UM ESTUDO QUANTITATIVO FORMAL SOBRE PRIVACIDADE NA PUBLICAÇÃO DOS CENSOS EDUCACIONAIS OFICIAIS NO BRASIL

GABRIEL HENRIQUE LOPES GOMES ALVES NUNES

UM ESTUDO QUANTITATIVO FORMAL SOBRE PRIVACIDADE NA PUBLICAÇÃO DOS CENSOS EDUCACIONAIS OFICIAIS NO BRASIL

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Orientador: Mário Sérgio Ferreira Alvim Júnior

Coorientadora: Annabelle McIver

Belo Horizonte, Minas Gerais

Maio de 2021

GABRIEL HENRIQUE LOPES GOMES ALVES NUNES

A FORMAL QUANTITATIVE STUDY OF PRIVACY IN THE PUBLICATION OF OFFICIAL EDUCATIONAL CENSUSES IN BRAZIL

Thesis presented to the Graduate Program in Computer Science of the Federal University of Minas Gerais in partial fulfillment of the requirements for the degree of Master in Computer Science.

Advisor: Mário Sérgio Ferreira Alvim Júnior

Co-Advisor: Annabelle McIver

Belo Horizonte, Minas Gerais

May 2021

Nunes, Gabriel Henrique Lopes Gomes Alves

N972f A formal quantitative study of privacy in the publication of official educational censuses in Brazil [manuscrito] Gabriel Henrique Lopes Gomes Alves Nunes. — 2021.
xxvii, 258 f. : il. ; 29cm

Orientador: Mário Sérgio Ferreira Alvim Júnior.
Coorientadora: Annabelle McIver.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação.
Referências: f.251-258

1. Computação – Teses. 2. Teoria da Informação – Teses. 3. Fluxo de Informação Quantitativo – Teses. 4. Dados estatísticos – Direito a privacidade – Teses. 5. Censo escolar – Brasil – Teses. I. Alvim Júnior, Mário Sérgio Ferreira. II. McIver, Annabelle. III. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação. IV. Título.

CDU 519.6*44(043)

Ficha catalográfica elaborada pela bibliotecária Irénquer Vismeg Lucas Cruz CRB 6ª nº 819.

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

A formal quantitative study of privacy in the publication of official educational censuses in Brazil

GABRIEL HENRIQUE LOPES GOMES ALVES NUNES

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

Prof. Mário Sérgio Ferreira Alvim Júnior - Orientador
Departamento de Ciência da Computação - UFMG

A McIver

Profa. Annabelle McIver - Coorientadora
Department of Computing - Macquarie University

Diego de Freitas Aranha

Prof. Diego de Freitas Aranha
Department of Computer Science - Aarhus University

Prof. Gabriel de Morais Coutinho
Departamento de Ciência da Computação - UFMG

Prof. Jeroen Antonius Maria van de Graaf
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 28 de Abril de 2021.

Abstract

Privacy preservation in the release of statistical data has been a concern of the scientific community for decades. This preoccupation has been gradually expanding to outside of academia, and has been reflected in the widespread enactment and reinforcement of privacy-protection legislation around the world. In Brazil, the new privacy law enacted in 2018 (LGPD) establishes, among other provisions, mandatory restrictions on governmental agencies that publicly release data on individuals, and prescribes sanctions in case of non-compliance. In this context, it is paramount for those agencies to thoroughly review and, if necessary, adapt their current methods of data publishing. However, it is well known that any disclosure control method applied to the release of statistical data may present deleterious effects on data utility, i.e. on the quality of information provided to legitimate consumers, such as analysts and society as a whole. A fine balance between privacy and utility must be achieved, taking into consideration the interests of several stakeholders, including data owners, legitimate data consumers, and the government.

In this thesis, we provide a thorough quantitative study of privacy risks in the release of the official Brazilian Educational Censuses provided annually by INEP, which is Brazil's governmental agency responsible for the development and maintenance of educational statistics systems. More precisely, we formally analyze privacy risks in databases released as microdata, i.e. data at each individual's record level, and protected by the technique of de-identification, i.e. the removal of direct identifying information such as the individuals' names or personal identification numbers.

In order to do so, we propose a unified classification system for attacks, which allows us to properly cover and formalize the landscape of privacy risks in the Educational Censuses. Our first contribution are models of attacks rigorously formalized in the framework of quantitative information flow, defined along three orthogonal dimensions: (i) risk of re-identification vs. risk of attribute-inference; (ii) attacks on a single database vs. attacks on longitudinal databases, i.e. those that are updated and extended frequently, as in the case of INEP's Censuses; and (iii) deterministic vs. probabilistic measures of privacy risk.

As a second contribution, we employ our formal models to obtain extensive quantitative evaluations of privacy risks on INEP's Educational Census databases, which account for more than fifty million students, or around 25% of the country's current population. Those experiments unequivocally show that INEP's current disclosure control methods are insufficient to guarantee individuals' privacy at any acceptable level, and therefore may be in contempt of Brazil's new privacy legislation. For instance, 81.13% of students in the School Census of 2019, corresponding to approximately 39 085 531 individuals, may be subject to complete re-identification under reasonably modest attacks. We argue, therefore, that INEP should abandon current practices and consider stricter disclosure control methods.

As a third contribution, we formally evaluate the trade-off between privacy and utility in two variants of differential privacy –the golden standard disclosure control technique in the literature– as the method to be employed to INEP's Educational Censuses releases. Our results confirm that global differential privacy tends to favor utility over privacy, whereas local differential privacy tends to act in the opposite way.

To the best of our knowledge, our analyses are the most extensive of its kind in the literature. Furthermore, our results provide INEP with solid empirical evidence to guide well-informed future decisions when complying with Brazil's new privacy legislation, and have the potential to positively impact a significant fraction of the Brazilian population.

Keywords: Quantitative Information Flow, Disclosure Control, Microdata, Differential Privacy, Privacy, Utility

Resumo

A preservação da privacidade na divulgação de dados estatísticos tem sido uma preocupação da comunidade científica há décadas. Essa preocupação tem se expandido gradualmente para fora da academia e tem se refletido na promulgação e no reforço generalizado da legislação de proteção à privacidade em todo o mundo. No Brasil, a nova lei de privacidade promulgada em 2018 (LGPD) estabelece, dentre outras providências, restrições obrigatórias aos órgãos governamentais que divulgam publicamente dados sobre pessoas físicas e prescreve sanções em caso de não conformidade. Nesse contexto, é fundamental que essas agências revisem minuciosamente e, se necessário, adaptem seus métodos atuais de publicação de dados. No entanto, é bem conhecido que qualquer método de controle de divulgação aplicado à liberação de dados estatísticos pode apresentar efeitos deletérios na utilidade dos dados, ou seja, na qualidade da informação fornecida aos consumidores legítimos, como analistas e a sociedade como um todo. Um equilíbrio preciso entre privacidade e utilidade deve ser alcançado, levando em consideração os interesses de várias partes, incluindo proprietários de dados, consumidores legítimos de dados e o governo.

Nesta dissertação, fornecemos um estudo quantitativo completo dos riscos à privacidade na divulgação dos Censos Educacionais Brasileiros oficiais fornecidos anualmente pelo INEP, que é o órgão governamental brasileiro responsável pelo desenvolvimento e manutenção de sistemas de estatísticas educacionais. Mais precisamente, analisamos formalmente os riscos de privacidade em bancos de dados divulgados como microdados, i.e. dados no nível de registro de cada indivíduo, e protegidos pela técnica de desidentificação, i.e. a remoção de informações de identificação direta, como nomes de indivíduos ou números de identificação pessoal.

Para tanto, propomos um sistema unificado de classificação de ataques, que nos permite cobrir e formalizar adequadamente o panorama de riscos à privacidade nos Censos Educacionais. Nossa primeira contribuição são modelos de ataques rigorosamente formalizados no *framework* de fluxo de informação quantitativa, definidos ao longo de três dimensões ortogonais: (i) risco de reidentificação vs. risco de inferência de atributos; (ii) ataques a uma única base de dados vs. ataques a bases de dados longitudinais, i.e. aquelas que são atualizadas e ampliadas com frequência, como no caso dos Censos do INEP; e (iii) medidas determinísticas vs. probabilísticas de risco de privacidade.

Como uma segunda contribuição, empregamos nossos modelos formais para obter avaliações quantitativas extensas de riscos de privacidade nas bases de dados dos Censos Educacionais do INEP, que respondem por mais de cinquenta milhões de alunos, ou cerca de 25% da população atual do país. Esses experimentos mostram inequivocamente que os métodos atuais de controle de divulgação do INEP são insuficientes para garantir a privacidade dos indivíduos em qualquer nível aceitável e, portanto, podem estar em desacordo com a nova legislação de privacidade do Brasil. Por exemplo, 81,13% dos alunos no Censo Escolar de 2019, correspondendo a aproximadamente 39 085 531 indivíduos, podem estar sujeitos a reidentificação completa sob ataques razoavelmente modestos. Argumentamos, portanto, que o INEP deve abandonar as práticas atuais e considerar métodos de controle de divulgação mais rígidos.

Como uma terceira contribuição, avaliamos formalmente o *trade-off* entre privacidade e utilidade em duas variantes de privacidade diferencial –a técnica de controle de divulgação padrão-ouro na literatura– como o método a ser empregado para divulgação dos Censos Educacionais do INEP. Nossos resultados confirmam que a privacidade diferencial global tende a favorecer a utilidade em relação à privacidade, enquanto a privacidade diferencial local tende a agir de forma oposta.

Até onde sabemos, nossas análises são as mais extensas desse tipo na literatura. Além disso, nossos resultados fornecem ao INEP evidências empíricas sólidas para orientar decisões futuras bem informadas ao cumprir a nova legislação de privacidade do Brasil e têm o potencial de impactar positivamente uma fração significativa da população brasileira.

Palavras-chave: Fluxo de Informação Quantitativo, Controle de Divulgação, Microdados, Privacidade Diferencial, Privacidade, Utilidade

To sanity check.

Acknowledgement

I dedicate this thesis to my parents, grandparents, sister, and friends, for always supporting me in my pursuit for knowledge.

I would like to thank my advisors, Professor Mário Sérgio Alvim and Professor Annabelle McIver, for the opportunity of working with them. Their ethics, passion for science, and hard-work are always inspiring.

I would also like to thank Prof. Diego Aranha, Prof. Jeroen van de Graaf, and Prof. Gabriel Coutinho. It was an honor to have them in my thesis defense jury.

I am grateful for all the help I received from my friends. Particularly, for the technical discussions with Natasha, Ramon, and Igor.

I would also like to thank all the support from the Program's staff, particularly from Sônia, Linda, and Ludmila.

Finally, I am grateful for both CNPq and the Brazilian people for financing this work.

“We can only see a short distance ahead,
but we can see plenty there that needs to be done.”

(Alan Turing)

List of Definitions

List of Examples

List of Experiments

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Statistics are everywhere. This is particularly evident in challenging times such as those of economic depression or, more sadly, of a pandemic such as the one our societies are currently experiencing. The importance of accurate and up-to-date data has been demonstrated on a daily basis since the beginning of the *Coronavirus disease 2019* (COVID-19) pandemic, declared as such by the *World Health Organization* (WHO) on 11 March 2020 [79, 80]. In times like this, most of the released data on confirmed cases, deaths, and now vaccination, is expected to come from official governmental agencies. [1]

So is the case in less challenging times, when governments collect information on their citizens in order to keep track of health, educational, and other demographic information. By doing so and properly using this data, governments can better fulfill their duty by correctly assigning resources to areas and populations most in need.

Similarly, private companies have also been collecting vast amounts of data on individuals for the past few decades. This information has allowed the development of new technologies, from targeted advertisement to personalized services. But as the amount of data collected increases, so does the need to guarantee individuals' privacy. For instance, Garfinkel described in his 2000 book *Database Nation* several scenarios in which privacy breaches could happen given the technology in use at the time [25].

Therefore, it is also the responsibility of those collecting this data, whether governments or private companies, to guarantee that information on individuals are kept safe against exploitation. This task, however, is not easy. For instance, past attempts to release *microdata*, i.e. data at the record level, have had disastrous consequences in terms of privacy breaches, even when there was the intention of keeping the data safe, e.g. by

[1] Except when governments arbitrarily decide to *not* be transparent [57, 58].

applying disclosure avoidance methods such as *de-identification*, i.e. the removal of direct identifying information such as names or government-issued unique numbers.

Even so, microdata de-identification can still be found in releases of databases nowadays, for example in the official Brazilian Educational Censuses provided annually by INEP, which is Brazil's agency responsible for the development and maintenance of educational statistics systems and assessment projects. However, the use of de-identification as the sole disclosure control method by INEP may have to change soon, given Brazil's new privacy legislation enacted in 2018 (LGPD) that establishes, among other provisions, mandatory restrictions to governmental agencies that publicly release data on individuals and prescribes sanctions in case of non-compliance. Hence, it is paramount for those agencies to thoroughly review and adapt their current methods of data publishing.

Anyhow, it is well known that any disclosure control method applied to statistical data releases also have effects on data utility and must consider possibly opposite interests from different stakeholders. For instance, when it comes to national educational censuses, those stakeholders include the data holders (e.g. students, teachers, lecturers, and professors), the data analysts (e.g. demographers, and civil and governmental entities), the government (responsible for regulating both privacy and transparency), and the agency responsible for the data release (which must realistically balance requests from other stakeholders and its own operational capacity).

In this thesis, we provide a thorough quantitative study of privacy risks in the databases released by INEP for the official Brazilian Educational Censuses. More precisely, we propose models for analyzing existent privacy risks in databases released as microdata and protected by the technique of de-identification.

In order to do so, we propose a new classification of attacks against releases of databases and a database model, both of which have allowed us to better explore the possible existent vulnerabilities and to develop a series of new attack models. Thus, we consider both re-identification and attribute-inference attacks, i.e. the complete disclosure of a data holder's identity or the inference of an attribute's value; both single and longitudinal releases of databases, i.e. the privacy risks posed by the release of only one database or by continuously releasing microdata with annual frequency, as is the case for the Educational Censuses databases; and both deterministic and probabilistic measures of privacy risk, i.e. how many individuals are vulnerable to a given attack with absolute certainty or the average privacy risk for data holders.

As a second contribution, we employ our models to provide extensive quantitative evaluations of privacy risks on INEP's Educational Censuses databases, which account for more than fifty million students, or around 25% of the country's current population. Those experiments have allowed us to conclude unequivocally that INEP's current disclosure control methods are insufficient to guarantee individuals' privacy at any acceptable level, and therefore may be in contempt of Brazil's new privacy legislation. For instance, 81.13% of students in the School Census of 2019, corresponding to approximately 39 085 531 individuals, may be subject to complete re-identification under reasonably modest attacks. We argue, therefore, that INEP should abandon current practices and consider stricter disclosure control methods, particularly differential privacy, a state-of-the-art method with formal privacy guarantees.

As a third contribution, we formally evaluate the trade-off between privacy and utility in two variants of differential privacy as possible disclosure control method for the Educational Censuses. Those analyses are particularly important given that possible correlations between attributes in a database may expose sensitive information to unexpected inference attacks. Our results confirm that global differential privacy tends to favor utility over privacy, whereas local differential privacy tends to act in the opposite way. Furthermore, our results provide INEP with empirical evidence to guide future decisions when complying with Brazil's new privacy legislation.

1.1 Brief historical review

The possibility of unintended disclosure of information from releases of databases has been discussed in the literature since before Tore Dalenius' 1977 paper proposing a methodology for statistical disclosure control [10]. Given the societal benefits that may come from statistical studies, *elimination* of disclosure was unfeasible, so Dalenius proposed its control. As described by Dalenius, the scientific community's main concern at the time could be summarized as the search for "a reasoned balance between the right to privacy and the need to know". Since Dalenius' work, several attempts at controlling information disclosure from databases were suggested and implemented, but usually followed by the discovery of some serious vulnerabilities.

