal{A} \uplus \mathcal{B} \subseteq \mathcal{A}$,

$$\sum_{i=1}^{n} c_i N_i(\mathcal{A}) \geq |\mathcal{A}|/2.$$

Poonen's theorem provides a (theoretically computable) necessary and sufficient condition for $\mathcal{B}$ to be an FC family; then each known FC family $\mathcal{B}$ yields an infinite class of UC families satisfying the conjecture, namely the UC families containing $\mathcal{B}$. On the other hand, any family for which the conjecture fails, cannot contain any FC families; then each known FC family supplies limiting conditions on any family in which the conjecture fails. It is to be hoped that some special approach would be applicable to UC families not containing any FC families.

Condition (P) involves solving a very large number of linear inequalities, and furthermore it is not easy to identify the families $\mathcal{A} \in \mathcal{P}(n)$ satisfying $\mathcal{A} \uplus \mathcal{B} \subseteq \mathcal{A}$. However, if one can find a suitable set of numbers c_i, it is possible to check whether or not Condition (P) holds without actually listing all the pertinent families.

Any set of cardinality 1 or 2 is an FC-family, and Poonen shows in [5] (using condition (P)) that a 4-set together with three of its 3-subsets is an FC family, and a 4-set together with just two of its 3-subsets is not an FC family. In [8], we used the special case when the c_i are all equal, to prove that a few families in $\mathcal{P}(n)$ for $n = 5, 6, 7$, were FC families.

Identifying all the FC families in $\mathcal{P}(5)$ is a tremendous computational problem; in this paper we use a short-cut method, which does not give full information, but which does allow the identification of most of the FC families in $\mathcal{P}(5)$ (and two FC families in $\mathcal{P}(6)$).

In Section 2, we describe a heuristic method for finding a reasonable candidate for the numbers c_i, supposing there are any, for a given family $\mathcal{B} \in \mathcal{P}(n)$. This method amounts to finding the solution $x = (x_1, x_2, \ldots, x_n)$ for a particular set of n equations in n unknowns, and is very easily programmable. If the solution x has every $x_i \geq 0$, and $\sum x_i \leq 1$, then we construct the numbers c_i using the x_i; otherwise the method gives no information.

Let $M(\mathcal{B}) = M$ denote the coefficient matrix of the above system of equations. In all our calculations, for different families $\mathcal{B}$, the matrices $M(\mathcal{B})$ all share a rather long list of common properties (e.g. they are all diagonalizable). It is tempting to conjecture that most of these common properties are generally true for all such $M(\mathcal{B})$. We list these among the questions in Section 6.

In the computations I have done so far, whenever the method produces a suitable set of numbers c_i for a family $\mathcal{B}$, it has always been possible to verify that $\mathcal{B}$ is in fact an FC-family. I have not been able to prove that this is generally true.

In Section 3, we apply the method above to families $\mathcal{B} \in \mathcal{P}(5)$, and also to two families in $\mathcal{P}(6)$.

From this work, for example, we can say that if a UC family contains three 3-sets which contain a common 2-set, then the conjecture holds for that family; and in any UC family for which the conjecture fails, a 5-set (in the family) can contain no more than two 3-sets (in the family), and a 6-set can contain no more than three 3-sets. The data also suggest that a 7-set can contain no more than three 3-sets, and an 8-set can contain no more than four 3-sets; however we have not proved these results.

In Section 4, we work out one case (in $\mathcal{P}(5)$) in detail, and in Section 5 we outline a proof for one case in $\mathcal{P}(6)$.

In Section 6, we give a list of questions suggested by our calculations.

2 The Method and the Matrix M

In this section we describe a heuristic procedure which, for a given UC family $\mathcal{B} \in \mathcal{P}(n)$, either produces a set $\{c_1, c_2, \ldots, c_n\}$ suitable for checking to see if Poonen's condition (P) is satisfied, or else gives no information.

This procedure was developed almost accidentally. In [8] using the special case when all the c_i are equal, I was able to prove that a few families were in fact FC families. In all these proofs, in trying to show that $\mathcal{B}$ is an FC family, by far the most difficult verifications for the families $\mathcal{A}$ such that $\mathcal{A} \uplus \mathcal{B} \subseteq \mathcal{A}$, involved the families $\mathcal{A}$ which were generated by $\mathcal{B}$, all but one of the singletons, and the empty set. The difference in difficulty was really remarkable. Furthermore, whenever the attempt to prove that $\mathcal{B}$ was an FC family (using equal c_i) was unsuccessful, there was always one of these particular families that did not work. This led to the idea of using these families to help in determining the values of c_i to use.

If s is a finite collection of finite sets, then the UC family generated by s is the smallest UC family containing s; this family consists of s together with all unions of members of s. It is proved in [4] that every UC family has a unique minimal generating set.

Let $\mathcal{B} \in \mathcal{P}(n)$ be a UC family, where we suppose that $\cup\mathcal{B} = \{1, 2, \ldots, n\} = S$. For each $1 \leq i \leq n$, let A_i be the UC family generated by $\mathcal{B}$ and $\mathcal{P}(S - \{i\})$. Put $M_{ij} = N_j(A_i)$, and let M be the matrix (M_{ij}), and let b be the column vector $b = \text{col}(b_1, b_2, \ldots, b_n)$ where $b_i = |A_i|/2$.

