Distance of Closures

G.O.H. Katona A. Sali

Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences

Tuesday, June 19$^{\text{th}}$ 2012.

1 Introduction

Motivating questions

Motivating questions

- When are two databases the same?

Motivating questions

- When are two databases the same?
- Are two databases similar?

Motivating questions

- When are two databases the same?
- Are two databases similar?
- If a row is added or deleted in the database,
  - is only the instance changed?

Motivating questions

- When are two databases the same?
- Are two databases similar?
- If a row is added or deleted in the database,
  - is only the instance changed?
  - Does this change alter the dependency structure, so a new database is obtained?

Motivating questions

- When are two databases the same?
- Are two databases similar?
- If a row is added or deleted in the database,
  - is only the instance changed?
  - Does this change alter the dependency structure, so a new database is obtained?
- A new attribute is added to the database, all rows are completed with the value of the new attribute. Is this another database?

Motivating questions

- When are two databases the same?
- Are two databases similar?
- If a row is added or deleted in the database,
  - is only the instance changed?
  - Does this change alter the dependency structure, so a new database is obtained?
- A new attribute is added to the database, all rows are completed with the value of the new attribute. Is this another database?
- If two databases should be considered different, how different they are?

Motivating questions

- When are two databases the same?
- Are two databases similar?
- If a row is added or deleted in the database,
  - is only the instance changed?
  - Does this change alter the dependency structure, so a new database is obtained?
- A new attribute is added to the database, all rows are completed with the value of the new attribute. Is this another database?
- If two databases should be considered different, how different they are?

A notion of *distance* between databases is needed.

Our approach

Our approach

Previous approaches are based on algorithmic conflict resolution.

Our approach

Previous approaches are based on algorithmic conflict resolution.

Present talk: combinatorial approach, interested in the apparent dependency structure of a database.

Our approach

Previous approaches are based on algorithmic conflict resolution.

Present talk: combinatorial approach, interested in the apparent dependency structure of a database. Leads to the concept of *distance of closures*.

Our approach

Previous approaches are based on algorithmic conflict resolution.

Present talk: combinatorial approach, interested in the apparent dependency structure of a database. Leads to the concept of *distance of closures*.

Two database are considered the same if they have the same number of attributes and the system of functional dependencies are identical. The distance is introduced only between two databases having the same number of attributes.

2 Definition of distance of databases/closures

Closures

Closures

A schema $\mathcal{R}$ is considered fixed and every instance $\mathbf{r}$ is considered together with *all* functional dependencies $A \rightarrow B$ such that $\mathbf{r} \models A \rightarrow B$.

Closures

A schema $\mathcal{R}$ is considered fixed and every instance $\mathbf{r}$ is considered together with *all* functional dependencies $A \to B$ such that $\mathbf{r} \models A \to B$.

For a set of attributes $A \subseteq \mathcal{R}$ the *closure* of A is given by $\ell(\mathbf{r})(A) = \{a \in \mathcal{R} \colon \mathbf{r} \models A \to a\}$. It is well known that the function $\ell(\mathbf{r}) \colon 2^{\mathcal{R}} \longrightarrow 2^{\mathcal{R}}$ is a *closure* that is it satisfies the properties

Closures

A schema $\mathcal{R}$ is considered fixed and every instance $\mathbf{r}$ is considered together with *all* functional dependencies $A \to B$ such that $\mathbf{r} \models A \to B$.

For a set of attributes $A \subseteq \mathcal{R}$ the *closure* of A is given by $\ell(\mathbf{r})(A) = \{a \in \mathcal{R} \colon \mathbf{r} \models A \to a\}$. It is well known that the function $\ell(\mathbf{r})\colon 2^{\mathcal{R}} \longrightarrow 2^{\mathcal{R}}$ is a *closure* that is it satisfies the properties

$$A \subseteq \ell(\mathbf{r})(A)$$

Closures

A schema $\mathcal{R}$ is considered fixed and every instance $\mathbf{r}$ is considered together with *all* functional dependencies $A \to B$ such that $\mathbf{r} \models A \to B$.

For a set of attributes $A \subseteq \mathcal{R}$ the *closure* of A is given by $\boldsymbol{\ell}(\mathbf{r})(A) = \{a \in \mathcal{R} \colon \mathbf{r} \models A \to a\}$. It is well known that the function $\boldsymbol{\ell}(\mathbf{r}) \colon 2^{\mathcal{R}} \longrightarrow 2^{\mathcal{R}}$ is a *closure* that is it satisfies the properties

$$A \subseteq \boldsymbol{\ell}(\mathbf{r})(A)$$
$$A \subset B \Longrightarrow \boldsymbol{\ell}(\mathbf{r})(A) \subseteq \boldsymbol{\ell}(\mathbf{r})(B)$$

Closures

A schema $\mathcal{R}$ is considered fixed and every instance $\mathbf{r}$ is considered together with *all* functional dependencies $A \to B$ such that $\mathbf{r} \models A \to B$.

For a set of attributes $A \subseteq \mathcal{R}$ the *closure* of A is given by $\boldsymbol{\ell}(\mathbf{r})(A) = \{a \in \mathcal{R} \colon \mathbf{r} \models A \to a\}$. It is well known that the function $\boldsymbol{\ell}(\mathbf{r}) \colon 2^{\mathcal{R}} \longrightarrow 2^{\mathcal{R}}$ is a *closure* that is it satisfies the properties

$$\begin{aligned} &A \subseteq \boldsymbol{\ell}(\mathbf{r})(A) \\ &A \subset B \Longrightarrow \boldsymbol{\ell}(\mathbf{r})(A) \subseteq \boldsymbol{\ell}(\mathbf{r})(B) \\ &\boldsymbol{\ell}(\mathbf{r})(\boldsymbol{\ell}(\mathbf{r})(A)) = \boldsymbol{\ell}(\mathbf{r})(A). \end{aligned} \tag{1}$$

Closures

A schema $\mathcal{R}$ is considered fixed and every instance $\mathbf{r}$ is considered together with *all* functional dependencies $A \to B$ such that $\mathbf{r} \models A \to B$.

For a set of attributes $A \subseteq \mathcal{R}$ the *closure* of A is given by $\ell(\mathbf{r})(A) = \{a \in \mathcal{R} \colon \mathbf{r} \models A \to a\}$. It is well known that the function $\ell(\mathbf{r}) \colon 2^{\mathcal{R}} \longrightarrow 2^{\mathcal{R}}$ is a *closure* that is it satisfies the properties

$$\begin{aligned} & A \subseteq \ell(\mathbf{r})(A) \\ & A \subset B \Longrightarrow \ell(\mathbf{r})(A) \subseteq \ell(\mathbf{r})(B) \\ & \ell(\mathbf{r})(\ell(\mathbf{r})(A)) = \ell(\mathbf{r})(A). \end{aligned} \tag{1}$$

Attribute set A is *closed* if $A = \ell(\mathbf{r})(A)$. Since constant columns are not really interesting, we assume that $\ell(\mathbf{r})(\emptyset) = \emptyset$.

Poset of closures

Poset of closures

Fact

If $\mathbf{r} \models A \to B$, *then* $\mathbf{r}' \models A \to B$ *holds for any* $\mathbf{r}' \subset \mathbf{r}$.

Poset of closures

Fact

If $\mathbf{r} \models A \to B$*, then* $\mathbf{r}' \models A \to B$ *holds for any* $\mathbf{r}' \subset \mathbf{r}$*.*

Closure ℓ_1 is said to be *richer* than or equal to ℓ_2, $\ell_1 \geq \ell_2$ in notation, iff $\ell_1(A) \subseteq \ell_2(A)$ for all attribute sets.

Poset of closures

Fact

If $\mathbf{r} \models A \to B$*, then* $\mathbf{r}' \models A \to B$ *holds for any* $\mathbf{r}' \subset \mathbf{r}$*.*

Closure ℓ_1 is said to be *richer* than or equal to ℓ_2, $\ell_1 \geq \ell_2$ in notation, iff $\ell_1(A) \subseteq \ell_2(A)$ for all attribute sets.