For instance, back when the United States Census of 1990 was published, *de-identification*, i.e. the removal of obvious identifying attributes from databases, such as name and government-issued identification numbers, was the main disclosure control method in use. In this context, Sweeney was able to show in 2000 that 87% of the United States population in that census could be uniquely identified by using only a combination of the ZIP code, gender, and date of birth [67]. The work done by

Sweeney highlighted an expected vulnerability in the then current disclosure control methods, drawing attention to the extension of re-identification risks individuals were exposed to. This result created a whole new era for disclosure control methods, which has gained traction given the ever increasing volume and detail of data produced.

Another famous data disclosure in the literature came from the *Netflix Prize* and was published by Narayanan and Shmatikov in 2008 [56]. For that contest, the company released a database containing 100 480 507 movie ratings of 480 189 Netflix subscribers from December 1999 to December 2005, which it claims was properly anonymized (even though no details on this were provided). By using the *Internet Movie Database* (IMDb) as auxiliary information and by applying their de-anonymization algorithm for sparse data, i.e. data with high dimensionality, Narayanan and Shmatikov were able to de-anonymize the released database and recover movie history and ratings for individual users of the Netflix service.

The Netflix Prize database had vulnerabilities similar to those found by Sweeney in the United States Census of 1990, in addition to its own high dimensionality that could also be used as an attack vector. But the most important lesson from those and other disclosures is the need of formal and accurate definitions for both the adversary, i.e. the entity interested in learning sensitive information, and their context. Only by doing so, proper disclosure avoidance methods can be developed in such a way that their privacy guarantees are well-established.

Hence the importance of our work, built upon a firm theoretical foundation provided by the theory of *Quantitative Information Flow* (QIF) and careful work on our newly proposed definitions and extensive quantitative analyses.

1.2 Brief legal review

We now present a brief legal review of the privacy legislation around the world and in Brazil. This illustrates the increasing international concern for individual privacy protection, particularly given the technological developments and globalisation in the last few decades. We also present a more detailed review in Appendix A.

1.2.1 Overview of foreign privacy legislation

In Australia. The *Privacy Act* of 1988 [32], revised in November 2015, establishes the *Australian Privacy Principles*, which guarantees the use of *anonymization*, i.e. the disassociation of individuals from their respective records, or *pseudonymization*, i.e.

the attribution of an individual code to each record in a database to replace other direct identifiers, for the treatment of personal information.

In the European Union. Regulation (EU) 2016/679, known as *The General Data Protection Regulation* (GDPR) [72], establishes fines to businesses of up to 20 million Euros or up to 4% of the annual worldwide turnover of the preceding financial year in case of an enterprise, whichever is greater, in case of violations to the GDPR.

In the United States. The *Confidential Information Protection and Statistical Efficiency Act* (CIPSEA), approved in 2002 [43], establishes that personally identifiable information provided to federal agencies for statistical purposes under the promise of confidentiality cannot be intentionally disclosed for non-statistical purposes or without consent, which became a federal crime.

1.2.2 Overview of privacy legislation in Brazil

The Constitution of the Federative Republic of Brazil from 1988 [35], in its Article 5, guarantees to all Brazilians and foreigners residing in the country that:

> X the privacy, private life, honour and image of persons are inviolable, and the right to compensation for property or moral damages resulting from their violation is ensured;
>
> XXXIII all persons have the right to receive, from the public agencies, information of private interest to such persons, or of collective or general interest, which shall be provided within the period established by law, subject to liability, except for the information whose secrecy is essential to the security of society and of the State;

Hence, item X of Article 5 sets the constitutional, individual right to privacy, whereas item XXXIII of the same Article sets the constitutional right, individual and collective, to transparency by the State. However, there is no definition on how to balance those two principles or on what would be the legal limits of each one of them. Therefore, to regulate the rights to privacy and transparency determined in those entrenched clauses of the Constitution, Law 12 527 of 2011 was sanctioned to regulate access to information and Law 13 709 of 2018 was sanctioned to regulate the protection of personal data.

Law 12 527 of 2011 [39], known as the Access to Information Law (*Lei de Acesso à Informação*, or LAI, in Portuguese), regulates item XXXIII of Article 5 of the Constitution. Articles 6 and 8 of LAI determine to the public authority the duty to guarantee broad access to information, particularly that considered to be of collective or general interest, which must be made available via the Internet regardless of requirement. In addition, Article 7 guarantees the right of access to other information not immediately available via the Internet. However, according to Article 22, access to information may be denied in whole or in part if the information is considered to be confidential. Particularly for the treatment of personal information, Article 31 establishes that it must be done in a transparent manner and with respect to individual freedoms and guarantees, according to item X of Article 5 of the Constitution. However, provisions on the handling of personal information are open to subsequent regulation.

Law 13 709 of 2018 [40], known as the General Personal Data Protection Law (*Lei Geral de Proteção de Dados Pessoais*, or LGPD, in Portuguese), as per its Article 1, aims to protect the fundamental rights of freedom and privacy. Article 7 determines in which cases the processing of personal data is allowed, and Article 11 does the same specifically for "sensitive personal data". In both cases, processing is permitted with the consent of the data subject or with the consent of the legal guardian for sensitive data. One important definition comes in Article 12 of the LGPD for "anonymous data", which is not to be considered personal data, except when the anonymization process can be reversed with reasonable efforts. Therefore, objective factors such as the cost and time needed to reverse the anonymization process should be considered given the available technologies and disregarding the use of third party means. But again, the proper definition of what would be considered a reasonable effort, or which anonymization methods should be used, were left to subsequent regulation. [2]

The LGPD is to be regulated by the National Data Protection Authority (*Autoridade Nacional de Proteção de Dados*, or ANPD, in Portuguese), which is still being implemented as of 2021. The ANPD is expected to face several challenges in harmonizing the constitutional principle of transparency and the LAI with the LGPD. In this context, academic work such as the one we present here can help both society and the ANPD to understand the existent trade-offs in the case of INEP's Educational Censuses.

[2] From the LGPD Article 5 [40]:

II Sensitive personal data: personal data on racial or ethnic origin, religious belief, political opinion, union membership or affiliation to organizations of a religious, philosophical, or political nature, data relating to health or sexual life, genetic or biometric data, when linked to a natural person.

III Anonymous data: data relating to an unidentifiable holder, considering the use of reasonable technical means available at the time of processing.

1.3 The INEP databases

The National Institute of Educational Studies and Research (*Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*, or INEP, in Portuguese) [3] is responsible for the development and maintenance of educational statistics systems and assessment projects, as well as the dissemination of this information, according to Law 9 448 of 1997 [36] and Decree 6 317 of 2007 [37]. As per the Decree, both Basic and Higher Education establishments, whether public or private, are required to provide the information requested by INEP, while also ensuring the confidentiality of personal data collected and prohibiting its use for other purposes. The databases released by INEP constitute our case study in this thesis and can be found on the agency's website, [4] but also on Brazil's Open Data Portal [5] and on Google Dataset. [6]

In this thesis we will focus on two of those studies, i.e. the School Census (*Censo Escolar*, in Portuguese) and the Higher Education Census (*Censo da Educação Superior*, in Portuguese), both released as microdata, i.e. data at the record level.

- **School Census**: with annual frequency, it is the main educational statistical survey in the country, which covers both Basic and Professional Education. It includes information on students of all ages, usually from 0 to 18 years old, on adults who have previously abandoned formal education and have joined a Youth and Adult Education Program, and on teachers and schools. Microdata releases are currently available from 2007 through 2020.

- **Higher Education Census**: with annual frequency, it is the most complete statistical survey on Higher Education Institutions (*Instituições de Educação Superior*, or IES, in Portuguese) in the country. It includes information on students of undergraduate and graduate levels, on lecturers and professors, and on the institutions. Microdata releases are currently available from 2009 through 2019.

A preliminary analysis of the databases on students and instructors resulted in the identification of two disclosure control methods implemented by INEP:

- *de-identification*, by which direct identifiers are removed from the records, e.g. name, Social Security Number (*Cadastro de Pessoa Física*, or CPF, in Por-

[3] https://www.gov.br/inep

[4] https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados

[5] https://dados.gov.br/

[6] https://datasetsearch.research.google.com

tuguese) and other government-assigned unique-numbers, or addresses at more detailed levels than cities;

- *pseudonymization*, by which INEP assigns to each record a unique, artificially-created identification code, e.g. a non-changing number across releases of databases that allows following an individual through different years.

Even so, several other attributes are available, including date and city of birth, gender, ethnicity, and the unique school or Higher Education Institution code. As we will discuss in Section 2.2, it is known from the literature that combining such attributes can constitute enough information to uniquely re-identify individuals, rendering those disclosure control methods adopted by INEP insufficient [11, 56, 65, 67].

This vulnerability in the microdata released by INEP was already observed by two Brazilian researchers back in 2015 [63]. In that work, Queiroz and Motta were able to re-identify one of the authors among 383 683 records of lecturers from the Higher Education Census of 2013 by using only the date of birth, gender, and the Higher Education Institution name. However, they did not provide an analysis of how common this re-identification risk was in that database, i.e. how vulnerable were data holders other than the considered author.

In the same work, Queiroz and Motta suggested the use of two other disclosure control methods to increase the privacy of individuals in the INEP databases, the “syntactic methods” known as “k-anonymity” [65] and “distinct l-diversity” [53], both discussed in Section 2.2.2. Those techniques were applied by them with assistance of the ARX Data Anonymization Tool [62] and the results were significant, at least at first sight.

For instance, before applying any anonymization technique, the ARX tool reported that 100% of the records in the database were subject to a *high* risk of re-identification. After anonymizing the database using the parameter values of $k = 2$ and $l = 2$, the ARX tool reported that 50% of the records were then at a *high* risk of re-identification. By further increasing the parameters values to $k = 100$ and $l = 2$ and then to $k = 100$ and $l = 9$, the new proportions of records at a *high* risk of re-identification dropped to 0.99010% and to 0.00412%, respectively. However, the absolute number of records subject to a *high* risk of re-identification is still considerable even for the most strict parameters values.

Furthermore, as we will discuss in Section 2.2.2.1, any syntactic method, including both k-anonymity and distinct l-diversity, is subject to composition attacks, which is particularly problematic for the studies published by INEP given their annual frequency. Also, the paper did not discuss the high costs on utility, i.e. the quality of the information provided to legitimate consumers of the data, caused by such gains in privacy, a well-known trade-off in the literature.

Given the recent enactment of Brazil's LGPD privacy legislation, mitigating such vulnerabilities is necessary and urgent for INEP. Currently available work, such as the one by Queiroz and Motta, provides only very limited evidence of the actual privacy risks and how widespread they are. Moreover, no analysis exists for the longitudinal aspect of INEP's databases, which are released with annual frequency and hence particularly susceptible to composition attacks, a known vulnerability of syntactic methods. Therefore, in this thesis we analyze "semantic methods", particularly "differential privacy". Furthermore, Brazil's LAI transparency legislation, enacted ten years ago, has developed an expectation of access to microdata in the country, both from researchers and policy makers, which increases the difficulty in convincing stakeholders of any changes in the utility that should be provided by the published databases.

We now present our objectives and contributions in Sections 1.4 and 1.5, respectively.

1.4 Objectives

Our main goal in this thesis is to provide a formal, extensive quantitative evaluation of privacy risks relative to INEP's current disclosure control (DC) methods, grounded on solid, formal theoretical frameworks. Moreover, we evaluate the application of differential privacy, the current state-of-the-art DC method, and analyze the balance between privacy and utility in two of its variants.

Particularly, we have the following research questions:

RQ1 What are the data holders' risks in a single database given the current disclosure control (DC) methods applied by INEP?

(a) What are the re-identification risks, i.e. the risks of a complete disclosure of a data holder's identity?

(b) What are the attribute-inference risks, i.e. the risks of inference for an attribute's value of a data holder?

(c) Would the removal of attributes from the databases be effective as an additional DC method?

RQ2 What are the data holders' risks when considering the longitudinal aspect of the databases released, i.e. the continuous release of new databases, given the current DC methods applied by INEP?

(a) What are the re-identification risks?

(b) What are the attribute-inference risks?

(c) Would the removal of attributes from the databases be effective as an additional DC method?

RQ3 What is the effectiveness of differential privacy to mitigate data holders' risks and to maintain data utility?

(a) What is the effectiveness for the "oblivious" differential privacy model, in which a trustworthy party with access to the original, raw data from respondents is responsible for controlling the queries performed on the database?

(b) What is the effectiveness for the "local" differential privacy model, in which the data is changed at the record level independently of the existence of a trustworthy party?

1.5 Contributions

The contributions of this thesis are summarized as follows.

- We propose a new classification of attacks against releases of databases in Section 3.1 that better covers the space of possible attacks in comparison to the literature. We also develop a database model for single databases in Section 3.2 that we further extend to account for longitudinal databases in Section 5.1, in the educational censuses' context. Those database models and the following attack models were developed on the same framework, which has allowed us to directly compare their respective results;

- We formally model collective-target attacks on single databases in Section 4.1 for both re-identification and attribute-inference privacy risks and considering both deterministic and probabilistic metrics for the adversary success, in the educational censuses' context. We also provide experimental results for all of those scenarios in Section 4.2. Those experiments were performed on databases released by INEP consisting of tens of millions of individuals and demonstrate the existent vulnerabilities in those databases. This contribution supports our conclusions for **RQ1**;

- We formally model collective-target attacks on longitudinal databases in Section 5.1 for both re-identification and attribute-inference privacy risks and considering both deterministic and probabilistic metrics for the adversary success, in the educational censuses' context. We also provide experimental results for all of those scenarios in Section 5.2. Those experiments were performed on databases released by INEP and demonstrate the existent vulnerabilities in those databases. This contribution supports our conclusions for **RQ2**;

- We formally model global and local differential privacy mechanisms in Section 6.1 to analyze how each of those methods affect both privacy and utility, particularly in the context of possibly correlated-databases. We also implement those models and provide experimental results for those analyses in Section 6.2. Those experiments were performed on a sample from one of INEP's databases and demonstrate the impossibility of having optimal noise-adding mechanisms with the best possible privacy or utility in all scenarios. This contribution supports our conclusions for **RQ3**.

Furthermore, we provide the largest, most thorough study of actual privacy threats in official government data releases in Brazil. In doing so, our results demonstrate the privacy risks to which more than fifty million current Basic Education students are subject. Also, our work has contributed to INEP's current efforts to tackle privacy issues given the recent enactment of Brazil's LGPD privacy legislation. Particularly, our work was fundamental to Reports 1 and 2, and contributed to Report 5, of the Decentralized Execution Term 8750 (*Termo de Execução Descentralizada 8750*, in Portuguese) signed between INEP and the Federal University of Minas Gerais (*Universidade Federal de Minas Gerais*, or UFMG, in Portuguese), whose goal was to improve INEP's public release of microdata in the context of the new privacy legislation. [7]

Finally, our work presented in Chapter 6 was published in the *31st International Conference on Concurrency Theory* (CONCUR) [6], where we demonstrated our theoretical results with experiments performed on data from the *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS) tool released by *ProPublica*.

1.6 Outline

Here we summarize how the following chapters are organized.

In Chapter 2, we define some basic concepts from the literature on *Disclosure Control* (DC) and review the literature on DC methods and on *Quantitative Information Flow*

[7] This partnership between INEP and UFMG was initially sought by the former given the new privacy legislation in Brazil and the previous awareness of existent privacy vulnerabilities in the released databases [63].