For a UC family $\mathcal{B}$ generated by a collection s of sets, and the corresponding matrix equation $Mx = b$, put $NUM(\mathcal{B}) = NUM(s) = \sum x_i$. This number is easily computable (e. g. by a Maple program) and we use it to identify possible FC families: If $NUM(s) \leq 1$, then $\mathcal{B}$ is a candidate.

Now, suppose the matrix equation $Mx = b$ has a solution $x = \text{col}(x_1, x_2, \ldots, x_n)$ such that $\sum x_i \leq 1$, and all the entries x_i are non-negative. Then we define the numbers c_i as follows: Put $u = 1 - \sum x_i$, and for each i, $c_i = x_i + u/n$. Then $\sum c_i = 1$ and the c_i are all non-negative; using these c_i, it remains to verify that Poonen's condition (P) is satisfied.

Obviously, it would be possible to construct the numbers c_i in other ways; for instance, if one (or more) of the x_i were negative, and the sum of the negative x_i were smaller in magnitude than $1 - \sum x_i$, we could add on positive quantities to the negative x_i to produce non-negative c_i. However, the method described above has the advantage of being easily programmable; it also preserves any regularity there may be in the numbers x_i, which in practice is sometimes convenient.

3 Applications

In this section, we compute $NUM(s)$ for some particular cases, using the method described in Section 2.

Two UC families in $\mathcal{P}(n)$ are isomorphic if there is some permutation of $\{1, 2, \ldots, n\}$ which transforms one family into the other. Since each UC family has a unique minimal generating set, we may classify UC families according to the isomorphism class of their generating sets.

If s and t are collections of sets in $\mathcal{P}(n)$, and if there is a permutation σ of $1, 2, \ldots, n$ such that $\sigma(s) = t$, then $NUM(s) = NUM(t)$; in the table below we list representatives of non-isomorphic collections s ($n = 5, 6$) containing no sets of cardinality one or two. For simplicity, we use the notation $abc\ldots$ for the set $\{a, b, c, \ldots\}$; e.g. 14 represents the 2-set $\{1, 4\}$.

We are ultimately interested in FC families, and so, having found that some configuration s generates an FC family, we have not listed any configuration t containing a permutation-isomorphic copy of s. For example, any five 3-sets in $\mathcal{P}(6)$ will have at least three 3-sets in a 5-set, so we do not list any such collection. There are just two isomorphism types of four 3-sets in a 6-set which do not contain three 3-sets from a 5-set; these are $123, 124, 356, 456$

and 123, 345, 246, 156; since 123, 345, 246 generates an FC family, we include only 123, 124, 356, 456 as four 3-sets in a 6-set. There are, up to isomorphism, only three types of three 3-sets in a 6-set and not in a 5-set: one type with no repeated pair; one type with one repeated pair and a common element in all three; and one type with one repeated pair and no common element.

Generating set s	NUM(s)
123, 124, 125	1
123, 124, 135	.99230
123, 124, 345	.99234
123, 124, 1235	1.0119
123, 124, 1345	.99623
123, 145, 1234	1.0108
123, 145, 2345	1
123, 124, 1235, 1245	1.01138
123, 145, 1234, 1235	1.00692
123, 145, 1234, 1245	1.00514
123, 145, 1234, 1235, 1245	1.000848
123, 145, 1234, 1235, 1245, 1345	.99627
123, 1234, 1235, 1245, 1345	1.00251
234, 1234, 1235, 1245, 1345	.98805
1234, 1235, 1245, 1345	1.011508
1234, 1235, 1245, 1345, 2345	.990566
123, 124, 356, 456	.99363
123, 124, 156	.99724
123, 345, 246	.99378
123, 124, 356	1.0035

For the generating sets s listed above, we have been able to show that $NUM(s) \leq 1$ implies that $\mathcal{B}$ is an FC family. We give the details of the proof for the case $s = \{123, 124, 135\}$ in Section 4, and an outline of the proof for the case $s = \{123, 124, 356, 456\}$ in Section 5. All the proofs are similar: a case-by-case analysis according to the numbers of sets of given cardinality, in which the details depend on the actual values of the numbers c_i produced by the method.

We have not been able to prove that $NUM(s) > 1$ implies that the family generated by s is not FC (see Question 2 in Section 6). For $\mathcal{P}(5)$, if this were true, we could then give a complete list of all the FC families in $\mathcal{P}(5)$, namely, any family which contains either a singleton, a doubleton, three 3-sets in a 4-set, or one of the 5-configurations s listed above with $NUM(s) \leq 1$.

The list above yields restrictions which must be satisfied by any UC family $\mathcal{A}$ for which the conjecture fails: such a family cannot contain any FC family, and therefore cannot contain any configuration s listed above for which $NUM(s) \leq 1$. For example, $\mathcal{A}$ can have no more than two 3-sets in any 5-set and no more than three 3-sets in any 6-set.