Proposition (Burosch, Demetrovics, Katona)

Let $\mathcal{F}(\ell)$ *denote the collection of closed attribute sets for closure* ℓ*.*

Poset of closures

Fact

If $\mathbf{r} \models A \to B$*, then* $\mathbf{r}' \models A \to B$ *holds for any* $\mathbf{r}' \subset \mathbf{r}$*.*

Closure ℓ_1 is said to be *richer* than or equal to ℓ_2, $\ell_1 \geq \ell_2$ in notation, iff $\ell_1(A) \subseteq \ell_2(A)$ for all attribute sets.

Proposition (Burosch, Demetrovics, Katona)

Let $\mathcal{F}(\ell)$ *denote the collection of closed attribute sets for closure* ℓ*.*

$$\ell_1 \leq \ell_2 \textit{ iff } \mathcal{F}(\ell_1) \subseteq \mathcal{F}(\ell_2).$$

Poset of closures

Fact

If $\mathbf{r} \models A \to B$*, then* $\mathbf{r}' \models A \to B$ *holds for any* $\mathbf{r}' \subset \mathbf{r}$*.*

Closure ℓ_1 is said to be *richer* than or equal to ℓ_2, $\ell_1 \geq \ell_2$ in notation, iff $\ell_1(A) \subseteq \ell_2(A)$ for all attribute sets.

Proposition (Burosch, Demetrovics, Katona)

Let $\mathcal{F}(\ell)$ *denote the collection of closed attribute sets for closure* ℓ*.*

$$\ell_1 \leq \ell_2 \text{ iff } \mathcal{F}(\ell_1) \subseteq \mathcal{F}(\ell_2).$$

Interpretation

Interpretation

If $\mathbf{r_1}$ and $\mathbf{r_2}$ are two instances of schema $\mathcal{R}$, then the statement "$\ell(\mathbf{r_2})$ is richer than $\ell(\mathbf{r_1})$" can be interpreted as follows.

Interpretation

If $\mathbf{r_1}$ and $\mathbf{r_2}$ are two instances of schema $\mathcal{R}$, then the statement "$\ell(\mathbf{r_2})$ is richer than $\ell(\mathbf{r_1})$" can be interpreted as follows.

In $\mathbf{r_2}$ there are more subsets of attributes that only determine attributes inside them, that is in $\mathbf{r_2}$ we need more attributes to determine some attribute, so $\mathbf{r_2}$ conveys "more information" in the sense that the values of tuples are more abitrary.

Height of the poset of closures

Height of the poset of closures

ℓ_2 *covers* ℓ_1, if $\ell_1 \leq \ell_2$ and for all ℓ' such that $\ell_1 \leq \ell' \leq \ell_2$ either $\ell' = \ell_1$ or $\ell' = \ell_2$.

Proposition (BDK)

ℓ_2 covers ℓ_1 iff $\mathcal{F}(\ell_1) \subseteq \mathcal{F}(\ell_2)$ and $|\mathcal{F}(\ell_2) \setminus \mathcal{F}(\ell_1)| = 1$.

Height of the poset of closures

ℓ_2 *covers* ℓ_1, if $\ell_1 \leq \ell_2$ and for all ℓ' such that $\ell_1 \leq \ell' \leq \ell_2$ either $\ell' = \ell_1$ or $\ell' = \ell_2$.

Proposition (BDK)

ℓ_2 covers ℓ_1 iff $\mathcal{F}(\ell_1) \subseteq \mathcal{F}(\ell_2)$ and $|\mathcal{F}(\ell_2) \setminus \mathcal{F}(\ell_1)| = 1$.

The poset of all closures over a given schema $\mathcal{R}$, $\mathbf{P}(\mathcal{R})$, is *ranked*: its elements are distributed in *levels* and if ℓ_2 covers ℓ_1, then ℓ_2 is in the next level above ℓ_1's one.

Height of the poset of closures

ℓ_2 *covers* ℓ_1, if $\ell_1 \leq \ell_2$ and for all ℓ' such that $\ell_1 \leq \ell' \leq \ell_2$ either $\ell' = \ell_1$ or $\ell' = \ell_2$.

Proposition (BDK)

ℓ_2 covers ℓ_1 iff $\mathcal{F}(\ell_1) \subseteq \mathcal{F}(\ell_2)$ and $|\mathcal{F}(\ell_2) \setminus \mathcal{F}(\ell_1)| = 1$.

The poset of all closures over a given schema $\mathcal{R}$, $\mathbf{P}(\mathcal{R})$, is *ranked*: its elements are distributed in *levels* and if ℓ_2 covers ℓ_1, then ℓ_2 is in the next level above ℓ_1's one.

Let $|\mathcal{R}| = n$. Since $\emptyset$ and $\mathcal{R}$ are both closed for any closure considered, the *height* of $\mathbf{P}(\mathcal{R})$ is $2^n - 2$.

Distance

Distance

Definition

Let $\mathbf{r}$ and $\mathbf{r}'$ be two instances of schema $\mathcal{R}$. Their distance $d(\mathbf{r}, \mathbf{r}')$ is defined to be the graph theoretic distance of $\ell(\mathbf{r})$ and $\ell(\mathbf{r}')$ in the Hasse diagram of $\mathbf{P}(\mathcal{R})$. That is, the length of the shortest path between points $\ell(\mathbf{r})$ and $\ell(\mathbf{r}')$ using only covering edges.

Distance

Definition

Let $\mathbf{r}$ and $\mathbf{r}'$ be two instances of schema $\mathcal{R}$. Their distance $d(\mathbf{r}, \mathbf{r}')$ is defined to be the graph theoretic distance of $\ell(\mathbf{r})$ and $\ell(\mathbf{r}')$ in the Hasse diagram of $\mathbf{P}(\mathcal{R})$. That is, the length of the shortest path between points $\ell(\mathbf{r})$ and $\ell(\mathbf{r}')$ using only covering edges.

Satisfies the triangle inequality.

Distance

Definition

Let $\mathbf{r}$ and $\mathbf{r}'$ be two instances of schema $\mathcal{R}$. Their distance $d(\mathbf{r}, \mathbf{r}')$ is defined to be the graph theoretic distance of $\ell(\mathbf{r})$ and $\ell(\mathbf{r}')$ in the Hasse diagram of $\mathbf{P}(\mathcal{R})$. That is, the length of the shortest path between points $\ell(\mathbf{r})$ and $\ell(\mathbf{r}')$ using only covering edges.

Satisfies the triangle inequality.
Since every closure $\mathcal{L}$ on a finite underlying set is in the form $\mathcal{L} = \ell(\mathbf{r})$, the *distance of closures* is defined.

3 Possible applications

Approximate semantic sampling of existing databases

Introduced by De Marchi and Petit.

Approximate semantic sampling of existing databases

Introduced by De Marchi and Petit.
Given a database relation $\mathbf{r}$, which has normally many tuples, the goal is to find a subrelation $\mathbf{r}' \subseteq \mathbf{r}$ with a "small" number of tuples such that $\mathbf{r}'$ satisfies the same set of functional dependencies as $\mathbf{r}$.

Approximate semantic sampling of existing databases

Introduced by De Marchi and Petit.
Given a database relation $\mathbf{r}$, which has normally many tuples, the goal is to find a subrelation $\mathbf{r}' \subseteq \mathbf{r}$ with a "small" number of tuples such that $\mathbf{r}'$ satisfies the same set of functional dependencies as $\mathbf{r}$.
Sometimes any $\mathbf{r}'$ that satisfies the same set of functional dependencies as $\mathbf{r}$ has still "too many" tuples for a database administrator to understand what functional dependencies $\mathbf{r}$ currently satisfies.

Approximate semantic sampling of existing databases

Introduced by De Marchi and Petit.

Given a database relation $\mathbf{r}$, which has normally many tuples, the goal is to find a subrelation $\mathbf{r}' \subseteq \mathbf{r}$ with a "small" number of tuples such that $\mathbf{r}'$ satisfies the same set of functional dependencies as $\mathbf{r}$.