(QIF). We also present some related work on *re-identification* and *attribute-inference attacks* and some proposed classifications of attacks on releases of databases.

In Chapter 3, we propose a new classification of attacks against releases of databases that better covers the space of possible attacks in comparison to previous propositions, and we introduce our model of databases used for the experiments performed.

In Chapters 4 and 5, we formalize collective-target attacks against single and longitudinal databases, respectively. Those chapters also include the respective experimental results for attacks performed against the Basic Education databases released by INEP and described in Section 1.3. Chapter 4 explores the **RQ1** and Chapter 5 explores the **RQ2**, as proposed in our objectives presented in Section 1.4.

In Chapter 6, we formalize privacy and utility analyses for two different implementations of differential privacy, a state-of-the-art DC method described in Chapter 2. Due to the challenge of extending differential privacy to longitudinal databases, we consider only scenarios for single databases. We also present the corresponding experimental results for analyses performed on a small sample from one of the Basic Education databases. Chapter 6 explores the **RQ3**, as proposed in our objectives.

We present our conclusions in Chapter 7, which is followed by some appendices.

- Appendix A presents an overview of the privacy legislation around the world, which expands on the brief legal review presented in Section 1.2.
- Appendix B presents an overview on why to consider "min-entropy" instead of "Shannon entropy" for some scenarios, including the concept of "vulnerability".
- Appendix C presents illustrative individual-target attack models and some experimental results on databases released by INEP. Single databases are presented in Section C.1 and longitudinal databases in Section C.2.
- Appendix D presents detailed examples of each individual and collective-target attack performed. Single databases are presented in Section D.1 and longitudinal databases in Section D.2.
- Appendix E presents additional experimental results for collective-target attacks on single databases performed on the School Census of 2019.
- Appendix F presents additional experimental results for collective-target attacks performed on the Higher Education Census, on single databases in Section F.1 and longitudinal databases in Section F.2.

- Appendix G presents some remarks, propositions, and proofs on the theoretical model developed in Chapter 6.

For readers interested only in the experimental results, we suggest skipping Sections 4.1, 5.1, and 6.1 on the theoretical foundations in favor of Appendix D Sections D.1 and D.2, and Section 6.1.4, respectively, in which we present numerical examples. Furthermore, for illustrative individual-target experimental results, in which we target either famous people or our acquaintances selected *a priori*, see Appendix C.

Chapter 2

Background

The contributions in this thesis are built upon two main areas of knowledge, namely *Disclosure Control* (DC) and *Quantitative Information Flow* (QIF). We review the most important aspects of those areas in this chapter, beginning with the definition of some basic concepts from the DC literature in Section 2.1. Next, we review the literature on DC methods for databases release in Section 2.2, on QIF in Section 2.3, and on some related work on "re-identification" and "attribute-inference" attacks, including some proposed attack classifications, in Section 2.4.

2.1 Basic concepts

In this section, we formally introduce some basic concepts from the DC literature that are used throughout this thesis. Although some of those terms may have an everyday use, in the literature their meaning is precise [5, 14, 15, 23, 29, 61, 66].

2.1.1 Disclosure Control

Disclosure Control (DC) is the area of research that focus on how to publicly share the information contained in databases while preserving the privacy of "data holders". Its main goal is to guarantee that statistical patterns are revealed while the sensitive information of "data holders" is kept safe, i.e. to balance the information's usefulness, or "utility", for legitimate users with the privacy of "data holders". We now introduce the main concepts of the area.

- **Data holder**: the person or entity to whom the data collected, processed, or disclosed directly refers.

- **Data curator**: the person or entity responsible for collecting, integrating, organizing, publishing, and presenting data. A data curator is usually responsible for disclosure control, which includes maintaining the data over time for update, reuse, and preservation.

- **Data analyst**: the person or entity considered to be a legitimate user of a database release, including researchers, civil and governmental entities, and even lay people.

- **Utility**: the property of a database release to satisfy the legitimate purposes of data analysts. It is usually linked to a data analyst's ability to infer statistical information on data holders.

- **Sensitive information**: any kind of information considered worthy of protection. It may be (even if not necessarily) defined by law. Some common examples of sensitive information are those belonging to individuals, such as credit and debit card numbers, medical conditions, unique identification numbers, geolocation data, political or religious affiliation, ethnicity, sexual orientation, and genetic or biometric data. [1]

  Here we consider only the following categories of sensitive information: [2]

  - the linkage of a record to its holder, which can be determined by the "re-identification" of a data holder;
  - the value of an individual's sensitive attribute, which can be determined by the "re-identification" of a data holder or by "attribute-inference".

- **Adversary**: the person or entity from which sensitive information must be protected. In contrast to the everyday use of the term, an adversary may be anyone, regardless of intentions, as long as their access to the sensitive information was not planned in the design of the database release. For example, a legitimate user of a database, such as an academic researcher, could inadvertently act as an adversary if infer sensitive information on individuals in the database.

[1] Brazil's LGPD privacy legislation explicitly defines "sensitive personal data" as [40]:

> Personal data on racial or ethnic origin, religious belief, political opinion, union membership or affiliation to organizations of a religious, philosophical, or political nature, data relating to health or sexual life, genetic or biometric data, when linked to a natural person.

[2] As discussed in Section 1.3, we can always assume for our databases of interest that every individual of concern is also a data holder.

- **Attack**: an action or a set of actions by an adversary that leads to the obtention of sensitive information. In contrast with the everyday use of the term, an attack may be the result of legitimate actions that inadvertently reveal sensitive information, such as unintended inferences during scientific research.
- **Auxiliary information**: any sort of information, apart from the database of interest, that could provide information about an individual. For instance, the value of an individual's date of birth could be retrieved from a social network public profile or other web-based resources.
- **Plausible deniability**: the ability of a data holder to credibly deny any property about their sensitive information to an adversary. This includes the denial of a value for a certain attribute or even of participation in a database, before or after an attack by the adversary. The possibility of plausible deniability is essential to persuade individuals to honestly participate in research involving the provision of sensitive information [16, 78]. Furthermore, privacy metrics are usually based on quantifying the level of plausible deniability enjoyed by data holders.

The concepts defined above are valid for the following DC subareas: *Privacy-Preserving Data Publishing* (PPDP), in which the information is released in a database as microdata, i.e. data at the record level; *Privacy-Preserving Data Mining* (PPDM), in which the information is released as aggregated data tailored for specific data mining purposes; and *Statistical Disclosure Control* (SDC), in which the information is released as statistical tables for general use or with the indirect access to data through query-based systems [23]. Note that, for both PPDM and SDC, the microdata is kept from data analysts. Furthermore, all those subareas are concerned with guaranteeing that statistical patterns are revealed while the sensitive information of data holders is kept safe, but their differences on how the information is published results in the use of different DC methods to achieve this goal.

Even though data analysts cannot access the microdata in PPDM or SDC, privacy-preserving techniques must also be carefully applied due to what is known as *database reconstruction attacks*. As demonstrated by Dinur and Nissim in 2003 [13], even a database accessed via a query-based system that reports a randomly perturbed, or noisy, result can be reconstructed if that noise's magnitude is much lower than the square root of the database size. This result was the main reason why the United States Census Bureau initiated its transition to formally proven methods of random noise addition in order to protect their statistical tables releases, starting with the 2020 Census [26, 76, 77].

However, this thesis focus on INEP's databases released as microdata, which are specified in Section 1.3. Therefore, we will consider henceforth only the Privacy-Preserving Data Publishing (PPDP) area, starting with commonly used classifications for attributes and privacy risks in Sections 2.1.2 and 2.1.3, respectively.

2.1.2 Attributes classification

The literature on PPDP often considers the attributes present in a database publication in some specific classes. Particularly, "sensitive attributes" are of great interest in the area, since protecting the data holders' values for those attributes is usually the main goal in disclosure control. We now introduce a commonly used classification of attributes [23, 29, 61].

- **Direct identifiers**: attributes that can uniquely identify an individual, such as full name, unique identification numbers, mobile phone number, and genetic or biometric data.
- **Quasi-identifiers**: attributes that cannot be used to uniquely identify individuals on their own, but could be used for this purpose if combined. Differently from direct identifiers, quasi-identifiers are attributes that are sufficiently correlated with an individual such that a set of quasi-identifier attributes could be used to re-identify the individual. For instance, an individual's ZIP code, gender, and date of birth form a set of quasi-identifiers capable of uniquely identifying 87% of the United States population in the Census of 1990, as demonstrated by Sweeney in 2000 [67].
- **Sensitive attributes**: those whose values are considered worthy of protection, such as credit and debit card numbers, medical conditions, unique identification numbers, geolocation data, political or religious affiliation, ethnicity, sexual orientation, and genetic or biometric data.
- **Non-sensitive attributes**: those whose values are not considered worthy of protection, such as an individual's height.

It is important to note that this classification of attributes depends directly on the scenario of interest, i.e. some attributes considered to be sensitive in a context may not be so in another, and an attribute may not be a viable quasi-identifier at one point in time, but be so at another. For this reason, more recent DC methods are formulated independently of quasi-identifiers altogether, e.g. "differential privacy".

2.1.3 Privacy risk classification

Given a specific scenario of interest, privacy risks can be classified as follows [23, 29, 61].

- **Re-identification risk**: also known as *record-linkage* or *de-anonymization*, it is the risk that an adversary determines the correspondence between database records and their respective data holders.
- **Attribute-inference risk**: also known as *attribute-linkage*, it is the risk that an adversary infers the sensitive attribute value for data holders, regardless of being successful in re-identifying their respective data holders. The re-identification of a record implies the disclosure of values for every attribute of that record.
- **Membership-inference risk**: also known as *table-linkage*, it is the risk that an adversary determines whether an individual or a group of individuals are represented in a database or not. This is of particular concern if being a data holder in said database is already considered to be sensitive, e.g. a database containing only individuals affected by a medical condition.

In this thesis we focus on re-identification and attribute-inference risks, since we can always assume for our databases of interest that every individual of concern is also a data holder, as discussed in Section 1.3.

2.1.4 Privacy degradation

In order to properly determine whether there is privacy degradation for data holders in a given context or not, it is first necessary to precisely measure privacy. Here we introduce the theoretical foundation for privacy as proposed in the area of *Quantitative Information Flow* (QIF) [2–6, 66]. We further formalize the concepts proposed in the QIF literature in Section 2.3.

- **Maximum entropy principle**: it states that, given the known information, the probability distribution that best represents a state of knowledge is that in which the entropy has its maximum value. Therefore, in the absence of any information, i.e. a state in which nothing is assumed from the adversary's knowledge, the best possible distribution would be the uniform distribution and the best possible action for the adversary would be to choose a random value for the secret [49].

- ***A priori* knowledge**: even before the attack is performed, we assume the adversary may have some sort of *a priori* knowledge on the secret. Such knowledge may come, for instance, from previous attacks, from auxiliary information, or even from information on how the secret was sampled and therefore how it may be distributed. As an example, consider an individual's weekday of birth as the sensitive information and the adversary knowledge that, for the given population, the distribution of birthdays in a week is close to uniform.

- ***A priori* success**: it is a measure of how much knowledge on the sensitive information is available to the adversary even before performing the attack, i.e. by relying solely on the *a priori* knowledge. Consider again the adversary whose prior knowledge on date of birth is uniform. His best course of action would be to randomly select any day of the week as the weekday of birth for the target, according to the maximum entropy principle. Therefore, the *a priori* success of the adversary could be quantified as $1/7 \approx 14\%$.

- ***A posteriori* knowledge**: it is the blend of the adversary's *a priori* knowledge with the information obtained from observing the results of the performed attack. As an example, consider again an individual's weekday of birth as the sensitive information and the adversary knowledge that, for the given population, the distribution of birthdays in a week is close to uniform. Only now, after performing an attack, the adversary was able to verify that the target was born on a weekend and that both Saturday and Sunday are equally likely.

- ***A posteriori* success**: it is a measure of how much knowledge on the sensitive information is available to the adversary after performing the attack and updating the *a priori* knowledge. Therefore, the *a posteriori* success of the adversary we have been considering could be quantified as $1/2 \approx 50\%$.

- **Privacy degradation**: it is a comparison between the *a priori* and *a posteriori* knowledge of the adversary such that a greater privacy degradation implies a better outcome for the adversary from their attack. [3]

[3] Instead of being defined as the adversary's increase in success, one could define privacy degradation as *a posteriori* success alone. However, this alternative definition would not capture by how much the adversary's chance of success has increased due to the performed attack on itself, which is our object of study. For instance, consider again an individual's weekday of birth as the sensitive information and two adversaries: adversary A knows *a priori* that the target was born on a Sunday, but adversary B only knows *a priori* that the target was born on a weekend and that both Saturday and Sunday are equally likely. Suppose that after the attack both adversaries managed to re-identify the target's record, i.e. both adversaries achieve maximum *a posteriori* success. If the privacy degradation only consisted of the *a posteriori* success, both attacks would have caused the same amount

The comparison between the *a priori* and *a posteriori* knowledge could be made in terms of a ratio or of a difference. In the case of a ratio, we have a *multiplicative measure of privacy degradation* [2]:

$$\begin{array}{c}\textit{multiplicative privacy}\\ \textit{degradation}\end{array} \overset{\text{def}}{=} \frac{\begin{array}{c}\textit{adversary's a posteriori}\\ \textit{success}\end{array}}{\begin{array}{c}\textit{adversary's a priori}\\ \textit{success}\end{array}}\,.$$

In the case of a difference between the *a priori* and *a posteriori* knowledge, we have an *additive measure of privacy degradation*:

$$\begin{array}{c}\textit{additive privacy}\\ \textit{degradation}\end{array} \overset{\text{def}}{=} \begin{array}{c}\textit{adversary's a posteriori}\\ \textit{success}\end{array} - \begin{array}{c}\textit{adversary's a priori}\\ \textit{success}\end{array}.$$

Consider again the adversary presented when defining the *a priori* and *a posteriori* successes. For this adversary, the multiplicative measure of privacy degradation would be of ${}^{({}^1/_2)}/_{({}^1/_7)} = {}^7/_2 = 3.5$, therefore a relative gain of 250% in the chance of successfully inferring the secret's value. But then, the additive measure of privacy degradation would be of ${}^1/_2 - {}^1/_7 = {}^9/_{14} \approx 0.36$, therefore an absolute gain of about 36% in the chance of successfully inferring the secret's value.

We finalize the introduction of basic concepts from the DC literature with an overview of privacy metrics in the next section.

2.1.5 Privacy metrics

Continuing on the task of precisely measuring privacy, we can devise two main groups for both *a priori* and *a posteriori* successes when it comes to quantifying the privacy degradation: the deterministic and probabilistic metrics of success.

- **Deterministic metrics of success**: qualitatively measure the adversary's absolute success, i.e. the returned value encompasses whether the sought after information was retrieved or not with absolute certainty. As examples of deterministic metrics of success, we have the number of uniquely re-identified individuals in a database or the number of individuals whose value for a sensitive attribute could be inferred with absolute certainty.

of privacy degradation for the target. Although in fact, the attack performed by adversary B has caused more harm to the target since it has started from a state of less knowledge than was the case for adversary A. Therefore, in order to quantify the harm caused by an attack, one should consider both *a priori* and *a posteriori* knowledge of the adversary, not only the latter.

- **Probabilistic metrics of success**: quantitatively measure the degree of certainty in which a sensitive value could be retrieved by the adversary. As examples of probabilistic metrics of success, we have the probability that a randomly selected individual in a database could be re-identified or the probability that such individual could have the value for a sensitive attribute inferred, even without absolute certainty.

Furthermore, privacy metrics can be classified according to the adversary's degree of success, which allows us to consider both the average case and worse case metrics.

