Counting Unions of Sets

Tomás Feder[*] Carlos Subi[†]

Abstract

1 Introduction

Let $[n] = \{1, 2, \ldots, n\}$ for some positive integer n. Let $S_1, S_2, \ldots, S_k$ be $k \geq 1$ distinct nonempty subsets of $[n]$, whose union is $[n]$. Let T be one of the $2^k - 1$ nonempty unions of sets S_i.

We ask how many such sets T can be obtained in this way for a given k and some n. If we denote this numer by t, then clearly $k \leq t \leq 2^k - 1$. We show that all such t can be achieved.

We also ask how many such sets T can be obtained for a given n and some k. The t achievable are those that satisfy $n \leq t \leq 2^n - 1$ by a dual argument.

We show that counting the number t of such T given the sets S_i is $\#P$-complete, but polynomial in the bounded tree width case.

Frankl's conjecture is as follows. If the sets S_i give t nonempty unions T, then some $j \in [n]$ is in at least $t/2$ of these sets T.

We show that testing whether j satisfies Frankl's conjecture is $\#P$-hard, but polynomial in the bounded tree width case.

2 Duality

The description of sets T that can be viewed as unions of sets S_i can be described as choosing sets $S_i \subseteq T$ and elements $j \in T$ The valid choices are described as follows (including for this T as the empty set):

(1) Suppose $j \in S_i$. If S_i is chosen, then j is chosen. (Equivalently, if j is not chosen, then S_i is not chosen.)

(2) If S_i is not chosen, then some $j \in S_i$ is not chosen.

(3) If j is chosen, then some S_i with $j \in S_i$ is chosen.

The dual problem has sets T_j and elements $i \in T_j$ whenever $j \in T_i$. The conditions for the sets T_j can be obtained from the above (1),(2),(3) can be obtained by interexchanging chosen and not chosen, replacing $j \in S_i$ with $i \in T_j$, S_i with i, and j with T_j. Therefore we get the following.

Theorem 1 *The two dual problems interexchange n, k and give the same t.*

There is an alternative way to look at the duality. For each element j, consider a vertex u_j. For each set S_i, consider a vertex v_j. If $j \in S_i$, consider the edge $u_j v_i$. Interexchange chosen and

[*]268 Waverley Street, Palo Alto, CA 94301. Email: `tomas@theory.stanford.edu`

[†]`carlos.subi@hotmail.com`

not chosen only for the u_j. The conditions for defining T are then precesily those of a maximal independent set in the bipartite graph, and the symmetry of this notion with respect to the two sides of the bipartition proves the duality theorem.

3 The Number of Unions

The proof is by induction on k. Suppose t can be achieved for k and n.

Lemma 1 *$t+1$ can be achieved for $k+1$ and $n+1$.*

Proof. Given sets $S_1, \ldots, S_k$, let $S'_i = S_i \cup \{n+1\}$ and $S'_{k+1} = \{n+1\}$. Each T generated by S_i gives $T \cup \{n+1\}$ generated by S'_i, plus the additional set S'_{k+1}. Alternatively, we may just add $S_{k+1} = [n+1]$ to the S_i. □

Lemma 2 *$2t+1$ can be achieved for $k+1$ and $n+1$.*

Proof. Let $S'_i = S_i$ fo $1 \le i \le k$ and let $S'_{k+1} = \{n+1\}$. Then the S_i generate t sets and the union of these sets with S'_{k+1} generates t other sets as well, plus S_{k+1} alone. □

Lemma 3 *$2t$ can be achieved for $k+1$ and $2n$.*

Proof. Let $S'_i = S_i \cup (S_i + n)$, where $S_i + n = \{j+n, j \in S_i\}$. Let $S'_{k+1} = [n]$. Then there are t unions of the first k S'_i, $t-1$ more if we take the union of all these with S'_{k+1} (as S'_{k+1} is contained in the union of the previous S'_i) and one more for S'_{k+1} alone. □

Theorem 2 *For all $k \ge 1$, all $k \le t \le 2^k - 1$ can be achieved.*

Proof. If $k=1$, a single set S_1 gives $t=1$ as desidered. Suppose the result holds for k. The operation $t+1$ gives all $k+1 \le t' \le 2^k$, and the operations $2t, 2t+1$ give all $2k \le t' \le 2^{k+1}-1$, so combining both cases gives all $k+1 \le t' \le 2^{k+1}-1$, completing the induction. □

The dual result is as follows.

