Discrimination- and Privacy-aware Data Mining

Sara Hajian
Eurecat Technology Center, Yahoo! labs, Barcelona, Spain

FAT ML 2015

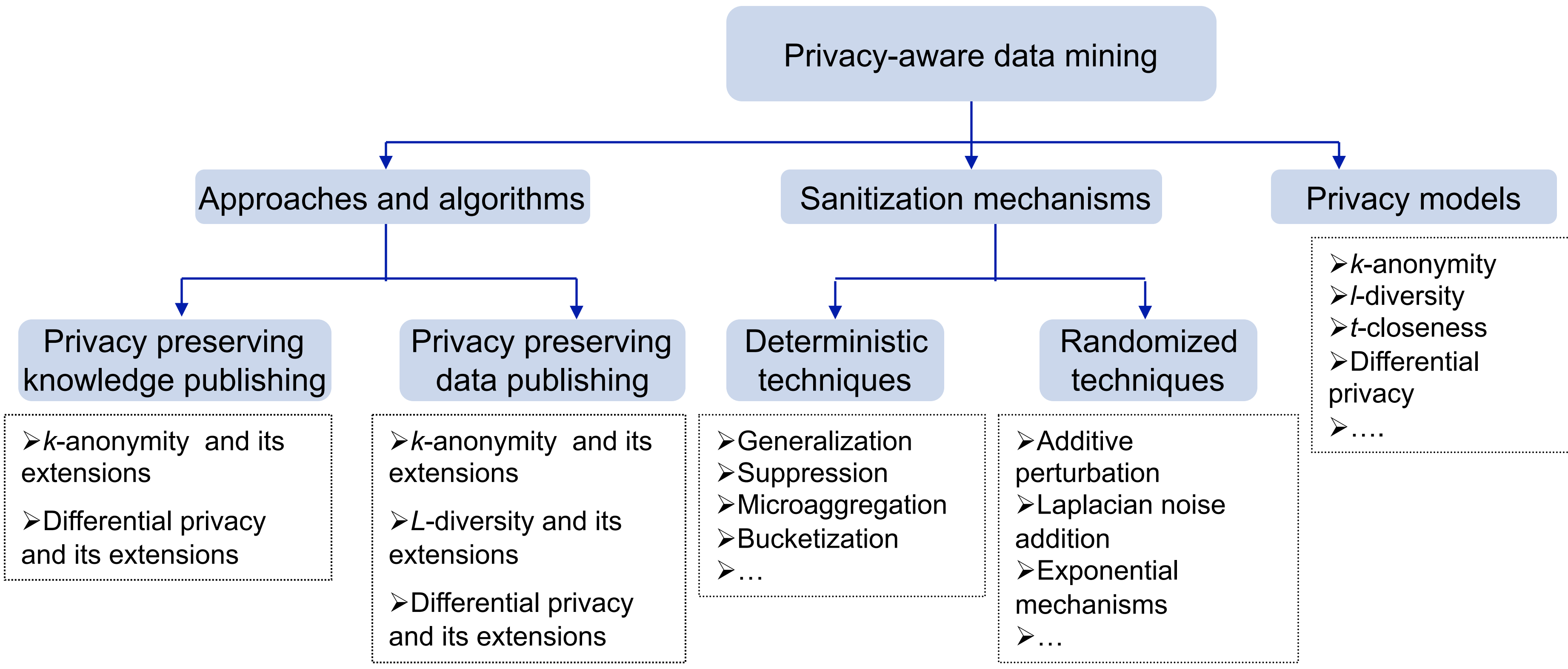
Motivation

- Google's algorithm shows prestigious job ads to men, but not to women.
- Flicker debuted image recognition tools in May, users noticed the tool sometimes tagged black people as "apes" or "animals".
- Google image search for "C.E.O." produced 11 percent women, even though 27 percent of United States chief executives are women.

Outline

- A framework for direct and indirect discrimination prevention in data mining
- Simultaneous discrimination prevention and privacy protection
  - Data publishing
  - Pattern publishing

Privacy-aware data mining (PADM)



Discrimination-aware data mining (DADM)

Discrimination-aware data mining

Discrimination prevention

Challenging task: Automatic decision-making software and recommendation systems

Pre-processing approach

Controlled distortion of the training set

- Direct discrimination
- Genuine occupational requirement

Kamiran and Calders KAIS 2011

Zliobaitye et al. ICDM 2012

In-processing approach

Modify the classification learning algorithm, by integrating it with anti-discrimination criteria

- Naïve Bayes models
- Decision tree induction
- Logistic regression

Calders and Verwer DAMI 2010

Kamiran et al. ICDM 2010

Kamishima et al. PKDD 2012

Post-processing approach

Post-process the classification model once it has been extracted

- Alter the confidence of classification rules
- Probabilities of a naïve Bayes model
- Re-label the class predicted at the leaves of a decision tree

Pedreschi et al. SDM 2009

Calders and Verwer DAMI 2010

Kamian et al. ICDM 2010

Discrimination discovery

Difficult task: Control authority

- Legal principle of under-representation
- Legal experimental procedure of situation testing
- Exploratory DADM

Pedreschi et al. KDD 2008

Luong et al. KDD 2011

Berendt and Preibusch ICDM 2012

On the relationship between PPDM and DPDM

- Privacy and anti-discrimination are two intimately intertwined concepts:
  - Share common challenges
  - Share common methodological problems to be solved
  - In certain contexts, directly interact with each other.

PPDM	DPDM
Measuring disclosure risk	Measuring potential discrimination
Data/model anonymization to protect privacy	Data/model transformation to prevent discrimination
Measuring data/model utility	Measuring data/model utility
.....	....

Open problems

- There is an evident gap between the large body of research in PPDM and the recent early results in DPDM.
- Research questions
  - Can we adapt and use some of the approaches from PPDM for DPDM?
  - What is the relationship between PPDM and DPDM?
  - Is it enough to tackle only privacy or discrimination?
    - If not, how can we design a holistic method capable of addressing both threats together in significant data mining processes?
- Need for simultaneous privacy and anti-discrimination by design.

- **A Methodology for Direct and Indirect Discrimination Prevention in Data Mining**

Sara Hajian[1]
Josep Domingo-Ferrer[2]

[1,2] Universitat Rovira I Virgili, Spain

IEEE Transactions on Knowledge and Data Engineering, 25(7): 1445-1459, 2013.

Roadmap: DADM

Discrimination-aware data mining

Discrimination prevention

Challenging task: Automatic decision making software and recommendation systems

Pre-processing approach

Controlled distortion of the training set

- Direct and indirect discrimination
- Genuine occupational requirement

Kamiran and Calders KAIS 2011

Zliobaitye et al. ICDM 2012

In-processing approach

Modify the classification learning algorithm, by integrating it with anti-discrimination criteria

- Naïve Bayes models
- Decision tree induction
- Logistic regression

Calders and Verwer DAMI 2010

Kamiran et al. ICDM 2010

Kamishima et al. PKDD 2012

Post-processing approach

Post-process the classification model once it has been extracted

- Alter the confidence of classification rules
- Probabilities of a naïve Bayes model
- Re-label the class predicted at the leaves of a decision tree

Pedreschi et al. SDM 2009

Calders and Verwer DAMI 2010

Kamian et al. ICDM 2010

Discrimination discovery

Difficult task: Control authority

- Legal principle of under-representation
- Legal experimental procedure of situation testing
- Direct and indirect discrimination

Pedreschi et al. KDD 2008

Luong et al. KDD 2011

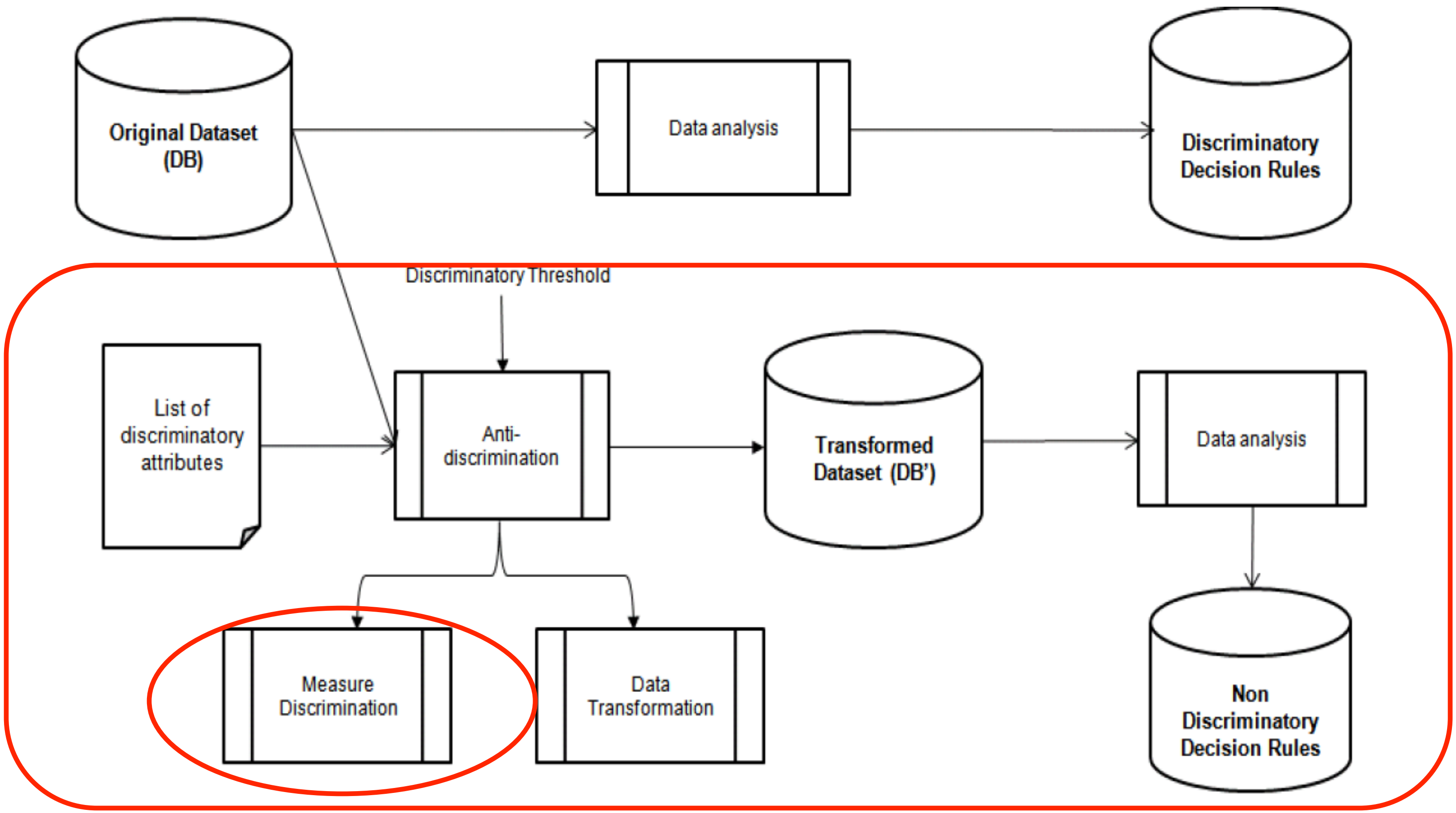
Berendt and Preibusch ICDM 2012

A framework for direct and indirect discrimination prevention in data mining

- Limitations of previous approaches
  - Prevent only direct discrimination
  - Deal with only one protected group
  - Deal with discrimination only at top level while discrimination may occur in some subsets

