WARSAW UNIVERSITY OF TECHNOLOGY
DEVELOPMENT PROGRAMME

Functional and Approximate Dependencies

Marzena Kryszkiewicz

HUMAN CAPITAL HUMAN – BEST INVESTMENT!
EUROPEAN UNION EUROPEAN SOCIAL FUND
Project is co-financed by European Union within European Social Fund

Functional Dependencies (FDs)

Let D = (O, *AT*) and $A, B \subseteq AT$.

- $A \rightarrow B$ is defined as a *functional dependency* if $\forall x \in O, [x]_A \subseteq [x]_B$.
- A functional dependency $A \rightarrow B$ is called *minimal* if $\forall C \subset A$, $C \rightarrow B$ is not functional.
- A dependency $A \rightarrow B$ is called *trivial* if $B \subseteq A$.

Example: FDs

D = (O, *AT*), where O = {1,2,3,4,5,6}, AT = {*abcd*}.

Id	*a*	*b*	*c*	*d*
1	1	1	2	2
2	1	1	2	2
3	2	2	2	2
4	3	3	3	3
5	5	5	6	6
6	5	5	6	6

- {*cd*} → {*ab*} is not a functional dependency.
- {*ab*} → {*cd*} is a functional dependency, but is not minimal.
- {*a*} → {*cd*} is a minimal functional dependency.
- {*ab*} → {*a*} is trivial.

Functional Dependencies (FDs)

Let D = (O, *AT*) and $A, B \subseteq AT$.

- $A \rightarrow B$ is defined as a *functional dependency* if $\forall x \in O, [x]_A \subseteq [x]_B$.
- A functional dependency $A \rightarrow B$ is called *minimal* if $\forall C \subset A$, $C \rightarrow B$ is not functional.
- A dependency $A \rightarrow B$ is called *trivial* if $B \subseteq A$.

Example: FDs…

D = (O, *AT*), where O = {1,2,3,4,5,6}, AT = {*abcd*}.

Id	*a*	*b*	*c*	*d*
1	1	1	2	2
2	1	1	2	2
3	2	2	2	2
4	3	3	3	3
5	5	5	6	6
6	5	5	6	6

- {*cd*} → {*ab*} is not a functional dependency.
- {*ab*} → {*cd*} is a functional dependency, but is not minimal.
- {*a*} → {*cd*} is a minimal functional dependency.
- {*ab*} → {*a*} is trivial.

Example: FDs

D = (O, *AT*), where O = {1,2,3,4,5,6}, AT = {*abcd*}.

Id	*e*	*f*	*g*	*h*
1	1	1	1	2
2	1	1	1	2
3	1	1	2	2
4	1	2	3	3
5	5	5	5	5
6	5	5	5	5

- {*e*} → {*h*} is not a functional dependency.
- {*f*} → {*h*} is a minimal functional dependency.
- {*ef*} → {*h*} is a functional dependency, but is not minimal.
- {*efg*} → {*h*} is a functional dependency, but is not minimal.

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Properties of FDs

- **Theorem (Basic rules for the TANE algorithm).** Let $a \in X$ and $b \notin X$.
  - **r1:** If $X \setminus \{a\} \to \{a\}$ is functional, then $\forall Y \supset X$, $Y \setminus \{a\} \to \{a\}$ is a functional, but not minimal, dependency.
  - **r2:** If $X \setminus \{a\} \to \{a\}$ and $X \to \{b\}$ are functional, then $\forall Y \supset X$, $Y \setminus \{b\} \to \{b\}$ is a functional, but not minimal, dependency.

7

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Ilustration of property r1

Theorem. Let $Y \supset X$, $a \in X$ and $b \notin X$.

- **r1:** If $X \setminus \{a\} \to \{a\}$ is functional then $Y \setminus \{a\} \to \{a\}$ is functional, but not minimal.