The list above is useful in considering 3-sets in $\mathcal{P}(7)$ and $\mathcal{P}(8)$. Suppose that s is a collection of four 3-sets in $\mathcal{P}(7)$. Then one of the elements $1, 2, \ldots 7$ must be contained in just one of the members of s; say this element is 7. If the three 3-sets not containing 7 do not generate an FC family, then they must be isomorphic to 123, 124, 356. So, if the family $\mathcal{B}$ generated by s is not already an FC family, then (up to isomorphism) s has the form $123, 124, 356, ab7$. Then we compute that $NUM(s) \leq 1$ for all possible values of ab. This suggests that $\mathcal{B}$ would be an FC family; if this is true, applying the same idea to a set s of five 3-sets in $\mathcal{P}(8)$ implies that s generates an FC family.

For all cases with $n = 2, 3, 4$, it is true that $NUM(s) \leq 1$ implies that $\mathcal{B}$ is an FC family.

Computing $NUM(s)$ for assorted collections of $1 + \lfloor n/2 \rfloor$ 3-sets in $\mathcal{P}(n)$, for $n = 8, 9, 10, 11$, in every case we find $NUM(s) \leq 1$.

Applying the Maple program to the sets T_n consisting of all the (n-1)-subsets of $\{1, 2, \ldots, n\}$, for $3 \leq n \leq 12$, the values of $NUM(T_n)$ are (to four decimals): $.9545, .9600, .9906, 1.0270, 1.0574, 1.0772, 1.0870, 1.0898, 1.0882, 1.0845$.

4 One case in $\mathcal{P}(5)$

Let $\mathcal{B} = \{123, 124, 135, 1234, 1235, 12345\}$. In this section we give the details of the proof that $\mathcal{B}$ is an FC family. Observe that $\mathcal{B}$ has just one non-trivial automorphism, namely $(2,3)(4,5)$. We first give some general definitions, and for convenience recall some notation: For a UC family $\mathcal{A}$,

$S(\mathcal{A}) = S = \cup\mathcal{A}$;
$N_i(\mathcal{A}) = N_i$ is the number of sets in $\mathcal{A}$ containing i;
$n_i(\mathcal{A}) = n_i$ is the number of sets in $\mathcal{A}$ of cardinality i;
A_i is the UC family generated by $\mathcal{A}$ and $\mathcal{P}(S - \{i\})$;
$M_{ij}(\mathcal{A}) = M_{ij} = N_j(A_i)$ and $M = (M_{ij})$;
$b_i(\mathcal{A}) = b_i = |A_i|/2$ and $b = \mathrm{col}(b_1, b_2, \ldots)$;
$x = \mathrm{col}(x_1, x_2, \ldots)$ is a solution to the matrix equation $Mx = b$.

Suppose that $\mathcal{A}$ is a UC family in $\mathcal{P}(n)$, with $\cup\mathcal{A} = \{1, 2, \ldots, n\}$, for which the method of Section 2 produces non-negative numbers c_i with $\sum c_i = 1$.

Definition 4.1 *For any set X in $\mathcal{P}(n)$, define $f(X) = \sum_{i \in X} c_i$ and $h(X) = f(X) - 1/2$. Then, if T is any collection of sets in $\mathcal{P}(n)$, let $f(T) = \sum_{X \in T} f(X)$ and $h(T) = \sum_{X \in T} h(X)$.*

Note that if $\mathcal{A}$ is a UC family in $\mathcal{P}(n)$, then $h(\mathcal{A}) = \sum_{i=1}^{n} c_i N_i(\mathcal{A}) - |\mathcal{A}|/2$; $h(X) + h(S - X) = 0$ for every $X \in \mathcal{P}(n)$, and $h(\mathcal{P}(n)) = 0$. The first result is an easy consequence of this observation.

Theorem 4.2 *If $\mathcal{A} \subseteq \mathcal{P}(n)$, and $h(\mathcal{A}) \geq 0$, then $\mathcal{A}$ satisfies the Frankl Conjecture.*

Proof: We have $h(\mathcal{A}) = \sum_{i=1}^{n} c_i N_i(\mathcal{A}) - |\mathcal{A}|/2 \geq 0$, that is $\sum_{i=1}^{n} c_i N_i(\mathcal{A}) \geq |\mathcal{A}|/2$. Since $\sum c_i = 1$, at least one of the N_i must be greater than or equal to $|\mathcal{A}|/2$, i.e. i is contained in at least half of the members of $\mathcal{A}$.

The converse is certainly not true; in $\mathcal{P}(5)$, the UC family consisting of one singleton will always have $h < 0$.

Definition 4.3 *Let $\mathcal{A}$ be any family of sets in $\mathcal{P}(n)$. For $0 \leq k \leq n$, let E_i be the collection of sets of cardinality i in $\mathcal{A}$, $n_i = |E_i|$, and $e_i = h(E_i)$ (if $E_i = \emptyset$, we put $e_i = 0$).*