- We propose new utility metrics
  - Discrimination removal

A framework for direct and indirect discrimination prevention in data mining



Measures of discrimination

- On the legal side, different measures are adopted worldwide.

> ➢ ***Selection lift (slift)*** is the ratio of the proportions of benefit denial between the protected and unprotected groups in the given context.
>
> ➢ ***Election lift (elift)*** is the ratio of the proportions of benefit denial between the protected groups and all people who were not granted the benefit in the given context.

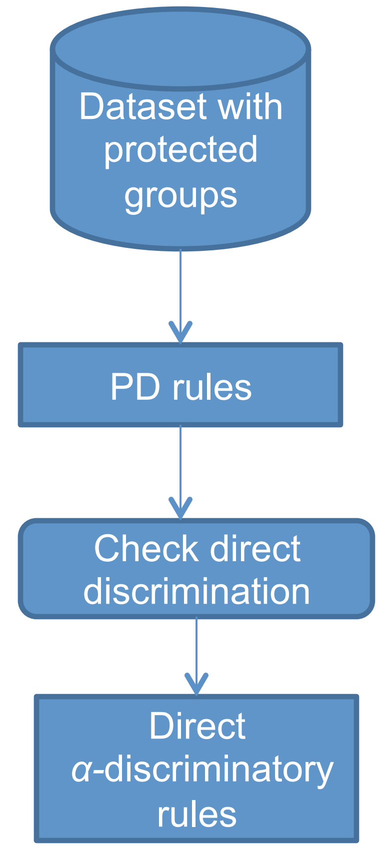
[Pedreschi2009] D. Pedreschi, S. Ruggieri and F. Turini. Measuring discrimination in socially-sensitive decision records. In SDM 2009, pp. 581-592. SIAM, 2009.

Direct discrimination measurement

- The purpose of direct discrimination discovery is to identify α-discriminatory rules that are directly inferred from protected groups

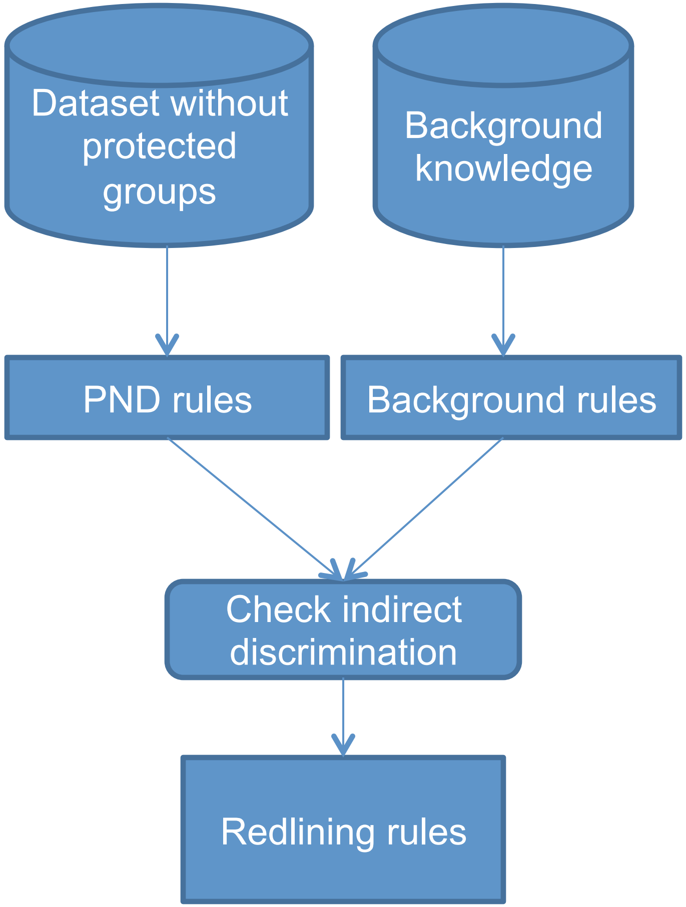
- Note that α states an acceptable level of discrimination according to laws and regulations
  - *e.g.* U.S. Equal Pay Act: This amounts to using *slift* with = 1.25.

- Based on direct discriminatory measures *f*, a PD classification rule *r* is:
  - α-discriminatory if $f(r) \geq \alpha$; *or*
  - α-protective if $f(r) < \alpha$



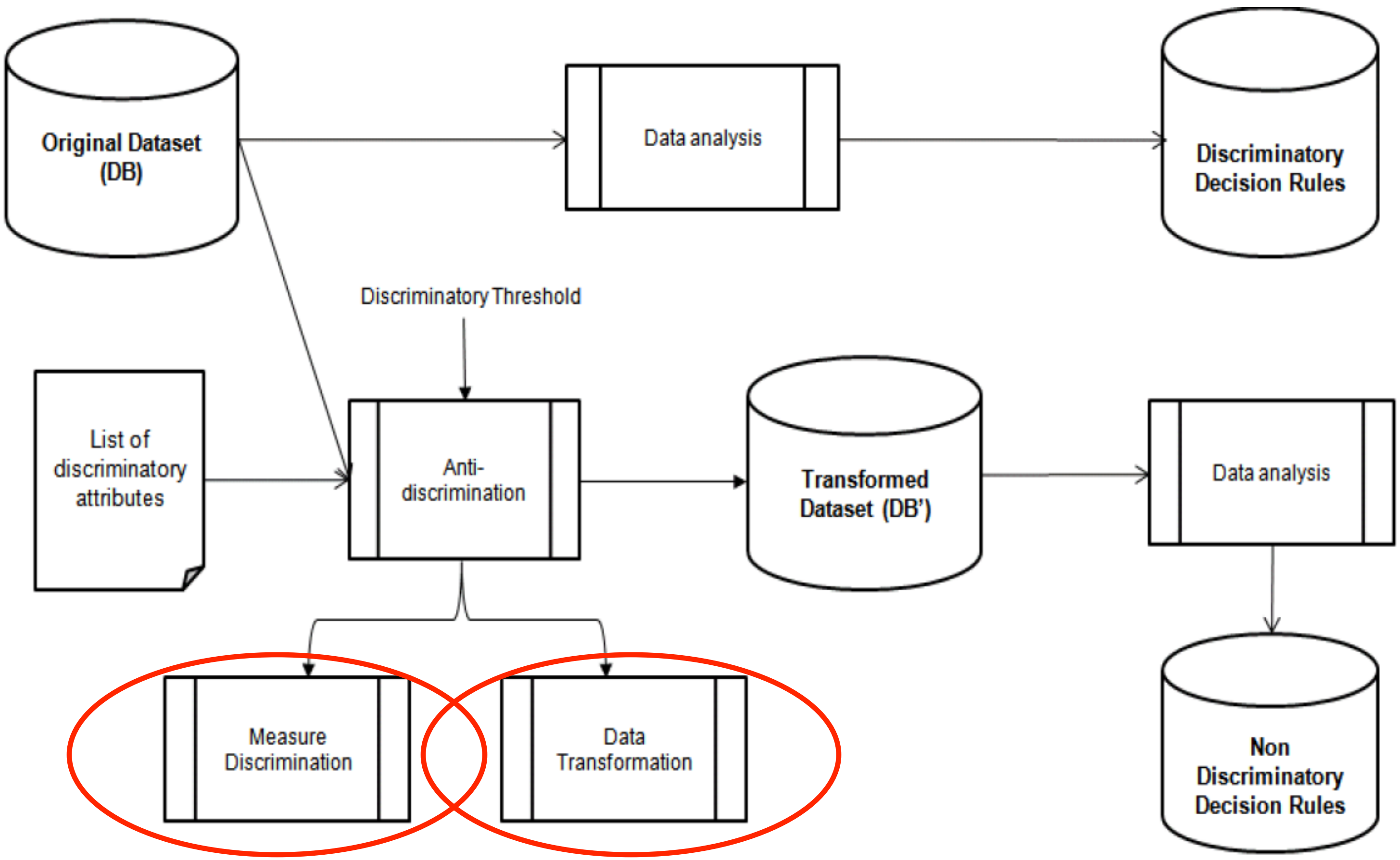
Indirect discrimination measurement

- The purpose of indirect discrimination discovery is to indicate α-discriminatory rules that are indirectly inferred from non-protected groups
  - e.g. Zip = 10451
- Based on indirect discriminatory measures *elb* [Pedreschi2009], a PND classification rule *r* is:
  - *Redlining*
    - e.g., {Zip=10451, City=NYC} → Hire=No.
  - *Non-redlining* or legitimate



[Pedreschi2009] D. Pedreschi, S. Ruggieri and F. Turini. Measuring discrimination in socially-sensitive decision records. In SDM 2009, pp. 581-592. SIAM, 2009.