Example (illustrating a method of deriving rule r1).
Let $X = \{acde\}$ and $Y = \{abcdef\}$. Then, $a \in X$, $b \notin X$, $Y \supset X$.

- **r1:** $X \setminus \{a\} \to \{a\}$ is functional $\Rightarrow$
  $\{cde\} \to \{a\}$ is functional $\Rightarrow$
  $\{bcde\} \to \{a\}$ is functional, but not minimal $\Rightarrow$
  $\{bcdef\} \to \{a\}$ is functional, but not minimal $\Rightarrow$
  $Y \setminus \{a\} \to \{a\}$ is functional, but not minimal.

8

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Proof of Properties of FDs…

- **Theorem (Basic rules for TANE).** Let $a \in X$.
  - **r1:** If $X \setminus \{a\} \to \{a\}$ is functional, then $\forall Y \supset X$, $Y \setminus \{a\} \to \{a\}$ is a functional, but not minimal, dependency.

Proof:

Since $a \in X$, then $Y \supset X \Leftrightarrow Y \setminus \{a\} \supset X \setminus \{a\}$.
Hence, as $X \setminus \{a\} \to \{a\}$ is functional,
then $\forall Y \setminus \{a\} \supset X \setminus \{a\}$, $Y \setminus \{a\} \to \{a\}$ is functional.
Thus, $\forall Y \supset X$, $Y \setminus \{a\} \to \{a\}$ is functional, but not minimal.

9

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Ilustration of property r2

Theorem. Let $Y \supset X$, $a \in X$ and $b \notin X$.

- **r2:** If $X \setminus \{a\} \to \{a\}$ and $X \to \{b\}$ are functional, then $Y \setminus \{b\} \to \{b\}$ is functional, but not minimal.

Example (illustrating a method of deriving rule r2).
Let $X = \{acde\}$ and $Y = \{abcdef\}$. Then, $a \in X$, $b \notin X$, $Y \supset X$.

- **r1:** $X \setminus \{a\} \to \{a\}$ and $X \to \{b\}$ are functional $\Rightarrow$
  $\{cde\} \to \{a\}$ i $\{acde\} \to \{b\}$ are functional $\Rightarrow$
  $\{cde\} \to \{acde\}$ i $\{acde\} \to \{b\}$ are functional $\Rightarrow$
  $\{acde\} \to \{b\}$ is functional, but not minimal $\Rightarrow$
  $\{acdef\} \to \{b\}$ is functional, but not minimal $\Rightarrow$
  $Y \setminus \{b\} \to \{b\}$ j is functional, but not minimal.

10

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Proof of Properties of FDs

- **Theorem (Basic rules for TANE).** Let $Y \supset X$, $a \in X$, $b \notin X$.
  - **r2:** If $X \setminus \{a\} \to \{a\}$ and $X \to \{b\}$ are functional, then $\forall Y \supset X$, $Y \setminus \{b\} \to \{b\}$ is a non-minimal fun. dependency.

Proof:

Since $X \setminus \{a\} \to \{a\}$ and $X \to \{b\}$ are functional dependencies,
then $X \setminus \{a\} \to X$ and $X \to \{b\}$ are functional dependencies,
so $X \to \{b\}$ is a functional dependency, but not minimal.
As $b \notin X$, then $X \setminus \{b\} = X$.
Hence, $X \setminus \{b\} \to \{b\}$ is a non-minimal functional dependency.
Therefore, $\forall Y \supset X$, $Y \setminus \{b\} \to \{b\}$ is a a non-minimal functional dependency.

11

WARSAW UNIVERSITY OF TECHNOLOGY DEVELOPMENT PROGRAMME

Discovering FDs with TANE

- TANE discovers *minimal functional non-trivial dependencies with one dependent attribute*.
- In iteration i, TANE builds candidate dependencies $X \setminus \{a\} \to \{a\}$ from attribute sets X of cardinality i.
- New attribute sets of cardinality $i + 1$ are built from attribute sets of cardinality i in an Apriori-like way.
- Basic rules r1 and r2 are applied to avoid creation of attribute sets of cardinality $i + 1$ from which no minimal functional dependency could be created.