Now let $\mathcal{B} = \{123, 124, 135, 1234, 1235, 12345\}$ (the UC family generated by $s = \{123, 124, 135\}$). The family A_1, for example, is generated by $\mathcal{B}$ and $\mathcal{P}(\{2,3,4,5\})$. Thus A_1 contains $\mathcal{B}$, $\mathcal{P}(\{2,3,4,5\})$, and two extra 4-sets, 1345 and 1245, and so $|A_1| = 16 + 6 + 2 = 24$; by counting, we find $n_0 = 1$, $n_1 = 4$, $n_2 = 6$, $n_3 = 7$, $n_4 = 5$ and $n_5 = 1$; $N_1(A_1) = M_{11} = 8$, $N_2(A_1) = M_{12} = N_3(A_1) = M_{13} = 14$, $N_4(A_1) = N_5(A_1) = 13$, giving the first row of the matrix M. We get

$$M = \begin{pmatrix} 8 & 14 & 14 & 13 & 13 \\ 14 & 6 & 12 & 12 & 11 \\ 14 & 12 & 6 & 11 & 12 \\ 12 & 12 & 10 & 4 & 10 \\ 12 & 10 & 12 & 10 & 4 \end{pmatrix}$$

and $b = [|A_1|/2, |A_2|/2, |A_3|/2, |A_4|/2, |A_5|/2] = [12, 11, 11, 10, 10]$. The solution x to $Mx = b$ is [173/650, 14/65, 14/65, 48/325, 48/325], and the sum of the x_i is 129/130. The corresponding vector c is found by adding (1/5)(1/130) to each entry of x: c = [87/325, 141/650, 141/650, 97/650, 97/650], and the sum of the entries of c is 1. In decimal notation, truncated here to four digits, we get c = [.2676, .2169, .2169, .1492, .1492].

Now, for each $1 \leq k \leq 4$, arrange the k-sets of $\mathcal{P}(5)$ in lexicographic order $X_1, X_2, \ldots$ (e.g. for 2-sets, $X_1 = 12$ is on the far left, and $X_{10} = 45$ is on the far right), and compute the entries of the vectors $v_k = [h(X_1), h(X_2), \ldots]$. We get

$$v_1 = [-.2323, -.2830, -.2830, -.3507, -.3507]$$

$$v_2 = [-.0153, -.0153, -.0830, -.0830, -.0661, -.1338, -.1338, -.1338, -.1338, -.2015]$$

$$v_3 = [.2015, .1338, .1338, .1338, .1338, .0661, .0830, .0830, .0153, .0153]$$

$$v_4 = [.3507, .3507, .2830, .2830, .2323].$$

We also have $h(\emptyset) = -.5$ and $h(12345) = .5$. It is convenient that the members of $\mathcal{B}$ have the largest possible h-values (but this is not generally true).

Now suppose that $\mathcal{A}$ is a UC family in $\mathcal{P}(5)$ satisfying $\mathcal{A} \uplus \mathcal{B} \subseteq \mathcal{A}$. To avoid trivialities, we assume that $\mathcal{A}$ contains some non-empty sets. Evidently if $X \in \mathcal{B}$

and $X \notin \mathcal{A}$, then $\mathcal{A}$ cannot contain any of the subsets of X, and in particular $\mathcal{A}$ does not contain $\emptyset$.

Put $e = h(\mathcal{A}) = e_5 + e_4 + e_3 + e_2 + e_1 + e_0$; we will show that $e \geq 0$. Note that if $e_4 + e_3 + e_2 + e_1 \geq 0$ then also $e \geq 0$.

We observe that if X is a 3-set or a 4-set, then $h(X) > 0$; if X is a 1-set or a 2-set, then $h(X) < 0$, and if X is a 4-set and Y is a 2-set, then $h(X)+h(Y) > 0$. We also have $h(X) + h(S - X) = 0$ for every X, and then the following Lemma is obvious.

Lemma 4.4 *(a) If $n_1 = n_2 = 0$, then $e \geq 0$.*
(b) If $n_4 = 5$ and $n_3 = 10$, then $e \geq 0$.
(c) If $\mathcal{A} \subseteq \mathcal{C}$, $h(\mathcal{C}) \geq 0$, and $n_i(\mathcal{A}) = n_i(\mathcal{C})$ for $i = 3, 4$, then $e \geq 0$.
(d) If $\mathcal{A} \subseteq \mathcal{C}$, $h(\mathcal{A}) \geq 0$, and $n_i(\mathcal{A}) = n_i(\mathcal{C})$ for $i = 1, 2$, then $e \geq 0$.

If $n_1 = 5$, then $h(\mathcal{A})$ is either 0 or .5, so we assume that $n_1 \leq 4$. We first prove that $e(\mathcal{A}) \geq 0$ if $n_1 = 4$, and then, assuming $n_1 \leq 3$, we consider cases according to the value of n_4.

Theorem 4.5 *If $n_1 = 4$, then $e(\mathcal{A}) \geq 0$.*

Proof: Since $n_1 = 4$, then $\mathcal{A}$ contains one of the families A_i. Suppose that $\mathcal{A}$ contains A_5. The vector Mc gives the values of $h(A_i)$ (which are all positive), and we observe that $h(A_5) = e(A_5)$ is approximately .07.