A framework for direct and indirect discrimination prevention in data mining



Data transformation

- The purpose is transform the original data *D* in such a way to remove direct and/or indirect discriminatory biases, with minimum impact
  - On the data, and
  - On legitimate decision rules

- We have developed metrics that specify
  - Which records (and in which order) should be changed?
  - How many records should be changed?
  - How those records should be changed during data transformation?

Data transformation for direct discrimination prevention

- Direct rule protection (DRP)
  - A suitable data transformation with minimum information loss to make each α-discriminatory rule α-protective.

Data transformation methods for direct rule protection

	Direct Rule Protection
DTM 1	$\neg A, B \rightarrow \neg C \Rightarrow A, B \rightarrow \neg C$
DTM 2	$\neg A, B \rightarrow \neg C \Rightarrow \neg A, B \rightarrow C$

Data transformation for direct discrimination prevention

- Rule generalization (DRP)
  - A suitable data transformation with minimum information loss to make each α-discriminatory rule an instance of a non-redlining PND rule.

Data transformation method for rule generalization

	Rule Generalization
DTM	$A, B, \neg D \rightarrow C \Rightarrow A, B, \neg D \rightarrow \neg C$

Data transformation for indirect discrimination prevention

- Indirect rule protection (IRP)
  - A suitable data transformation with minimum information loss to make each redlining rule non-redlining.

Data transformation methods for indirect rule protection

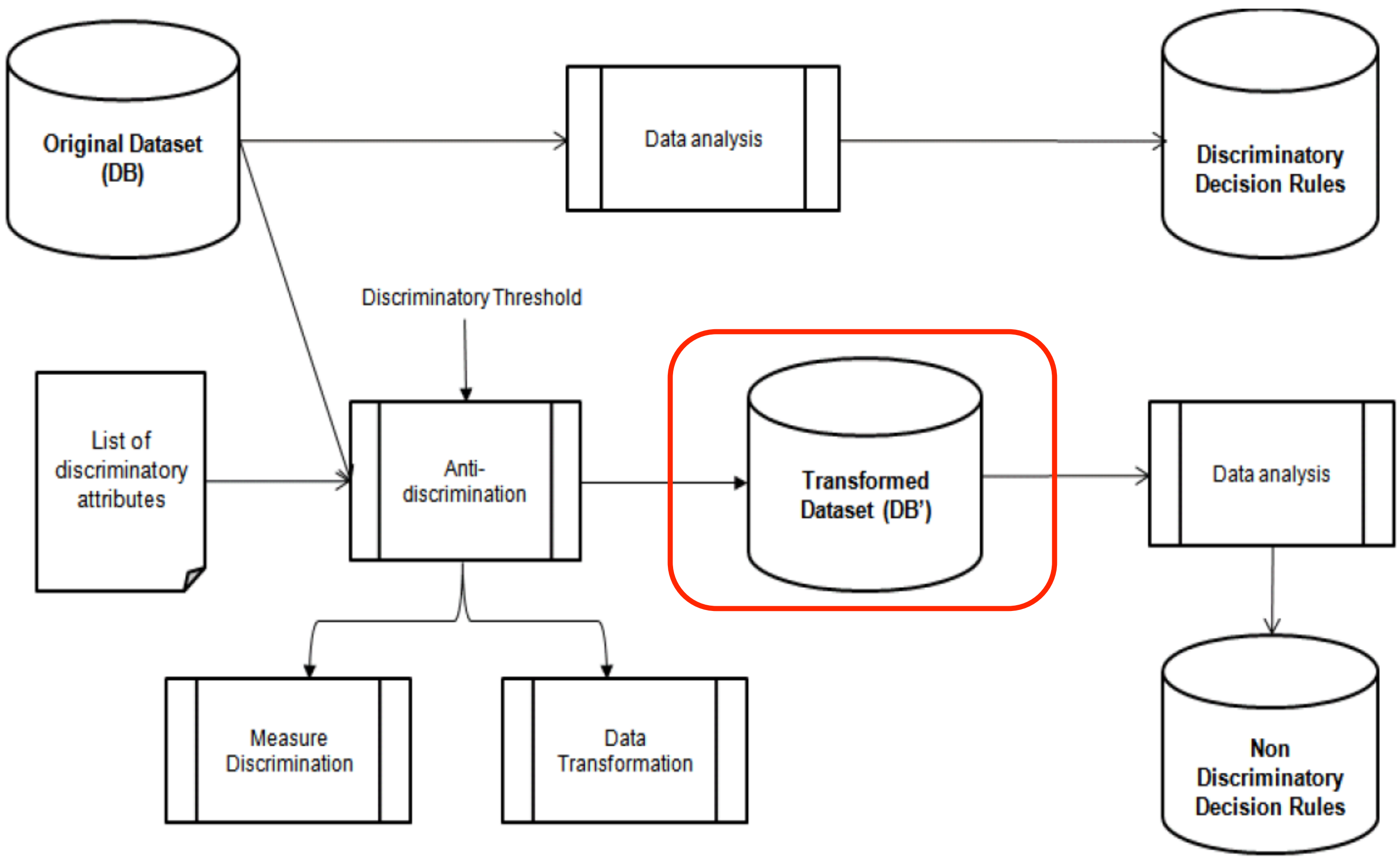
	Indirect Rule Protection
DTM 1	$\neg A, B, \neg D \rightarrow \neg C \Rightarrow A, B, \neg D \rightarrow \neg C$
DTM 2	$\neg A, B, \neg D \rightarrow \neg C \Rightarrow \neg A, B, \neg D \rightarrow C$

Data transformation for both direct and indirect discrimination

- Direct and indirect rule protection
  - Lemma. Method 2 for IRP is beneficial for Method 2 for DRP. On the other hand, Method 2 for DRP is at worst neutral for Method 2 for IRP.

	Method 1	Method 2
Direct Rule Protection	$\neg A, B \to \neg C \Rightarrow A, B \to \neg C$	$\neg A, B \to \neg C \Rightarrow \neg A, B \to C$
Indirect Rule Protection	$\neg A, B, \neg D \to \neg C \Rightarrow A, B, \neg D \to \neg C$	$\neg A, B, \neg D \to \neg C \Rightarrow \neg A, B, \neg D \to C$

A framework for direct and indirect discrimination prevention in data mining



Experiments

- Discrimination removal
- Data quality

Utility measures

- Measuring direct/indirect discrimination removal
  - **Direct/indirect discrimination prevention degree (DDPD)** quantifies the percentage of *α-discriminatory* rules that are no longer *α-discriminatory* in the transformed dataset.
  - **Direct/indirect discrimination protection preservation (DDPP)** quantifies the percentage of the *α-protective* rules in the original dataset that remain *α-protective* in the transformed dataset.

Utility measures

- Measuring Data Quality

  - **Misses Cost** (MC) quantifies the percentage of original rules that cannot be extracted from the transformed dataset.

  - **Ghost Cost** (GC) quantifies the percentage of the rules that were not extractable from the original dataset.

Datasets

- Adult dataset
  - Number of records: 48,842
    - **“*train*” part with 32,561 records**
    - “*test*” part with 16,281 records
  - Number of attributes: 14 attributes (without class attribute)

- German Credit dataset
  - Number of records: 1,000 records
  - Number of attributes: 20 attributes (without class attribute)

Empirical evaluation

- Adult dataset for minimum support 2% and confidence 10%
  - Protected groups ={Sex=Female, Age=Young}

Methods	α	p	No. Redlining Rules	No. Indirect α-Disc. Rules	No. Direct α-Disc. Rules	Discrimination Removal Direct DDPD	Direct DDPP	Indirect IDPD	Indirect IDPP	Data Quality MC	Data Quality GC
Removing. Disc. Attributes	n.a.	n.a	n.a.	n.a	n.a.	n.a.	n.a.	n.a.	n.a.	66.08	0
DRP (Method 1)	1.2	n.a	n.a.	n.a	274	100	100	n.a.	n.a.	4.16	4.13
DRP (Method 2)	1.2	n.a.	n.a.	n.a.	274	100	100	n.a.	n.a.	0	0
DRP (Method 1) + RG	1.2	0.9	n.a.	n.a.	274	100	100	n.a.	n.a.	4.1	4.1
DRP (Method 2) + RG	1.2	0.9	n.a.	n.a.	274	91.58	100	n.a.	n.a.	0	0
IRP (Method 1)	1.1	n.a.	21	30	n.a.	n.a.	n.a.	100	100	0.54	0.38
IRP (Method 2)	1.1	n.a.	21	30	n.a.	n.a.	n.a.	100	100	0	0
DRP(Method 2) + IRP(Method 2)	1.1	n.a.	21	30	280	100	100	100	100	0	0
No of Freq. Class. Rules: 5,092						No. of Back. Know. Rules: 2089					

1) We get very good results for all methods in terms of discrimination removal.
2) In terms of data quality, the best results for direct discrimination prevention are obtained with Method 2 for DRP or Method 2 for DRP combined with Rule Generalization.
3) The best results for indirect discrimination prevention are obtained with Method 2 for IRP.