Sometimes any $\mathbf{r}'$ that satisfies the same set of functional dependencies as $\mathbf{r}$ has still "too many" tuples for a database administrator to understand what functional dependencies $\mathbf{r}$ currently satisfies.

Approximation of the problem: Given a database relation $\mathbf{r}$ and an upper bound b, find a subrelation $\mathbf{r}' \subseteq \mathbf{r}$ such that $|\mathbf{r}'| \leq b$ and $d(\mathbf{r}, \mathbf{r}') = min\{d(\mathbf{r}, \mathbf{r}''): |\mathbf{r}''| \leq b,\ \mathbf{r}'' \subseteq \mathbf{r}\}$.

Approximate integrity enforcement

Approximate integrity enforcement

The usual update operations on a database system only allow updates on a given relation $\mathbf{r}$ that result in a modified relation $\mathbf{r}'$ that satisfies the set Σ of functional dependencies specified over the given relation schema. Instead of this, we may allow more updates that keep the integrity constraints "approximately intact", as follows.

Approximate integrity enforcement

The usual update operations on a database system only allow updates on a given relation $\mathbf{r}$ that result in a modified relation $\mathbf{r}'$ that satisfies the set Σ of functional dependencies specified over the given relation schema. Instead of this, we may allow more updates that keep the integrity constraints "approximately intact", as follows.
Given $\mathbf{r}$, a time-window t and an upper bound b, only allow updates within the time-window t that result in a modified relation $\mathbf{r}'$ whose distance $d(\mathbf{r}', \mathbf{r})$ from $\mathbf{r}$ is at most b. Here it is an important special case if we enforce that $\mathbf{r}'$ must satisfy the same keys as $\mathbf{r}$.

Semantic schema matching

The definition of distance of databases can be extended to relational schemata.

Semantic schema matching

The definition of distance of databases can be extended to relational schemata.
For given $(\mathcal{R}, \Sigma)$ and $(\mathcal{R}', \Sigma')$ where $\mathcal{R}$ and $\mathcal{R}'$ have the same number of attributes and Σ and Σ' are sets of functional dependencies on $\mathcal{R}$ and $\mathcal{R}'$, respectively, $d(\Sigma, \Sigma')$ denotes the number of elements in the symmetric difference of the closed set systems generated by Σ and Σ', respectively.

Semantic schema matching

The definition of distance of databases can be extended to relational schemata.
For given $(\mathcal{R}, \Sigma)$ and $(\mathcal{R}', \Sigma')$ where $\mathcal{R}$ and $\mathcal{R}'$ have the same number of attributes and Σ and Σ' are sets of functional dependencies on $\mathcal{R}$ and $\mathcal{R}'$, respectively, $d(\Sigma, \Sigma')$ denotes the number of elements in the symmetric difference of the closed set systems generated by Σ and Σ', respectively. Note that if $\mathbf{r}$ is an Armstrong instance of $(\mathcal{R}, \Sigma)$ and $\mathbf{r}'$ is an Armstrong instance of $(\mathcal{R}', \Sigma')$, then $d(\Sigma, \Sigma') = d(\mathbf{r}, \mathbf{r}')$.

Semantic schema matching

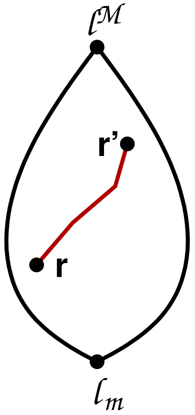
The definition of distance of databases can be extended to relational schemata.
For given $(\mathcal{R}, \Sigma)$ and $(\mathcal{R}', \Sigma')$ where $\mathcal{R}$ and $\mathcal{R}'$ have the same number of attributes and Σ and Σ' are sets of functional dependencies on $\mathcal{R}$ and $\mathcal{R}'$, respectively, $d(\Sigma, \Sigma')$ denotes the number of elements in the symmetric difference of the closed set systems generated by Σ and Σ', respectively. Note that if $\mathbf{r}$ is an Armstrong instance of $(\mathcal{R}, \Sigma)$ and $\mathbf{r}'$ is an Armstrong instance of $(\mathcal{R}', \Sigma')$, then $d(\Sigma, \Sigma') = d(\mathbf{r}, \mathbf{r}')$.
We want to match the two schemata as much as possible, so we want to find a bijection β between $\mathcal{R}$ and $\mathcal{R}'$ such that
$d(\beta(\Sigma), \Sigma') = \min\{d(\beta'(\Sigma), \Sigma') \colon \beta' \text{ is a bijection between} \mathcal{R} \text{ and } \mathcal{R}'\}$.
Here $\beta(\Sigma) = \{\beta(a_1)...\beta(a_n) \to \beta(a) \colon a_1 \ldots a_n \to a \in \Sigma\}$.

4 General observations on the distance of closures

Maximum distance

Proposition

Let $|\mathcal{R}| = n$. Then $d(\mathbf{r}, \mathbf{r}') \leq 2^n - 2$ for any two instances of the schema $\mathcal{R}$.



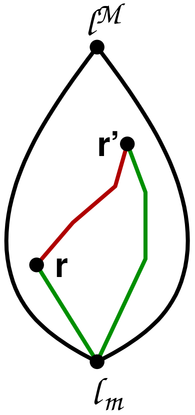
Maximum distance

Proposition

Let $|\mathcal{R}| = n$. Then $d(\mathbf{r}, \mathbf{r}') \leq 2^n - 2$ for any two instances of the schema $\mathcal{R}$.

Proof

$$d(\mathbf{r}, \mathbf{r}') \leq d(\mathbf{r}, \ell_m) + d(\ell_m, \mathbf{r}')$$



Maximum distance

Proposition

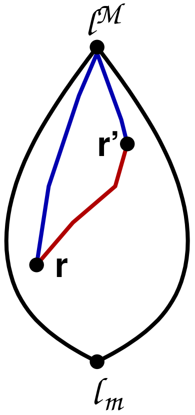
Let $|\mathcal{R}| = n$. Then $d(\mathbf{r}, \mathbf{r}') \leq 2^n - 2$ for any two instances of the schema $\mathcal{R}$.

Proof

$$d(\mathbf{r}, \mathbf{r}') \leq d(\mathbf{r}, \ell_m) + d(\ell_m, \mathbf{r}')$$

and

$$d(\mathbf{r}, \mathbf{r}') \leq d(\mathbf{r}, \ell^M) + d(\ell^M, \mathbf{r}')$$



Maximum distance

Proposition

Let $|\mathcal{R}| = n$. Then $d(\mathbf{r}, \mathbf{r}') \leq 2^n - 2$ for any two instances of the schema $\mathcal{R}$.

Proof

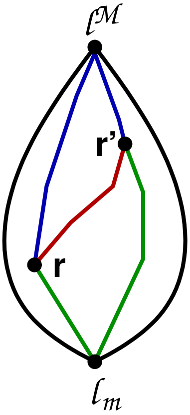
$$d(\mathbf{r}, \mathbf{r}') \leq d(\mathbf{r}, \boldsymbol{\ell}_m) + d(\boldsymbol{\ell}_m, \mathbf{r}')$$

and

$$d(\mathbf{r}, \mathbf{r}') \leq d(\mathbf{r}, \boldsymbol{\ell}^M) + d(\boldsymbol{\ell}^M, \mathbf{r}')$$

Thus,

$$2\, d(\mathbf{r}, \mathbf{r}') \leq d(\mathbf{r}, \boldsymbol{\ell}_m) + d(\mathbf{r}, \boldsymbol{\ell}^M) + d(\boldsymbol{\ell}_m, \mathbf{r}') + \\ + d(\boldsymbol{\ell}^M, \mathbf{r}') = 2^n - 2 + 2^n - 2.$$



Distance of two instances

Theorem

For any two instances $\mathbf{r}$ *and* $\mathbf{r}'$ *of the schema* $\mathcal{R}$ *we have*

$$d(\mathbf{r},\mathbf{r}') = |\mathcal{F}(\ell(\mathbf{r})) \,\Delta\, \mathcal{F}(\ell(\mathbf{r}'))|, \tag{2}$$

where $A \,\Delta\, B$ *denotes the symmetric difference of the two sets, i.e.,* $A \,\Delta\, B = A \setminus B \cup B \setminus A$.