- **Average-case metrics**: quantitatively measure the expected, or average, success of the adversary, considering a probability distribution over all the possible attack scenarios. For instance, consider the expected re-identification risk for a randomly selected targeted-individual in a database. In this case, all the individuals that can be chosen in the database have their respective re-identification risks computed and the results are used to compute the weighted average risk considering a probability distribution over individuals, e.g. the uniform distribution. A downside of average case metrics is that they can hide higher values of privacy risk for few individuals.
- **Worst-case metrics**: quantitatively measure the adversary's success according to the most advantageous attack scenario for them, regardless of how unlikely it is. In this case, all the individuals that can be chosen in the database have their respective re-identification risks computed and the worst result is taken as the final one. A downside of worse case metrics is that they may not represent the general case. But then it is the most appropriate in the case of skewed distributions [4, 5].

2.2 Disclosure Control methods

As we have discussed, the publication of databases may subject data holders to different privacy risks. This possibility is of great concern when it comes to the feasibility of statistical studies. First, because the degradation of individuals' privacy may be illegal, which comes as a great trouble for data curators. Second, because this possibility may diminish individuals' willingness to participate in such studies, or when mandatory, encourage the share of inaccurate information.

Due to the societal benefits that may come from statistical studies, it is important to keep the general public safely engaged with them, a goal widely accepted by the international community. However, since anyone might act as an adversary, even if inadvertently, it is necessary to go beyond legislation and consider means to decrease an adversary's chances of success without jeopardizing the databases' usefulness.

An initial attempt to overcome those concerns could be to *de-identify* the data, i.e. to remove any direct identifiers, such as individuals' names or government-issued ID numbers. This approach has been widely used and can still be found in databases publications nowadays. However, de-identification of data is known to be an insufficient measure since at least the 1980s, according to Dalenius [11]. In fact, it was demonstrated to be unsuccessful as early as 1998 [65], when Samarati and Sweeney first observed that just a few quasi-identifying attributes were enough to enable the re-identification of individuals. For instance, Sweeney was able to demonstrate in 2000 that 87% of the United States population in the Census of 1990 could be uniquely identified by using only a combination of the ZIP code, gender, and date of birth [67], information that can be found in public voter lists. Hence, de-identification of data does not preserve privacy nor guarantee anonymity.

Those results made it clear that new DC methods were necessary. We can nowadays devise two main distinct approaches to data privacy that have been developed since then: "deterministic anonymization methods", or "syntactic privacy methods", and "probabilistic anonymization methods", or "semantic privacy methods".

2.2.1 Classification of Disclosure Control methods

Disclosure Control (DC) methods are sets of actions that can be performed by a data curator, and sometimes by a data holder, in order to mitigate the data holder's privacy degradation. Such methods can be classified according to the conditions they impose on the data to be released and to how and at what level their actions change the data.

We start with the dimension regarding the conditions imposed, which classify DC methods as either "syntactic" or "semantic".

- **Syntactic methods**: they establish syntactic constraints on the data to be published, i.e. they are concerned with the released data structure, which includes constraints on how many individuals should hold the same values for certain attributes in order to keep them safe.

- **Semantic methods**: they establish semantic constraints on the data to be published, i.e. their privacy guarantees are defined according to the amount of information an adversary may gain from the released data, being independent of data structure.

The second dimension, which regards how the data is changed, classify the methods as either "perturbative" or "non-perturbative".

- **Perturbative methods**: they may change the data by adding noise to it in a controlled way, i.e. they may introduce untrue information to the released data.
- **Non-perturbative methods**: they change the data by tweaking its granularity, i.e. they apply actions such as generalization or suppression of information. However, they do not introduce untrue information.

Finally, the third dimension, which regards where the data is changed, classify the methods as either "global" or "local".

- **Global methods**: they account for a data curator responsible for the collection of individuals' data and subsequent application of DC methods, i.e. the data curator is assumed to be a trustworthy entity for handling sensitive information and DC methods are applied at once to the data as a whole.
- **Local methods**: they usually account for the data holder to be responsible for applying a pre-defined DC method to the data prior to submitting it to the data curator, i.e. the raw private information is only accessible by the data holder. [4]

2.2.2 Syntactic methods

As discussed, syntactic or deterministic methods apply syntactic restrictions on their outputs to avoid the unintended disclosure of information. Among the restrictions usually applied we have "generalization", "suppression", and "swapping".

- **Generalization**: values of an attribute are replaced by a more generic category, e.g. age ranges instead of exact values.
- **Suppression**: values of an attribute are replaced by a special value, e.g. hiding of some final numbers of the ZIP code by using an asterisk.

[4] Local methods can also be applied by a trustworthy data curator in case microdata, i.e. data at the record level, is to be released.

- **Swapping**: values of an attribute for different records are swapped. Differently from generalization and suppression, which are non-perturbative techniques, swapping can be seen as a perturbative technique since it may introduce untrue information to the released data.

In the same paper published by Samarati and Sweeney in 1998, when they first observed that just a few quasi-identifying attributes were enough to enable the re-identification of individuals in de-identified data, they also proposed a novel anonymization method called *k-anonymity* [65]. Given a database, it is considered to be k-anonymous if any attempt to link quasi-identifying attributes to sensitive information creates a many-to-one mapping in which at least k records map to the same sensitive information, i.e. the database would be partitioned in blocks of at least k records in each. In order to achieve this result, Samarati and Sweeney proposed the use of generalization and suppression.

However, as shown by Aggarwal [1] in 2005, k-anonymity robustness is inversely dependent on the dimensionality of the database, i.e. on the number of attributes that can be used as quasi-identifiers. Furthermore, Machanavajjhala et al. unveiled in 2007 two severe weaknesses in k-anonymity [53]. First, they showed that a low diversity in the values assumed by a sensitive attribute allows an attacker to discover those values within a block of records. Second, k-anonymity does not provide any privacy guarantee against the use of background knowledge by an adversary, i.e. the use of auxiliary information.

In that same paper describing k-anonymity's weaknesses, Machanavajjhala et al. proposed an alternative approach called *l-diversity*, which improves on k-anonymity by also requiring at least l "well-represented" values for the sensitive attribute within each block. By "well-represented" values, they suggested some different interpretations, the most common being *distinct l-diversity*, which requires at least l distinct values for the sensitive attribute within each block, automatically satisfying k-anonymity with k equal to l. This interpretation does not prevent probabilistic inference attacks, though, since it does not prevent a value from appearing much more frequently in a block of records than other values, which increases the likelihood of that value for any individual who may hold a record withing the given block.

Nevertheless, it did not take long before l-diversity would also have its limitations identified. In the same year it was proposed, Li et al. showed that l-diversity would fail if the distribution of values for the sensitive attribute in the whole database was skewed. In fact, Li et al. demonstrated that l-diversity is neither necessary nor sufficient to prevent sensitive attributes from being disclosed and proposed another approach called *t-closeness* [52]. Again improving on k-anonymity, t-closeness requires that the distribution of values for a sensitive attribute in any of the blocks of at least k records to be at least as close to the distribution of values for that attribute in the whole database as a threshold value t.[5]

Differently from other syntactic methods, t-closeness also has some characteristics found in semantic methods. Because it tries to limit how much information an adversary can gather from the published database, it applies the "uninformative principle". [6] Furthermore, Domingo-Ferrer and Soria-Comas showed in 2015 that, under certain conditions, t-closeness and "ϵ-differential privacy" can be equivalent, particularly if the "uninformative principle" holds.

2.2.2.1 Composition attacks

Despite the improvements provided by t-closeness, Ganta et al. [24] showed in 2008 that any technique based on the definition of blocks of equivalent records is susceptible to *composition attacks*. In such attacks, the adversary relies on independent databases released with some attributes in common and protected by only DC methods based on blocks of equivalent records. By combining similar blocks between those databases, the adversary can reduce the size of those blocks and increase their knowledge of the sensitive attribute value for records within those groups. Interestingly, Ganta et al. also showed that privacy mechanisms based on the addition of random noise, such as "ϵ-differential privacy", are resistant to composition attacks and to the use of auxiliary information.

Still, de-identification, pseudonymization, and some syntactic methods such as k-anonymity continue to be used nowadays, particularly when it comes to the anonymization of medical records. For instance, the United States *Health Insurance Portability and Accountability Act* (HIPAA) only requires in its *Safe Harbor* rules that a set of 18 predetermined data attributes, the *Protected Health Information* (PHI), are removed in order to achieve compliance [29].

Similarly, both Brazil's LGPD [40] and the European Union's GDPR [72] explicitly define pseudonymization and incentive its use, even though they still consider pseudonymized data to be personal data. Furthermore, both laws loosely define

[5] In their original work, Li et al. used the *Earth Mover Distance* measure as their t-closeness requirement.

[6] The uninformative principle holds if the adversary's *a posteriori* knowledge on any data holder is not much larger than their *a priori* knowledge.

anonymized data, which is not to be considered personal data and hence is not protected as such, leaving the decision of which DC method to be used either to the data curator or to future legislation.

Several other syntactic methods have been proposed since k-anonymity and were not presented here. Most of them up to 2010 were compiled in a comprehensive review by Fung et al. [23].

2.2.3 Semantic methods

As we have discussed in the previous section, syntactic methods are not robust enough to ensure privacy in realistic scenarios. Particularly, those methods are concerned with the released data structure instead of information measurements that could quantify how much knowledge an adversary could obtain from that data. Hence, new methods that account for the adversary's acquired knowledge have been proposed and are currently considered the state-of-the-art in disclosure control.

Semantic or probabilistic methods do not depend on the data itself, but instead on how much information an adversary may gain from it. They are built upon the concept of plausible deniability, i.e. the ability of a data holder to credibly deny any property about their sensitive information to an adversary, and upon the *uninformative principle*, i.e. the adversary's *a posteriori* knowledge on any data holder should not be much larger than their *a priori* knowledge. [7]

A foundational work for the semantic methods was the demonstration by Dinur and Nissim in 2003 [13] that there is a minimum amount of noise that must be added to the results of sum queries performed on a database in order to guarantee privacy. Otherwise, the adversary can reconstruct the underlying database in almost its entirety, hence potentially violating the privacy of all data holders.

Building upon that result, Dwork et al. proposed in 2006 a novel semantic definition of privacy, which would be latter known as ϵ-*differential privacy*, capable of guaranteeing to individuals that being a data holder in a differentially private database would not

[7] The uninformative principle in a more strict version is known as the *Dalenius' Desideratum*, after the statistician Tore Dalenius, who formally defined statistical disclosure and proposed the development of a methodology for statistical disclosure control in 1977 [10]:

> If the release of the statistics S makes it possible to determine the value D more accurately than is possible without access to S, a disclosure has taken place.

The Dalenius' Desideratum was proposed in the context of both macro and micro-statistics, i.e. aggregated data and microdata, respectively. However, Dalenius knew it is unachievable, what was latter demonstrated by Dwork et al. in 2006 [15].

increase an adversary's knowledge in any way [15]. This new definition of privacy would hence respect the uninformative principle and provide the data holder with plausible deniability even against an adversary with arbitrary background knowledge.

In order to guarantee ϵ-differential privacy, where ϵ is a positive real value, we must consider a query mechanism $\mathcal{M}$ and two adjacent databases D, D', i.e. that differ exactly in one record, be it because of this record's removal or addition, or due to a change in one of its attributes' value. The query mechanism $\mathcal{M}$, also responsible for perturbing the results, receives a database and outputs a response in $img(\mathcal{M})$. This mechanism is considered to be ϵ-differentially private iff for every pair of adjacent databases D, D' and for all $S \subseteq img(\mathcal{M})$,

$$Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon} \cdot Pr[\mathcal{M}(D') \in S], \tag{2.1}$$

i.e. the probability that the mechanism's reported result belongs to a set S regardless of the input being D or D' is given by a multiplicative constant value equal to Euler's constant e to the power ϵ.

Differently from the syntactic methods discussed in Section 2.2.2, ϵ-differential privacy satisfies some important operational properties[14, 16]. Among them, the most interesting are the *automatic mitigation of linkage attacks*, including past and future databases; the definition of a *quantitative privacy loss*; the possibility of both *sequential* and *parallel compositions*, i.e. the control of privacy loss over multiple computations; and the *closure under post-processing*, i.e. there is no computation the adversary can perform on the ϵ-differentially private output in order to increase their knowledge.

We now present the two main frameworks for differentially private mechanism.

2.2.3.1 Oblivious differential privacy

The *oblivious differential privacy* framework considers a trustworthy data curator with access to the original, raw data from respondents and is responsible for controlling the queries performed on the database. For each query passed by an external data analyst, the data curator performs the required computations on the original database and gets the real answer. However, the data curator perturbs the real answer according to a chosen differentially-private mechanism and value of ϵ before publication, making those query results ϵ-differentially private, as schematized in Figure 2.1. Only then the randomized answer is provided to the data analyst. This framework is based on the original work by Dwork et al. [15], which allowed for the use of general queries, not only summations as was the case for the work by Dinur and Nissim [13].

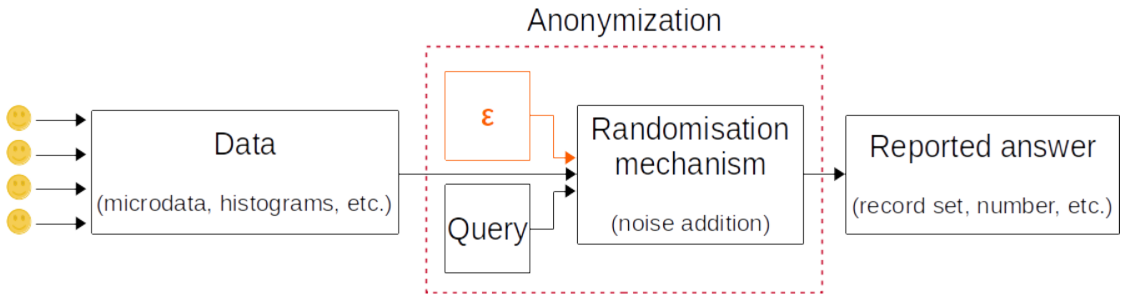


Figure 2.1: Schema of oblivious differential privacy. The data curator maintains the original, raw data provided by the data holders and adds noise to the result of each query made on the database. The amount of noise is set by the value of ϵ, which controls how much privacy is guaranteed to data holders.

Nonetheless, care must be taken when it comes to the amount of queries allowed by the data curator. Since, for practical purposes, there is an upper limit for noise addition given by ϵ, an adversary with indiscriminate access to the query system could repeatedly perform the same query until the desired response accuracy is achieved. This corresponds to the idea of a *privacy budget* defined by the value of ϵ that should be respected in order to guarantee the disclosure control protections provided by ϵ-differential privacy.

2.2.3.2 Local differential privacy

The *local differential privacy* framework considers either a trustworthy data curator with access to the original raw data from respondents or a possibly untrustworthy data curator who should not have access to that data. In both cases, the chosen mechanism perturbs the data at the record level, making the microdata itself ϵ-differentially private. This allows any data analyst to directly perform queries on the database without worrying about possible disclosures or even the publication of the whole database for widespread use, as schematized in Figure 2.2. This framework was first proposed by Raskhodnikova et al. in 2008 [64] and is based on the concept of “randomized responses”, first proposed by Warner in 1965 [78].