The sets not in A_5 are: 5, 15, 25, 35, 45, 125, 145, 235, 245, 345, 1245, 2345. If $\mathcal{A}$ contains 15 then it also has 125, 145, 1245; if it has 25, then it also has 125, 235, 245, 1245, 2345; if it has 35, it also has 235, 345, 2345; if it has 45, then it also has 145, 245, 345, 1245, 2345. Suppose $\mathcal{A}$ has just one 2-set which is not in A_5, say 15. Put $X = \{15, 125, 145, 1245\}$, and observe that $e(X) > 0$ since it has just one 2-set, and a 4-set. Then $e(A_5 \cup X) = e(A_5) + e(X) \geq 0$, and it follows from Lemma 1(d) that $e(\mathcal{A}) \geq 0$. If $\mathcal{A}$ has exactly two 2-sets not in A_5, then it will also have two 4-sets not in A_5, and by the same reasoning, $e(\mathcal{A}) \geq 0$. If $\mathcal{A}$ has three or four of the 2-sets not in A_5, then $\mathcal{A}$ has all the 3-sets and 4-sets of $\mathcal{P}(n)$, and $e(\mathcal{A}) \geq 0$. The arguments are similar for the remaining families A_i, $i = 1, 2, 3, 4$.

Lemma 4.6 *Suppose that $\mathcal{A}$ does not contain one of the 4-sets of $\mathcal{B}$. Then $e \geq 0$.*

Proof: Without loss of generality, suppose that 1234 is not in $\mathcal{A}$. Then $\mathcal{A}$ cannot contain any subset of 1234, so $e_0 = 0$, $E_1 \subseteq \{5\}$, and $E_2 \subseteq \{15, 25, 35, 45\}$. We compute $h(\{5, 15, 25, 35, 45\}) = -.9028$; so $e_1 + e_2 \geq -.9028$. If $\mathcal{A}$ contains either 5 or 45, then it also contains 1245, 1345 and then $e_4 + e_5 \geq 1.566$, and $e \geq 0$. If $\mathcal{A}$ does not contain either 5 or 45, then $e_1 + e_2 \geq h(\{15, 25, 35\}) = -.3507$ and then $e_5 + e_1 + e_2 > 0$ and so $e > 0$.

From now on, we assume that $n_1 \leq 3$, and organize the work according to the value of n_4.

If $n_4 = 0, 1$ then $\mathcal{A}$ does not contain both of the 4-sets of $\mathcal{B}$ and then $e \geq 0$, by Lemma 4.6.

If $n_4 = 2$, we assume, in view of Lemma 4.6, that $\mathcal{A}$ contains both 4-sets of $\mathcal{B}$, so $e_4 = .7014$; $E_1 \subseteq \{1,2,3\}$ ($e_1 \geq -.7983$), and $E_2 \subseteq \{12,13,23,24,35\}$; furthermore E_2 can contain at most one of 24, 35 (else we get 2345, a contradiction), so $e_2 \geq -.2305$. If $n_1 = 2, 3$, or if $n_2 = 4$, then $\mathcal{B} \subseteq \mathcal{A}$ and then $e_3 \geq .4691$, and $e(\mathcal{A}) > 0$. If $n_1 = 1$ ($e_1 \geq -.2830$), then $\mathcal{A}$ contains at least two of the 3-sets of $\mathcal{B}$ so that $e_3 \geq .2676$ and $e(\mathcal{A}) > 0$. If $n_1 = 0$, then $e_4 + e_2 > 0$ and $e > 0$.

Now suppose that $n_4 = 3$ (and $\mathcal{A}$ contains both 4-sets of $\mathcal{B}$). Up to isomorphism there are two possibilities for the missing 4-sets: 1245, 1345 and 1245, 2345.

If $\mathcal{A}$ misses 1245, 1345 ($e_4 = .9337$) then $E_1 \subseteq \{1,2,3\}$ ($e_1 \geq -.7983$), and $E_2 \subseteq \{12,13,23,24,35\}$ ($e_2 \geq -.3643$). If $n_1 \geq 2$, then $\mathcal{A}$ contains all the 3-sets of $\mathcal{B}$, and then $e_3 \geq .4691$ and $e(\mathcal{A}) > 0$. If $n_1 \leq 1$, then $e_1 \geq -.2830$ and $e_4 + e_2 + e_1 > 0$, so $e(\mathcal{A}) > 0$.

If $\mathcal{A}$ misses 1245, 2345 then $E_1 \subseteq \{1,2,3,4\}$, $E_2 \subseteq \{12,13,23,24,35,14,34\}$, and $\mathcal{A}$ can have at most one of 24, 35 ($e_2 \geq -.4473$). If $n_1 = 2, 3$ ($e_1 \geq -.9167$), then $\mathcal{A}$ contains all the 3-sets of $\mathcal{B}$, and $e_3 \geq .4691$, so that $e(\mathcal{A}) > 0$. If $n_1 \leq 1$, then $e_1 \geq -.3507$ and $e_4 + e_2 + e_1 > 0$, so $e(\mathcal{A}) > 0$.

Now suppose $n_4 = 4$. If $\mathcal{A}$ misses 1245 ($e_4 = 1.217$), then $E_1 \subseteq \{1,2,3,4\}$ and $E_2 \subseteq \{12,13,23,14,24,34,35\}$ ($e_2 \geq -.5811$). If $n_1 = 3$ ($e_1 \geq -.9167$), then $\mathcal{A}$ contains all the 3-sets of $\mathcal{B}$, and $e_3 \geq .4691$ and $e(\mathcal{A}) > 0$. If $n_1 = 2$ ($e_1 \geq -.6337$), then $\mathcal{A}$ contains at least two of the 3-sets of $\mathcal{B}$, and $e_3 \geq .2676$ and $e(\mathcal{A}) > 0$. If $n_1 \leq 1$ ($e_1 \geq -.3507$), then $e_4 + e_2 + e_1 > 0$, so $e(\mathcal{A}) > 0$.