Distance of two instances

Theorem

For any two instances $\mathbf{r}$ *and* $\mathbf{r}'$ *of the schema* $\mathcal{R}$ *we have*

$$d(\mathbf{r},\mathbf{r}') = |\mathcal{F}(\ell(\mathbf{r})) \,\Delta\, \mathcal{F}(\ell(\mathbf{r}'))|, \tag{2}$$

where $A \,\Delta\, B$ *denotes the symmetric difference of the two sets, i.e.,* $A \,\Delta\, B = A \setminus B \cup B \setminus A$.

The proof uses the following well-known result.

Distance of two instances

Theorem

For any two instances $\mathbf{r}$ *and* $\mathbf{r}'$ *of the schema* $\mathcal{R}$ *we have*

$$d(\mathbf{r},\mathbf{r}') = |\mathcal{F}(\ell(\mathbf{r})) \,\Delta\, \mathcal{F}(\ell(\mathbf{r}'))|, \tag{2}$$

where $A \,\Delta\, B$ *denotes the symmetric difference of the two sets, i.e.,* $A \,\Delta\, B = A \setminus B \cup B \setminus A$.

The proof uses the following well-known result.

Theorem (Demetrovics, Katona)

A collection $\mathcal{F}$ *of subsets of* $\mathcal{R}$ *is the collection of closed sets of some closure* $\ell(\mathbf{r})$ *for an appropriate instance* $\mathbf{r}$ *of* $\mathcal{R}$ *iff* $\emptyset, \mathcal{R} \in \mathcal{F}$ *and* $\mathcal{F}$ *is closed under intersection.*

Distance of two instances

Theorem

For any two instances $\mathbf{r}$ *and* $\mathbf{r}'$ *of the schema* $\mathcal{R}$ *we have*

$$d(\mathbf{r},\mathbf{r}') = |\mathcal{F}(\ell(\mathbf{r}))\,\Delta\,\mathcal{F}(\ell(\mathbf{r}'))|, \tag{2}$$

where $A\,\Delta\,B$ *denotes the symmetric difference of the two sets, i.e.,* $A\,\Delta\,B = A\setminus B \cup B\setminus A$.

The proof uses the following well-known result.

Theorem (Demetrovics, Katona)

A collection $\mathcal{F}$ *of subsets of* $\mathcal{R}$ *is the collection of closed sets of some closure* $\ell(\mathbf{r})$ *for an appropriate instance* $\mathbf{r}$ *of* $\mathcal{R}$ *iff* $\emptyset, \mathcal{R} \in \mathcal{F}$ *and* $\mathcal{F}$ *is closed under intersection.*

$|\mathcal{F}(\ell(\mathbf{r}))|$ changes by one traversing along a covering edge in the Hasse diagram of $\mathbf{P}(\mathcal{R})$, so $LHS \geq RHS$.

Upper bound (sketch)

Upper bound (sketch)

First, we "peel off" sets of $\mathcal{F}(\boldsymbol{\ell}(\mathbf{r})) \setminus \mathcal{F}(\boldsymbol{\ell}(\mathbf{r}'))$ one by one, taking always a maximal element.

Upper bound (sketch)

First, we “peel off” sets of $\mathcal{F}(\boldsymbol{\ell}(\mathbf{r})) \setminus \mathcal{F}(\boldsymbol{\ell}(\mathbf{r}'))$ one by one, taking always a maximal element.

We arrive at $\mathcal{F}(\boldsymbol{\ell}(\mathbf{r})) \cap \mathcal{F}(\boldsymbol{\ell}(\mathbf{r}'))$.

Upper bound (sketch)

First, we "peel off" sets of $\mathcal{F}(\boldsymbol{\ell}(\mathbf{r})) \setminus \mathcal{F}(\boldsymbol{\ell}(\mathbf{r}'))$ one by one, taking always a maximal element.

We arrive at $\mathcal{F}(\boldsymbol{\ell}(\mathbf{r})) \cap \mathcal{F}(\boldsymbol{\ell}(\mathbf{r}'))$.

Add successively a minimal element of $\mathcal{F}(\boldsymbol{\ell}(\mathbf{r}')) \setminus \mathcal{F}(\boldsymbol{\ell}(\mathbf{r}))$ with respect to set containment.

3 Diameter of collection of databases with the same set of minimal keys

Keys, minimal keys

Keys, minimal keys

$K \subseteq \mathcal{R}$ is a *key* in instance $\mathbf{r}$ if $\mathbf{r} \models K \to \mathcal{R}$, and it is *minimal* if $\forall K' \subsetneqq K : \mathbf{r} \not\models K' \to \mathcal{R}$.

Keys, minimal keys

$K \subseteq \mathcal{R}$ is a *key* in instance $\mathbf{r}$ if $\mathbf{r} \models K \to \mathcal{R}$, and it is *minimal* if $\forall K' \subsetneqq K : \mathbf{r} \not\models K' \to \mathcal{R}$. $\iff$ K is a minimal key for $\mathbf{r}$ iff there are no two rows of $\mathbf{r}$ that agree on K, but K is minimal with respect to this property.

Keys, minimal keys

$K \subseteq \mathcal{R}$ is a *key* in instance $\mathbf{r}$ if $\mathbf{r} \models K \to \mathcal{R}$, and it is *minimal* if $\forall K' \subsetneqq K : \mathbf{r} \not\models K' \to \mathcal{R}$. $\iff$ K is a minimal key for $\mathbf{r}$ iff there are no two rows of $\mathbf{r}$ that agree on K, but K is minimal with respect to this property.

$\mathcal{K}(\mathbf{r})$ denotes the system of minimal keys of $\mathbf{r}$. It is an inclusion-free family.

Keys, minimal keys

$K \subseteq \mathcal{R}$ is a *key* in instance $\mathbf{r}$ if $\mathbf{r} \models K \to \mathcal{R}$, and it is *minimal* if $\forall K' \subsetneqq K : \mathbf{r} \not\models K' \to \mathcal{R}$. $\iff$ K is a minimal key for $\mathbf{r}$ iff there are no two rows of $\mathbf{r}$ that agree on K, but K is minimal with respect to this property.

$\mathcal{K}(\mathbf{r})$ denotes the system of minimal keys of $\mathbf{r}$. It is an inclusion-free family.

$\boldsymbol{\ell}(\mathbf{r})$ uniquely determines $\mathcal{K}(\mathbf{r})$, since $\mathbf{r} \models A \to B \iff B \subseteq \boldsymbol{\ell}(\mathbf{r})(A)$.

Keys, minimal keys

$K \subseteq \mathcal{R}$ is a *key* in instance $\mathbf{r}$ if $\mathbf{r} \models K \to \mathcal{R}$, and it is *minimal* if $\forall K' \subsetneq K : \mathbf{r} \not\models K' \to \mathcal{R}$. $\iff$ K is a minimal key for $\mathbf{r}$ iff there are no two rows of $\mathbf{r}$ that agree on K, but K is minimal with respect to this property.

$\mathcal{K}(\mathbf{r})$ denotes the system of minimal keys of $\mathbf{r}$. It is an inclusion-free family.

$\ell(\mathbf{r})$ uniquely determines $\mathcal{K}(\mathbf{r})$, since $\mathbf{r} \models A \to B \iff B \subseteq \ell(\mathbf{r})(A)$.

$\mathcal{K}(\mathbf{r})$ does not determine $\ell(\mathbf{r})$.

An example

An example

Let $\mathcal{R} = \{a, b, c, d\}$, and the family of keys be
$\mathcal{K} = \{\{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}\}$.

An example

Let $\mathcal{R} = \{a, b, c, d\}$, and the family of keys be $\mathcal{K} = \{\{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}\}$.

$\mathcal{K}$ is system of keys, when $\ell(\mathbf{r_1})$-closed sets are $\emptyset, \{a, b\}, \{c, d\}, \{a, b, c, d\}$.

An example

Let $\mathcal{R} = \{a, b, c, d\}$, and the family of keys be $\mathcal{K} = \{\{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}\}$.