Empirical evaluation

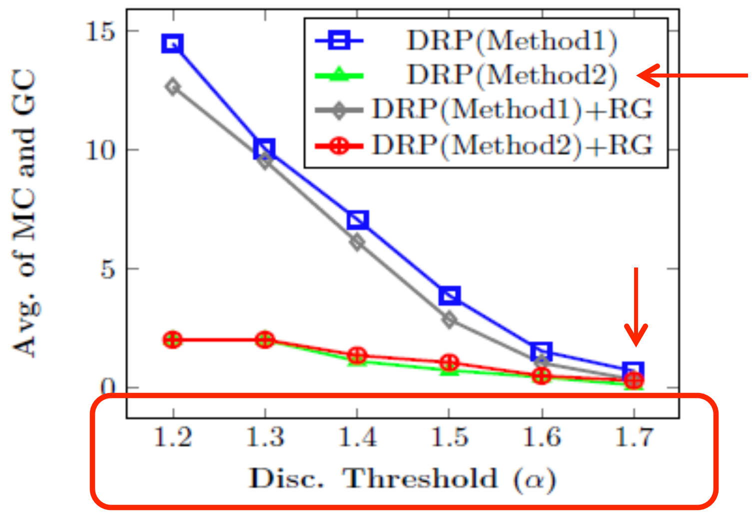
- German credit dataset for minimum support 5% and confidence 10%
- Protected groups = {Personal Status=Female and not Single, Age=Old, Foreign worker=Yes}

Methods	α	p	No. Redlining Rules	No. Indirect α-Disc. Rules	No. Direct α-Disc. Rules	Discrimination Removal Direct DDPD	Discrimination Removal Direct DDPP	Discrimination Removal Indirect IDPD	Discrimination Removal Indirect IDPP	Data Quality MC	Data Quality GC
Removing. Disc. Attributes	n.a.	n.a	n.a.	n.a	n.a.	n.a.	n.a.	n.a.	n.a.	64.35	0
DRP (Method 1)	1.2	n.a	n.a.	n.a	991	100	100	n.a.	n.a.	15.44	13.52
DRP (Method 2)	1.2	n.a.	n.a.	n.a.	991	100	100	n.a.	n.a.	0	4.06
DRP (Method 1) + RG	1.2	0.9	n.a.	n.a.	991	100	100	n.a.	n.a.	13.34	12.01
DRP (Method 2) + RG	1.2	0.9	n.a.	n.a.	991	100	100	n.a.	n.a.	0.01	4.06
IRP (Method 1)	1	n.a.	37	42	n.a.	n.a.	n.a.	100	100	1.62	1.47
IRP (Method 2)	1	n.a.	37	42	n.a.	n.a.	n.a.	100	100	0	0.96
DRP(Method 2) + IRP(Method 2)	1	n.a.	37	42	499	99.97	100	100	100	0	2.07
No of Freq. Class. Rules: 32,340						No. of Back. Know. Rules: 22,763					

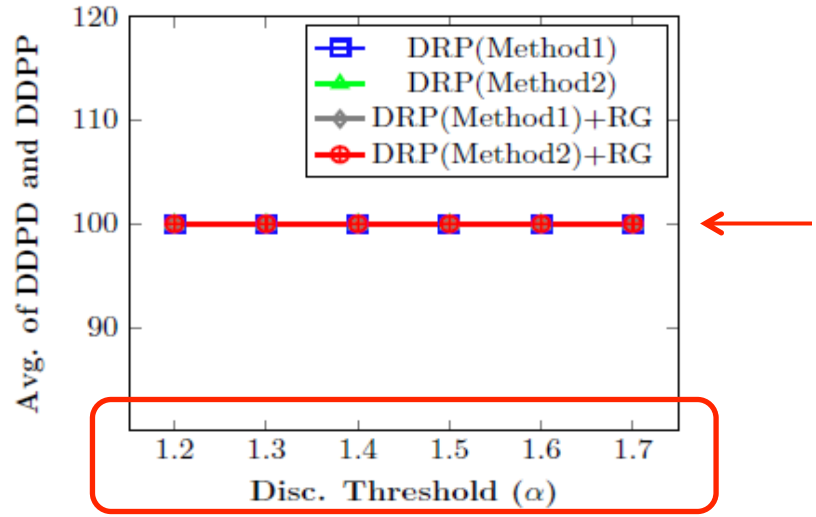
We obtained lower information loss in terms of MC and GC in the Adult dataset than in the German Credit dataset.

Empirical evaluation

- German Credit dataset
- Direct discrimination prevention
  - Information loss

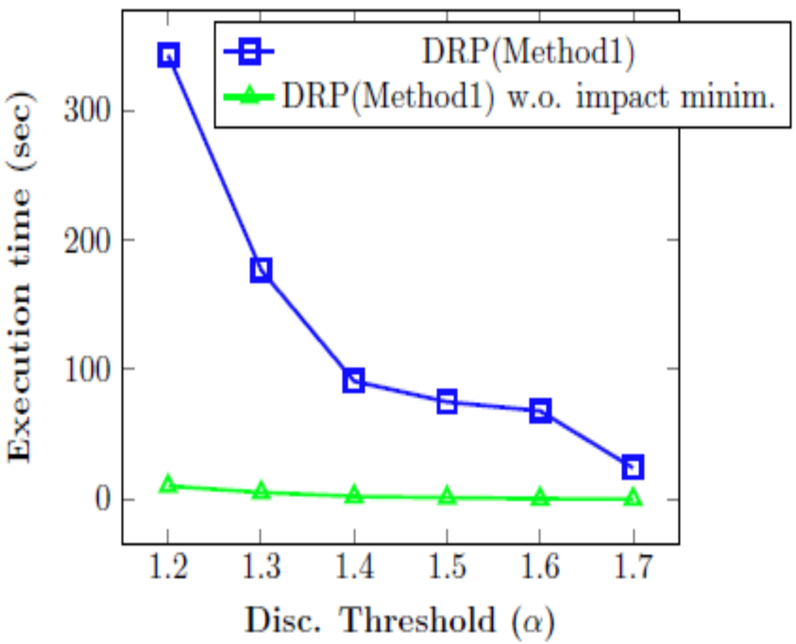


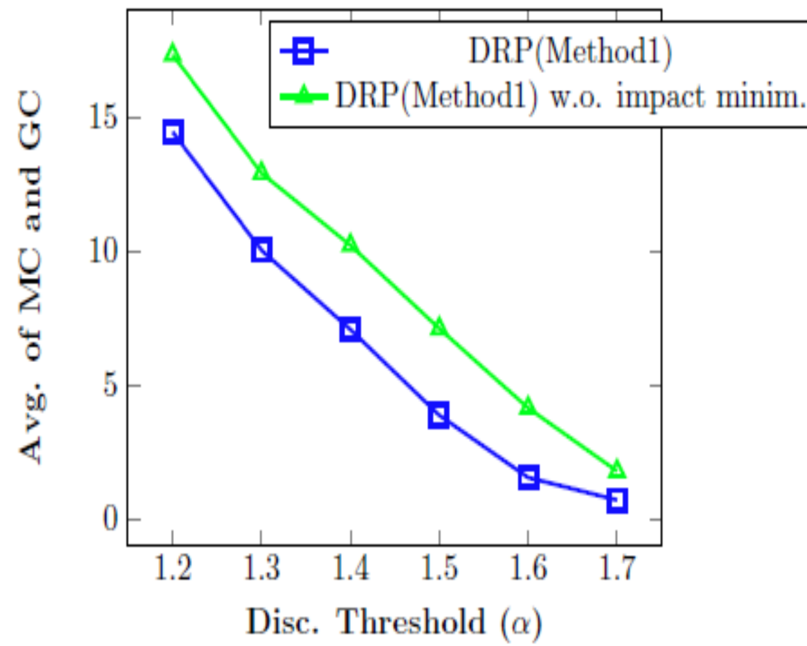
  - Discrimination removal degree



Empirical evaluation

- German Credit dataset
- Impact minimization procedure
  - Execution times
  - Information loss degree





Impact minimization procedure substantially increases the execution time of the algorithm.
Impact minimization procedure has a noticeable effect on information loss (decreasing MC and GC)

Summary

- We developed a new pre-processing discrimination prevention framework to prevent direct discrimination, indirect discrimination or both of them at the same time.
- The experimental results showed that the proposed methods are successful to provide a proper trade-off between discrimination removal and data quality.
- We showed that indirect discrimination removal can help direct discrimination removal.

- **Discrimination- and Privacy-aware Patterns**

Sara Hajian[1]
Josep Domingo-Ferrer[2]
Anna Monreale[3]
Dino Pedreschi[4]
Fosca Giannotti[5]

[1,2] Universitat Rovira I Virgili, Spain
[3,4] Dipartimento di Informatica, Università di Pisa, Italy
[5] ISTI-CNR, Pisa, Italy

Data Mining and Knowledge Discovery Journal

Roadmap: PADM

- Privacy-aware data mining
  - Approaches and algorithms
    - Privacy preserving knowledge publishing
      - ➢ *k*-anonymity and its extensions
      - ➢ Differential privacy and its extensions
    - Privacy preserving data publishing
      - ➢ *k*-anonymity and its extensions
      - ➢ *L*-diversity and its extensions
      - ➢ Differential privacy and its extensions
  - Sanitization mechanisms
    - Deterministic techniques
      - ➢ Generalization
      - ➢ Suppression
      - ➢ Microaggregation
      - ➢ Bucketization
      - ➢ …
    - Randomized techniques
      - ➢ Additive perturbation
      - ➢ Laplacian Noise Addition
      - ➢ Exponential Mechanisms
      - ➢ …
  - Privacy models
    - ➢ k-anonymity
    - ➢ l-diversity
    - ➢ t-closeness
    - ➢ Differential privacy
    - ➢ ….

Motivating example

- Question: Is it sufficient to focus on either privacy or anti-discrimination only and ignore the other property?

Sex	Job	Credit_history	Salary	Credit_approved
Male	Writer	No-taken	... €	Yes
Female	Lawyer	Paid-duly	... €	No
Male	Veterinary	Paid-delay	...€	Yes
...	...	...	...	...

- Privacy protection only
  - sex=female, credit-history=no-taken →credit-approved=no
- Discrimination protection only
  - job =veterinarian, salary =low → credit-approved=no
  - job = veterinarian → credit-approved=no

- Answer: This example shows that protecting both privacy and non-discrimination is needed when disclosing a set of patterns.

Privacy-aware frequent pattern discovery

- *k*-Anonymous frequent pattern set [Atzori2008]
  - Given $F(D, \sigma)$, obtain a *k*-anonymous version of it.