12

Example: FDs…

Id	a	b	c	d
1	1	A	$	F
2	1	A	€	T
3	2	A	$	D
4	2	A	$	F
5	2	B	€	L
6	3	B	$	O
7	3	C	€	F
8	3	C	£	R

i=1	{a}	{b}	{c}	{d}
C+	{abcd}	{abcd}	{abcd}	{abcd}
cand.	∅→{a}	∅→{b}	∅→{c}	∅→{d}

i=2	{ab}	{ac}	{ad}	{bc}	{bd}	{cd}
C+	{abcd}	{abcd}	{abcd}	{abcd}	{abcd}	{abcd}
cand.	{a}→{b}	{a}→{c}	{a}→{d}	{b}→{c}	{b}→{d}	{c}→{d}
	{b}→{a}	{c}→{a}	{d}→{a}	{c}→{b}	{d}→{b}	{d}→{c}

Example: FDs…

Id	a	b	c	d
1	1	A	$	F
2	1	A	€	T
3	2	A	$	D
4	2	A	$	F
5	2	B	€	L
6	3	B	$	O
7	3	C	€	F
8	3	C	£	R

i=2	{ab}	{ac}	{ad}	{bc}	{bd}	{cd}
C+	{abcd}	{abcd}	{abcd}	{abcd}	{abcd}	{abcd}
cand.	{a}→{b}	{b}→{a}	{a}→{d}	{b}→{c}	{b}→{d}	{c}→{d}
	{b}→{a}	{c}→{a}	{d}→{a}	{c}→{b}	{d}→{b}	{d}→{c}

i=3	{abc}	{abd}	{acd}	{bcd}
C+	{abcd}	{abcd}	{abcd}	{abcd}
cand.	{ab}→{c}	{ab}→{d}	{ac}→{d}	{bc}→{d}
	{ac}→{b}	{ad}→{b}	{ad}→{c}	{bd}→{c}
	{bc}→{a}	{bd}→{a}	{cd}→{a}	{cd}→{b}

Example: FDs…

Id	a	b	c	d
1	1	A	$	F
2	1	A	€	T
3	2	A	$	D
4	2	A	$	F
5	2	B	€	L
6	3	B	$	O
7	3	C	€	F
8	3	C	£	R

i=3	{abc}	{abd}	{acd}	{bcd}
C+	{abcd}	{abcd}	{abcd}	{abcd}
cand.	{ab}→{c}	{ab}→{d}	{ac}→{d}	{bc}→{d}
	{ac}→{b}	{ad}→{b}	{ad}→{c}	{bd}→{c}
	{bc}→{a}	{bd}→{a}	{cd}→{a}	{cd}→{b}

delete *b* by rule r1

delete *d* by rule r2

Example: FDs

Id	a	b	c	d
1	1	A	$	F
2	1	A	€	T
3	2	A	$	D
4	2	A	$	F
5	2	B	€	L
6	3	B	$	O
7	3	C	€	F
8	3	C	£	R

i=3	{abc}	{abd}	{acd}	{bcd}
C+	{ac}	{ad}	{ad}	{d}
cand.	{ab}→{c}	{ab}→{d}	{ac}→{d}	{bc}→{d}
	{ac}→{b}	{ad}→{b}	{ad}→{c}	{bd}→{c}
	{bc}→{a}	{bd}→{a}	{cd}→{a}	{cd}→{b}

i=4	{abcd}
C+	{ac}∩{ad} ∩{ad}∩{d} =∅
cand.	-

Example: FDs and Partitions

D = (O, *AT*), where
O = {1,2,3,4,5,6},
AT = {*abcd*}.