If $\mathcal{A}$ misses 2345 ($e_4 = 1.2677$), then $\mathcal{A}$ must also miss either 23, 24, 25 or 23, 34, 35 or 24, 34, 45 or 25, 35, 45; so $e_2 \geq -.6663$. The rest of the argument is similar to the case when $\mathcal{A}$ misses 1245.

Now suppose that $n_4 = 5$ ($e_4 = 1.5$). If $n_1 = 3$ ($e_1 \geq -.9844$) then $\mathcal{A}$ contains all the 3-sets of $\mathcal{B}$, and $e_3 \geq .4691$; then $e_4+e_3+e_1 \geq .9847 = 1-.0153$. If $n_2 \leq 9$, then $e_2 \geq -(1-.0153)$, and then $e(\mathcal{A}) > 0$. If $n_2 = 10$, then $\mathcal{A}$ contains all the 3-sets and 4-sets, so $e(\mathcal{A}) > 0$. If $n_1 = 2$, then $\mathcal{A}$ contains at least two of the 3-sets of $\mathcal{B}$ and $e_4 + e_3 + e_1 > 1 \geq |e_2|$; if $n_1 \leq 1$ then $e_4 + e_1 > 1 \geq |e_2|$. In all cases, $e \geq 0$.

5 One case in $\mathcal{P}(6)$

In this section, we give an outline of the proof that $\mathcal{B}$ is an FC family, where $\mathcal{B}$ is

$$\{123, 124, 356, 456, 1234, 3456, 12356, 12456, 123456\}$$

The automorphism group G of $\mathcal{B}$ is generated by: $(3,4)$, $(1,2)$, $(5,6)$, $(1,5,2,6)$.

From the method, we get $c = [73/471, 73/471, 179/942, 179/942, 73/471, 73/471]$, and there are comparatively few distinct values for $h(X)$; we compute them separately.

For 1-sets, 5-sets: put $x_1 = .3450106157$, $x_2 = .3099787686$. Then $h(\{i\}) = -x_1$ for $i = 1,2,5,6$, and the other two are $-x_2$. Also $h(12356) = x_2 = h(12456)$ and the others are x_1.

For 2-sets, 4-sets: Put $y_1 = .1900212314$, $y_2 = .1549893843$, and $y_3 = .1199575372$. Then $h(ij) = -y_1$ for $12, 15, 16, 25, 26, 56$, and $h(ij) = -y_2$ for $13, 14, 23, 24, 35, 36, 45, 46$, and $h(ij) = -y_3$ for 34 only, and $h(ijkm) = y_1$ for $1234, 3456, 1345, 1346, 2345, 2346$, $h(ijkm) = y_3$ for 1256 only, and all the others have $h(ijkm) = y_2$. Note that $ky_2 > (k-1)y_1$ for $k = 2,3,4,5$.

For 3-sets: Put $z = .0350318471$. Then $h(ijk) = -z$ for $125, 126, 156, 256$; $h(ijk) = z$ for $134, 234, 345, 346$, and $h(ijk) = 0$ for all other 3-sets. Notice that $4z = .1401 < y_2$.

If $\mathcal{A}_1$ and $\mathcal{A}_2$ are isomorphic via a permutation from the automorphism group G of $\mathcal{B}$, then $\mathcal{A}_1 \uplus \mathcal{B} \subseteq \mathcal{A}_1$ if and only if $\mathcal{A}_2 \uplus \mathcal{B} \subseteq \mathcal{A}_2$, and $h(\mathcal{A}_1) = h(\mathcal{A}_2)$. Thus, in carrying out an argument as in Section 4, we can group together sets which are transformed into each other by G, e.g. the 2-sets 12, 56. Since (except for sign) the h-values match for 2-sets and 4-sets, and for 1-sets and 5-sets, this gives an additional convenient grouping. For instance, 12, 56, 1234, 3456 is fixed by G, and the h-values are $-y_1, -y_1, y_1, y_1$, and we will consider these four sets as belonging to one group. To simplify notation, we use colors.

Color the 2-sets and 4-sets: $12, 56, 1234, 3456$ are colored red, $34, 1256$ are colored yellow, $15, 16, 25, 26, 1345, 1346, 2345, 2346$ are colored green, and all the others are colored blue.

Color the 1-sets and 5-sets: $3, 4, 12356, 12456$ are colored purple, and all the rest are colored white.

For a given set X $(1 \leq |X| \leq 5)$, let $g(X)$ be the collection of 4-sets and 5-sets in the UC family $\{X\} \uplus \mathcal{B}$. E.g., $g(1) = \{1234, 1356, 1456, 12356, 12456, 13456\}$. We will describe these by colors: $g(1)$ has one red,two blues, two purples, one white, or for short, r, 2b, 2p, w. Observe that if X is a blue 2-set, then $g(X)$ contains just one blue 4-set; each red (green) 2-set is contained in a red (green) 4-set, and the yellow 2-set is contained in all the red and green 4-sets, but not in any of the blue 4-sets. Detailed listings of the sets $g(X)$ produce the following observations, which we list as a Lemma.