Theorem 3 *For all $n \ge 1$, all $n \le t \le 2^n - 1$ can be achieved.*

The above construction gives $n = 2^{k-1}$ for $t = 3 \cdot 2^{k-2}$, so we wonder if a subsexponential bound is achievable for n here. Combinining the two operations that increase k and n by one, one may get an operation $2t+2$ that only increases k and n by two,

Theorem 4 *We obtain all $2 \le t < 2^r$ with $k \le 2r-2$ generators S_i and $n = k$.*

Theorem 5 *For all $n \ge 1$, all $n \le t \le 2^n - 1$ can be achieved with $k \le \frac{(n+1)^2}{4}$.*

Proof. Suppose $2^{n-1} \leq t \leq 2^n - 1$. Write t in binary as $\sum_{i=1}^n b_i 2^{i-1}$ with $b_i \in \{0,1\}$ and $b_n = 1$. We use two kinds of sets S_i. First the singletons $\{i\}$ if $b_i = 1$, for s of them. Then if $i' < i$ with $b_{i'} = 0$ and $b_i = 1$ we include the pair $\{i, i'\}$, for $s(n-s)$ of them.

Note that the largest i to be chosen for T must have $b_i = 1$, since a 0 must be chosen with a larger 1. Then all $i' < i$ can be chosen or not, either via a pair with i if $b_{i'} = 0$ or as a singleton if $b_{i'} = 1$. This gives $b_i 2^{i-1}$ for each choice of a rargest i, which add up to t.

We have $k \leq s + s(n-s) \leq \frac{(n+1)^2}{4}$ by maximizing at $s = \frac{n+1}{2}$. In the remaining case, we have $n - 1 \leq t - 1 \leq 2^{n-1} - 1$ and the result follows by induction by adding 1 to n, k, t. □

Let g be the largest S_i size. The above construction shows that for large t we may use $g \leq 2$, but larger g seems needed for smaller t.

Theorem 6 *$t \geq 2^{\lceil \frac{n}{g} \rceil} - 1$ always.*

Proof. Repeatedly remove sets S_i until each S_i is needed to cover some $j = j(i)$. Then if there are k sets S_i with k elements $j(i)$, we have $t \geq 2^k - 1$ and $n \leq kg$, proving the bound. □

The worst case of the above result is k disjoint S_i of size g, with $n = kg$. This case is redundant in the sense that two elements belong to the same set. To achieve irredundancy, we require $k = rg$, group the S_i into g disjoint sets of size g, and add g more disjoint sets S'_i of size g to this group. Now we have $2k$ sets giving $t \geq 2^{2k} - 1$ and $n = kg$.

Theorem 7 *$t < 2^{\lceil \frac{2n}{g} \rceil} - 1$ irredudandtly infinitely often.*

4 The Complexity of Counting Unions

Theorem 8 *If all S_i have $|S_i| = 2$, then counting the number of sets T to evaluate t is $\#P$-complete.*

Proof. Let G be the graph with $V(G) = n$ whose edges are the S_i. Assume G is bipartite, $G = (U, V, E)$. We must consider sets of vertices that are induced by a set of edges in G. Suppose we restrict our attention to sets of vertices that contain all of U. Then we must select vertices in V that cover as neighbors all of U. This is then the problem of counting set cover solutions, which is $\#P$-complete.

If we replace each vertex of U by n^3 vertices, then counting solutions that contain at least one of these n^3 vertices per vertex of U amplifies the count by $A = (2^{n^3} - 1)^{|U|}$, and the remaining solutions are in total less than A, so if there are s solutions here, the number of solutions to the preceding $\#P$-complete problem is $t = \lfloor s/A \rfloor$. □

Theorem 9 *Suppose sets S_i on $\{1, \ldots, n\}$ generate t sets T, and sets S'_i on $\{n+2, \ldots, n+1+n'\}$ generate t' sets T'. Then sets S_i and $[n+1] \cup S'_i$ generate $t + t'$ sets, with t not containing $n+1$ and t' containing $n+1$. The question of whether $t \leq t'$ is NP-hard.*

Proof. If T contains $n+1$, it is given by sets containing $[n+1]$, thus t' unions of S'_i with $[n+1]$. If T does not contain $n+1$, it is given by t unions of S_i.

