

## MultiRelational $k$ -Anonymity

M. Ercan Nergiz & Chris Clifton & A. Erhan Nergiz

### Single Table $k$ -Anonymity

- A table is  $k$ -anonymous if every tuple over the identifying attributes (age, sex, address, ...) occurs at least  $k$  times
- $k$ -Anonymity protects sensitive info (GPA) against adversaries asking:

"I know that Chris is 'male', '19' years old and from 'W.Lafayette'. Then what is his GPA?"

#### Original Dataset

Age	Sex	Address	GPA
19	M	W. Lafayette	3.72
18	M	Lafayette	2.34
23	F	Lafayette	3.12
25	M	Indianapolis	4.00

#### 2-Anonymized Dataset

Age	Sex	Address	GPA
10-20	M	G. Lafayette	3.72
10-20	M	G. Lafayette	2.34
20-30	*	Indiana	3.12
20-30	*	Indiana	4.00

### MultiRelational $k$ -Anonymity

- A database is  $k$ -anonymous if every query over the identifying attributes (course, book) returns at least  $k$  tuples
- MultiR Anonymity protects sensitive info (GPA, Grade, Price) against adversaries asking:

"I know that Chris is taking 'Math', 'Physics', and 'History' courses; he is using 'Discrete' book for 'Math', using 'Calculus', 'Dynamics' books for 'Physics', ...What is his GPA, grades and prices for books?"

#### MultiR Database

Sid	GPA
S1	3.72
S2	2.34
S3	3.12
S4	4.00

SCid	Sid	Course	Grade
SC1	S1	Math	93
SC2	S1	Physics	91
SC3	S1	History	85
SC4	S2	CS	78
SC5	S2	Physics	62
SC6	S2	Religion	42
SC7	S3	History	85
SC8	S3	Religion	75
SC9	S3	Physics	77
SC10	S4	History	98
SC11	S4	Religion	96

SCid	Book	Price
SC1	Discrete	\$63
SC2	Calculus	\$89
SC2	Dynamics	\$42
SC3	Relg. H.	\$33
SC4	Discrete	\$65
SC5	Dynamics	\$51
SC6	Yodaism	\$38
SC7	Ottomans	\$49
SC8	Yodaism	\$39
SC9	Calculus	\$84
SC10	Am. Hist	\$54

Single table algorithms either fail to protect privacy, or overly reduce the utility of the data, in a multiR setting.

### Single Table Approach: Bitmap Anonymizations

#### Bitmap Format

Sid	Math		Physics		CS		History			Religion		GPA	
	T	Di	T	Ca	T	Di	T	RH	Ot	AH	T		Yo
S1	1	1	1	1	0	0	1	1	0	0	0	0	3.72
S2	0	0	1	0	1	1	0	0	0	0	1	1	2.34
S3	0	0	1	1	0	0	1	0	1	0	1	1	3.12
S4	0	0	0	0	0	0	1	0	0	1	1	0	4.00

#### 2-Anonymized Bitmap

S1	*	*	1	*	1	*	*	*	*	0	0	*	*	3.72
S2	*	*	1	*	1	*	*	*	*	0	0	*	*	2.34
S3	0	0	*	*	0	0	0	1	0	*	*	1	*	3.12
S4	0	0	*	*	0	0	0	1	0	*	*	1	*	4.00

### MultiRelational CLustEring (MiRaCle) Anonymization Algorithm

- Clustering based anonymity algorithm for snowflake databases
- Every private entity in snowflake db can be abstracted by a tree.
- Anonymize two trees:
  - Anonymize the roots.
  - Make pairings among siblings.
  - Recurse to anonymize subtrees rooted by paired siblings.
- Distance of two entities (trees) from each other can be calculated as the cost of the anonymization.