$\mathcal{K}$ is system of keys, when $\ell(\mathbf{r_1})$-closed sets are $\emptyset, \{a, b\}, \{c, d\}, \{a, b, c, d\}$.

$$\mathbf{r_1} = \begin{array}{cccc} a & b & c & d \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 2 & 2 & 0 & 0 \end{array}$$

An example

Let $\mathcal{R} = \{a, b, c, d\}$, and the family of keys be
$\mathcal{K} = \{\{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}\}$.

$\mathcal{K}$ is system of keys, when $\ell(\mathbf{r_1})$-closed sets are $\emptyset, \{a, b\}, \{c, d\}, \{a, b, c, d\}$.

$$\mathbf{r_1} = \begin{array}{cccc} a & b & c & d \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 2 & 2 & 0 & 0 \end{array}$$

On the other hand, $\ell(\mathbf{r_2}) > \ell(\mathbf{r_1})$ has the same key system, where additionally the one-element subsets are $\ell(\mathbf{r_2})$-closed, too.

An example

Let $\mathcal{R} = \{a, b, c, d\}$, and the family of keys be $\mathcal{K} = \{\{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}\}$.

$\mathcal{K}$ is system of keys, when $\ell(\mathbf{r_1})$-closed sets are $\emptyset, \{a, b\}, \{c, d\}, \{a, b, c, d\}$.

$$\mathbf{r_1} = \begin{array}{cccc} a & b & c & d \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 2 & 2 & 0 & 0 \end{array}$$

On the other hand, $\ell(\mathbf{r_2}) > \ell(\mathbf{r_1})$ has the same key system, where additionally the one-element subsets are $\ell(\mathbf{r_2})$-closed, too.

$$\mathbf{r_2} = \begin{array}{cccc} a & b & c & d \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 2 & 2 & 0 & 0 \\ 0 & 3 & 3 & 3 \\ 4 & 0 & 4 & 4 \\ 5 & 5 & 0 & 5 \\ 6 & 6 & 6 & 0 \end{array}$$

Antikeys

A subset $A \subset \mathcal{R}$ is a *maximal antikey* if it does not contain any key, and maximal with respect to this property. The collection of antikeys for a minimal key system $\mathcal{K}$ is denoted by $\mathcal{K}^{-1}$.

Antikeys

A subset $A \subset \mathcal{R}$ is a *maximal antikey* if it does not contain any key, and maximal with respect to this property. The collection of antikeys for a minimal key system $\mathcal{K}$ is denoted by $\mathcal{K}^{-1}$.

Minimal keys and aximal antikeys determine each other, respectively:

Antikeys

A subset $A \subset \mathcal{R}$ is a *maximal antikey* if it does not contain any key, and maximal with respect to this property. The collection of antikeys for a minimal key system $\mathcal{K}$ is denoted by $\mathcal{K}^{-1}$.

Minimal keys and aximal antikeys determine each other, respectively:

Maximal antikeys are maximal sets that do not contain any key and keys are minimal sets that are not contained in any antikey.

Antikeys

A subset $A \subset \mathcal{R}$ is a *maximal antikey* if it does not contain any key, and maximal with respect to this property. The collection of antikeys for a minimal key system $\mathcal{K}$ is denoted by $\mathcal{K}^{-1}$.

Minimal keys and aximal antikeys determine each other, respectively:

Maximal antikeys are maximal sets that do not contain any key and keys are minimal sets that are not contained in any antikey.

Both minimal key systems and maximal antikey systems form inclusion-free families of subsets of $\mathcal{R}$, that is no minimal key/antikey can contain another minimal key/antikey.

Antikeys

A subset $A \subset \mathcal{R}$ is a *maximal antikey* if it does not contain any key, and maximal with respect to this property. The collection of antikeys for a minimal key system $\mathcal{K}$ is denoted by $\mathcal{K}^{-1}$.

Minimal keys and aximal antikeys determine each other, respectively:

Maximal antikeys are maximal sets that do not contain any key and keys are minimal sets that are not contained in any antikey.

Both minimal key systems and maximal antikey systems form inclusion-free families of subsets of $\mathcal{R}$, that is no minimal key/antikey can contain another minimal key/antikey.

In the previous example $\mathcal{K} = \{\{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}\}$ and $\mathcal{K}^{-1} = \{\{a, b\}, \{c, d\}\}$.

The theorem

For a set system $\mathcal{A}$ of subsets of $\mathcal{R}$ let $\mathcal{A}\!\downarrow = \{B \subseteq \mathcal{R} \colon \exists A \in \mathcal{A} \text{ with } B \subseteq A\} \cup \{\mathcal{R}\}$ and let $\mathcal{A}_\cap = \{B \subseteq \mathcal{R} \colon \exists i \geq 1, A_1, A_2, \ldots A_i \in \mathcal{A} \text{ with } B = A_1 \cap A_2 \cap \ldots \cap A_i\} \cup \{\mathcal{R}\} \cup \{\emptyset\}$.

The theorem

For a set system $\mathcal{A}$ of subsets of $\mathcal{R}$ let
$\mathcal{A}\!\downarrow = \{B \subseteq \mathcal{R}\colon \exists A \in \mathcal{A} \text{ with } B \subseteq A\} \cup \{\mathcal{R}\}$ and let $\mathcal{A}_\cap = \{B \subseteq \mathcal{R}\colon \exists i \geq 1, A_1, A_2, \dots A_i \in \mathcal{A} \text{ with } B = A_1 \cap A_2 \cap \dots \cap A_i\} \cup \{\mathcal{R}\} \cup \{\emptyset\}$.
That is, $\mathcal{A}\!\downarrow$ is the down-set generated by $\mathcal{A}$ appended with $\mathcal{R}$ and $\mathcal{A}_\cap$ is the set system closed under intersection generated by $\mathcal{A}$.

The theorem

For a set system $\mathcal{A}$ of subsets of $\mathcal{R}$ let $\mathcal{A}{\downarrow} = \{B \subseteq \mathcal{R} \colon \exists A \in \mathcal{A} \text{ with } B \subseteq A\} \cup \{\mathcal{R}\}$ and let $\mathcal{A}_{\cap} = \{B \subseteq \mathcal{R} \colon \exists i \geq 1, A_1, A_2, \ldots A_i \in \mathcal{A} \text{ with } B = A_1 \cap A_2 \cap \ldots \cap A_i\} \cup \{\mathcal{R}\} \cup \{\emptyset\}$. That is, $\mathcal{A}{\downarrow}$ is the down-set generated by $\mathcal{A}$ appended with $\mathcal{R}$ and $\mathcal{A}_{\cap}$ is the set system closed under intersection generated by $\mathcal{A}$.

Theorem

Let $\mathcal{K}$ be an inclusion-free family of subsets of $\mathcal{R}$. Then the closures whose minimal key system is $\mathcal{K}$ form an interval in the poset of closures $\mathbf{P}(\mathcal{R})$ whose smallest element is the closure with closed sets $\mathcal{K}_{\cap}^{-1}$ and largest element is the closure with closed sets $\mathcal{K}^{-1}{\downarrow}$

The theorem

For a set system $\mathcal{A}$ of subsets of $\mathcal{R}$ let $\mathcal{A}\downarrow = \{B \subseteq \mathcal{R} \colon \exists A \in \mathcal{A} \text{ with } B \subseteq A\} \cup \{\mathcal{R}\}$ and let $\mathcal{A}_\cap = \{B \subseteq \mathcal{R} \colon \exists i \geq 1, A_1, A_2, \dots A_i \in \mathcal{A} \text{ with } B = A_1 \cap A_2 \cap \dots \cap A_i\} \cup \{\mathcal{R}\} \cup \{\emptyset\}$. That is, $\mathcal{A}\downarrow$ is the down-set generated by $\mathcal{A}$ appended with $\mathcal{R}$ and $\mathcal{A}_\cap$ is the set system closed under intersection generated by $\mathcal{A}$.