Id	a	b	c	d
1	1	1	2	2
2	1	1	2	2
3	2	2	2	2
4	3	3	3	3
5	5	5	6	6
6	5	5	6	6

- $\Pi_{\{ab\}}$ = {{1,2}, {3}, {4}, {5,6}}
- $\Pi_{\{cd\}}$ = {{1,2,3}, {4}, {5,6}}

- {*ab*} → {*cd*} is a functional dependency.
- {*cd*} → {*ab*} is not a functional dependency.

FDs and Partitions

Theorem.

$X \rightarrow Y$ is functional iff $\Pi_X = \Pi_{X \cup Y}$ iff $|\Pi_X| = |\Pi_{X \cup Y}|$.

Id	a	b	c	d
1	1	1	2	2
2	1	1	2	2
3	2	2	2	2
4	3	3	3	3
5	5	5	6	6
6	5	5	6	6

- $\Pi_{\{ab\}}$ = {{1,2}, {3}, {4}, {5,6}}.
- $\Pi_{\{cd\}}$ = {{1,2,3}, {4}, {5,6}}.
- $\Pi_{\{abcd\}}$ = {{1,2}, {3}, {4}, {5,6}}.

Hence:

- $|\Pi_{\{ab\}}| = 4 = |\Pi_{\{abcd\}}|$, so {*ab*} → {*cd*} is a functional dependency.
- $|\Pi_{\{cd\}}| = 3 \neq 4 = |\Pi_{\{abcd\}}|$, so {*cd*} → {*ab*} is not functional.

Calculation of Partitions in TANE

- **Property.**

$$\Pi_{X\cup Y} = \Pi_X \cap \Pi_Y.$$

- In TANE:

If $X = \{x_1, \dots, x_{n-1}, x_n\}$, then

$$\Pi_X = \Pi_{X\setminus\{x_n\}} \cap \Pi_{X\setminus\{x_{n-1}\}}.$$

19

TANE: Determination of Partitions in Detail…

Id	*a*	*b*	*c*	*d*
1	1	A	$	F
2	1	A	€	T
3	2	A	$	D
4	2	A	$	F
5	2	B	€	L
6	3	B	$	O
7	3	C	€	F
8	3	C	£	R

- $\Pi_{\{c\}}$ = {{1,3,4,6}, {2,5,7}, {8}}.
- $\Pi_{\{d\}}$ = {{1,4,7}, {2}, {3} , {5} , {6} , {8}}.
- $\Pi_{\{cd\}}$ = ?

T	1	2	3	4	5	6	7	8
$\Pi_{\{c\}}$	I	II	I	I	II	I	II	III

S	I	II	III					
	1, 4	7						

- $\Pi_{\{cd\}}$ = {{1,4}, {7}

20

TANE: Determination of Partitions in Detail…

Id	*a*	*b*	*c*	*d*
1	1	A	$	F
2	1	A	€	T
3	2	A	$	D
4	2	A	$	F
5	2	B	€	L
6	3	B	$	O
7	3	C	€	F
8	3	C	£	R

- $\Pi_{\{c\}}$ = {{1,3,4,6}, {2,5,7}, {8}}.
- $\Pi_{\{d\}}$ = {{1,4,7}, {2}, {3}, {5}, {6}, {8}}.
- $\Pi_{\{cd\}}$ = ?

T	1	2	3	4	5	6	7	8
$\Pi_{\{c\}}$	I	II	I	I	II	I	II	III

S	I	II	III					
		2						

- $\Pi_{\{cd\}}$ = {{1,4}, {7}, {2}

21

TANE: Determination of Partitions in Detail

Id	*a*	*b*	*c*	*d*
1	1	A	$	F
2	1	A	€	T
3	2	A	$	D
4	2	A	$	F
5	2	B	€	L
6	3	B	$	O
7	3	C	€	F
8	3	C	£	R

- $\Pi_{\{c\}}$ = {{1,3,4,6}, {2,5,7}, {8}}.
- $\Pi_{\{d\}}$ = {{1,4,7}, {2}, {3}, {5}, {6}, {8}}.
- $\Pi_{\{cd\}}$ = ?

