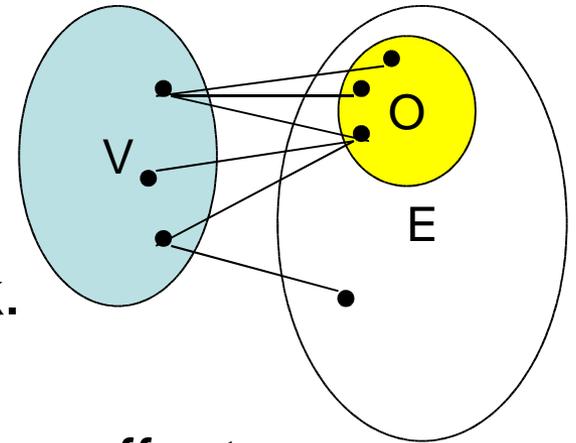


The Union of Minimal Hitting Sets: Parameterized Combinatorial Bounds and Counting

Peter Damaschke
Chalmers, Göteborg

Inferring Causes from Effects

V : set of n causes. E : set of effects.
 R , subset of $V \times E$: cause-effect relation.
 O : set of observed effects.
Infer set C of present causes, at most k .



Each present cause v generates all/some effects e with (v,e) in R . No interference.

C is *hitting set* in hypergraph with vertex set V and hyperedges $V(e) := \{v : (v,e) \text{ in } R\}$, for e in O .
(If all effects produced, some vertices are excluded.)

Example: Protein Mixture Reconstruction

Peptide mass fingerprinting:

- digest proteins by enzyme,
- search database for mass spectrum.



Applied to protein mixtures:

vertices are proteins, hyperedges are peptide masses.

In in-silico experiments with 192,000 proteins from Swissprot, digested by trypsin, up to 30 proteins always gave unique minimal hitting set, equal to mixture.

Several experimental errors make hypergraphs larger.

But many peptide masses appear in only a few proteins.

We may ignore large hyperedges. Small ranks.

Role of Minimal Hitting Sets

- Want summary of all consistent solutions.
- True solution found by further experiments, knowledge from the domain, expert's opinion, etc.
- Enough to know the *minimal* hitting sets.
- Support likelihood calculations, e.g.:
 - For each v , how many k -hitting sets do (not) contain v ?

$U(k)$: union of all minimal k -hitting sets (i.e., size $\leq k$).

k^* : minimum size of a hitting set.

Summary of Results:

- $U(k)$ kernel for counting problems.
- New bounds on $|U(k)|$, for fixed rank r .

Hypergraphs of Bounded Rank

Rank r : maximum size of hyperedges.

Enumerating the minimal k -hitting sets is FPT in r, k : Search tree, time $O^*(r^k)$

$$|U(k)| \leq (r-1)k^r + k$$

and hypergraphs with $|U(k)| = \Theta(k^r)$ exist.

[Dam. IWPEC 2004/TCS 351 (2006)]

What is the true worst-case factor?

$U(k)$ is a Kernel for Counting Small Hitting Sets

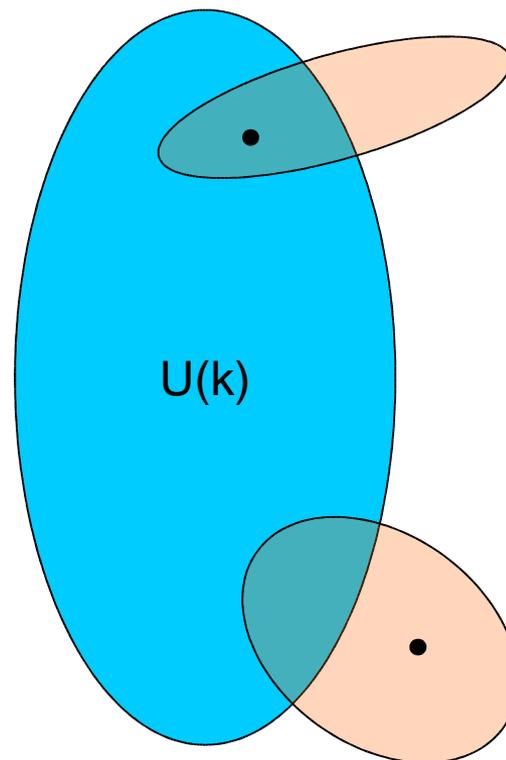
$s(h)$: number of hitting sets D in
 $U(k)$ with $|D|=h$.

$s(v,h)$: same, for D containing v .

Number of k -hitting sets
containing any v
is easy to compute
from the $s(h), s(v,h)$.

The $s(v,h)$ are

- equal for all v outside $U(k)$,
- larger for all v in $U(k)$.



Union of Bounded Minimal Vertex Covers

Previous results for $r=2$:

Bounds on $|U(k^*)|$. Computing $U(k^*)$ is NP-hard.
[Boros, Golumbic, Levit, 2002],
[Chlebik, Chlebikova, 2004].

New results for $r=2$:

Computing $U(k)$ through minimum vertex covers,
with $O(kk^*)$ subgraphs of size $O(kk^*)$.