Two steps:

- Step 1: Detecting non-*k*-anonymous patterns.
  - Example: *k*=3
    - p_1: { Job=veterinarian, Credit_approved=no }, $supp(p_1)$=41
    - p_2: {Job=veterinarian, Salary =low, Credit_approved=no}, $supp(p_2)$=40
    - p_x: {Job=veterinarian, Salary=high, Credit_approved=no}, $supp(C_{p1}^{p2})$=41-40=1
- Step 2: Privacy pattern sanitization
    - p_1: { Job=veterinarian, Credit_approved=no }, $supp(p_1)$=41+*k*=44

[Atzori2008] M. Atzori, F. Bonchi, F. Giannotti and D. Pedreschi. Anonymity preserving pattern discovery. VLDB Journal, 17(4): 703-727, 2008.

Discrimination-aware frequent pattern discovery

- Discrimination protected frequent pattern set
  - Given $F(D, \sigma)$, obtain α-protective version of it.

Has two steps:

- Step 1: Detecting α-discriminatory patterns
  - Example: discrimination threshold α =1.2, protected groups: {sex=female}
    - $\underbrace{p\text{: \{sex=female,credit-history=no-taken, credit-approved=no\},}supp(p)\text{=20}}_{slift(p) = 1.45}$
- Step 2: Anti-discrimination pattern sanitization
    - p_s: {sex=female,credit-history=no-taken}, $supp(p_s)$=34+Δ=40
  - **Theorem:** Anti-discrimination pattern sanitization for making $F(D, \sigma)$ α-protective does not generate new discrimination as a result of its transformation.

Simultaneous discrimination-privacy awareness in frequent pattern discovery

- We need to generate a discrimination- and privacy-protected version of $F(D, \sigma)$.
  - Definition (α-protective k-anonymous pattern set).
  - How making $F(D, \sigma)$ k-anonymous impacts on the α-protectiveness of $F(D, \sigma)$?
  - How making $F(D, \sigma)$ α-protective impacts on the k-anonymity of $F(D, \sigma)$?

Simultaneous discrimination-privacy awareness in frequent pattern discovery

- Question: How making $F(D,\sigma)$ k-anonymous impacts on the α-protectiveness of $F(D,\sigma)$?
  - First scenario

Patterns	*Support*
p_s:{female, veterinarian}	45
p_2 :{female, veterinarian, salary > 15000}	42
p_1 :{female, veterinarian, No}	32
p_n :{male, veterinarian, No}	16
p_{ns} :{male, veterinarian}	58

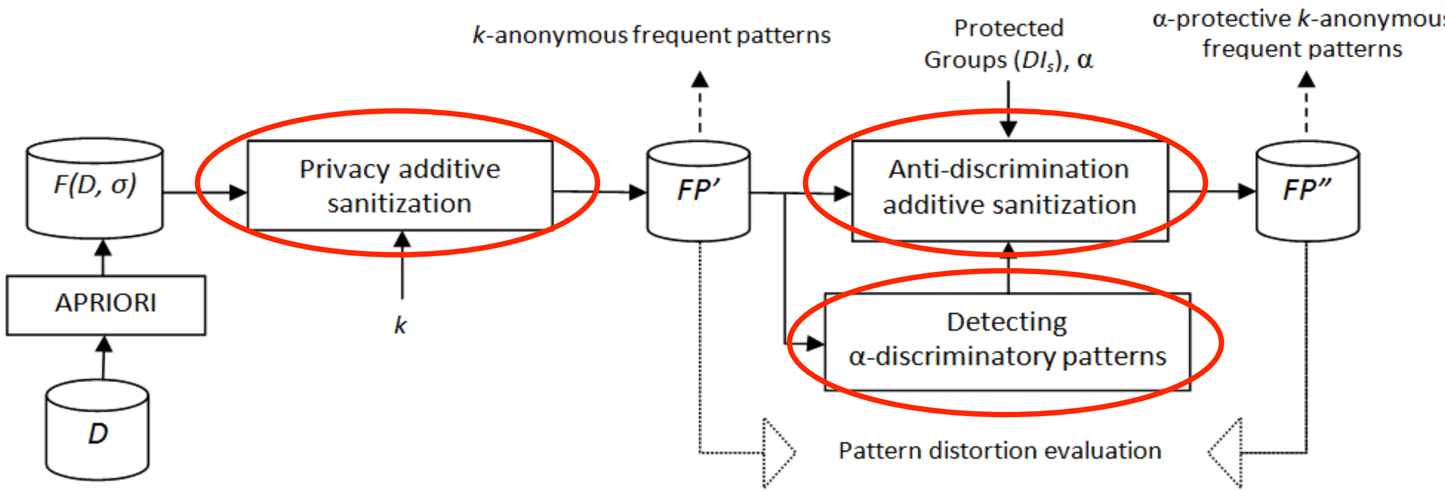
    - Using privacy pattern sanitization for making $F(D,\sigma)$ k-anonymous can make $F(D,\sigma)$ more α-protective.
  - Second scenario

Patterns	*Support*
p_s :{male, veterinarian}	58
p_2 :{male, veterinarian, salary > 15000}	56
p_1 :{female, veterinarian, No}	23
p_n :{male, veterinarian, No}	26
p_{ns} :{female, veterinarian}	45

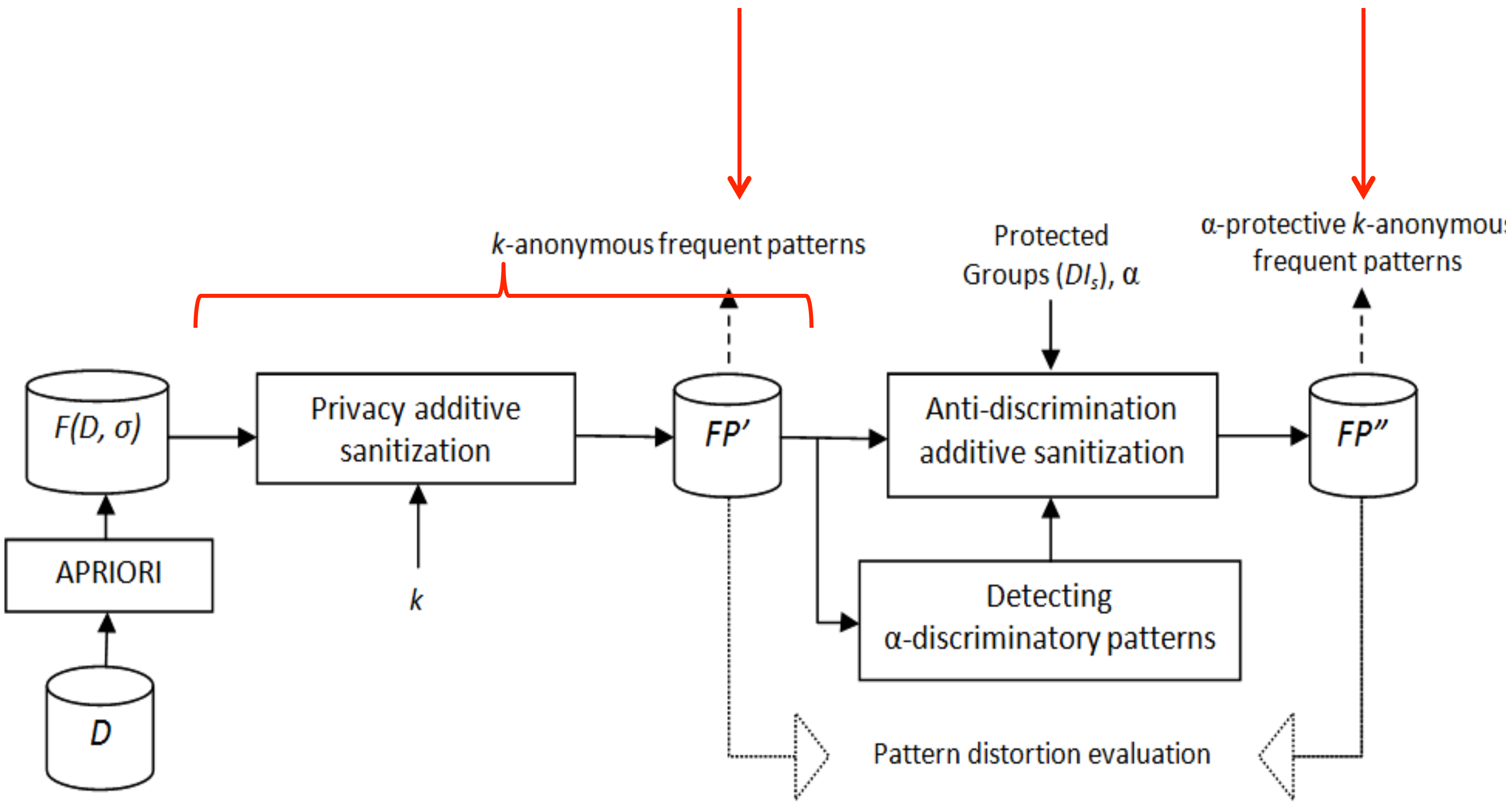
    - Using privacy pattern sanitization for making $F(D,\sigma)$ k-anonymous can make $F(D, \sigma)$ less α-protective.