Theorem

Let $\mathcal{K}$ be an inclusion-free family of subsets of $\mathcal{R}$. Then the closures whose minimal key system is $\mathcal{K}$ form an interval in the poset of closures $\mathbf{P}(\mathcal{R})$ whose smallest element is the closure with closed sets $\mathcal{K}_\cap^{-1}$ and largest element is the closure with closed sets $\mathcal{K}^{-1}\downarrow$

Corollary

The diameter, that is the largest distance between any two elements of the collection of closures with given key system $\mathcal{K}$ is $|\mathcal{K}^{-1}\downarrow| - |\mathcal{K}_\cap^{-1}|$.

Proof (or something like that) of the Theorem

Let A be a maximal antikey. For any $b \in \mathcal{R} \setminus A$, $A \cup \{b\}$ is a key, thus $A \cup \{b\} \to \mathcal{R}$ holds.

Proof (or something like that) of the Theorem

Let A be a maximal antikey. For any $b \in \mathcal{R} \setminus A$, $A \cup \{b\}$ is a key, thus $A \cup \{b\} \to \mathcal{R}$ holds.

Thus $A \not\to b$, so $\ell(\mathbf{r})(A) = A$ for every antikey $A \in \mathcal{K}^{-1}$.

Proof (or something like that) of the Theorem

Let A be a maximal antikey. For any $b \in \mathcal{R} \setminus A$, $A \cup \{b\}$ is a key, thus $A \cup \{b\} \to \mathcal{R}$ holds.

Thus $A \not\to b$, so $\boldsymbol{\ell}(\mathbf{r})(A) = A$ for every antikey $A \in \mathcal{K}^{-1}$.

$\mathcal{F}(\boldsymbol{\ell}(\mathbf{r}))$ is closed under intersection, so $\mathcal{K}_{\cap}^{-1} \subseteq \mathcal{F}(\boldsymbol{\ell}(\mathbf{r}))$.

Proof (or something like that) of the Theorem

Let A be a maximal antikey. For any $b \in \mathcal{R} \setminus A$, $A \cup \{b\}$ is a key, thus $A \cup \{b\} \to \mathcal{R}$ holds.

Thus $A \not\to b$, so $\boldsymbol{\ell}(\mathbf{r})(A) = A$ for every antikey $A \in \mathcal{K}^{-1}$.

$\mathcal{F}(\boldsymbol{\ell}(\mathbf{r}))$ is closed under intersection, so $\mathcal{K}_{\cap}^{-1} \subseteq \mathcal{F}(\boldsymbol{\ell}(\mathbf{r}))$.

If $\boldsymbol{\ell}(\mathbf{r})(X) = X$ holds for some $X \subsetneq \mathcal{R}$, then X is contained in maximal antikey $A \supset X$, hence $X \in \mathcal{K}^{-1}\downarrow$.

4.1 Non-uniform minimal key system

The Theorem

The Theorem

Let $\mathcal{M}$ be a non-empty, inclusion-free family. Define

$$\mathcal{D}(\mathcal{M}) = \{H : \exists M \in \mathcal{M} \text{ such that } H \subseteq M\}, \tag{3}$$

$$\mathcal{U}(\mathcal{M}) = \{H : \exists M \in \mathcal{M} \text{ such that } H \supseteq M\}. \tag{4}$$

The Theorem

Let $\mathcal{M}$ be a non-empty, inclusion-free family. Define

$$\mathcal{D}(\mathcal{M}) = \{H : \exists M \in \mathcal{M} \text{ such that } H \subseteq M\}, \tag{3}$$

$$\mathcal{U}(\mathcal{M}) = \{H : \exists M \in \mathcal{M} \text{ such that } H \supseteq M\}. \tag{4}$$

Theorem

Let $\mathcal{K}$ be a non-empty inclusion-free family of subsets of $[n]$, where $|\mathcal{K}| \geq n$ is fixed. Furthermore, let $S(\mathcal{K})$ denote the set of all closures in which the family of minimal keys is exactly $\mathcal{K}$. Then

$$\mathrm{diam}(S(\mathcal{K})) \leq 2^n - |\mathcal{U}(\mathcal{K}^*)|, \tag{5}$$

where $\mathcal{K}^*$ consists of some lexicographically last sets of size s and all the $s+1$-element sets not containing the selected s-element ones, for some $0 \leq s \leq n-2$ and $|\mathcal{K}^*| = |\mathcal{K}|$.

Tools of the proof

Tools of the proof

Define the (r, ℓ)-*shadow* of a family of r-element sets $\mathcal{A} \subseteq \binom{[n]}{r}$ for $\ell < r$:

$$\sigma_{r,\ell}(\mathcal{A}) = \{H : \ |H| = \ell, \exists A \in \mathcal{A} \text{ such that } H \subset A\}. \tag{6}$$

Tools of the proof

Define the (r,ℓ)-*shadow* of a family of r-element sets $\mathcal{A} \subseteq \binom{[n]}{r}$ for $\ell < r$:

$$\sigma_{r,\ell}(\mathcal{A}) = \{H : \ |H| = \ell, \exists A \in \mathcal{A} \text{ such that } H \subset A\}. \tag{6}$$

Theorem (Shadow Theorem, Kruskal, Katona)

If $\mathcal{A} \subseteq \binom{[n]}{r}, |\mathcal{A}| = m$ then $|\sigma_{r,\ell}(\mathcal{A})|$ is at least as large as the (r,ℓ)-shadow of the family of the lexicographically first m members of $\binom{[n]}{r}$, that is, the size of the (r,ℓ)-shadow attains its minimum for the lexicographically first r-element sets.

Tools of the proof II

Tools of the proof II

For $\mathcal{A} \subseteq 2^{[n]}$ let

$$\mathcal{A}_r = \mathcal{A} \cap \binom{[n]}{r}. \tag{7}$$

Tools of the proof II

For $\mathcal{A} \subseteq 2^{[n]}$ let

$$\mathcal{A}_r = \mathcal{A} \cap \binom{[n]}{r}. \tag{7}$$

The *profile vector* of the family $\mathcal{A} \subseteq 2^{[n]}$ is $p = (p_0, p_1, \ldots, p_n)$ where $p_r = p_r(\mathcal{A}) = |\mathcal{A}_r|$.

Tools of the proof II

For $\mathcal{A} \subseteq 2^{[n]}$ let

$$\mathcal{A}_r = \mathcal{A} \cap \binom{[n]}{r}. \tag{7}$$

The *profile vector* of the family $\mathcal{A} \subseteq 2^{[n]}$ is $p = (p_0, p_1, \ldots, p_n)$ where $p_r = p_r(\mathcal{A}) = |\mathcal{A}_r|$.

Lemma

Let $\mathcal{M}$ be a non-empty inclusion-free family of subsets of $[n]$ with fixed $|\mathcal{M}| \geq n$. Then $|\mathcal{D}(\mathcal{M})|$ attains its minimum for a family satisfying the following conditions with some $2 \leq r \leq n$.

$$p_n = \ldots = p_{r+1} = p_{r-2} = \ldots = p_1 = p_0 = 0, \tag{8}$$

$$\mathcal{M}_r \text{ consists of the lexicographically first } p_r \ r-\text{element subsets,} \tag{9}$$

$$\mathcal{M}_{r-1} = \binom{[n]}{r-1} \setminus \sigma_{r,r-1}(\mathcal{M}_r). \tag{10}$$

4.2 Uniform minimal key system

Unique minimal key

Unique minimal key

Theorem

The diameter of the set of closures having exactly one minimal key A where $0 < |A| = r < n$ is $2^n - 2^r - 2^{n-r}$.

Unique minimal key

Theorem

The diameter of the set of closures having exactly one minimal key A where $0 < |A| = r < n$ is $2^n - 2^r - 2^{n-r}$.

