Graphs, Probability, and Separating Families

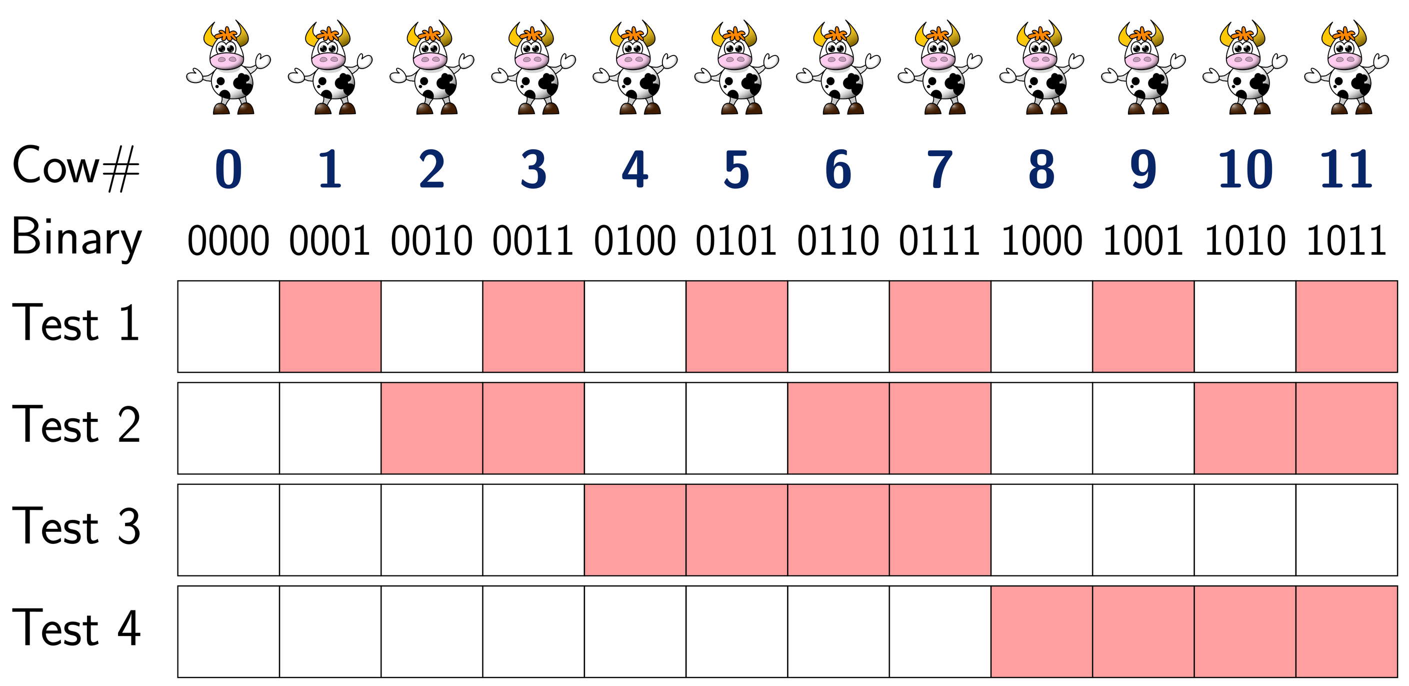
Daniel Condon, Samuel Coskey, Luke Serafin, and Cody Stockdale

Georgia Institute of Technology, Boise State University, Carnegie Mellon University, Bucknell University

Motivating Problem: Sick Cows

Suppose we have a herd of cows, and one is sick. We have a blood test that can determine whether a mixed sample of blood taken from multiple cows contains any blood that is infected. Thus we can test a group of cows at the same time, and the results will tell us whether or not any of those cows were sick. How can we minimize the number of simultaneous tests needed to identify the sick cow?

- Assign each cow a number starting with 0
- Write the binary expansions of those numbers
- Cows with a 1 as the i^{th} binary digit will be included in the i^{th} test
- This guarantees each cow goes through a unique set of tests

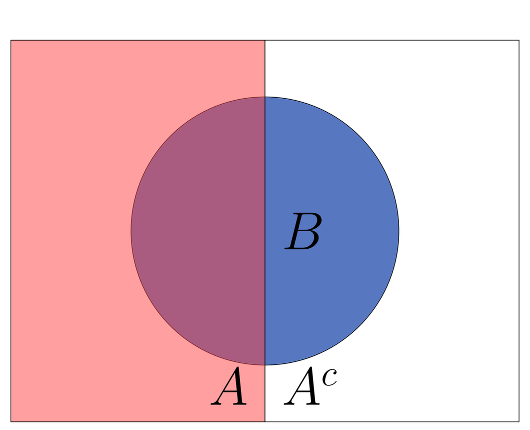


The test results give an index that shows which cow was sick. For example, if only test 2 and test 3 come back positive, then cow 6 was sick. Since each cow goes through a unique set of tests, no two sets of results correspond to the same cow.