Warner was concerned with the problem of convincing individuals to truthfully respond to statistical studies even if the responses would reveal sensitive information. In such cases, Warner argues that respondents could either refuse to answer or purposely provide wrong answers. As a solution, Warner proposed the use of “randomized responses”, in which the respondent would be truthful or not in their answer according to the result of a predetermined probability mechanism. Because the mechanism is known to the interviewer, they are capable of estimating the likelihood and of establishing the con-

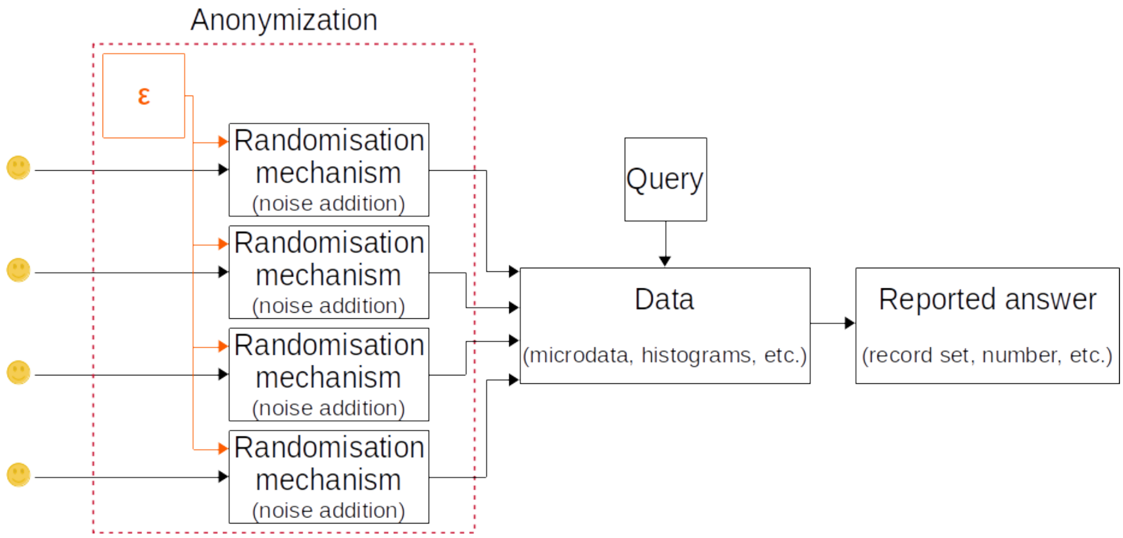

Anonymization
ε
Randomisation mechanism (noise addition)
Randomisation mechanism (noise addition)
Randomisation mechanism (noise addition)
Randomisation mechanism (noise addition)
Query
Data
(microdata, histograms, etc.)
Reported answer
(record set, number, etc.)

Figure 2.2: Schema of local differential privacy. Noise is added to the record provided by each data holder before all records are gathered to create the final, ϵ-differentially private database. The amount of noise is set by the value of ϵ, which controls how much privacy is guaranteed to data holders. After that, the data curator can either allow queries to be made on the private database or publicly release the whole database for independent use by data analysts.

fidence intervals and variance, due to sampling and due to the probability mechanism, which guarantees the data is useful [78].

A similar approach was proposed by Raskhodnikova et al. in 2008 [64], when they demonstrated the equivalence between the differential privacy statistical query model, i.e. the global differential privacy framework, and their newly proposed local differential privacy framework.

Even though semantic methods have not been as popular as their syntactic counterparts, differential privacy has been gaining some traction in the past decade. Particularly, it is currently being deployed by the United States Census Bureau for the 2020 Census [26] to circumvent the drawbacks inherent to syntactic methods. Also, it has already been implemented by some major technology companies mainly to collect telemetry from their most used software, including Microsoft's Windows 10 [12], Apple's iOS [69, 70], and Google's Chrome Browser [20].

We finalize the introduction of DC methods in the next section by discussing their impact on data utility and how to properly measure it.

2.2.4 Utility metrics

Another important aspect of any DC method is its impact on the utility, i.e. usefulness, of the published database for the intended data analysts. The difficulty in balancing data holders' privacy and data utility is widely known in the literature [6, 7, 15, 50], particularly because distinct data analysts may have different goals [27].

However, achieving a *universally optimal mechanism*, i.e. one that would preserve data utility in an optimal level for all, or at least for most data analysts, has been possible only in a few and strict scenarios [7, 50].

Notwithstanding, we can still classify the utility metrics from the literature in at least two categories: those of "specific" and those of "general" purposes.

- **Specific-purpose utility-metrics**: they focus on a predetermined, specific task for the published data and measure its usefulness only for that task. For example, data released for a specific machine learning classification task would have its utility measured according to the intended classification.

- **General-purpose utility-metrics**: they try to diminish the distortion between the original and processed data as much as possible. For example, it is usual to measure the distance between the probability distribution induced by both the original and processed data using distance metrics, e.g. *total variation distance* or *Kullback-Leibler divergence*, with a focus on keeping both distributions as close to each other as possible.

2.3 The Quantitative Information Flow framework

Semantic DC methods have greatly improved on their syntactic counterparts mainly because of the uninformative principle, which requires that the adversary's *a posteriori* knowledge on any data holder should not be much larger than their *a priori* knowledge. This requirement clearly depends on the definition of quantitative measures of knowledge, not only qualitative ones, hence the need to go beyond syntactic restrictions to databases publications.

Here we introduce some of the main concepts from *Quantitative Information Flow* (QIF) that will help us define our metrics in the following chapters. We start by defining "secrets", which are used to model the protected information, followed by "channels", which model the systems through which the secrets flow. Next, we define "gain functions", which are metrics used to quantify the channels' "vulnerability" and "leakage", hence also used to measure the knowledge gained by an adversary that observes the channels as secrets flow through them. The definitions presented in this section are directly taken from the work by Alvim, Chatzikokolakis, McIver, Morgan, Palamidessi, and Smith [2, 5].

Definition 1 (Secrets [5])**.** A secret, called X, has a set $\mathcal{X}$ of possible values, which we call the *type* of X and which we assume to be *finite*; moreover we assume that the knowledge that some adversary has about X is given by a *probability distribution* π on $\mathcal{X}$ that specifies the probability π_x of each possible value x of X. (Thus the *secrecy* of X is defined relative to a *particular* adversary, leaving open the possibility that a *different* adversary might have different knowledge of X, and so require a different distribution to reflect that.) ◁

The probability distribution π is also called the *prior distribution*, which means that it corresponds to the adversary's knowledge before the acquisition of any further information related to the secret.

Considering the secret will be kept secret, i.e. will not be purposely revealed to the adversary, further information related to it can only be acquired by observing the outputs, intentional or not, of computations performed on it. Such outputs can be actual results displayed on a screen, but also any sort of variation on physical attributes related to the system responsible for the computations, e.g. the total time spent, the heat produced by the silicon chips, or the amount of memory allocated. Here we consider only the direct outputs from computational systems.

Hence, it is useful to model the system that may process the secret, denoted here as a "channel". One important assumption derived from *Kerckhoffs's Principle*, often stated as "no security through obscurity", is that the adversary knows how the channel works, i.e. given an input secret x, the adversary is able to determine the output y returned by the channel or at least the probability of that output being returned [5].

Definition 2 (Channels [2])**.** A *channel* is a triple $(\mathcal{X}, \mathcal{Y}, C)$ where $\mathcal{X}$ and $\mathcal{Y}$ are finite sets (typically of secret input values and observable output values resp.) and C is an $|\mathcal{X}| \times |\mathcal{Y}|$ *channel matrix* whose entries are between 0 and 1 and whose rows each sum to 1. Typically we use upper-case Roman letters (like C) for channels, and calligraphic letters (like $\mathcal{X}, \mathcal{Y}$) for the sets over which they operate. We write $C_{x,y}$ for the element of C at row x in $\mathcal{X}$ and column y in $\mathcal{Y}$. For the (entire) row x we write $C_{x,-}$, and for column y we write $C_{-,y}$.

The value $C_{x,y}$ is the conditional probability of output y's being produced by C from input x. A channel is *deterministic* just when all its elements are either 0 or 1, implying that each input row contains a single 1 which identifies that input's unique corresponding output. $\lhd$

Other important definitions are those for "joint distribution", "marginal probabilities", and "conditional probabilities".

Definition 3 (Joint distribution, and marginal and conditional probabilities [2])**.** The *joint distribution* on $\mathcal{X} \times \mathcal{Y}$ determined by input distribution π and channel C is typically written $\Pi_{x,y} = \pi_x C_{x,y}$; when Π is understood, we can use a Π-implicit convention to write that as just $p(x, y)$. The jointly distributed random variables X, Y have *marginal* probabilities that (again Π-implicitly) we write $p(x) = \sum_y p(x, y)$ (which of course is just π_x again) and $p(y) = \sum_x p(x, y)$; occasionally we also write X, Y informally for "the (unnamed) distribution on $\mathcal{X}$, $\mathcal{Y}$" resp. The *conditional probabilities* are then given by $p(y|x) = {}^{p(x,y)}\!/_{p(x)}$ (if $p(x)$ is non-zero) and $p(x|y) = {}^{p(x,y)}\!/_{p(y)}$ (if $p(y)$ is non-zero). Note that Π is the *unique* joint distribution that recovers π and C, in that $p(x) = \pi_x$ and $p(y|x) = C_{x,y}$ (if $p(x)$ is non-zero). When necessary to avoid ambiguity, we write these Π-implicit distributions with subscripts, e.g. p_X or p_{XY} or p_Y.

For example given $|\mathcal{X}| = 3$ and $|\mathcal{Y}| = 4$, we could consider a channel C (at left) which, when applied to (the uniform) prior $\pi = ({}^1\!/_3, {}^1\!/_3, {}^1\!/_3)$ produces the joint matrix Π (at right):

$$\begin{array}{c|cccc} C & y_1 & y_2 & y_3 & y_4 \\ \hline x_1 & 1 & 0 & 0 & 0 \\ x_2 & 0 & {}^1\!/_2 & {}^1\!/_4 & {}^1\!/_4 \\ x_3 & {}^1\!/_2 & {}^1\!/_3 & {}^1\!/_6 & 0 \end{array} \quad \overset{\pi}{\to} \quad \begin{array}{c|cccc} \Pi & y_1 & y_2 & y_3 & y_4 \\ \hline x_1 & {}^1\!/_3 & 0 & 0 & 0 \\ x_2 & 0 & {}^1\!/_6 & {}^1\!/_{12} & {}^1\!/_{12} \\ x_3 & {}^1\!/_6 & {}^1\!/_9 & {}^1\!/_{18} & 0 \end{array}$$

By summing Π's columns we get the $\mathcal{Y}$-marginal distribution $p_Y = ({}^1\!/_2, {}^5\!/_{18}, {}^5\!/_{36}, {}^1\!/_{12})$, and by normalizing those columns we obtain the four posterior distributions $p_{X|y_1} = ({}^2\!/_3, 0, {}^1\!/_3)$, $p_{X|y_2} = (0, {}^3\!/_5, {}^2\!/_5)$, $p_{X|y_3} = (0, {}^3\!/_5, {}^2\!/_5)$, $p_{X|y_4} = (0, 1, 0)$. $\lhd$

However, it is also possible, and often useful, to model a channel regardless of the particular output values, i.e. the set $\mathcal{Y}$. This can be done because only the correlation between input and output is important for an adversary, and this is useful in order to avoid accounting for impossible outputs, i.e. those for which $p(y) = 0$, and to avoid dealing with outputs that result in equivalent posterior distributions. In order to do so, we use the definition of *hyper-distributions*, or just "hyper".

Definition 4 (Hyper [2])**.** A hyper on the input space $\mathcal{X}$ is of type $\mathbb{D}^2\mathcal{X}$. Consider a prior π on $\mathcal{X}$ and a channel $C : \mathcal{X} \to \mathcal{Y}$ which, as usual, determines a joint distribution $\Pi : \mathbb{D}(\mathcal{X} \times \mathcal{Y})$. As above, with Π understood we have a distribution p_Y on the outputs and, for each y, a corresponding posterior distribution $p_{X|y}$ on the inputs. If instead of considering p_Y to be a distribution on $\mathcal{Y}$ we consider it to be a distribution on the corresponding posteriors (i.e. on the normalised columns of Π themselves, rather than on their $\mathcal{Y}$-labels), then we have a hyper which we write $[\pi \triangleright C]$. [8] For the example above, that hyper would assign probabilities $({}^1\!/_2, {}^{15}\!/_{36}, {}^1\!/_{12})$ to the posteriors $({}^2\!/_3, 0, {}^1\!/_3)$, $(0, {}^3\!/_5, {}^2\!/_5)$, and $(0, 1, 0)$, respectively. [9] $\lhd$

In order to measure how much information can be obtained from a channel, there are several entropy-like functions that can be used, including the "min-entropy leakage".

Definition 5 (Min-entropy leakage [2])**.** *Min-entropy leakage* [66] is based on the prior *vulnerability* $V[\pi] = \max_x \pi_x$, the probability of the secret's being guessed in one try, and the expected vulnerability of the posterior distributions $V[\pi \triangleright C] = \sum_y p(y)V[p_{X|y}]$.

The min-entropy leakage $\mathcal{L}(\pi, C)$ is the logarithm of the ratio of the posterior- and prior vulnerabilities:

$$\mathcal{L}(\pi, C) = \lg(V[\pi \triangleright C]/V[\pi]).$$

Note that vulnerability implicitly assumes an operational scenario in which the adversary gains only by guessing the secret *exactly*, and in *one try*. $\lhd$

However, there is also a more general framework to model how much value an adversary may gain from a secret. Hence, we introduce the concept of "gain functions", the foundation of the "g-vulnerability" decision-theoretic framework.

Definition 6 (Gain functions, g-vulnerability, and g-leakage [2])**.** Formally $g : \mathcal{W} \times \mathcal{X} \to [0, 1]$, where $\mathcal{W}$ is a finite set. Given a gain function g, the prior *g-vulnerability* is defined as the maximum expected gain over all possible guesses:

$$V_g[\pi] = \max_w \sum_x \pi_x g(w, x).$$

[8] Here we consider the notation for a hyper according to [5]. We read $[\pi \triangleright C]$ as "π through C".

[9] There might be fewer posteriors in the support of hyper $[\pi \triangleright C]$ than there are columns in the joint distribution Π from which it derives, because if several columns of Π normalise to the same posterior then the hyper will automatically coalesce them [54]: columns y_2 and y_3 were coalesced in this case.

The posterior g-vulnerability and the g-leakage are then defined as for min-entropy leakage above, so that we have:

$$V_g[\pi \triangleright C] = \sum_y p(y) V_g[p_{X|y}],$$

and then $\mathcal{L}(\pi, C) = \lg(V_g[\pi \triangleright C]/V_g[\pi])$. $\triangleleft$

Furthermore, it is possible to define g-leakage additively and multiplicatively as follows.

Definition 7 (Multiplicative and additive g-leakages [5])**.** Given prior distribution π, gain function $g : \mathbb{G}\mathcal{X}$, and channel C, the *multiplicative g-leakage* is:

$$\mathcal{L}_g^{\times}(\pi, C) := \frac{V_g[\pi \triangleright C]}{V_g(\pi)},$$

and *additive g-leakage* is:

$$\mathcal{L}_g^{+}(\pi, C) := V_g[\pi \triangleright C] - V_g(\pi).$$

$\triangleleft$

It can be shown that both *prior* and *posterior* g-vulnerabilities are always non-negative, which means that $V_g = 0$ models the scenario in which the secret is safe. Furthermore, notice that no leakage occurs whenever both vulnerabilities are equal, what happens when the multiplicative g-leakage equals 1 and the additive g-leakage equals 0 [5].

As mentioned when introducing min-entropy, there are several entropy-like functions that can be used. In fact, each gain function can be used to define a corresponding g-vulnerability, covering all the reasonable measures with respect to fundamental information-theoretic axioms [3]. For instance, g_{id} is the gain function responsible for modeling the scenario in which the adversary has only one try and must guess the secret's value exactly.