The family of closed sets $\mathcal{F}$ satisfies the following conditions.

$$\text{If } F \supseteq \mathcal{R} \setminus A \text{ then } F \in \mathcal{F}, \tag{11}$$

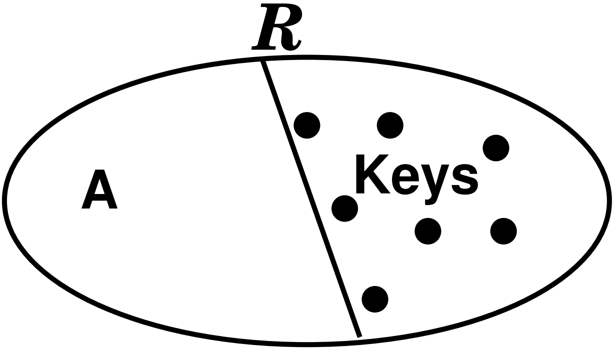
$$\text{if } F \supseteq A, F \neq R \text{ then } F \notin \mathcal{F}, \tag{12}$$

$$\emptyset \in \mathcal{F}. \tag{13}$$

Why not 1-element keys?

Why not 1-element keys?

If all keys are one-element sets, then $\mathcal{K}^{-1}$ consists of a single set A, thus $\mathcal{K}^{-1}\downarrow$ consists of all subsets of A and $\mathcal{R}$, while $\mathcal{K}_{\cap}^{-1}$ consists of three sets, $\emptyset$, A and $\mathcal{R}$, i.e. the diameter is $2^{|A|} - 2$.



2-element keys

2-element keys

Let $G = ([n], E)$ be the graph where $\{i, j\} \in E (i \neq j)$ iff $\{i, j\} \notin \mathcal{K}$.

2-element keys

Let $G = ([n], E)$ be the graph where $\{i, j\} \in E (i \neq j)$ iff $\{i, j\} \notin \mathcal{K}$. The set of closures having $\binom{[n]}{2} - E$ as the set of minimal kys is denoted by $S_2(G)$. We give upper bound for

$$s_2(e) = \max_{\{G=([n],E):\ |E|=e\}} \mathrm{diam} S_2(G). \tag{14}$$

2-element keys

Let $G=([n],E)$ be the graph where $\{i,j\}\in E(i\neq j)$ iff $\{i,j\}\notin\mathcal{K}$. The set of closures having $\binom{[n]}{2}-E$ as the set of minimal kys is denoted by $S_2(G)$. We give upper bound for

$$s_2(e)=\max_{\{G=([n],E):\ |E|=e\}}\operatorname{diam}S_2(G). \tag{14}$$

Theorem

If $e=\binom{t}{2}+r$*, where* $0<r\leq t$*, then*

$$\operatorname{diam}S_2(G)\leq\begin{cases}2^t+2^r-4 \ \textit{if}\ r<t\\ 2^{t+1}-2\ \textit{if}\ r=t\end{cases} \tag{15}$$

for a graph G whose connected components are isolated vertices except for one component. Furthermore, this bound is sharp.

Antikeys and cliques

Antikeys and cliques

If the family of minimal keys is $\mathcal{K} = \binom{[n]}{2} - E$, then the members of $\mathcal{K}^{-1}$ are maximal complete subgraphs in G.

Antikeys and cliques

If the family of minimal keys is $\mathcal{K} = \binom{[n]}{2} - E$, then the members of $\mathcal{K}^{-1}$ are maximal complete subgraphs in G.
$\mathcal{K}^{-1}\downarrow$ consists of all complete subgraphs of G, while $\mathcal{K}_{\cap}^{-1}$ consists of those complete subgraphs that are intersections of cliques.

Antikeys and cliques

If the family of minimal keys is $\mathcal{K} = \binom{[n]}{2} - E$, then the members of $\mathcal{K}^{-1}$ are maximal complete subgraphs in G.
$\mathcal{K}^{-1}\downarrow$ consists of all complete subgraphs of G, while $\mathcal{K}_{\cap}^{-1}$ consists of those complete subgraphs that are intersections of cliques.

$$|\mathcal{K}^{-1}\downarrow| \leq 2^t + 2^r + n - t - 1. \tag{16}$$

Antikeys and cliques

If the family of minimal keys is $\mathcal{K} = \binom{[n]}{2} - E$, then the members of $\mathcal{K}^{-1}$ are maximal complete subgraphs in G.
$\mathcal{K}^{-1}\downarrow$ consists of all complete subgraphs of G, while $\mathcal{K}_{\cap}^{-1}$ consists of those complete subgraphs that are intersections of cliques.

$$|\mathcal{K}^{-1}\downarrow| \leq 2^t + 2^r + n - t - 1. \tag{16}$$

Follows from the following theorem of Erdős.

Theorem (Erdős, 1962)

Let $G = (V, E)$ be a connected graph of e edges. Assume, that $e = \binom{t}{2} + r$, where $0 < r \leq t$. Then the number of complete k-subgraphs $C_k(G)$ of G is at most

$$C_k(G) \leq \binom{t}{k} + \binom{r}{k-1}. \tag{17}$$

r-uniform key system

r-uniform key system

Let D be a closure whose minimal keys have exactly $r(\geq 2)$ elements. $H = ([n], \mathcal{E})$ be the hypergraph where $\mathcal{K} = \binom{[n]}{r} \setminus \mathcal{E}$, with $|\mathcal{E}| = e$. The set of closures having $\binom{[n]}{r} \setminus \mathcal{E}$ as the set of minimal keys will be denoted by $S_r(H)$.

r-uniform key system

Let D be a closure whose minimal keys have exactly $r(\geq 2)$ elements. $H = ([n], \mathcal{E})$ be the hypergraph where $\mathcal{K} = \binom{[n]}{r} \setminus \mathcal{E}$, with $|\mathcal{E}| = e$. The set of closures having $\binom{[n]}{r} \setminus \mathcal{E}$ as the set of minimal keys will be denoted by $S_r(H)$. Want upper estimate on

$$\max_{\{H=([n],\mathcal{E}):\ |\mathcal{E}|=e\}} \mathsf{diam} S_r(H). \tag{18}$$

r-uniform key system

Let D be a closure whose minimal keys have exactly $r(\geq 2)$ elements. $H = ([n], \mathcal{E})$ be the hypergraph where $\mathcal{K} = \binom{[n]}{r} \setminus \mathcal{E}$, with $|\mathcal{E}| = e$. The set of closures having $\binom{[n]}{r} \setminus \mathcal{E}$ as the set of minimal keys will be denoted by $S_r(H)$. Want upper estimate on

$$\max_{\{H=([n],\mathcal{E}):\ |\mathcal{E}|=e\}} \mathrm{diam} S_r(H). \tag{18}$$

Theorem

If $e \leq \binom{a}{r}$ *then* $\mathrm{diam}(S_r(H)) \leq 2^a + e2^r$.

Proof

Proof sketch

Proof sketch

$\mathcal{K} = \binom{[n]}{r} \setminus \mathcal{E}$ implies that
$\mathcal{K}^{-1} = \{B \subset [n] \colon \binom{B}{r} \subset \mathcal{E} \wedge \forall B' \supsetneq B \ \binom{B'}{r} \setminus \mathcal{E} \neq \emptyset\}$. These are called the *(hyper)cliques* of H.

Proof sketch

$\mathcal{K} = \binom{[n]}{r} \setminus \mathcal{E}$ implies that
$\mathcal{K}^{-1} = \{B \subset [n] \colon \binom{B}{r} \subset \mathcal{E} \wedge \forall B' \supsetneq B \ \binom{B'}{r} \setminus \mathcal{E} \neq \emptyset\}$. These are called the *(hyper)cliques* of H.

Need: the number of sets of the vertices of H which are subsets of at least one hyperclique and are not intersections of those is at most $2^a + e2^r$.

Proof sketch

$\mathcal{K} = \binom{[n]}{r} \setminus \mathcal{E}$ implies that
$\mathcal{K}^{-1} = \{B \subset [n] \colon \binom{B}{r} \subset \mathcal{E} \wedge \forall B' \supsetneq B \ \binom{B'}{r} \setminus \mathcal{E} \neq \emptyset\}$. These are called the *(hyper)cliques* of H.
Need: the number of sets of the vertices of H which are subsets of at least one hyperclique and are not intersections of those is at most $2^a + e2^r$. $|\mathcal{K}^{-1}\downarrow| - |\mathcal{K}^{-1}_{\cap}| \leq 2^a + e2^r$.