Simultaneous discrimination-privacy awareness in frequent pattern discovery

- Question: How making $F(D, \sigma)$ α-protective impacts on the k-anonymity of $F(D, \sigma)$?
  - **Theorem:** Using anti-discrimination pattern sanitization for making $F(D,\sigma)$ α-protective cannot make $F(D,\sigma)$ non-k-anonymous



Framework



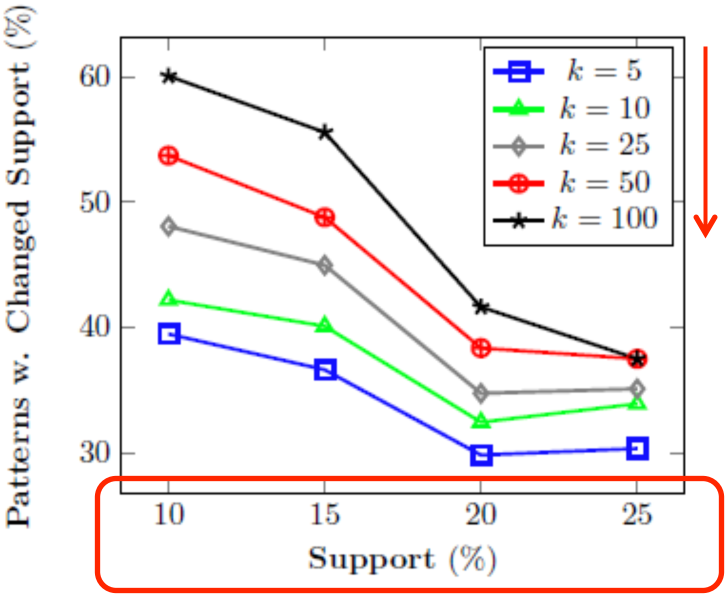
k-anonymous frequent patterns
Protected Groups (DI_s), α
α-protective k-anonymous frequent patterns
F(D, σ)
Privacy additive sanitization
FP'
Anti-discrimination additive sanitization
FP''
APRIORI
k
Detecting α-discriminatory patterns
D
Pattern distortion evaluation

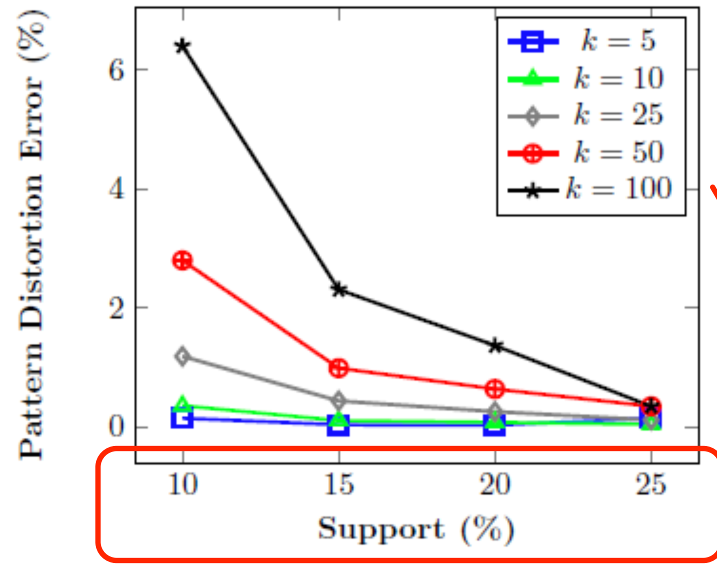
Experiments

- Pattern distortion
  - Patterns with changed support
  - Pattern distortion error
- The accuracy of classification
  - Using the CMAR (i.e. classification based on multiple association rules) approach.

Pattern distortion

- Pattern distortion scores to make the Adult dataset *k*-anonymous

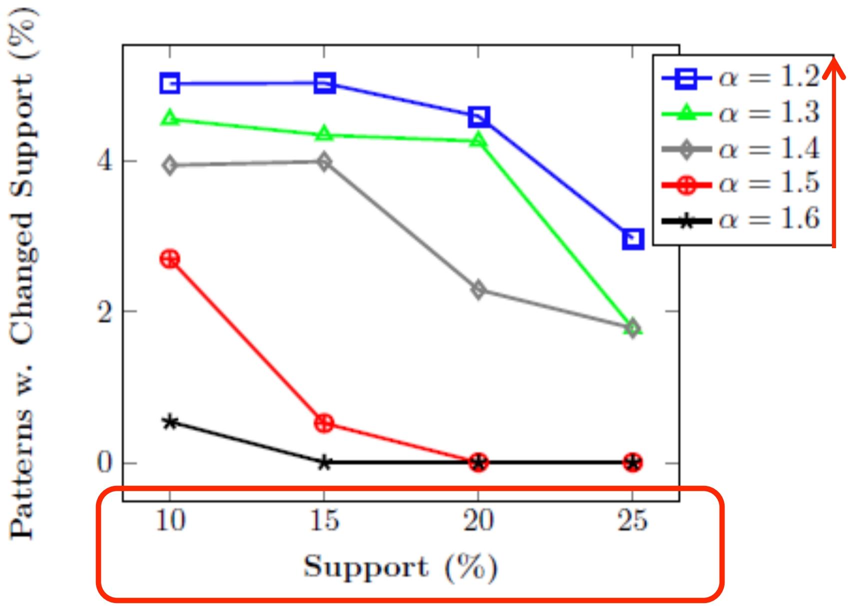


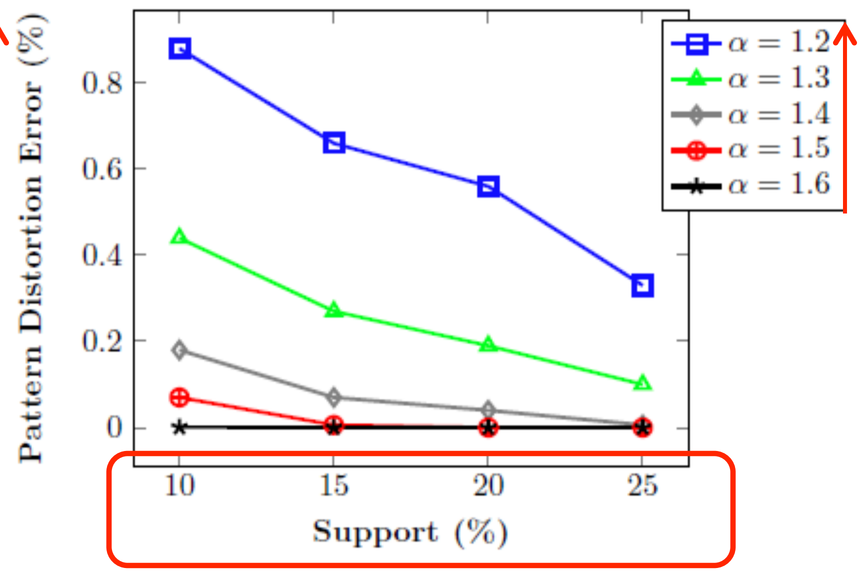


It can be seen that the percentage of patterns whose support has changed and the average distortion introduced) increase with larger *k* and with smaller support σ, due to the increasing number of inference channels.

Pattern distortion

- Pattern distortion scores to make the Adult dataset α-protective

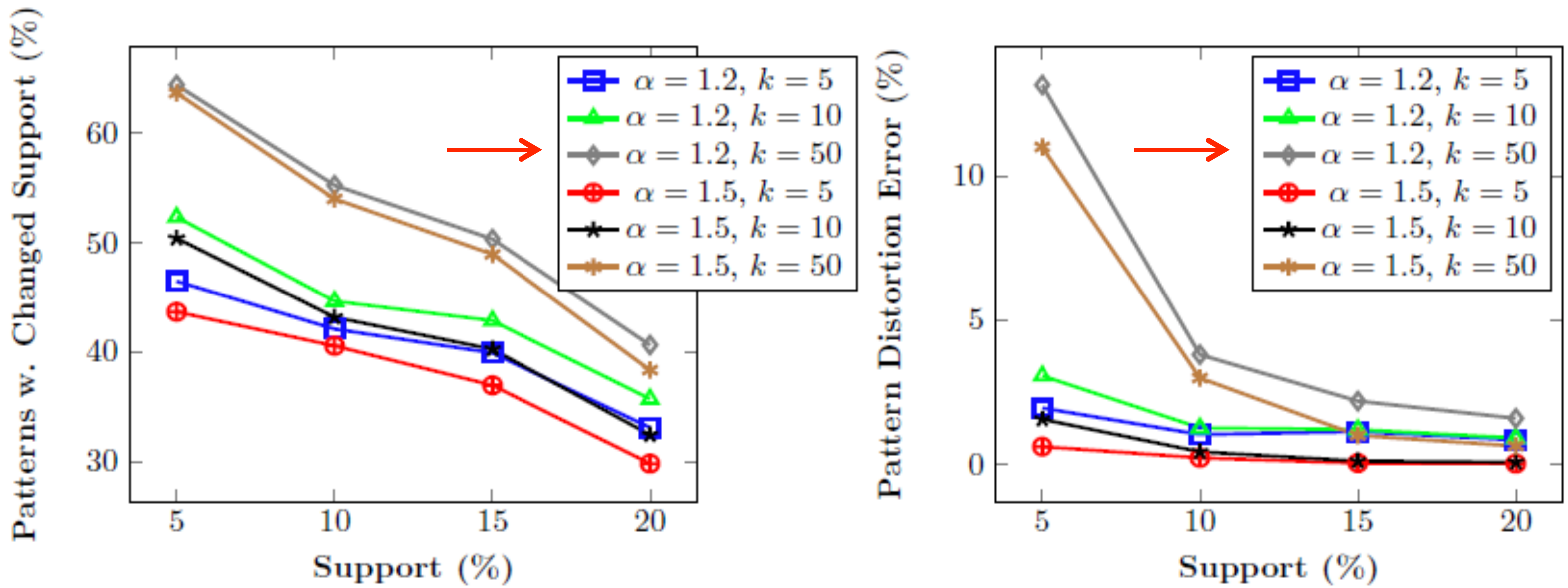




It can be seen that distortion scores increase with smaller σ and smaller α, because the number of α-discriminatory patterns increases.

Pattern distortion

- Pattern distortion scores to make the Adult dataset *α*-protective *k*-anonymous



We can (empirically) conclude that we provide protection against both the privacy and discrimination threats with a marginally higher distortion w.r.t. providing protection against the privacy threat only.

Accuracy of classification

- Preservation of the classification task

Adult dataset: accuracy of classifiers

k	α	$\mathcal{FP}$	$\mathcal{FP}'$	$\mathcal{FP}''$	$\mathcal{FP}^*$
5	1.2	0.744	0.763	0.724	0.691
5	1.5	0.744	0.763	0.752	0.739
50	1.2	0.744	0.751	0.682	0.691
50	1.5	0.744	0.751	0.746	0.739

German dataset: accuracy of classifiers

k	α	$\mathcal{FP}$	$\mathcal{FP}'$	$\mathcal{FP}''$	$\mathcal{FP}^*$
3	1.2	0.7	0.645	0.582	0.572
3	1.8	0.7	0.645	0.624	0.615
10	1.2	0.7	0.583	0.561	0.572
10	1.8	0.7	0.583	0.605	0.615

We do not observe a significant difference between the accuracy of the classifier obtained from an α-protective k-anonymous version of the original pattern set and the accuracy of the classifier obtained from either a k-anonymous or an α-protective version.