T	1	2	3	4	5	6	7	8
$\Pi_{\{c\}}$	I	II	I	I	II	I	II	III

S	I	II	III					
			8					

- $\Pi_{\{cd\}}$ = {{1,4}, {7}, {2}, {3}, {5}, {6}, {8}}

22

TANE: Alternative Determination of Partitions in Detail…

Id	*a*	*b*	*c*	*d*
1	1	A	$	F
2	1	A	€	T
3	2	A	$	D
4	2	A	$	F
5	2	B	€	L
6	3	B	$	O
7	3	C	€	F
8	3	C	£	R

- $\Pi_{\{d\}}$ = {{1,4,7}, {2}, {3}, {5}, {6}, {8}}.
- $\Pi_{\{c\}}$ = {{1,3,4,6}, {2,5,7}, {8}}.
- $\Pi_{\{cd\}}$ = ?

T	1	2	3	4	5	6	7	8
$\Pi_{\{d\}}$	I	II	III	I	IV	V	I	VI

S	I	II	III	IV	V	VI		
	1, 4		3		6			

- $\Pi_{\{cd\}}$ = {{1,4}, {3}, {6}

23

TANE: Alternative Determination of Partitions in Detail…

Id	*a*	*b*	*c*	*d*
1	1	A	$	F
2	1	A	€	T
3	2	A	$	D
4	2	A	$	F
5	2	B	€	L
6	3	B	$	O
7	3	C	€	F
8	3	C	£	R

- $\Pi_{\{d\}}$ = {{1,4,7}, {2}, {3}, {5}, {6}, {8}}.
- $\Pi_{\{c\}}$ = {{1,3,4,6}, {2,5,7}, {8}}.
- $\Pi_{\{cd\}}$ = ?

T	1	2	3	4	5	6	7	8
$\Pi_{\{d\}}$	I	II	III	I	IV	V	I	VI

S	I	II	III	IV	V	VI		
	7	2		5				

- $\Pi_{\{cd\}}$ = {{1,4}, {3}, {6}, {2}, {5}, {7}

24

TANE: Alternative Determination of Partitions in Detail

Id	*a*	*b*	*c*	*d*
1	1	A	$	F
2	1	A	€	T
3	2	A	$	D
4	2	A	$	F
5	2	B	€	L
6	3	B	$	O
7	3	C	€	F
8	3	C	£	R

- $\Pi_{\{d\}} = \{\{1,4,7\}, \{2\}, \{3\}, \{5\}, \{6\}, \{8\}\}$.
- $\Pi_{\{c\}} = \{\{1,3,4,6\}, \{2,5,7\}, \{8\}\}$.
- $\Pi_{\{cd\}}$ = ?

T	1	2	3	4	5	6	7	8
$\Pi_{\{d\}}$	I	II	III	I	IV	V	I	VI

S	I	II	III	IV	V	VI		
						8		

- $\Pi_{\{cd\}} = \{\{1,4\}, \{3\}, \{6\}, \{2\}, \{5\}, \{7\}, \{8\}\}$

25

Stripped Partitions

- π_X is called a *stripped partition* if it contains all and only non-singleton blocks in Π_X.
- **Example.**
  For $\Pi_{\{ab\}} = \{\{1,2\}, \{3\}, \{4\}, \{5,6\}\}$,
  $\pi_{\{ab\}} = \{\{1,2\}, \{5,6\}\}$.

26

FunctionaL Dependencies and Stripped Partitions

Theorem. $X \rightarrow Y$ is a functional dependency $\Leftrightarrow$

$|\Pi_X| = |\Pi_{X\cup Y}| \Leftrightarrow ||\pi_X|| - |\pi_X| = ||\pi_{X\cup Y}|| - |\pi_{X\cup Y}|$.

