- *k*-anonymity Defined
- *k*-anonymity-test is in P
- 3-anonymity is NP-hard (Meyerson & Williams)
- 2-anonymity is in P (Blocki & Williams)

k-anonymity Defined

A database V is a multiset of m-degree attribute tuples with each tuple $\in \Sigma^m$ where Σ is the domain of attributes. $v[i]$ projects ith attribute in v.

A database V is k-anonymous iff, for all $v \in V$, there exist a subset $S \subseteq V$ of size k such that, for all $s, s' \in S$, $s = s'$.

A suppressor t on a database V is a mapping $V \to \{\Sigma \cup \{*\}\}^m$ such that, for all $v \in V$, for all $j \in \{1, ..., m\}$, $t(v)[j] \in \{v[j], *\}$. $t(v)[j]$ is suppressed iff $t(v)[j] = *$.

For a suppressor t and a database V, $t(V) = \{t(v) : v \in V\}$. If $t(V)$ is k-anonymous, t is called a k-anonymizer on V.

k-anonymity-test is in P

A database V is k-anonymous iff, for all $v \in V$, there exist a subset $S \subseteq V$ of size k such that, for all $s, s' \in S$, $s = s'$.

INSTANCE: k-anonymity-test

Given a databse V with column names D, is $\Pi_Q V$ k-anonymous for all $Q \subseteq D$.

It's in P.

k-anonymity-test is in P

(INSTANCE) Given a databse V with column names D, is $\Pi_Q V$ k-anonymous for all $Q \subseteq D$.

Sweeney's Lemma: remove columns can only make it k-anonymous

For relational table RT of m-tuples with column names D, if $\Pi_H RT$ is k-anonymous, then $\Pi_K RT$ is k-anonymous, for $K \subseteq H \subseteq D$.

Naive Algorithm Check the case of QI$=D$. It runs in $\mathcal{O}(n^2 mk)$.

Proof

- If it checks out: then it is k-anonymous for all possible QIs, with Sweeney's Lemma.
- If it fails: then it is not k-anonymous for the QI of all attributes.

3-anonymity is NP-hard

NP-hardness is shown by reduction from 3-dimensional perfect matching to 3-anonymity.

INSTANCE: 3-dimensional perfect matching (NP-complete)

Given a simple 3-uniform hypergraph $H = (U, E)$, is there a subset of hyperedges $S \subseteq E$ of size $|U|/3$ such that each vertex of U is contained in exactly one hyperedge of S?

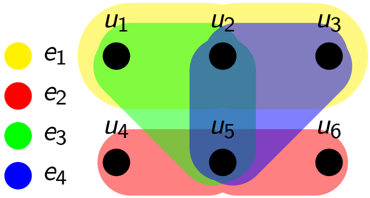
INSTANCE: 3-anonymity

Given $V \subseteq \Sigma^m$ and $l \in \mathbb{N}$, is there a suppressor t such that $t(V)$ is 3-anonymous and the total number of suppressions is at most l?

3-anonymity is NP-hard

Given a 3-uniform hypergraph $H = (U, E)$ where $U = \{u_1, \ldots, u_n\}$ and $E = \{e_1, \ldots, e_m\}$, construct a database $V = \{v_1, \ldots, v_n\}$ of m-degree tuples such that, for all $v_i \in V$, for all $j \in \{1, \ldots, m\}$:

$$v_i[j] = \begin{cases} 0, & \text{if } u_i \in e_j \\ 1, & \text{otherwise} \end{cases}$$

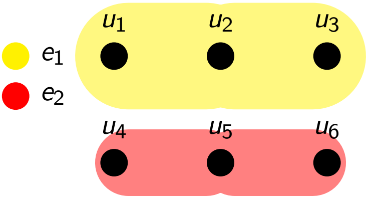


$$\begin{array}{lccccl} & e_1 & e_2 & e_3 & e_4 & \\ V = \{ & (0, & 1, & 0, & 1), & u_1 \\ & (0, & 1, & 0, & 0), & u_2 \\ & (0, & 1, & 1, & 0), & u_3 \\ & (1, & 0, & 1, & 1), & u_4 \\ & (1, & 0, & 0, & 0), & u_5 \\ & (1, & 0, & 1, & 1)\} & & & & u_6 \end{array}$$

3-anonymity is NP-hard

(Goal) H has a 3-dimensional perfect matching iff there exists a t such that $t(V)$ is 3-anonymous with $l = n(m-1)$.
(Left to right) Suppose H has a 3-dimensional perfect matching M, construct a suppressor t such that, for all $v_i \in V$, for all $j \in \{1, ..., m\}$:

$$t(v_i)[j] = \begin{cases} 0, & \text{if } e_j \in M \\ *, & \text{otherwise} \end{cases}$$



$$\begin{array}{lllllll} & e_1 & e_2 & e_3 & e_4 & & \\ V = \{ & (0, & *, & *, & *), & & u_1 \\ & (0, & *, & *, & *), & & u_2 \\ & (0, & *, & *, & *), & & u_3 \\ & (*, & 0, & *, & *), & & u_4 \\ & (*, & 0, & *, & *), & & u_5 \\ & (*, & 0, & *, & *) & \} & u_6 \end{array}$$

3-anonymity is NP-hard

(Goal) H has a 3-dimensional perfect matching iff there exists a t such that $t(V)$ is 3-anonymous with $l = n(m-1)$.
(Left to right) Suppose H has a 3-dimensional perfect matching M, construct a suppressor t such that, for all $v_i \in V$, for all $j \in \{1, ..., m\}$:

$$t(v_i)[j] = \begin{cases} 0, & \text{if } e_j \in M \\ *, & \text{otherwise} \end{cases}$$

Since, for all $e_j \in M$, $|e| = 3$, there are exactly 3 identical tuples v, v', v'' in $t(V)$ such that $v[j] = v'[j] = v''[j] = 0$ and $v[k] = v'[k] = v''[k] = *$ for all $k \neq j$. Thus, $t(V)$ is k-anonymous. Since the matching is perfect, each tuple has exactly one 0 and $m-1$ *s, and there are n tuples. Thus, the number of suppression is exactly $n(m-1)$.

3-anonymity is NP-hard

(Goal) H has a 3-dimensional perfect matching iff there exists a t such that $t(V)$ is 3-anonymous with $l = n(m-1)$.
(Right to left) Select as hyperedges the identical tuples in $t(V)$ and we're done.

Note This is the case where $|\Sigma| \leq n$, but it's also NP-hard for the cases of $|\Sigma| = 3$ and $|\Sigma| = 2$; see REU Summer 2007 slides for an overview.

2-anonymity is in P

Reduction to polynomial time "Simplex Matching" algorithm.