Extensions

- Alternative privacy models
  - Differential privacy
    - Similar to *k*-anonymity, achieving differential privacy in frequent pattern discovery can achieve α-protection or work against it.
    - We propose an algorithm to obtain an α-protective ϵ-differentially version of the original pattern set.
- Alternative anti-discrimination legal concepts
  - Genuine occupational requirement
    - We propose an algorithm to make the frequent pattern protected against unexplainable discrimination only.

Summary

- Simultaneous DADM and PADM in frequent pattern discovery
  - We found that privacy pattern sanitization methods based on either *k*-anonymity or differential privacy can work against anti-discrimination.
  - We found that our anti-discrimination pattern sanitization methods do not interfere with a privacy-preserving sanitization based on either *k*-anonymity or differential privacy.
  - The utility loss caused by simultaneous anti-discrimination and privacy protection is only marginally higher than the loss caused by each of those protections separately.

- **Generalization-based Privacy Preservation and Discrimination Prevention in Data Publishing and Mining**

Sara Hajian[1]
Josep Domingo-Ferrer[2]
Oriol Farràs[3]

Universitat Rovira I Virgili, Spain[1,2,3]

Data mining and Knowledge Discovery Journal (ECML/PKDD 2014 -- Journal Track), 28(5-6): 1158-1188 (2014)

Roadmap: PADM

- Privacy-aware data mining
  - Approaches and algorithms
    - Privacy preserving knowledge publishing
      - *k*-anonymity and its extensions
      - Differential privacy and its extensions
    - Privacy preserving data publishing
      - *k*-anonymity and its extensions
      - *L*-diversity and its extensions
      - Differential privacy and its extensions
  - Sanitization mechanisms
    - Deterministic techniques
      - Generalization
      - Suppression
      - Microaggregation
      - Bucketization
      - …
    - Randomized techniques
      - Additive perturbation
      - Laplacian Noise Addition
      - Exponential Mechanisms
      - …
  - Privacy models
    - k-anonymity
    - l-diversity
    - t-closeness
    - Differential privacy
    - ….

A study on the impact of data anonymization on anti-discrimination

Data Anonymization techniques	Achieve α-protection	Against α-protection	No impact
Global recoding generalizations	✓	✓	✓
Cell generalization/Cell suppression Type (1)	✓		✓
Cell generalization/Cell suppression Type (2)		✓	✓
Cell generalization/Cell suppression Type (3)	✓		
Cell generalization/Cell suppression Type (4)		✓	
Multidimensional generalization	✓	✓	✓
Record suppression Type (1)	✓		
Record suppression Type (2)		✓	
Record suppression Type (3)	✓		✓
Record suppression Type (4)		✓	
Value suppression	✓		✓

- We exploit the fact that some data anonymization techniques can protect data against discrimination.

Motivating example

- Raw customer credit data
  - Private data set with biased decision records

ID	Sex	Race	Hours	Salary	Credit_ approved
1	Male	White	40	High	Yes
2	Male	Asian-Pac	50	Medium	Yes
3	Male	Black	35	Medium	No
4	Female	Black	35	Medium	No
5	Male	White	37	Medium	Yes
6	Female	Amer-Indian	37	Medium	Yes
7	Female	White	35	Medium	No
8	Male	Black	35	High	Yes
9	Female	White	35	Low	No
10	Male	White	50	High	Yes

- The credit giver needs to eliminate two types of threats against her customers before publishing data:
  - Privacy threat
    - e.g., record linkage through QI attributes
  - Discrimination threat

PPDM via generalization

- To prevent record linkage attack
  - Model: *k*-anonymity.
  - Sanitization mechanism : Full-domain generalizations
  - Algorithm: Incognito [Lefevre2005]
    - Incognito is a well-known suite of optimal bottom-up generalization algorithms to generate all possible *k*-anonymous full-domain generalizations .

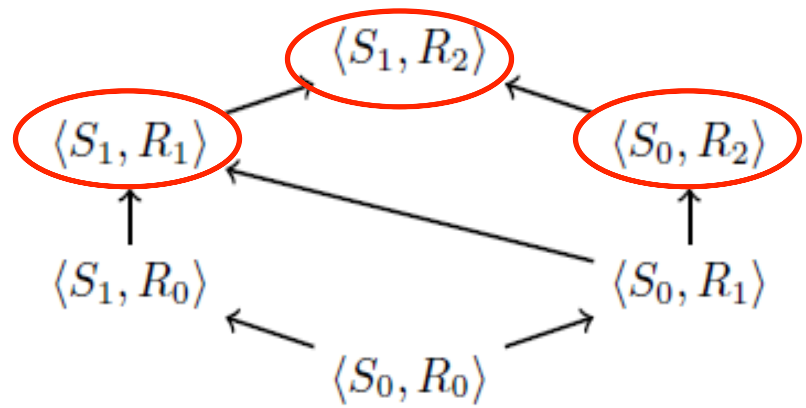
[Lefevre2005] K. Lefevre, D. J. Dewitt, and R. Ramakrishnan. Incognito: Efficient full-domain k-anonymity. In SIGMOD 2005, pp. 49-60. ACM, 2005.

PPDM via generalization

- Incognito is based on two main properties satisfied for *k*-anonymity
  - Subset property
  - Generalization property

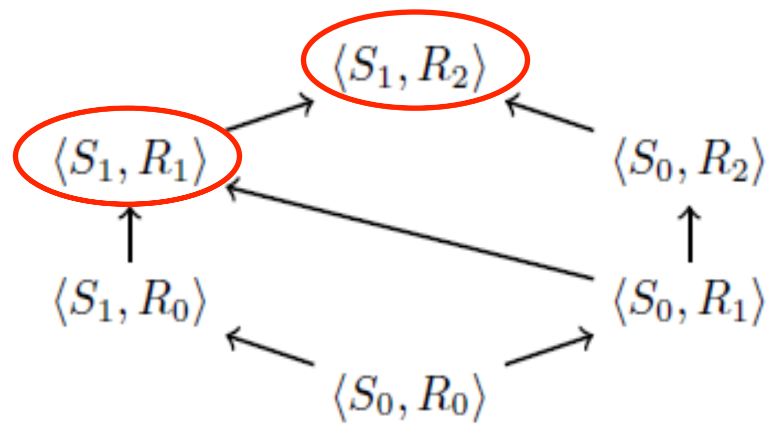
ID	Sex	Race	Hours	Salary	Credit_ approved
1	Male	White	40	High	Yes
2	Male	Asian-Pac	50	Medium	Yes
3	Male	Black	35	Medium	No
4	Female	Black	35	Medium	No
5	Male	White	37	Medium	Yes
6	Female	Amer-Indian	37	Medium	Yes
7	Female	White	35	Medium	No
8	Male	Black	35	High	Yes
9	Female	White	35	Low	No
10	Male	White	50	High	Yes

- Example
  - Consider the generalization lattice over *QI* attributes
  - *QI* = {Race, Sex} and *k* = 3



DPDM via generalization

- To prevent discrimination
  - Model: α-protection.
    - α-protective version of the original data table.
  - Sanitization mechanism: Full-domain generalizations?
    - Given the generalization lattice of *D* over *QI*, where *DA* $\subseteq$ *QI*, there are some candidate nodes for which *D* is α-protective
- Example
  - suppose *f* = *elift*
  - *QI* = {Race, Sex}
  - 1.2-protective with respect to *DA* = {Sex}.

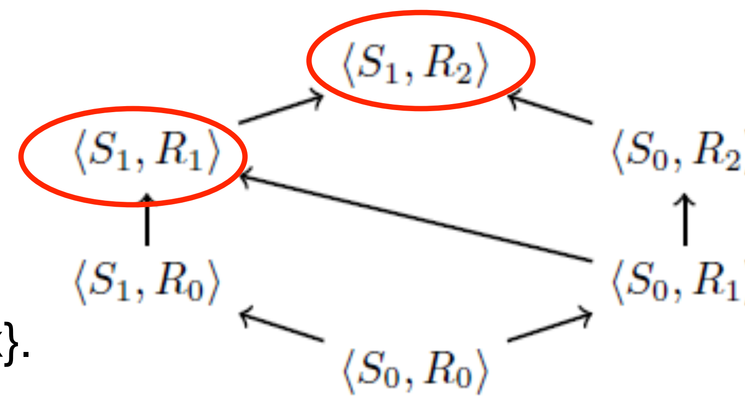


Simultaneous PPDM and DPDM via generalization

- Obtain anonymized data tables that are protected against record linkage and also free from discrimination
  - Definition (α-protective *k*-anonymous data table).

- Observation. *k*-Anonymity and α-protection can be achieved simultaneously in a data table by means of full-domain generalization.

- Example
  - suppose *f* = *elift*
  - 3-anonymous with respect to *QI* = {Race, Sex}.
  - 1.2-protective with respect to *DA* = {Sex}.