Proof: Let:
- $|D|$ - denote the number of objects,
- n – denote the number of singleton blocks in Π_X.

Then, $|\Pi_X| = |\pi_X| + n$ and $n = |D| - ||\pi_X||$.

Thus, $|\Pi_X| = |\pi_X| + |D| - ||\pi_X||$.

and $|\Pi_{X\cup Y}| = |\pi_{X\cup Y}| + |D| - ||\pi_{X\cup Y}||$.

Hence, $|\Pi_X| = |\Pi_{X\cup Y}| \Leftrightarrow$

$|\pi_X| + |D| - ||\pi_X|| = |\pi_{X\cup Y}| + |D| - ||\pi_{X\cup Y}|| \Leftrightarrow$

$||\pi_X|| - |\pi_X| = ||\pi_{X\cup Y}|| - |\pi_{X\cup Y}|$. □

27

FDs and Stripped Partitions

Theorem. $X \rightarrow Y$ is functional iff

$||\pi_X|| - |\pi_X| = ||\pi_{X\cup Y}|| - |\pi_{X\cup Y}|$, where $||\pi_X|| = \Sigma_{c\in\pi_X}|c|$.

Id	*a*	*b*	*c*	*d*
1	1	1	2	2
2	1	1	2	2
3	2	2	2	2
4	3	3	3	3
5	5	5	6	6
6	5	5	6	6

- $\pi_{\{ab\}} = \{\{1,2\}, \{5,6\}\}$, $\pi_{\{cd\}} = \{\{1,2,3\}, \{5,6\}\}$.
- $\pi_{\{abcd\}} = \{\{1,2\}, \{5,6\}\}$.

Hence:
- $||\pi_{\{ab\}}|| - |\pi_{\{ab\}}| = 4 - 2 = ||\Pi_{\{abcd\}}|| - |\Pi_{\{abcd\}}|$,
  so $\{ab\} \rightarrow \{cd\}$ is functional.
- $||\pi_{\{cd\}}|| - |\pi_{\{cd\}}| = 5 - 2 \neq$
  $4 - 2 = ||\Pi_{\{abcd\}}|| - |\Pi_{\{abcd\}}|$,
  so $\{cd\} \rightarrow \{ab\}$ is not functional.

28

TANE: Determination of Stripped Partitions in Detail…

Id	*a*	*b*	*c*	*d*
1	1	A	$	F
2	1	A	€	T
3	2	A	$	D
4	2	A	$	F
5	2	B	€	L
6	3	B	$	O
7	3	C	€	F
8	3	C	£	R

- $\Pi_{\{d\}} = \{\{1,4,7\}, \{2\}, \{3\}, \{5\}, \{6\}, \{8\}\}$.
- $\Pi_{\{c\}} = \{\{1,3,4,6\}, \{2,5,7\}, \{8\}\}$.
- $\pi_{\{d\}} = \{\{1,4,7\}\}$.
- $\pi_{\{c\}} = \{\{1,3,4,6\}, \{2,5,7\}\}$.
- $\pi_{\{cd\}}$ = ?

T	1	2	3	4	5	6	7	8
$\pi_{\{d\}}$	I			I			I	

S	I							
	1, 4							

- $\pi_{\{cd\}} = \{\{1,4\}$

29

TANE: Determination of Stripped Partitions in Detail

Id	*a*	*b*	*c*	*d*
1	1	A	$	F
2	1	A	€	T
3	2	A	$	D
4	2	A	$	F
5	2	B	€	L
6	3	B	$	O
7	3	C	€	F
8	3	C	£	R

- $\Pi_{\{d\}} = \{\{1,4,7\}, \{2\}, \{3\}, \{5\}, \{6\}, \{8\}\}$.
- $\Pi_{\{c\}} = \{\{1,3,4,6\}, \{2,5,7\}, \{8\}\}$.
- $\pi_{\{d\}} = \{\{1,4,7\}\}$.
- $\pi_{\{c\}} = \{\{1,3,4,6\}, \{2,5,7\}\}$.
- $\pi_{\{cd\}}$ = ?