We call such a system a separating family and our research investigates generalizations of this notion.

Separating Families

A family $\mathcal{F}$ of subsets of $[k]$ is a **separating family** if for all $B \subseteq [k]$, there exists $A \in \mathcal{F}$ such that $A \cap B \neq \emptyset$ and $A^c \cap B \neq \emptyset$.

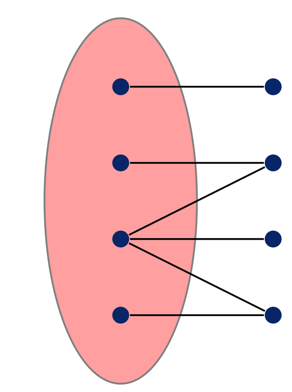


Equivalently, a family $\mathcal{F}$ of subsets of $[k]$ is separating if and only if for each pair $b = \{x, y\}$, there exists $A \in \mathcal{F}$ such that $|A \cap b| = 1$.

n-separating and n-separability

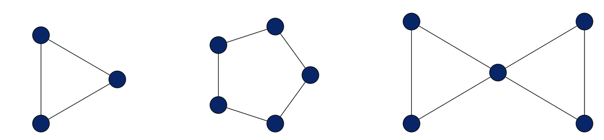
We generalize separating families in the following way.

A family $\mathcal{F}$ of subsets of $[k]$ is an **n-separating family** if for every separable collection of pairs $b_1, b_2, \ldots, b_n$, there exists $A \in \mathcal{F}$ which separates each pair.



In the figure at right, two dots connected by an line represent a pair, and the oval represents a separating set A.

Some collections cannot be separated. For example:

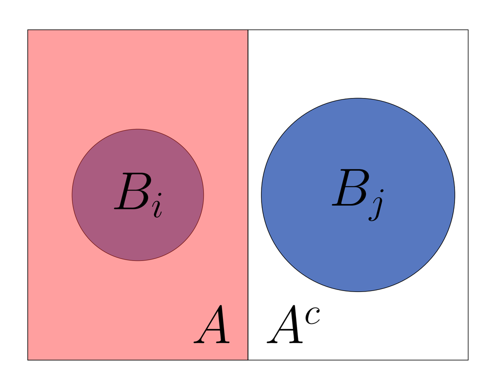


Proposition. A collection $B_1, B_2, \ldots, B_n$ of subsets of $[k]$ is separable if and only if there exist pairs $b_1 \subseteq B_1, b_2 \subseteq B_2, \ldots, b_n \subseteq B_n$ such that $\{b_1, b_2, \ldots, b_n\}$ is a bipartite graph.