Definition 8 (Identity gain function [5])**.** The *identity* gain function $g_{id} : \mathcal{X} \times \mathcal{X} \to \{0, 1\}$ is given by:

$$g_{id}(w, x) := \begin{cases} 1 & \text{if } w = x, \\ 0 & \text{if } w \neq x. \end{cases}$$

$\triangleleft$

The identity gain function g_{id} returns gain 1 if the adversary correctly guesses the secret or 0 otherwise. Furthermore, it can be shown that the vulnerability given by g_{id}, i.e. $V_{g_{id}}(\pi)$, coincides with that known nowadays as the *Bayes vulnerability*, i.e. $V_1(\pi) = \max_{x \in \mathcal{X}} \pi_x$, which in turn was used to define the min-entropy leakage in Definition 5. In fact, the subscript 1 was added to account for the single try allowed to the adversary in the context defined by g_{id}. In this thesis we will consider the Bayes vulnerability for our models.

Continuing on the example from Definition 3 and considering both the channel C and joint matrix Π defined in that example, the uniform prior $\pi = (1/3, 1/3, 1/3)$, and the gain function g_{id}, we can compute the corresponding vulnerabilities as follows.

The *a priori* Bayes vulnerability is given by:

$$V_{g_{id}}(\pi) = V_1(\pi) = \max_{x \in \mathcal{X}} \pi_x = 1/3,$$

i.e. the adversary has a $1/3$ chance of correctly guessing the secret in one try before observing any output from channel C.

The *a posteriori* Bayes vulnerability is given by:

$$\begin{aligned}
V_{g_{id}}[\pi \triangleright C] = V_1[\pi \triangleright C] &= \sum_y p(y) V_1[p_{X|y}] \\
&= p_{y_1} V_1[p_{X|y_1}] + p_{y_2} V_1[p_{X|y_2}] + p_{y_3} V_1[p_{X|y_3}] + p_{y_4} V_1[p_{X|y_4}] \\
&= 1/2 \cdot 2/3 + 5/18 \cdot 3/5 + 5/36 \cdot 3/5 + 1/12 \cdot 1 = 2/3,
\end{aligned}$$

i.e. the adversary has a $2/3$ chance of correctly guessing the secret in one try after observing the output from channel C.

Finally, given both *a priori* and *a posteriori* Bayes vulnerabilities, the corresponding multiplicative and additive g-leakages are given by:

$$\begin{aligned}
\mathcal{L}^{\times}_{g_{id}}(\pi, C) &= \mathcal{L}^{\times}_{1}(\pi, C) = \frac{V_1[\pi \triangleright C]}{V_1(\pi)} = \frac{2/3}{1/3} = 2, \\
\mathcal{L}^{+}_{g_{id}}(\pi, C) &= \mathcal{L}^{+}_{1}(\pi, C) = V_1[\pi \triangleright C] - V_1(\pi) = 2/3 - 1/3 = 1/3,
\end{aligned}$$

i.e. the adversary doubles his knowledge on the secret in multiplicative terms, or increases his knowledge on the secret by $1/3$ in additive terms.

We briefly discuss why min-entropy is better suited than Shannon entropy for some scenarios in Appendix B. In the following section, we discuss some related work.

2.4 Related work

Disclosure Control (DC) has been a topic of interest among statisticians even before Tore Dalenius' 1977 paper proposing a methodology for statistical disclosure control. Back then, statisticians were already worried about accidental disclosures caused by an ever increasing volume and detail of statistics produced, which has sparked the discussion on disclosure "in the context of a reasoned balance between the right to privacy and the need to know" [10].

Dalenius argued that the *elimination* of disclosure was unfeasible, since it would impose restrictions on what data could be published to the point of eliminating statistics altogether in the process. Hence the proposition of *controlling* instead of *eliminating* disclosures [10].

Data de-identification, i.e. the removal of any direct identifiers of individuals, is a natural first step towards disclosure control, particularly in the case of re-identification. However, as discussed in Section 2.2, Dalenius considered data de-identification to be a necessary but insufficient measure [11]. In fact, it was demonstrated to be unsuccessful by Samarati and Sweeney in 1998 [65], and exemplified by Sweeney's 2000 work that showed that 87% of the United States population in the Census of 1990 could be uniquely identified by using only a combination of the ZIP code, gender, and date of birth [67]. As a solution, Samarati and Sweeney proposed the k-anonymity DC method [65], which in turn is known nowadays to be vulnerable to composition attacks [24].

Another famous data disclosure in the literature came from the *Netflix Prize* and was published by Narayanan and Shmatikov in 2008 [56]. The data released by Netflix was de-identified and possibly treated using an undisclosed DC method. Nevertheless, by using the *Internet Movie Database* (IMDb) as auxiliary information and by applying their de-anonymization algorithm for sparse data, i.e. data with high dimensionality, Narayanan and Shmatikov were able de-anonymize the released database and recover individual users' movies history and ratings.

Given the known possible vulnerabilities in data releases and the increasingly complex legal landscape on privacy across different jurisdictions, as discussed in Section 1.2, non-experts in need of anonymization and data vulnerabilities analyses tools also increased. In this context, some DC software were proposed, from which we highlight two that are open source and continuously supported: the `sdcMicro` package [55] and the ARX Data Anonymization Tool [61].

The `sdcMicro` package for the R programming language, developed by Templ, Kowarik, and Meindl, and released in 2015 [68], implements a set of DC methods focused on the anonymization of microdata and on risk analyses. The available anonymization methods are mainly syntactic, such as k-anonymity and l-diversity, but there are also functions for noise addition and shuffling. Those semantic methods are not based on state-of-the-art ones such as differential privacy, though.

The ARX Data Anonymization Tool, developed by Prasser and Kohlmayer and released in 2014 [62], is a standalone software written in the Java language. It implements high-performance algorithms for syntactic methods, such as k-anonymity, l-diversity, and t-closeness [51, 59], but also semantic methods such as differential privacy [60].

Particularly, ARX implements privacy models for different threats and assumptions about the adversaries' goal and auxiliary information. It consider threats of membership, identity, and attribute disclosures, and adversaries modeled by El Emam's "prosecutor", "journalist", and "marketer" models [18, 19, 60].

According to El Emam, those adversary models are concerned only with re-identification of the data holder, i.e. identity disclosure [18] and are defined as follows.

- **Prosecutor model**: the adversary targets a specific individual who is known to be a data holder in the database of interest.
- **Journalist model**: the adversary targets a specific individual but is not sure whether the target is a data holder in the database of interest or not.
- **Marketer model**: the adversary's goal is to re-identify as many individuals as possible in the database of interest, i.e. it measures how many records, on average, could be re-identified.

Moreover, a known but less discussed DC problem relates to the continued release of data. Fung et al. discuss different approaches to this problem, including *multiple views publishing*, which consists of releasing databases with different sets of attributes from the same collection of data; *sequential releases with new attributes*, which consists of considering possible increments of attributes to the database release; and *incrementally updated data records*, which can be either a *continuous data publishing*, i.e. every subsequent database release contains all the previous releases in addition to the new records, or a *dynamic data republishing*, i.e. any records can be inserted, deleted, or updated as new databases are released [23]. In any case, all the proposed solutions presented by Fung et al. are syntactic in nature and hence vulnerable to composition attacks, as discussed in Section 2.2.2.1.

Here we are interested only on *incrementally updated data records*, particularly on its *dynamic data republishing* case, given the databases we have used in our experiments, as described in the next Section 1.3. The first work to identify this problem was published by Byun et al. in 2006 and proposed a new approach for updating incremental databases that would prevent an adversary's inferences. Their approach is based on l-diversity and on a newly defined *information loss* metric [9], but would only accept the insertion of new records.

In order to solve this limitation, Xiao and Tao proposed in 2007 the m-*invariance* privacy model, a generalization technique also used in k-anonymity and other syntactic methods, which would allow both the insertion and deletion of records. According to Xiao and Tao, a sequence of published databases is m-*invariant* if for a record that is published in more than one database, all the blocks of similar records, i.e. those with the same quasi-identifiers, that contain this record share the same sensitive values [81]. As a result, the intersection of databases from the published set of databases would not decrease the set of possible sensitive values for a given block of similar records. But m-invariance does not consider the update of quasi-identifying or sensitive values over time, a drawback for the dynamic data republishing scenario.

Therefore, Bu et al. proposed in 2008 the "HD-composition" privacy model, which considers not only that quasi-identifying and sensitive attributes may change over time, but also that some sensitive attribute values, once linked to a record, will never change [8]. For instance, an individual may move from one address to another and hence have their ZIP code updated. Similarly, an individual may be diagnosed with the flu in a hospital appointment, but such a disease is usually cured in a few days and that sensitive attribute value would be updated accordingly. Yet the same is not always true for other diseases, such as cancer, diabetes, and HIV infection. In summary, *HD-composition* is a model in which some records are *holders* of permanent sensitive values while other records are *decoys*, which are used to help mask those values in the partitioning of the databases.

Finally, regarding our contributions in Chapter 6 where we formalize privacy and utility analyses for two different implementations of differential privacy, we have Kifer and Machanavajjhala's *no-free-lunch* theorem demonstrated in 2011 [50]. Through this theorem, the authors showed that it is not possible to provide privacy and utility without considering how the data is generated, i.e. how it is collected and treated prior to release. This result debunks the idea that differential privacy does not consider assumptions about the data in order to guarantee privacy and was the first analysis of differential privacy limitations in databases where there is correlation between secrets.

Chapter 3

Models for attack classification and for databases

In this chapter, we introduce our initial contributions in the form of some new, foundational concepts used in the following chapters. Particularly, we propose a new classification of attacks against databases that better covers the space of possible attacks in comparison to previous propositions, and we develop our database model.

3.1 Attack classification

We have already presented in Chapter 2 some commonly used privacy and risk models from the literature. Particularly, they categorize the possible attacks against databases according to the privacy threat (membership, identity, or attribute disclosures, as presented in Section 2.1.3) [23, 29, 61] and the adversary (prosecutor, journalist, or marketer, as presented in Section 2.4) [18, 19, 60].

However, our analysis of attack classification along those dimensions has shown that:

1. they do not account for all possible scenarios, e.g. attribute-inference risk and longitudinal database releases are not covered by those dimensions;
2. some of them overlap with one another, e.g. both the prosecutor and journalist models target a specific individual.

Therefore, our first contribution was to refactor the possible categorization dimensions for attacks classification in order to better understand our domain of work. Specifically, we propose three new dimensions based on the type of information sought by an adversary, their focus, and their access to releases of databases. The proper definition of those dimensions has allowed us to define a series of different, non-overlapping attacks, which we discuss in Chapters 4 and 5 alongside experimental results that show some vulnerabilities we have found in the databases released by INEP, as introduced in Section 1.3.

Henceforth, we consider the following three dimensions to classify attacks on databases.

(I) **The type of information sought by the adversary**: [1] [2]

- **Re-identification attacks**: the adversary's goal is to ascertain a correspondence between records in the database and individuals to whom those records refer.
- **Attribute-inference attacks**: the adversary's goal is to infer the value of an attribute, usually considered to be sensitive, regardless of whether or not it was possible to re-identify the respective individuals.

(II) **The adversary's focus**:

- **Individual-target attacks**: the adversary's goal is to obtain sensitive information on a predefined, specific individual.
- **Collective-target attacks**: the adversary's goal is to obtain sensitive information on as many individuals as possible, no matter who they are.

(III) **The adversary's access to databases**:

- **Single database attacks**: the adversary can access only a single database corresponding to a specific point in time.
- **Longitudinal database attacks**: the adversary can access several versions of the database, each one corresponding to a distinct point in time. Longitudinal attacks allow the adversary to compound the information present in different databases to increase their chances of success.

[1] Note that success in a re-identification attack implies success in the inference of every attribute for the targeted-individual.

[2] One could also consider *membership attacks*, i.e. the adversary's goal is to determine whether individuals are represented in a database or not. We do not consider membership attacks here since we assume that in every scenario the adversary knows that all the individuals of interest are data holders in the considered databases. Since in our experiments we use educational censuses that account for all the students in Brazil, this assumption always holds.

	Single database (S)		**Longitudinal database (L)**	
	Individual-target (I)	**Collective-target (C)**	**Individual-target (I)**	**Collective-target (C)**
Re-identification (R)	IRS	CRS	IRL	CRL
Attribute-inference (A)	IAS	CAS	IAL	CAL

Table 3.1: Attacks classification and their respective acronyms.

Accounting for all the described classification dimensions, we have eight distinct types of attacks. We will reference each of those attacks henceforth by their acronyms, as defined in Table 3.1. Particularly, single database attacks are discussed in Chapter 4 and longitudinal database attacks in Chapter 5.

3.2 Database model

In this section, we define our database model, which will be used in the next chapters when we formalize each of those attacks on databases introduced in the previous section. We start by defining a "database", which we consider to be a collection of "records" or "rows", each consisting of a tuple of values corresponding to "attributes" or "columns". The new definitions are illustrated in Example 12 for attacks on single databases.

Definition 9 (Database and record)**.** Let $\mathcal{A} = \{a_1, \ldots, a_m\}$ be a finite set of *attributes* or *columns*. For every $1 \leq i \leq m$, we denote by $dom(a_i)$ the domain of all values the attribute a_i admits. Likewise, assume that a *record* or *row* x is a mapping from attributes to values, i.e. $x = \langle x[a_1], \ldots, x[a_m] \rangle$, where $x[a_i] \in dom(a_i)$ is the value taken on by the attribute a_i in record x. The domain of records is then $records(\mathcal{A}) = dom(a_1) \times \cdots \times dom(a_m)$.

Hence, a *database* on a set of attributes $\mathcal{A}$ is a finite multi-set of records $D = \{\{x_1, x_2, \ldots, x_n\}\}$, each one belonging to $records(\mathcal{A})$. [3]

Furthermore, we denote by $mult_D(x)$ the *multiplicity* of record x in database D, i.e. the number of distinct appearances of record x in D. Finally, we denote by

[3] A *multi-set* is a generalization of a set such that the repetition of elements is permitted. In our model, the use of multi-sets allows the same record to appear multiple times in a given database.

$|D| = \sum_{x \in \mathit{records}(\mathcal{A})} \mathit{mult}_D(x)$ the *cardinality* of D, i.e. the total number of records in D, including repetitions. ◁

From Definition 9, we can define the concept of a "sub-record".

Definition 10 (Sub-record)**.** Given a subset of attributes $\mathcal{A}' \subseteq \mathcal{A}$, we denote by $x[\mathcal{A}']$ the *sub-record* or *sub-row* of x consisting in the projection of x over $\mathcal{A}'$, i.e. the sub-tuple of x containing only the values corresponding to the attributes in $\mathcal{A}'$. The domain of sub-records with attributes in $\mathcal{A}' = \{a_{i_1}, \ldots, a_{i_k}\}$ is then $\mathit{dom}(\mathcal{A}') = \mathit{dom}(a_{i_1}) \times \ldots \times \mathit{dom}(a_{i_k})$. ◁

Finally, we can define a "partition on records", which will be a key concept for the theoretical framework to be developed in the next chapters.

Definition 11 (Partition on records)**.** We denote by $D/_{\sim_{\mathcal{A}'}}$ the *partition on records* of the database D induced by the attributes $\mathcal{A}' \subseteq \mathcal{A}$. [4] Intuitively, the partition breaks the database into blocks, each one with all the records with the same values for the attributes in $\mathcal{A}'$.