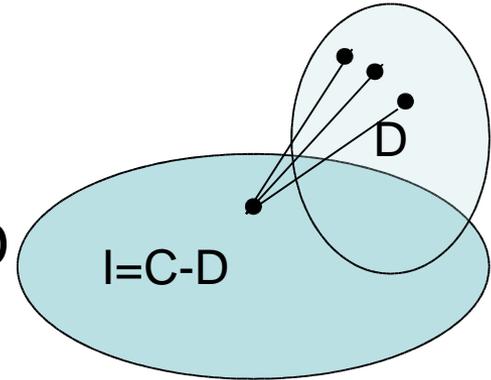
$|U(k)| \leq (k - k^* + 2)k^*$, and this bound is tight.

Union of Bounded Minimal Vertex Covers: Complexity

C : a fixed minimal vertex cover.

Lemma (proof is simple):

For minimal vertex cover D , and $I=C-D$ we have $D=(C-I) \cup N(I)$.



Take $|C|=k^*$, $|D|\leq k$. $|N(v)|\leq k$ for each v in some I .

Hence $|U(k)|\leq (k+1)k^*$.

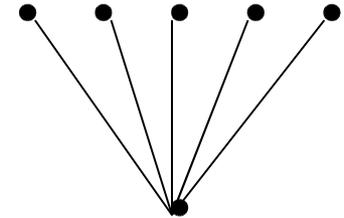
Theorem:

For computing $U(k)$ we need one minimum vertex cover (1) in G , and (2) in $O(kk^*)$ subgraphs of size $O(kk^*)$, each computable in polynomial time.

Union of Bounded Minimal Vertex Covers: Size

Theorem: $|U(k)| \leq (k - k^* + 2)k^*$, and this bound is tight.

Proof (lower bound): k^* stars with $x+1$ leaves, $x = k - k^*$. Minimal k -vertex covers involve all $(x+2)k^*$ vertices.



Proof (upper bound):

$|C| = k^*$. $D = (C - I) \cup N(I)$ is minimal vertex cover. $|N(I) - C| \leq |I|$.

For $|D| \leq k = k^* + x$, it must be $|N(I) - C| \leq |I| + x$.

We call such I replacement sets.

Claim: The union of $N(I) - C$ of all replacement sets I has at most $(x+1)k^*$ vertices. (Claim implies the upper bound.)

Proved by induction on the number of replacement sets.

Other Corollaries of Difference Lemma

For minimal vertex cover D , and $I=C-D$, we have $D=(C-I) \cup N(I)$.
Note that I determines D .

At most 2^{k^*} minimal vertex covers exist. (Tight: k^* disjoint edges.)

[Dam. IWPEC 2004/TCS 351 (2006)]:

Repetition-free, concise enumeration of all minimal k -vertex covers
in time $O^*(1.74^k)$

Counting hitting sets of given size h , faster than by trivial search tree?
Technique of [Chen, Kanj, et al., IWPEC 2006] gives base 1.47 for $r=2$.

Hypergraphs of Fixed Rank > 2

$|U(k)|$ for rank $r > 2$ apparently more difficult.

We use from

[Dam. IWPEC 2004/TCS 351 (2006)]:

$h \leq k^r$ hyperedges in a "reduced" rank- r hypergraph
with the same minimal k -hitting sets,
and all vertex degrees are at most k^{r-1}

This reduced hypergraph is computed in polynomial time by
an elimination process.

New Lower Bound on $|U(k)|$

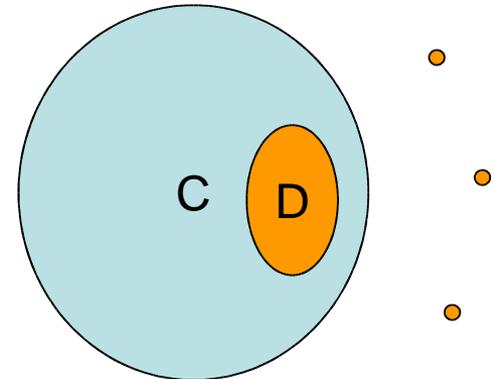
$$|C| = k(r-1)/r.$$

For each $(r-1)$ -subset D of C
create k/r hyperedges:

D plus k/r different single vertices.

Every vertex appears in some
minimal k -hitting set.

$$|U(k)| = \frac{(r-1)^{r-1}}{r! r^{r-1}} k^r \approx \frac{1}{er!} k^r$$



Relating Hyperedge and Vertex Number

We show $|U(k)| < h + o(h)$ for $h = \Theta(k^r)$
(Slightly better than in Proceedings.)



Vertex v called *d-thin* if v in $< d$ hyperedges.

At most rh/d vertices are *d-fat*.

From every hyperedge with *d-thin* vertices, delete some. Get rank $r-1$.

Assume that

each hyperedge still has d-thin vertices.