Evaluating t is NP-hard by the previous Theorem. But this can be done by a polynomial number of tests of $t \leq t'$ for various $t' \leq 2^n$, and these t' can be obtained as in the previous section with $n' \leq 2n-2$ and $g' = n'$ generators S'_i. $\square$

The following result also follows via duality.

Theorem 10 *If each j belongs to exactly two S_i, then counting the number of sets T to evaluate t is $\#P$-complete.*

Proof. Let G be the graph whose vertices are the sets S_i, and whose edges are the j appearing in the two endpoints S_i. Given a set T, let U be the set of edges excluded in T, and let V be the set of vertices in G induced by U. Clearly V cannot have isolated vertices, and the number of such V without isolated vertices counts the number of T, since these are unions of S_i in $V(G) \setminus V$. The $\#P$-completeness follows as in the previous theorem. $\square$

Theorem 11 *Consider the bipartite graph $G = (U, V, E)$, where $U = [n]$, V is the collection of sets S_i, and (j, S_i) is an edge if and only if $j \in S_i$. If G is a forest, then counting t is polynomial time.*

Proof. We shall include in the counting the empty set T as well. If G has more than one component, we multiply the count over each component, so we may assume G is a tree.

Pick $j \in U$, and consider the case where j is not chosen in T. Then j and all S_i with $j \in S_i$ (the neighbors of j in G) are removed and we proceed to the resultant forest by induction.

In the remaining case, j is chosen in T. Choose S_i with $j \in S_i$. If $S_i \subseteq T$, then remove S_i and its members from G, and proceed to the resultant forest by induction.

Finally j is chosen in T, and j' is the first member of S_i not chosen in T. Remove S_i, and require that all members of S_i before j', and j, be chosen in the resulting subtrees, that j' not be chosen, and no choice for the memvers of S_i is enforced one way or the other. Solve the various subtrees, and multiply. $\square$

As we show in the section on duality, the problem of counting T can be viewed as the problem of counting maximal independent sets in a biparite graph.

Given an n-vertex graph G of maximum degree d, the graph G has treewidth w if there exists a tree R with r nodes, such that each node corresponds to at most $w+1$ vertices in G, each vertex is so assigned to a nonempty subtree of G, and each edge uv of G has both u, v in some node of R.

Theorem 12 *For such a G and R, the maximal independent sets can be counted in $O(r(d+1)^{2w+2}) = O(n^{3w+3})$ time. Time $O(nr(d+1)^{2w+2}) = O(n^{3w+4})$ allows us to test Frankl's conjecture.*

Proof. For the at most $w+1$ vertices in a node of R, we consider the possible choices for these vertices being in the maximal independent set T or not, and if a vertex v is not in T, name the least neighbor of v in some ordering that is in T.

This gives at most $(d+1)^{w+1}$ possibilities per node, the square of this value to check consistency and count in a child node for each node, then multiply over all children to get the values for the node, in time $O(r)$. We can bound $r \leq n^{w+1}$ by requiring that no two nodes have exacly the same vertices.

Frankl's conjecture requires similar n additional runs of this algorithm, one per vertex. $\square$

5 Frankl's Conjecture

The following Lemma is known.

Lemma 4 *If some S_i satisfies $|S_i| \leq 2$ then Frankl's conjecture holds.*

Proof. Let $b_i = 1$ if $i \in T$ and $b_i = 0$ if $i \not\in nT$.

Let $S_1 = \{1\}$. Let t_{b_1} be the number of T having b_1. We have $t_0 \leq t_1$, since when T has $b_1 = 0$ then $T \cup S_1$ has as many sets with $b_1 = 1$.

Let $S_2 = \{1, 2\}$. Let $t_{b_1 b_2}$ be the number of T having b_1, b_2. We have $t_{00} \leq t_{11}$, since when T has $b_1 = 0$ and $b_2 = 0$ then $T \cup S_2$ has as many sets with $b_1 = 1$ and $b_2 = 1$. Assume by symmetry between 1 and 2 that $t_{01} \leq t_{10}$, and then $t_0 = t_{00} + t_{01} \leq t_{10} + t_{11} = t_1$. □

Theorem 13 *If each j belongs to exactly two S_i, which meet only at j, and $|S_i| \geq \log_2 k$, then Frankl's conjecture holds.*

Proof. Form a set T as a union of sets S_i, each one chosen or not with probability $\frac{1}{2}$. This choice is invalid if some S_i is not chosen and all $|S_i| = s_i$ sets $S_{i'}$ that meet S_i are chosen. This happens with probability $(\frac{1}{2})^{s_i+1} \leq (\frac{1}{2})^{1+\log_2 k} = \frac{1}{2k}$ for each S_i, thus with probability at most $\frac{1}{2}$ over all k choices of S_i.