Furthermore, we denote by $\mathit{block}_{D/_{\sim_{\mathcal{A}'}}}(a)$ the block on partition $D/_{\sim_{\mathcal{A}'}}$ consisting of all elements $x \in D$ with value a for the attributes in $\mathcal{A}'$:

$$\mathit{block}_{D/_{\sim_{\mathcal{A}'}}}(a) \stackrel{\text{def}}{=} \{x \in D \mid x[\mathcal{A}'] = a\} \ .$$

When $D/_{\sim_{\mathcal{A}'}}$ is clear from the context, we may write $\mathit{block}(a)$ instead of $\mathit{block}_{D/_{\sim_{\mathcal{A}'}}}(a)$. ◁

We now present the leading example for single database attacks.

Example 12 (Leading example for single database attacks)**.** Consider the set of attributes to be $\mathcal{A} = \{\mathit{id}, \mathit{age}, \mathit{gender}, \mathit{occupation}, \mathit{illness}\}$, specified as follows:

- $\mathit{dom}(\mathit{id}) = \{1, 2, 3, \ldots\}$, consisting of the unique identification number of an individual given by a positive integer;
- $\mathit{dom}(\mathit{age}) = \{0, 1, 2, \ldots, 120\}$, consisting of an individual's age, in whole years, given by a positive integer including zero;

[4] A *partition* P on a set $\mathcal{X}$ is a collection of *blocks* $P = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_k\}$ such that: (i) $\mathcal{X}_i \subseteq \mathcal{X}$ for all $1 \leq i \leq k$; (ii) $\mathcal{X}_i \cap \mathcal{X}_j = \emptyset$ for all $1 \leq i, j \leq k$ e $i \neq j$; (iii) $\bigcup_{i=1}^{k} \mathcal{X}_i = \mathcal{X}$.

- $\mathit{dom}(\mathit{gender}) = \{\mathtt{F}, \mathtt{M}\}$, denoting an individual's gender, considered here to be either female ($\mathtt{F}$) or male ($\mathtt{M}$) as a simplification for the current example;
- $\mathit{dom}(\mathit{occupation}) = \{\mathtt{1}, \mathtt{2}, \mathtt{3}, \mathtt{4}, \mathtt{5}\}$, denoting an individual's occupation category;
- $\mathit{dom}(\mathit{illness}) = \{\mathtt{yes}, \mathtt{no}\}$, denoting whether an individual is bearer of a given medical condition ($\mathtt{yes}$) or not ($\mathtt{no}$).

Thus the set of attributes $\mathcal{A}$ spans a domain consisting of every combination of the possible values for each attribute, i.e.

$$\mathit{records}(\mathcal{A}) = \mathit{dom}(\mathit{id}) \times \mathit{dom}(\mathit{age}) \times \mathit{dom}(\mathit{gender}) \times \mathit{dom}(\mathit{occupation}) \times \mathit{dom}(\mathit{illness}).$$

Table 3.2 represents a database D with ten records, i.e. with cardinality $|D| = 10$. Each record is a tuple on domain $\mathit{records}(\mathcal{A})$ of all possible records and represents an individual from a population of interest, e.g. subjects of a medical database in this case.

For instance, consider the subset of attributes $\{\mathit{gender}, \mathit{illness}\} \subseteq \mathcal{A}$ and an individual x represented by the tuple $\langle 1, 25, \mathtt{F}, \mathtt{1}, \mathtt{no}\rangle$. The sub-record representing this individual is then $x[\{\mathit{gender}, \mathit{illness}\}] = \langle \mathtt{F}, \mathtt{no}\rangle$. The corresponding partition of D induced by this subset of attributes, $D/_{\sim_{\{gender, illness\}}}$, consists of four blocks:

- one block for all records x such that $x[\{\mathit{gender}, \mathit{illness}\}] = \langle \mathtt{M}, \mathtt{yes}\rangle$, which includes the individual with *id* 4;
- one block for all records x such that $x[\{\mathit{gender}, \mathit{illness}\}] = \langle \mathtt{M}, \mathtt{no}\rangle$, which includes the individuals with *id* 5, 9, and 10;
- one block for all records x such that $x[\{\mathit{gender}, \mathit{illness}\}] = \langle \mathtt{F}, \mathtt{yes}\rangle$, which includes the individuals with *id* 2, 3, 6, and 7;
- one block for all records x such that $x[\{\mathit{gender}, \mathit{illness}\}] = \langle \mathtt{F}, \mathtt{no}\rangle$, which includes the individuals with *id* 1 and 8.

$\lhd$

We now define both the "*a priori* probability distribution on records" and the "marginal probability distribution on attributes", also key concepts for the theoretical framework to be developed in the next chapters.

id	*age*	*gender*	*occupation*	*illness*
1	25	F	1	no
2	25	F	1	yes
3	25	F	3	yes
4	25	M	2	yes
5	25	M	2	no
6	49	F	3	yes
7	49	F	3	yes
8	49	F	5	no
9	49	M	4	no
10	60	M	4	no

Table 3.2: Database D from the leading example for single database attacks, Example 12.

Definition 13 (*A priori* probability distributions on records)**.** From the database D, the adversary can derive an *a priori* probability distribution $\pi^D : \mathbb{D}(\mathit{records}(\mathcal{A}))$ on the records such that, for each record $x{\in}\mathit{records}(\mathcal{A})$, its probability π^D_x is its relative frequency in the database D:

$$\pi^D_x \quad \overset{\text{def}}{=} \quad \frac{\mathit{mult}_D(x)}{|D|} \ . \tag{3.1}$$

$\lhd$

Definition 14 (Marginal probability distributions on attributes)**.** From the probability distribution on records π^D, the adversary can derive the marginal probability distribution $\pi^{\mathcal{A}'|D}$ on values of any subset of attributes $\mathcal{A}'{\subseteq}\mathcal{A}$ as usual, i.e. the probability $\pi^{\mathcal{A}'|D}_a$ for each value $a{\in}\mathit{dom}(\mathcal{A}')$ is:

$$\pi^{\mathcal{A}'|D}_a \quad \overset{\text{def}}{=} \quad \sum_{\substack{x\in D,\\ x[\mathcal{A}']=a}} \pi^D_x = \sum_{\substack{x\in D,\\ x[\mathcal{A}']=a}} \frac{\mathit{mult}_D(x)}{|D|} = \sum_{x\in D} \frac{|\{x{\in}D \mid x[\mathcal{A}'] = a\}|}{|D|} \ . \tag{3.2}$$

$\lhd$

We now show how the adversary can derive the marginal probability distribution $\pi^{\mathcal{A}'|D}$ on values for any subset of attributes $\mathcal{A}'{\subseteq}\mathcal{A}$.

Example 15 (Marginal probability distribution on a subset of attributes)**.** Consider the database D on the set of attributes $\mathcal{A}$ from Example 12 and assume the probability distribution π^D on the records of D as defined, for each record $x \in \mathit{records}(\mathcal{A})$, by Equation 3.1. If we consider only the subset of attributes $\{\mathit{gender}, \mathit{illness}\} \subseteq \mathcal{A}$, the domain is:

$$dom(\{gender, illness\}) = \{\langle \texttt{F}, \texttt{yes}\rangle, \langle \texttt{F}, \texttt{no}\rangle, \langle \texttt{M}, \texttt{yes}\rangle, \langle \texttt{M}, \texttt{no}\rangle\} \ .$$

The corresponding marginal probability distribution $\pi^{\{gender,illness\}|D}$ derived from the frequency of records in D is then:

$$\pi^{\{gender,illness\}|D}_{\langle \texttt{F},\texttt{yes}\rangle} = \sum_{\substack{x \in D, \\ x[\{gender,illness\}]= \\ \langle \texttt{F},\texttt{yes}\rangle}} \pi^{D}_{x} = \sum_{\substack{x \in D, \\ x[\{gender,illness\}]= \\ \langle \texttt{F},\texttt{yes}\rangle}} \frac{mult_D(x)}{|D|} = \frac{4}{10} \ .$$

Analogously,

$$\pi^{\{gender,illness\}|D}_{\langle \texttt{F},\texttt{no}\rangle} = {}^{2}\!/_{10}, \qquad \pi^{\{gender,illness\}|D}_{\langle \texttt{M},\texttt{yes}\rangle} = {}^{1}\!/_{10}, \qquad \text{and} \qquad \pi^{\{gender,illness\}|D}_{\langle \texttt{M},\texttt{no}\rangle} = {}^{3}\!/_{10} \ .$$

◁

We leave the leading example for longitudinal database attacks, Example 20, to be presented after we introduce further assumptions necessary for the development of that model, which is done in Section 5.1.1.

Chapter 4

Collective-target attacks on single databases

As discussed in Chapter 2, publicly released databases may be subject to several vulnerabilities even if some Disclosure Control (DC) methods have been applied to them prior to release. Indeed, one of the main drawbacks found in most DC methods relate to imprecise definitions of an adversary model. For instance, techniques of de-identification and k-anonymity do not consider an adversary with access to auxiliary information, and therefore are vulnerable to linkage and composition attacks, respectively. Also, most DC methods have been designed to be applied to single databases, leaving longitudinal databases open to additional vulnerabilities.

Hence the importance of an accurate formalization of database attacks. Only by precisely defining the adversary model one can clarify to what vulnerabilities a given database publication is subject. Implicit models are not enough. [1]

In this chapter, we present the theoretical foundation on attacks against single databases for collective-targets in Section 4.1, followed by the respective experimental results for attacks on the School Census of 2018 released by INEP in Section 4.2. We conclude this chapter by emphasizing the most relevant contributions in Section 4.3.

[1] Brazil's LGPD explicitly defines "anonymization" as [40]:

> Use of reasonable and available technical means at the time of treatment, whereby data loses the possibility of association, directly or indirectly, with an individual.

Therefore, by using formal definitions for database attacks and adversaries, we can properly demonstrate the vulnerabilities and improve the discussion on what should be considered *reasonable*.

4.1 Theoretical foundation

In this section, we present our model for attacks performed on single databases for collective-targets. We consider our database model introduced in Section 3.2.

4.1.1 Assumptions

We have the following assumptions for all attacks on single databases:

AS0 There is a database of interest D on an attributes set $\mathcal{A}$.

AS1 Each individual of interest holds a record in the database D.

AS2 There is a set of *quasi-identifiers* $\mathcal{Q}_{ID} \subseteq \mathcal{A}$ corresponding to attributes whose values for some individuals can be known or learned by the adversary.

AS3 The adversary can compute marginal probability distributions on both records and attributes from the database D by using the functions presented in Equations 3.1 and 3.2, respectively.

AS4 The adversary knows *a priori* the whole database D, i.e. they have access to every record and attribute value. Also, the adversary knows *a priori* that every individual of interest holds a record in D. [2] Hence, the attack consists in the use of auxiliary information obtained by the adversary from sources other than the database D, which in turn are used by the adversary as quasi-identifier values in order to infer some sensitive information about the targeted-individuals represented in D. Therefore, the degradation of privacy is a comparative measure between the adversary's *a priori* and *a posteriori* successes, as schematized in Figure 4.1.

Based on the assumptions above, we present in the following section both collective-target re-identification and attribute-inference on single databases attack models.

[2] Assumption **AS4** may seem to provide too much power to the adversary, and for sure it cannot be applied to every database release. In our case, the INEP databases used for the experiments and described in Section 1.3 are real examples of scenarios in which the adversary would be able to rely on Assumption **AS4**, since every student has to be in the national census. Moreover, we are worried about worst-case scenarios, e.g. Sweeney's re-identification of 87% of the United States population in the Census of 1990 [67].

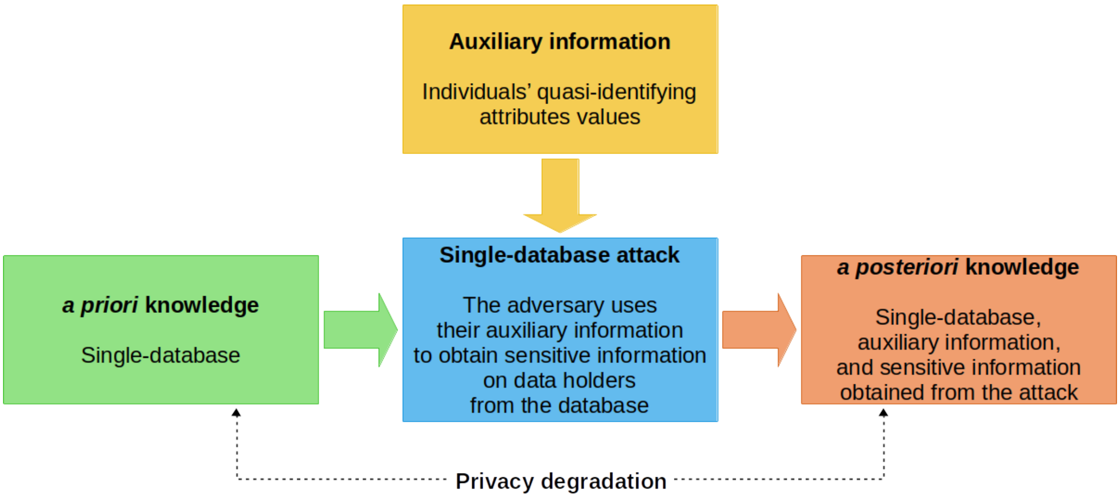

Auxiliary information
Individuals' quasi-identifying attributes values
***a priori* knowledge**
Single-database
Single-database attack
The adversary uses their auxiliary information to obtain sensitive information on data holders from the database
***a posteriori* knowledge**
Single-database, auxiliary information, and sensitive information obtained from the attack
Privacy degradation

Figure 4.1: Schema of single database attack. The degradation of privacy is a comparative measure between the adversary's *a priori* and *a posteriori* successes and accounts for how much the auxiliary information used as quasi-identifying attributes values can help the adversary on the tasks of re-identification or sensitive attribute-inference.

4.1.2 Attack models

In this section, we formally define both re-identification and attribute-inference attacks on single databases for collective-targets. Definition 16 accounts for an adversary interested in re-identifying as many individuals as possible, while Definition 17 accounts for an adversary interested in inferring the value of an attribute for as many individuals as possible. Detailed numeric examples of the execution of both attacks are presented in Sections D.1.1 and D.1.2, respectively, and illustrative individual-target attack models are presented in Section C.1.1.

Definition 16 (Collective-target Re-identification Single database (CRS) attack)**.** In a CRS attack, the adversary can completely access a single database, i.e. they have knowledge of every record in the database. The adversary can also gather auxiliary information on the values of a set of quasi-identifiers for every individual who holds a record in the database. The adversary's goal is to re-identify as many individuals as possible, no matter who they are, i.e. to precisely determine which record in the database corresponds to each individual of interest.

We now formally define the adversary's knowledge and their inflicted privacy degradation, both deterministic and probabilistic, after performing their attack.

- **Adversary's knowledge.** In a CRS attack, we assume the adversary:

(i) can completely access a (non-empty) database D on a set of attributes $\mathcal{A}$, according to Assumptions **AS0**–**AS2**;

(ii) has every individual x in the database D as individuals of interest for re-identification;

(iii) given a set of quasi-identifiers $\mathcal{Q}_{ID}\subseteq\mathcal{A}$, the adversary can gather auxiliary information on the sub-records $\{x[\mathcal{Q}_{ID}]\}_{x\in D}$ corresponding to the values of the quasi-identifiers for every individual in the database.

- **Deterministic degradation of privacy.** The deterministic degradation of privacy in a CRS attack accounts for the adversary's goal of re-identifying as many individuals as possible. Hence, the adversary's success is defined as the ratio of individuals in the database that are re-identified with absolute certainty.