T	1	2	3	4	5	6	7	8
$\pi_{\{d\}}$	I			I			I	

S	I							
	7							

- $\pi_{\{cd\}} = \{\{1,4\}\}$

30

Error of a Dependency

- $err(X \rightarrow Y) =$

$$\Sigma_{c \in \Pi_X} (|c| - \max\{|c'| \; c' \subseteq c \wedge c' \in \Pi_{X \cup Y}\}).$$

Id	*fName*	*gender*
1	Mary	F
2	Mary	F
3	Mary	M
4	Chris	M
5	Chris	M
6	Chris	F
7	Charlie	M
8	Charlie	F
9	Rob	M

$\Pi_{\{fName\}} = \{\{1,2,3\}, \{4,5,6\}, \{7,8\}, \{9\}\}.$

$\Pi_{\{fName,gender\}} = \{\{1,2\},\{3\},\{4,5\},\{6\},\{7\},\{8\},\{9\}\}.$

$err(fName \rightarrow gender) =$
$(|\{1,2,3\}| - \max\{|\{1,2\}|, |\{3\}|\}) +$
$(|\{4,5,6\}| - \max\{|\{4,5\}|, |\{6\}|\}) +$
$(|\{7,8\}| - \max\{|\{7\}|, |\{8\}|\}) +$
$(|\{9\}| - \max\{|\{9\}|\}) =$
$(3-2) + (3-2) + (2-1) + (1-1) = 3$

31

Approximate Dependencies (ADs)

- $X \rightarrow Y$ is defined as an *approximate dependency* if $err(X \rightarrow Y) \leq \varepsilon$.

Id	*fName*	*gender*
1	Mary	F
2	Mary	F
3	Mary	M
4	Chris	M
5	Chris	M
6	Chris	F
7	Charlie	M
8	Charlie	F
9	Rob	M

Let $\varepsilon = 3$. Then:
fName→gender is an approximate dependency as $err(fName \rightarrow gender) = 3 \leq \varepsilon$.

32

Error of a Dependency in Terms of Stripped Partitions

- **Property.** $err(X \rightarrow Y) =$

$$\Sigma_{c \in \pi_X} (|c| - \max(\{|c'| \; c' \subseteq c \wedge c' \in \pi_{X \cup Y}\} \cup \{1\})).$$

Id	*fName*	*gender*
1	Mary	F
2	Mary	F
3	Mary	M
4	Chris	M
5	Chris	M
6	Chris	F
7	Charlie	M
8	Charlie	F
9	Rob	M

$\pi_{\{fName\}} = \{\{1,2,3\}, \{4,5,6\}, \{7,8\}\}.$

$\pi_{\{fName,gender\}} = \{\{1,2\}, \{4,5\}\}.$

$err(fName \rightarrow gender) =$
$(|\{1,2,3\}| - \max\{|\{1,2\}|, 1\}) +$
$(|\{4,5,6\}| - \max\{|\{4,5\}|, 1\}) +$
$(|\{7,8\}| - \max\{1\}) =$
$(3-2) + (3-2) + (2-1) = 3$

33

Adapting TANE to Calculation of ADs

- Checking if $X \rightarrow Y$ is functional is replaced by checking if $err(X \rightarrow Y) \leq \varepsilon$.
- Rule r2 is used only when $err(X \rightarrow Y) = 0$; that is, when $X \rightarrow Y$ is functional.

34

References

– Ykä Huhtala, Juha Kärkkäinen, Pasi Porkka, Hannu Toivonen: TANE: An Efficient Algorithm for Discovering Functional and Approximate Dependencies. Comput. J. 42(2): 100-111 (1999)

35