Lemma 5.1 *If $|X| = 1$, and X is white, then $g(X)$ has r, 2b, 2p,w; and if X is purple then $g(X)$ has 2r, p.*

If $|X| = 2$ and X is red, then $g(X)$ has r,2p. If X is green, then $g(X)$ has 4b, 2p, 2w. If X is blue, then $g(X)$ has r,b,p,w. If X is yellow, then $g(X)$ has 2r, 2p.

The mapping from blue 2-sets to blue 4-sets given by assigning a blue 2-set X to the blue 4-set in $g(X)$, is 1-1.

The mapping from red (resp. green) 2-sets to the red (resp. green) 4-sets given by assigning a red (resp. green) 2-set to the red (resp. green) 4-set which contains it, is 1-1.

For a set of 2-sets of a given color pattern, P, we let $C(P)$ denote the color pattern of the (minimal collection of) 4-sets which must also be there. Using the Lemma above, and when necessary the listing of the sets $g(X)$, one can establish many relations between different patterns P and their corresponding $C(P)$, for instance, $C(y) = 2r$, $C(2r) = 2r, y$, $C(5b) = 2r, 2g, 5b$, $C(g) = 4b$, $C(2g) = 6b$, $C(3g) = 8b, y$, $C(g, y) == 2r, g, 4b$, $C(2g, y) = 2r, 2g, 6b$.

The idea is, if X and Y are two sets of the same color, or if X is red and Y is green, then $|h(X)| = |h(Y)|$; if X is a 4-set, $h(X) > 0$ and if X is a 2-set, $h(X) < 0$. Thus, the Lemma and the various relations of a color pattern P to $C(P)$, allow a comparison between the values of e_4 and e_2. For example, if E_2 contains just eight blue 2-sets, then E_4 contains 8 blue 4-sets, four green 4-sets, and two red 4-sets; from the blue sets alone we can say that $e_4 \geq e_2$. If X is blue, then $|h(X)| = y_2$; if X is red or green then $|h(X)| = y_1$, so we can use the fact that $ky_2 > (k-1)y_1$ in cases when there are more blue sets than red or green sets. This allows considerable streamlining of the arguments needed for the proof.

Overall, the proof is arranged like the proof in Section 4. The case when $n_1 = 5$ is very similar, although a little more delicate. Then assuming $n_1 \leq 4$, the work is organized according to the value of n_5. In a few cases, the color pattern idea alone is not sufficient, and then it is necessary to consider the sets $g(X)$ in more detail.

Theorem 5.2 *If $n_1 = 5$, then $e \geq 0$.*

Proof: If $n_1 = 5$, then $\mathcal{A}$ contains one of the families A_i. We give a summary of the argument for $A_6 \subseteq \mathcal{A}$. Let $X = \mathcal{A} - A_6$. We need to show that $h(X) \geq -h(A_6) = .14012\ldots$. It is easy to compute that if X contains two or more of $16, 26, 36, 46, 56$, then $\mathcal{A}$ contains all but at most one of the 4-sets of $\mathcal{P}(6)$ and $e > 0$. Suppose X contains just one of these 2-sets. If X contains one of $16, 26, 36, 46$, then X also contains at least three 4-sets and one 5-set, and $e > 0$. Suppose that X contains 56; then X also has $156, 256, 1256$. We compute that $h(56, 156, 256, 1256) = -h(A_6)$, and if X contains any 3-set other than $125, 256$, then it also has another 4-set. So in all cases, $h(X) \geq -h(A_6)$, i.e. $e \geq 0$.

Lemma 5.3 *If $\mathcal{A}$ does not contain one of the 5-sets of $\mathcal{B}$, then $e \geq 0$.*

Proof. Suppose $\mathcal{A}$ does not contain one of the 5-sets of $\mathcal{B}$, say 12456. Then $n_0 = 0$, $E_1 \subseteq \{3\}$, $E_2 \subseteq \{13, 23, 34, 35, 36\}$, and E_3 cannot contain any 3-sets X

with $h(X) < 0$, so $e_3 \geq 0$. Since $\{13, 23, 34, 35, 36\}$ contains four blue sets and one yellow set, it follows from Lemma 5.1 that if $|E_2| = k$, then $|E_4| \geq k + 1$ and $e_4 + e_2 > 0$. Since $12356 \in g(3)$, and 3 and 12356 are both purple, then $e_5 + e_1 \geq 0$, so $e > 0$.

From here on, the idea is to assume that $n_1 \leq 4$, and $\mathcal{A}$ contains both of the 5-sets of $\mathcal{B}$, and consider cases according to the value of n_5. The approach used is like Section 4, somewhat simplified by the use of the colors and the sets $g(X)$.

6 Questions

Question 1. If the equation $Mx = b$ does have a non-negative solution x with $\sum x_i \leq 1$, does it follow that $\mathcal{B}$ is an FC-family?

Question 2. If the equation $Mx = b$ does not have a non-negative solution x with $\sum x_i \leq 1$, does it follow that $\mathcal{B}$ is not an FC-family?

For $n = 3, 4, 5$, the answer to Question 1 is always Yes. I have not been able to prove anything general relative to Questions 1 and 2. I have not found any cases where any of the x_i were negative.