Simultaneous PPDM and DPDM via generalization

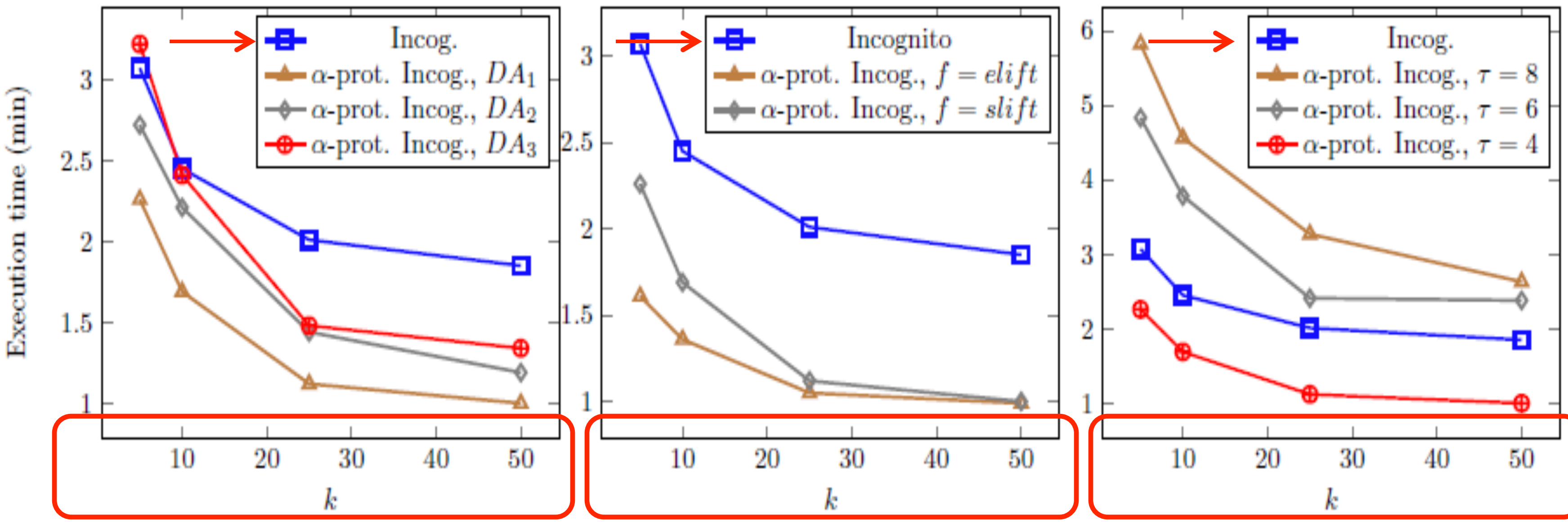
- Our task is to obtain α-protective *k*-anonymous full-domain generalizations.
  - The naïve approach is the sequential way.
    - It is a very expensive solution
  - Our proposal: we present a more efficient algorithm that takes advantage of the common properties of α-protection and *k*-anonymity
    - Foundations
      - Observation (Subset property of α-protection).
      - Proposition (Generalization property of α-protection).
      - Proposition (Roll-up property of α-protection).
    - Algorithm
      - α-protective Incognito

Experiments

- Evaluate the execution time of α-protective Incognito and compare it with Incognito.
- Evaluate the quality of unbiased anonymous data, compared to that of the anonymous data
  - Using general and specific data analysis metrics.

Execution time

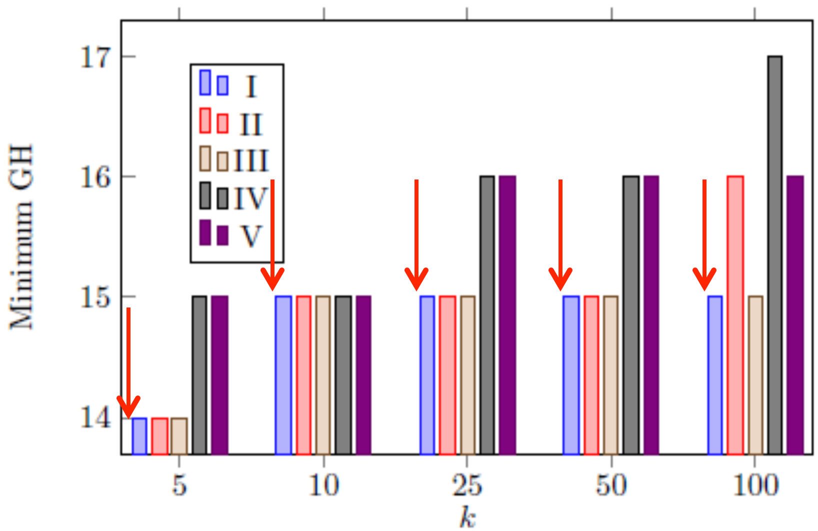
- Performance of Incognito and α-protective Incognito .

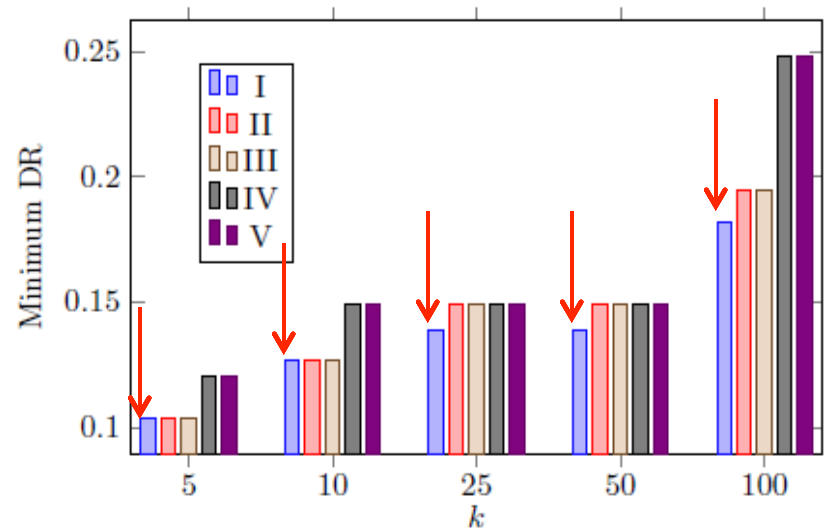


Since α-protective Incognito provides extra protection (i.e. against discrimination) in comparison with Incognito, the cost is sometimes a longer execution time.

General data quality

- General data quality metrics. Left, generalization height (GH). Right, discernibility ratio (DR).

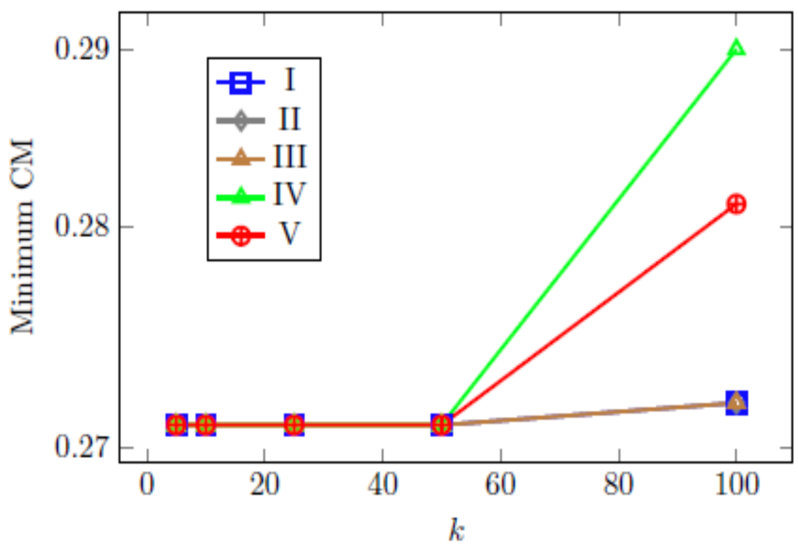


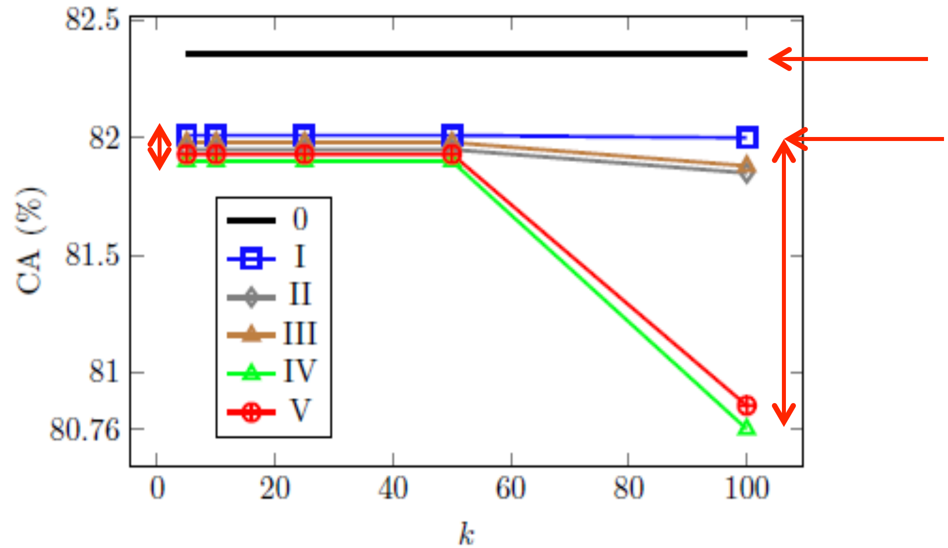


We found that the data quality of *k*-anonymous tables (i.e. in terms of GH and DR) without *α*-protection is equal or only slightly better than the quality of *k*-anonymous tables with *α*-protection.

Data quality for classification

- Data quality for classification analysis. Left, classification metric (CM). Right, classification accuracy, in percentage (CA).





We found that the data quality of *k*-anonymous tables (i.e. in terms of CM) without *α*-protection is equal or only slightly better than the quality of *k*-anonymous tables with *α*-protection.

Extensions

- Alternative privacy models
  - Attribute disclosure
  - Differential privacy
- Alternative anti-discrimination legal concepts
  - Indirect discrimination

Summary

- Simultaneous DADM and PADM in data publishing
  - We found that a subset of *k*-anonymous full-domain generalizations with the same or slightly higher data distortion than the rest are also α-protective.
  - We have adapted to α-protection two well-known properties of *k*-anonymity, namely the subset and the generalization properties.
  - We have sketched how our approach can be extended to satisfy alternative privacy models or anti-discrimination legal constraints.

Projects and future works

- H2020 EU innovation action project
  - TYPES: Towards transparencY and Privacy in the onlinE advertising businesS
- Unbiased recommendation and machine learning algorithms
- New types of data
  - Wikipedia
  - Social network data

Thank you for your attention!

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 INNOVATION ACTION PROGRAMME UNDER GRANT AGREEMENT NO 653449