Proof sketch

$\mathcal{K} = \binom{[n]}{r} \setminus \mathcal{E}$ implies that
$\mathcal{K}^{-1} = \{B \subset [n] \colon \binom{B}{r} \subset \mathcal{E} \wedge \forall B' \supsetneq B \ \binom{B'}{r} \setminus \mathcal{E} \neq \emptyset\}$. These are called the *(hyper)cliques* of H.
Need: the number of sets of the vertices of H which are subsets of at least one hyperclique and are not intersections of those is at most $2^a + e2^r$. $|\mathcal{K}^{-1}\downarrow| - |\mathcal{K}_{\cap}^{-1}| \leq 2^a + e2^r$.
We show that $|\mathcal{K}^{-1}\downarrow| \leq 2^a + e2^r$.

Proof sketch

$\mathcal{K} = \binom{[n]}{r} \setminus \mathcal{E}$ implies that
$\mathcal{K}^{-1} = \{B \subset [n] \colon \binom{B}{r} \subset \mathcal{E} \wedge \forall B' \supsetneq B \ \binom{B'}{r} \setminus \mathcal{E} \neq \emptyset\}$. These are called the *(hyper)cliques* of H.
Need: the number of sets of the vertices of H which are subsets of at least one hyperclique and are not intersections of those is at most $2^a + e2^r$. $|\mathcal{K}^{-1}\downarrow| - |\mathcal{K}_{\cap}^{-1}| \leq 2^a + e2^r$.
We show that $|\mathcal{K}^{-1}\downarrow| \leq 2^a + e2^r$.
For $0 < i \leq r$ let I be an i-element subset of a hyperclique.

Proof sketch

$\mathcal{K} = \binom{[n]}{r} \setminus \mathcal{E}$ implies that

$\mathcal{K}^{-1} = \{B \subset [n] \colon \binom{B}{r} \subset \mathcal{E} \wedge \forall B' \supsetneq B \ \binom{B'}{r} \setminus \mathcal{E} \neq \emptyset\}$. These are called the *(hyper)cliques* of H.

Need: the number of sets of the vertices of H which are subsets of at least one hyperclique and are not intersections of those is at most $2^a + e2^r$. $|\mathcal{K}^{-1}\downarrow| - |\mathcal{K}^{-1}_{\cap}| \leq 2^a + e2^r$.

We show that $|\mathcal{K}^{-1}\downarrow| \leq 2^a + e2^r$.

For $0 < i \leq r$ let I be an i-element subset of a hyperclique.

$\Rightarrow \exists |R| = r \colon I \subseteq R \in \mathcal{E}$.

Proof sketch

$\mathcal{K} = \binom{[n]}{r} \setminus \mathcal{E}$ implies that
$\mathcal{K}^{-1} = \{B \subset [n] \colon \binom{B}{r} \subset \mathcal{E} \wedge \forall B' \supsetneq B \ \binom{B'}{r} \setminus \mathcal{E} \neq \emptyset\}$. These are called the *(hyper)cliques* of H.
Need: the number of sets of the vertices of H which are subsets of at least one hyperclique and are not intersections of those is at most $2^a + e2^r$. $|\mathcal{K}^{-1}\!\downarrow| - |\mathcal{K}_{\cap}^{-1}| \leq 2^a + e2^r$.
We show that $|\mathcal{K}^{-1}\!\downarrow| \leq 2^a + e2^r$.
For $0 < i \leq r$ let I be an i-element subset of a hyperclique.
$\Rightarrow \exists |R| = r \colon I \subseteq R \in \mathcal{E}$. number of such I's is at most $e\binom{r}{i}$.

Proof sketch

$\mathcal{K} = \binom{[n]}{r} \setminus \mathcal{E}$ implies that
$\mathcal{K}^{-1} = \{B \subset [n] \colon \binom{B}{r} \subset \mathcal{E} \wedge \forall B' \supsetneq B \ \binom{B'}{r} \setminus \mathcal{E} \neq \emptyset\}$. These are called the *(hyper)cliques* of H.
Need: the number of sets of the vertices of H which are subsets of at least one hyperclique and are not intersections of those is at most $2^a + e2^r$. $|\mathcal{K}^{-1}{\downarrow}| - |\mathcal{K}_{\cap}^{-1}| \leq 2^a + e2^r$.
We show that $|\mathcal{K}^{-1}{\downarrow}| \leq 2^a + e2^r$.
For $0 < i \leq r$ let I be an i-element subset of a hyperclique.
$\Rightarrow \exists |R| = r \colon I \subseteq R \in \mathcal{E}$. number of such I's is at most $e\binom{r}{i}$.
$r < i$. Let $A_1, \ldots, A_m$ be the family of i-element subsets, whose all r-element subsets are in $\mathcal{E}$.

Proof sketch

$\mathcal{K} = \binom{[n]}{r} \setminus \mathcal{E}$ implies that
$\mathcal{K}^{-1} = \{B \subset [n] \colon \binom{B}{r} \subset \mathcal{E} \wedge \forall B' \supsetneq B \ \binom{B'}{r} \setminus \mathcal{E} \neq \emptyset\}$. These are called the *(hyper)cliques* of H.
Need: the number of sets of the vertices of H which are subsets of at least one hyperclique and are not intersections of those is at most $2^a + e2^r$. $|\mathcal{K}^{-1}\downarrow| - |\mathcal{K}_\cap^{-1}| \leq 2^a + e2^r$.
We show that $|\mathcal{K}^{-1}\downarrow| \leq 2^a + e2^r$.
For $0 < i \leq r$ let I be an i-element subset of a hyperclique.
$\Rightarrow \exists |R| = r \colon I \subseteq R \in \mathcal{E}$. number of such I's is at most $e\binom{r}{i}$.
$r < i$. Let $A_1, \ldots, A_m$ be the family of i-element subsets, whose all r-element subsets are in $\mathcal{E}$. If $m > \binom{a}{i}$ then by the Shadow Theorem (Lovász' version) the number of r-element subsets (hyperedges) is $> \binom{a}{r} \geq e$, so $m \leq \binom{a}{i}$.

Proof sketch

$\mathcal{K} = \binom{[n]}{r} \setminus \mathcal{E}$ implies that
$\mathcal{K}^{-1} = \{B \subset [n]\colon \binom{B}{r} \subset \mathcal{E} \land \forall B' \supsetneq B \; \binom{B'}{r} \setminus \mathcal{E} \neq \emptyset\}$. These are called the *(hyper)cliques* of H.
Need: the number of sets of the vertices of H which are subsets of at least one hyperclique and are not intersections of those is at most $2^a + e2^r$. $|\mathcal{K}^{-1}\!\downarrow| - |\mathcal{K}_{\cap}^{-1}| \le 2^a + e2^r$.
We show that $|\mathcal{K}^{-1}\!\downarrow| \le 2^a + e2^r$.
For $0 < i \le r$ let I be an i-element subset of a hyperclique.
$\Rightarrow \exists |R| = r\colon I \subseteq R \in \mathcal{E}$. number of such I's is at most $e\binom{r}{i}$.
$r < i$. Let $A_1, \ldots, A_m$ be the family of i-element subsets, whose all r-element subsets are in $\mathcal{E}$. If $m > \binom{a}{i}$ then by the Shadow Theorem (Lovász' version) the number of r-element subsets (hyperedges) is $> \binom{a}{r} \ge e$, so $m \le \binom{a}{i}$.
Add up these maximums:

$$e \sum_{i=1}^{r} \binom{r}{i} + \sum_{i=r+1}^{a} \binom{a}{i} \le 2^a + e2^r.$$

An interesting combinatorial question

An interesting combinatorial question

Given a hypergraph $H = (V, \mathcal{E})$, what is the number of complete subhypergraphs that are *not* intersections of *maximal complete subhypergraphs*?

An interesting combinatorial question

Given a hypergraph $H = (V, \mathcal{E})$, what is the number of complete subhypergraphs that are *not* intersections of *maximal complete subhypergraphs*?
We have given good upper bounds in the case of ordinary graphs.