The probability for each j to be in T is $\frac{3}{4}$ since one of two independent S_i must be chosen. Therefore the expected number of js in a chosen T is $\frac{3}{4}n$. Since at most half the sets T are invalid and may include all js, it follows that the remaing valid T have expected number of js at least $\frac{1}{2}n$, so some j belongs to at least 1/2 the valid T. □

Theorem 14 *If each j belongs to at most d sets S_i, which meet only at j, and $|S_i| \geq 2^{d-1} logk$, then Frankl's conjecture holds.*

Proof. Letting $s_i = |S_i|$, the probability that S_i is not chosen yet each $j \in S_i$ is chosen is at most $\frac{1}{2}(1 - \frac{1}{2^{d-1}})^s_i < \frac{1}{2}e^{-\log k} = \frac{1}{2k}$. The rest of the calculation is as in the previous Theorem. □

Theorem 15 *Supppose $[n]$ is partitioned by two collections of sets, namely k sets S_i and k' sets $S'_{i'}$. So each j belongs to exactly one S_i and one $S'_{i'}$, which meet only at j. If $|S_i| \geq \log_2(kn)$, but the $|S'_{i'}|$ are arbitrary, then Frankl's conjecture holds.*

Proof. Consider each of the 2^k possible choices of S_i. Choose in this case all S'_i contained in the union of chosen S_i, and choose each remaining $S'_{i'}$ with probability $\frac{1}{2}$. The probability that these chosen $S'_{i'}$ will cover an unchosen S_i is $(\frac{1}{2}^{s_i} \leq \frac{1}{kn}$, thus at most $\frac{1}{n}$ over the at most k unchosen S_i.

We may assume $|S'_i| \geq 2$ and $|S_i| \geq 4$. Putting together the cases where either no S_i or just the largest S_1 is chosen, gives probability $\frac{1}{2}$ to j not in S_1 and probability $\frac{3}{4}$ to j in S_1, for an average of at least $a = \frac{1}{2} + \frac{s_1}{4n}$. Subtracting the $\frac{1}{n}$ invalid fraction gives $\frac{1}{2}$. The remaining cases give $\frac{1}{2}$ similarly, since some subset S_i is assigned probability 1, giving probability $\frac{1}{2} + \frac{s_i}{2n} - \frac{1}{n} \geq \frac{1}{2}$. □

A special case of this result occurs when every S_i meets every $S'_{i'}$, so that $n = kk'$, $|S_i| = k'$, and $|S'_{i'}| = k$. The probability that the chosen $S'_{i'}$ will cover an unchosen S_i is $(\frac{1}{2}^{s_i} = \frac{1}{2^{k'}}$. The putting together over choosing S_1 or S_2, gives $a = \frac{1}{2} + \frac{2k'}{4n} = \frac{1}{2} + \frac{1}{2k}$ and finally $a - \frac{1}{2^{k'}} \geq \frac{1}{2}$ for $k \leq k'$. The case of choosing at least one set S_1 gives $a' = \frac{1}{2} + \frac{k'}{2n} = \frac{1}{2} + \frac{1}{2k}$ and works the same for $k \leq k'$.

Theorem 16 *Frankel's conjecture holds if the sets S_i are given by intervals $[a_i, b_i]$ of consecutive integers.*

Proof. Consider a $[a_i, b_i]$ whose b_i is the least. Suppose s sets T do not contain b_i. Then these T do not contain any element from $[1, b_i]$, and taking union with $[a_i, b_i]$ gives s sets containing b_i. Therefore $t - s \geq t/2$ as desired. □

Theorem 17 *Suppose that whenever a subcollection S of sets S_i pairwise intersect, then all S_i in S have a common element. Let G have $V(G)$ with all the S_i and $E(G)$ containing (S_i, S_j) precisely when S_i, S_j are distinct and intersect. If G is chordal, then Frankel's conjecture holds.*