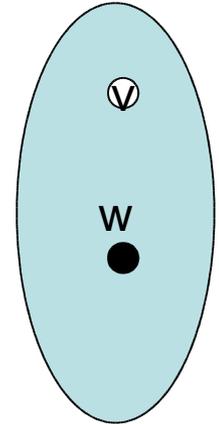
Diminish and Replace and ...

H: any minimal k-hitting set.

Construct minimal hitting set H'

in the diminished hypergraph:

- Replace each deleted v (which was in H) with d -thin vertices (fewer than d).
- Remove redundant vertices, until H' is minimal.



All H' are in a fixed set of size $O((dk)^{r-1})$

Vertices v in the $H-H'$ were deleted or redundant.

Each deleted/redundant v is assigned to some w in H' ,
in a common hyperedge with v . These w are d -thin.

All $H-H'$ are in a fixed set of size $O(rd(dk)^{r-1})$

Diminishing "most" Hyperedges

Delete each d-thin vertex independently with probability $\frac{1}{2}$.
A hyperedge losing either none or all d-thin vertices
is called *bad*.

Hyperedge with t d-thin vertices bad with probability $1/2^{t-1}$

Expected number of d-thin vertices in a bad hyperedge:

$$t / 2^{t-1} \leq 1$$

Expected number of d-thin vertices in all bad hyperedges
is at most h.

We can delete a set of d-thin vertices so that
at most h d-thin vertices were initially in all bad hyperedges.

Decomposition Lemma and Result

Given: a minimal hitting set H ,
a partitioning into s families of hyperedges.
Then, a minimal hitting set $H(i)$ of the i -th family
exists with $H = H(1) \cup \dots \cup H(s)$.

Apply it to bad and good hyperedges.

$$|U(k)| \leq rh/d + h + O(d^r k^r)/k$$

For $h = \Theta(k^r)$ we have $|U(k)| \leq (r/d + 1 + O(d^r)/k)h$

For large k we get the asymptotic result.

Current Upper Bounds

on constant factor in $|U(k)| = \Theta(k^r)$

1 for $r=1$ (tight)

0.25 for $r=2$ (tight)

0.5 for $r=3$

0.71 for $r=4$

1 for $r>4$

Another Hypergraph Decomposition

G : hypergraph of rank r . H : a fixed minimum hitting set.

All $H-H'$ (H' minimal hitting set) are called replacement sets.

I : any fixed replacement set.

$G(I)$: hypergraph with hyperedges $e-I$, where e are all hyperedges of G which intersect H in subsets J of I .

$G[J]$: hypergraph of rank $r-j$, $j=|J|$, with hyperedges $e-J$ from $G(I)$ such that e intersects H in J .

Lemma (proof straightforward):

Any minimal hitting set H' with $H-H'=I$ has vertices only from H and from minimal hitting sets of the $G[J]$, where J subset of I , $0<|J|<r$.

Better Upper Bound for Rank 3

For $r=3$ we have $|U(k)| \leq \frac{1}{4} k * (k^2 + (k^*)^2) + k * k$

Proof:

H: hitting set with $|H|=k^*$.

For v in H, let $I(v)$ be a maximum replacement set containing v . (W.l.o.g. $I(v)$ exists.)

$x(v)=|I(v)|$, $x=\max\{x(v): v \text{ in } H\}$.

For each v we mark $x(v)-1$ sets consisting of v and another vertex of $I(v)$.

All of $U(k)$ is in H and in minimal hitting sets of the $G[J]$ where $|J|=1,2$.

... Rank 3 (cont'd)

The $G[J]$, $J=\{v\}$ for all v in H have together
 $\leq (1/4) \sum_{v \in H} (k - k^* + x(v) + 2)^2$ vertices in $U(k)$.

Every vertex in $G[J]$, $|J|=2$, is in every hitting set that extends some $H-I$, (I superset of J).

All these $G[J]$ have together at most $k - k^* + |I|$ vertices.
Apply this to the $I(v)$.

All $G[J]$ of marked J have together at most
 $k^*(k - k^* + x) < k^*k$ vertices in $U(k)$.

Each $G[J]$, $|J|=2$, unmarked, has at most
 $k - k^* + x$ vertices in $U(k)$.

Summing up all bounds gives the result.

Upper Bounds for Small Ranks

Decomposition yields recursion for the constant factors $f(r)$:

$$f(r) \leq \sum_{j=1}^{r-1} \frac{f(r-j)}{j!}$$

Unfortunately, exponential in r . But:

$$f(3) < 1/2, f(4) < 19/24.$$

Conclusions and Open Problems

- Union of minimal k-hitting sets is interesting in combinatorial inference, e.g., for counting of hypotheses. We studied the size.
- Tight bound for rank 2 (graphs).
- Progress: new asymptotic bounds for any fixed rank r . Size related to number h of hyperedges in a reduced equivalent hypergraph. Decomposition techniques.
- Still large gap for $r > 2$. Reduce h further?
- Smaller bounds if k close to minimum k^* ?
- Nontrivial FPT algorithms for counting the k -hitting sets in rank- r hypergraphs. Recently, base $r-1+o(r)$ achieved?
- How many solutions in real data?