The matrix M associated with the family $\mathcal{B}$ displays many interesting properties. This matrix may be singular (for instance, if $\mathcal{B} = \mathcal{P}(n)$, the corresponding M has rank one). From the assumption that $\cup\mathcal{B} = \{1, 2, \ldots, n\}$, it follows that every entry of M is a positive integer, and in particular M has a positive eigenvalue and a matching non-negative eigenvector. In my calculations so far, the non-singular matrices M also have the following properties:

(i) All the eigenvalues are real. There is just one large positive eigenvalue, and all remaining eigenvalues are negative, rather small compared to the positive eigenvalue, and rather close together.

(ii) In the characteristic polynomial, all coefficients except the leading coefficient are negative.

(iii) If $\mathcal{B}$ contains no smaller (in a smaller powerset) FC family, then $Mx = b$ has a positive solution x with $\sum x_i$ very close to 1 (all within .05). In all cases where $Mx = b$ had a non-negative solution with some entry equal to zero, the family $\mathcal{B}$ contained a smaller FC family.

(iv) The minimum polynomial of M has distinct real roots (so M is diagonalizable), and all coefficients except the leading coefficient are negative. In all factorizations so far, all but one of the factors have all non-negative coefficients, and one factor has all but the leading coefficient negative (or zero).

(v) The determinant of M is large (in magnitude) and has (mostly) comparatively small prime factors. (The furthest off this was: $2^9 * 617 * 354973$.)

(vi) I found a couple of examples where the minimum and characteristic polynomials were not the same; the repeated factor was linear. I found a couple of examples where the characteristic polynomial was irreducible.

Question 3. Which (if any) of the properties (i)-(v) are generally true? If M is singular, what does this say about the corresponding UC family?

I have no explanation for the effectiveness of the method. I have not been able to find any meaningful connection between the matrix properties of M, and the UC family $\mathcal{B}$, although my calculations so far indicate that there must be such connections.

Suppose that $\mathcal{A}$ is a UC family with $\cup\mathcal{A} = \{1, 2, \ldots, n\}$.

Question 4. Suppose $NUM(\mathcal{A}) \leq 1$. Is it true that the entries of $x = [x_1, x_2, \ldots, x_n]$ are all non-negative? If not, is it true that the sum of the negative x_i is less in absolute value than $1 - \sum x_i$? I.e., if $NUM(\mathcal{A}) \leq 1$, is it always possible to find non-negative c_i with sum 1, by adding on appropriate amounts to the negative x_i? (This would allow some generalization of the method of Section 2.)

Question 5. Suppose the method of Section 3 produces non-negative numbers c_i with sum 1. Is it true that for $n \geq 4$, every $c_i < 1/2$? More generally, is it true that $h(X) > 0$ for every set X with $|X| > n/2$? What if the numbers c_i were produced by some alternate method?

Question 6. Suppose that $\mathcal{A} \subseteq \mathcal{C}$ where $\cup\mathcal{A} = \cup\mathcal{C} = \{1, 2, \ldots, n\}$. Is it true that $NUM(\mathcal{C}) \leq NUM(\mathcal{A})$? If so, is it true that the inequality holds if and only if the containment is proper?

Question 7. Suppose the method of Section 3 produces non-negative numbers c_i with sum 1. Suppose there is an automorphism σ of $\mathcal{A}$ with $\sigma(i) = j$. Is it true that $c_i = c_j$? (Probably). Is the converse true? (Probably not).

Question 8. It is easy to see that $\cap A_i = \mathcal{A}$, and in all calculations so far, $h(\mathcal{A})$ is not only positive, but is comparatively large. Is this generally true? What can be said about $h(\mathcal{C})$ if $\mathcal{C}$ is the intersection of two or more of the families A_i?

Question 9. If $\mathcal{A}$ is generated by more than $[n/2]$ 3-sets, does $Mx = b$ have a positive solution x with $\sum x_i \leq 1$?

References

[1] Tetsuya Abe and Bumpei Nakano, *Lower Semimodular Types of Lattices: Frankl's Conjecture Holds for Lower Quasi-Semimodular Lattices* Graphs and Combinatorics (2000) **16**, 1-16.

[2] Ian Anderson, *Combinatorics of Finite Sets* , Oxford Science Publications, Clarendon Press, Oxford, 1987

[3] Giovanni Lo Faro, *Union-Closed Sets Conjecture: Improved Bounds* , J. Comb. Math. and Comb. Computing **16** (1994), 97-102.

[4] Robert T. Johnson and Theresa P. Vaughan, *On Union-Closed Families, I*, J. Comb. Th. Series A 84, 242-249, 1998

[5] Bjorn Poonen, *Union-Closed Families* , J. Combinatorial Theory, Series A **59** (1992) 253-268.

[6] Jurgen Reinhold, *Frankl's Conjecture is True for Lower Semimodular Lattices*, Graphs and Combinatorics (2000) **16**, p. 115-116.

[7] D. G. Sarvate and J-C. Renaud, *On the Union-Closed Sets Conjecture*, Ars Combinatoria **27** (1989) 149-154.

[8] Theresa P. Vaughan, *Families Implying the Frankl Conjecture* (preprint)