Proof. Since G is chordal, it has a vertex S_1 whose neibhorhood K is a clique Let a be the common element of S_1 and all S_i in K. If a is not chosen for T, then neither S_1 nor all the sets S_i in K are chosen (as they contain a), so no element of S_1 is chosen (as the remaining S_j do not intersect S_1. If we denote by t_a with $a = 0, 1$ the number of T excluding or including a respectively, then $t_0 \leq t_1$, since the sets T counted in t_0 give distinct sets $T \cup S_1$, as T does not intersect S_1. So a satisfies the conjecture. □

Theorem 18 *Suppose one of the sets S_i, say S_1, contains two possibly equal elements a, b such that for each set S_j, if S_j intersects S_i then S_j contains a or b. Then Frankl's conjecture.*

Proof. Let t_{ab} for $a, b = 0, 1$ denote the number of sets T that exclude or include a, b for $0, 1$ value respectively. Then $t = t_{00} + t_{01} + t_{10} + t_{11}$. We may assume $t_{01} \leq t_{10}$ (the symmetric case is similar). Suppose T is counted in t_{00}. Then T excludes a, b, and therefore does not intersect S_1. Then the number of sets $T \cup S_1$ is at least as many as such sets T, therefore $t_{00} \leq t_{11}$ and $t_{00} + t_{01} \leq t_{10} + t_{11} \geq t/2$, so in this case a verifies the conjecture. □

Corollary 1 *Suppose that for each S_i, consecutive elements of S_i differ by at most two, and the element preceding the last element of S_i is also in S_i. Then Frankl's conjecture holds.*

Proof. Choose S_1 with the leftmost right enpoint b, and let $a = b - 1$. Define t_{ab} as before. If T counts t_{00}, then $T \cup S_1$ gives $t_{00} \leq t_{11}$ since T contains no element up to b. We may assume $t_{01} \leq t_{10}$ and choose a, or the reverse and choose b. □

Corollary 2 *Suppose R is a rooted tree with $n/2$ vertices, the jth vertex corresponding to the pair $\{2j - i, 2j\}$ out of $[n]$. Each S_i corresponds to a subtree R_i such that the jth vertex of R is in R_i if and only at least one of $2j - 1, 2j$ is in S_i, and if the root of R_i is the jth vertex, then both $2j - 1, 2j$ are in S_i. Then Frankl's conjecture holds.*

Proof. Consider j such an R_i is rooted at the jth vertex, and no $R_{i'}$ is rooted stricly farther away from the root of R. Let $t_{2j-1,2j}$ denote the number of T with $2j-1, 2j$ excluded or included denoted by $0, 1$. Then t_{00} counts T that are disjoint from S_i, and $T \cup S_i$ gives $t_{00} \leq t_{11}$. We may assume $t_{01} \leq t_{10}$ (the other direction is symmetric) and then $2j-1$ verifies the conjecture. □

We may generalize this result by assigning to the jth vertex a set of size s_j, so that all such sets are disjoint and their union is $[n]$. If the jth vertex is in R_i, then at least half of the s_j elements are in S_i. If the jth vertex is the root of R_i, then all of the s_j elements are in S_i.

After choosing j, R_i farthest from the root if R, set the s_i elements to get excluded, to count t_0, or all included, to count t_1. Then $t_0 \leq t_1$, since we can again go from T to $T \cup S_i$. For the remaining t_i, choosing an element from the s_i will hit a T with probabiliy at least 1/2 since at least half the elements of the s_i are in T. By averages, some element hits at least half of the T counted in t_i, and this element proves the conjecture.

Theorem 19 *Suppose there are five elements, a, b, c, d, e, such that $S_1 = \{a, b, c\}$, $S_2 = \{c, d, e\}$, possibly $b = d$, and c belongs to no S_i other than S_1, S_2. Then Frankl's conjecture holds.*

Proof. Consider as before numbers t_{abcde}, where 0,1 indicate if the element is chosen or not. We may assume

$$t_{00011} + t_{00111} \leq t_{11000} + t_{11100}.$$

Otherwise just revert the order of *abcde* to *edcba* and this inequality will hold. By adding S_1 and possibly S_2 to T we get

$$t_{00000} \leq t_{11111}, t_{00001} \leq t_{11101}, t_{00010} \leq t_{11110}.$$

Finally, we may assume $\sum t_{01cde} \leq \sum t_{10cde}$, otherwise exchange a, b. But then the number of T with $a = 1$ is at least the number with $a = 0$, so the result follows.

If $b = d$, we consider t_{abce} and observe that

$$t_{0000} \leq t_{1110}, t_{0001} \leq t_{1111},$$

so the result follows for one of a, b. □

References