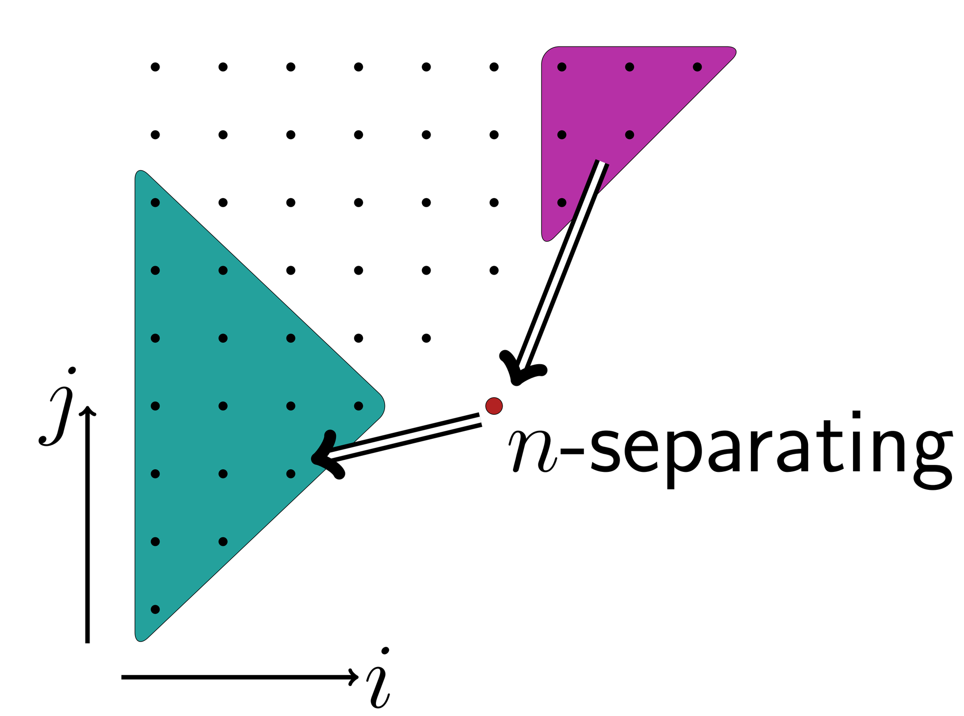
Relationship with (i, j)-separating

We compare n-separating families to the well-studied notion of (i, j)-separating families.

A family $\mathcal{F}$ of subsets of $[k]$ is an **(i, j)-separating family** if for all $B_i, B_j \subseteq [k]$ with $|B_i| \leq i$ and $|B_j| \leq j$, there exists $A \in \mathcal{F}$ such that $B_i \subseteq A$ and $A \cap B_j = \emptyset$ or $B_j \subseteq A$ and $A \cap B_i = \emptyset$.



Proposition.
- If $\mathcal{F}$ is an (n, n)-separating family, then $\mathcal{F}$ is an n-separating family.
- If $\mathcal{F}$ is an n-separating family, then $\mathcal{F}$ is an (i, j)-separating family whenever $i + j \leq n + 1$.



The Probabilistic Method

These families are only useful if they are not too big; we use the probabilistic method to obtain bounds on their minimal sizes.

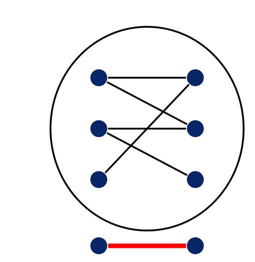
Here is how it works: if p is the probability that a random object does a thing, and there are N things to be done, then the probabilistic method gives an upper bound on the minimal size of a family $\mathcal{F}$ of objects which do all N things:

$$|\mathcal{F}| \leq \frac{\log(N)}{-\log(1-p)}$$

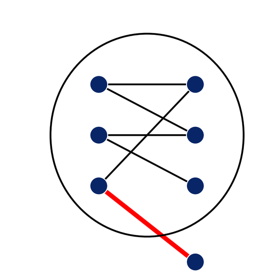
If the things to be done are separating collections of n pairs, this gives an upper bound on the minimal size of an n-separating family!

Proposition. The probability that a random $A \subseteq [k]$ simultaneously separates a fixed collection of pairs $b_1, b_2, \ldots, b_n$ is at least 2^{-n}.

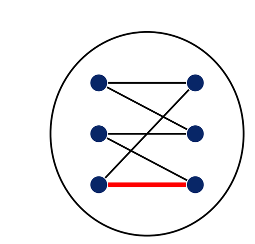
Proof. Call this probability p_n; we proceed by induction on n. The base case $n = 1$ is clear. Let G be the graph with edges $\{b_1, b_2, \ldots, b_n\}$.



Case 1: The event that b_{n+1} is separated is independent of the event that G is separated, and has probability $1/2$. Hence $p_{n+1} = \frac{1}{2}p_n \geq 2^{-(n+1)}$.



Case 2: Given that G is separated and b_{n+1} shares vertex x with G, the probability that b_{n+1} is separated is the probability that x is in A and y is not, or vice versa. Hence $p_{n+1} = \frac{1}{2}p_n \geq 2^{-(n+1)}$.



Case 3: If both vertices of b_{n+1} are in G, then b_{n+1} is separated, and so $p_{n+1} \geq 2^{-n} \geq 2^{-(n+1)}$. □

Theorem. There exists an n-separating family for subsets of $[k]$ of size at most

$$\frac{2n \log k}{-\log(1-2^{-n})}$$

Summary of Results

Our results contribute to the following table of bounds on the minimum sizes of separating and splitting families.

Notion	Lower Bound	Upper Bound
separating families	$\log(k)$	$\log(k)$
n-separating families	$2^n \log k$	$2^n n \log(k)$
$(n, 1)$-separating families	$n(\log k - \log n)$	$n^{-1}(n+1)^{n+2} \log k$
splitting families	$\sqrt{k}$	$k/2$
2-splitting families	$\sqrt{k}$	$k^2 \log k$