  - ***A priori* deterministic success.** Before performing the CRS attack, the adversary can rely only on the database itself, i.e. the quasi-identifiers are not used yet. Hence, the adversary can only precisely determine which record in the database corresponds to an individual of interest iff the database contains only a single record. Therefore, the *a priori* deterministic success in a CRS attack on a database D is defined as:

$$\mathit{prior\text{-}suc}^{CRS}_{det}(D) \quad \overset{\text{def}}{=} \quad \begin{cases} 1, & \text{if } |D|{=}1, \\ 0, & \text{if } |D|{\geq}2. \end{cases} \tag{4.1}$$

  - ***A posteriori* deterministic success.** By performing a CRS attack, the adversary uses the auxiliary information on the sub-records $\{x[\mathcal{Q}_{ID}]\}_{x\in D}$ corresponding to the values of quasi-identifiers for every individual in the database to support the task of re-identification. A CRS attack is then performed as follows: For each individual in the database, the adversary uses the corresponding quasi-identifier values to filter the database records, disregarding those that do not match the query. Then, given the remaining database, the adversary can only precisely determine which record corresponds to the individual of interest iff the remaining database contains only a single record. The final deterministic success is determined over the set of all individuals and defined as the ratio between the number of re-identified individuals and the total number of individuals in the database.

    Therefore, the *a posteriori* deterministic success in a CRS attack on a database D, given the adversary's knowledge of the values $\{x[\mathcal{Q}_{ID}]\}_{x\in D}$ for every individual's quasi-identifiers, is defined as:

$$\textit{post-suc}^{CRS}_{det}(D, \{x[\mathcal{Q}_{ID}]\}_{x\in D}) \quad \overset{\text{def}}{=} \quad \frac{\Big|\{\alpha{\in}D/_{\sim_{\mathcal{Q}_{ID}}} \mid |\alpha|{=}1\}\Big|}{|D|} \ , \tag{4.2}$$

where $D/_{\sim_{\mathcal{Q}_{ID}}}$ is the partition on set $\mathcal{Q}_{ID}$ of all individuals that share the same quasi-identifiers and α is a variable corresponding to each block within the partition.

- **Deterministic degradation of privacy.** The deterministic degradation of privacy in a CRS attack is defined as the difference between the adversary's *a posteriori* and *a priori* successes, i.e. by how much the attack increases the adversary's success:

$$\textit{priv-degrad}^{CRS}_{det}(D, \{x[\mathcal{Q}_{ID}]\}_{x\in D}) \quad \overset{\text{def}}{=} \quad \textit{post-suc}^{CRS}_{det}(D, \{x[\mathcal{Q}_{ID}]\}_{x\in D}) - \textit{prior-suc}^{CRS}_{det}(D) \ . \tag{4.3}$$

- **Probabilistic degradation of privacy.** The adversary's probabilistic success in a CRS attack does not rely on precisely determining which record in the database corresponds to each individual of interest. Rather, we compute the probability of correctly re-identifying a randomly selected target in the database, i.e. the greater this probability, the more successful the adversary is and greater the expected risk of re-identification for each individual in the database.

  - ***A priori* probabilistic success.** Before performing the CRS attack, the adversary can rely only on the database itself, i.e. the quasi-identifiers are not used yet. Hence, given a random target chosen by the adversary, their best course of action is to randomly select a record from the database, according to the maximum entropy principle. Since each individual only holds one record in the database, the probability of the adversary being successful is the inverse of the database size. Therefore, the *a priori* probabilistic success in a CRS attack on a database D is defined as: [3]

$$\textit{prior-suc}^{CRS}_{prob}(D) \quad \overset{\text{def}}{=} \quad \frac{1}{|D|} \ . \tag{4.4}$$

  - ***A posteriori* probabilistic success.** By performing a CRS attack, the adversary uses the auxiliary information on the sub-records $\{x[\mathcal{Q}_{ID}]\}_{x\in D}$

[3] This definition corresponds to the *a priori* Bayes vulnerability $V_1[\pi^D_x]$, given by Definitions 6 and 8 in Section 2.3, of the database's records on the *a priori* probability distribution π^D_x.

corresponding to the values of quasi-identifiers for every individual in the database to support the task of re-identification. Given a randomly selected target in the database, the adversary uses the corresponding quasi-identifier values to filter the database records, disregarding those that do not match the query. From there, their best course of action is to randomly select a record from the filtered database, according to the maximum entropy principle. Since each individual only holds one record in the original database, the probability of the adversary being successful is the inverse of the filtered database size.

Therefore, the *a posteriori* probabilistic success in a CRS attack on a database D, given the adversary's knowledge of the values $\{x[\mathcal{Q}_{ID}]\}_{x \in D}$ for every individual's quasi-identifiers, is defined as the expected value of the probability of success taken from the probability distribution for selecting each individual as the target: [4]

$$\begin{aligned}
\textit{post-suc}^{CRS}_{prob}(D, \{x[\mathcal{Q}_{ID}]\}_{x\in D}) &\overset{\text{def}}{=} \sum_{x\in D} \pi^D_x \cdot \frac{1}{|\{x'{\in}D \mid x'[\mathcal{Q}_{ID}]{=}x[\mathcal{Q}_{ID}]\}|} \\
&= \sum_{x\in D} \frac{1}{|D|} \cdot \frac{1}{|\{x'{\in}D \mid x'[\mathcal{Q}_{ID}] = x[\mathcal{Q}_{ID}]\}|} \\
&= \sum_{\alpha \in D/\sim_{\mathcal{Q}_{ID}}} |\alpha| \cdot \frac{1}{|D|} \cdot \frac{1}{|\alpha|} = \frac{\left|D/\sim_{\mathcal{Q}_{ID}}\right|}{|D|} \,,
\end{aligned} \tag{4.5}$$

i.e. the ratio between the number of blocks in the partition on set $\mathcal{Q}_{ID}$ and the database size.

– **Probabilistic degradation of privacy.** The probabilistic degradation of privacy in a CRS attack is defined as the ratio by which the attack increases

[4] This definition corresponds to the *a posteriori* Bayes vulnerability $V_1[\pi^D_x \triangleright C^{\mathcal{Q}_{ID}}]$, given by Definitions 6 and 8 in Section 2.3, of the database's records given the knowledge of the quasi-identifiers values, where:

* π^D_x is the *a priori* probability distribution on the database's records;
* $C^{\mathcal{Q}_{ID}} : D \rightarrow \mathbb{D}(\textit{dom}(\mathcal{Q}_{ID}))$ is the channel that deterministically maps each record to the respective values of quasi-identifiers and is defined, for all $x{\in}D$ and $q{\in}\textit{dom}(\mathcal{Q}_{ID})$, as:

$$C^{\mathcal{Q}_{ID}}_{x,q} = \begin{cases} 1, & \text{if } x[\mathcal{Q}_{ID}] = q, \\ 0, & \text{otherwise.} \end{cases}$$

the probabilistic success of the adversary:

$$\begin{aligned} \textit{priv-degrad}^{CRS}_{prob}(D, \{x[\mathcal{Q}_{ID}]\}_{x \in D}) &\overset{\text{def}}{=} \frac{\textit{post-suc}^{CRS}_{prob}(D, \{x[\mathcal{Q}_{ID}]\}_{x \in D})}{\textit{prior-suc}^{CRS}_{prob}(D)} \\ &= \left| D/_{\sim_{\mathcal{Q}_{ID}}} \right| , \end{aligned} \tag{4.6}$$

i.e. the number of blocks in the partition on set $\mathcal{Q}_{ID}$.

A detailed numeric example of a CRS attack is presented in Section D.1.1, Example 48.

◁

We now present the attribute-inference attack on single databases for collective-targets.

Definition 17 (Collective-target Attribute-inference Single database (CAS) attack)**.** In a CAS attack, the adversary can completely access a single database, i.e. they have knowledge of every record in the database. The adversary can also gather auxiliary information on the values of a set of quasi-identifiers for every individual who holds a record in the database. The adversary's goal is to infer the values of an attribute considered to be sensitive for as many individuals as possible, no matter who they are.

We now formally define the adversary's knowledge and their inflicted privacy degradation, both deterministic and probabilistic, after performing their attack.

- **Adversary's knowledge.** In an CAS attack, we assume the adversary:
  (i) can completely access a (non-empty) database D on a set of attributes $\mathcal{A}$;
  (ii) has every individual x in the database D as individuals of interest for attribute-inference;
  (iii) given a set of quasi-identifiers $\mathcal{Q}_{ID} \subseteq \mathcal{A}$, the adversary can gather auxiliary information on the sub-records $\{x[\mathcal{Q}_{ID}]\}_{x \in D}$ corresponding to the values of the quasi-identifiers for every individual in the database;
  (iv) given an attribute $a_{sens} \in \mathcal{A}$ such that $a_{sens} \notin \mathcal{Q}_{ID}$, considered to be sensitive, the adversary's goal is to infer the attribute's value $x[a_{sens}]$ for as many individuals as possible.
- **Deterministic degradation of privacy.** The deterministic degradation of privacy in a CAS attack accounts for the adversary's goal of precisely determining the value of the sensitive attribute for as many individuals as possible. Hence, the adversary's success is defined as the ratio of individuals in the database whose value for the sensitive attribute can be inferred with absolute certainty.

- ***A priori* deterministic success.** Before performing the CAS attack, the adversary can rely only on the database itself, i.e. the quasi-identifiers are not used yet. Hence, the adversary can only precisely determine the value of the sensitive attribute for any individual of interest iff all the records in the database have the same value for the sensitive attribute. Therefore, the *a priori* deterministic success in a CAS attack on a database D and with respect to the sensitive attribute a_{sens} is defined as:

$$\textit{prior-suc}_{det}^{CAS}(D, a_{sens}) \overset{\text{def}}{=} \begin{cases} 1, & \text{if for all } x, x' {\in} D, \\ & x[a_{sens}]{=}x'[a_{sens}], \\ 0, & \text{otherwise.} \end{cases} \tag{4.7}$$

- ***A posteriori* deterministic success.** By performing a CAS attack, the adversary uses the auxiliary information on the sub-records $\{x[\mathcal{Q}_{ID}]\}_{x\in D}$ corresponding to the values of quasi-identifiers for every individual in the database to support the task of inferring the value of the sensitive attribute. A CAS attack is then performed as follows: For each individual in the database, the adversary uses the corresponding quasi-identifier values to filter the database records, disregarding those that do not match the query. Then, given the remaining database, the adversary can only precisely determine the value of the sensitive attribute for the individual of interest iff all the records in the remaining database have the same value for the sensitive attribute. The final deterministic success is determined over the set of all individuals and defined as the ratio between the number of individuals whose value for the sensitive attribute could be precisely determined and the total number of individuals in the database.

  Therefore, the *a posteriori* deterministic success in a CAS attack on a database D and with respect to the sensitive attribute a_{sens}, given the adversary's knowledge of the values $\{x[\mathcal{Q}_{ID}]\}_{x\in D}$ for every individual's quasi-identifiers, is defined as:

$$\begin{aligned} &\textit{post-suc}_{det}^{CAS}(D, a_{sens}, \{x[\mathcal{Q}_{ID}]\}_{x\in D}) \overset{\text{def}}{=} \\ &\frac{\sum_{\alpha\in D/\sim_{\mathcal{Q}_{ID}}} |\alpha| \cdot \textit{unique-sens}(\alpha, a_{sens})}{|D|}, \end{aligned} \tag{4.8}$$

  where $D/_{\sim_{\mathcal{Q}_{ID}}}$ is the partition on set $\mathcal{Q}_{ID}$ of all individuals that share the same quasi-identifiers and α is a variable corresponding to each block within the partition. Also, the function *unique-sens*(α, a_{sens}) is defined as:

$$\mathit{unique\text{-}sens}(\alpha, a_{sens}) \overset{\text{def}}{=} \begin{cases} 1, & \text{if for all } x, x'{\in}\alpha, \\ & x[a_{sens}]{=}x'[a_{sens}], \\ 0, & \text{otherwise.} \end{cases} \tag{4.9}$$

- **Deterministic degradation of privacy.** The deterministic degradation of privacy in a CAS attack is defined as the difference between the adversary's *a posteriori* and *a priori* successes, i.e. by how much the attack increases the adversary's success:

$$\mathit{priv\text{-}degrad}^{CAS}_{det}(D, a_{sens}, \{x[\mathcal{Q}_{ID}]\}_{x\in D}) \overset{\text{def}}{=} \mathit{post\text{-}suc}^{CAS}_{det}(D, a_{sens}, \{x[\mathcal{Q}_{ID}]\}_{x\in D}) - \mathit{prior\text{-}suc}^{CAS}_{det}(D, a_{sens}) \ . \tag{4.10}$$

- **Probabilistic degradation of privacy.** The adversary's probabilistic success in a CAS attack does not rely on precisely determining the value of the sensitive attribute for each individual of interest. Rather, we compute the probability of correctly determining the value of the sensitive attribute for a randomly selected target in the database, i.e. the greater this probability, the more successful the adversary is and greater the expected risk of sensitive attribute-inference for each individual in the database.

  - ***A priori* probabilistic success.** Before performing the CAS attack, the adversary can rely only on the database itself, i.e. the quasi-identifiers are not used yet. Hence, given a random target chosen by the adversary, their best course of action is to randomly select a possible value for the sensitive attribute, according to the maximum entropy principle. Since each individual only holds one record in the database, the most probable value would be the most frequent one in the database. Therefore, the *a priori* probabilistic success in a CAS attack on a database D and with respect to the sensitive attribute a_{sens}, given the probability distribution of values for the sensitive attribute $\pi_s^{a_{sens}|D}$, according to Equation 3.2, is defined as: [5]

$$\mathit{prior\text{-}suc}^{CAS}_{prob}(D, a_{sens}) \overset{\text{def}}{=} \max_{s\in dom(a_{sens})} \pi_s^{a_{sens}|D}$$

[5] This definition corresponds to the *a priori* Bayes vulnerability $V_1[\pi_s^{a_{sens}|D}]$, given by Definitions 6 and 8 in Section 2.3, of the sensitive attribute on the *a priori* probability distribution $\pi_s^{a_{sens}|D}$.

$$= \max_{s \in dom(a_{sens})} \frac{|\{x{\in}D \mid x[a_{sens}]{=}s\}|}{|D|} . \tag{4.11}$$

– ***A posteriori* probabilistic success.** By performing a CAS attack, the adversary uses the auxiliary information on the sub-records $\{x[\mathcal{Q}_{ID}]\}_{x\in D}$ corresponding to the values of quasi-identifiers for every individual in the database to support the task of inferring the sensitive attribute value. Given a randomly selected target in the database, the adversary uses the corresponding quasi-identifier values to filter the database records, disregarding those that do not match the query. From there, their best course of action is to randomly select a possible value for the sensitive attribute among those available in the remaining database, according to the maximum entropy principle. Since each individual only holds one record in the database, the most probable value would be the most frequent one in the remaining database.

Therefore, the *a posteriori* probabilistic success in a CAS attack on a database D and with respect to the sensitive attribute a_{sens}, given the adversary's knowledge of the values $\{x[\mathcal{Q}_{ID}]\}_{x\in D}$ for every individual's quasi-identifiers, is defined as the expected value of the probability of success taken from the probability distribution for selecting each individual as the target: [6]

$$\textit{post-suc}^{CAS}_{prob}(D, a_{sens}, \{x[\mathcal{Q}_{ID}]\}_{x\in D}) \overset{\text{def}}{=}$$

$$\sum_{x\in D} \pi^D_x \max_{s \in dom(a_{sens})} \frac{|\{x'{\in}D \mid x'[a_{sens}]{=}s, x'[\mathcal{Q}_{ID}]{=}x[\mathcal{Q}_{ID}]\}|}{|\{x'{\in}D \mid x'[\mathcal{Q}_{ID}]{=}x[\mathcal{Q}_{ID}]\}|} =$$

[6] This definition corresponds to the *a posteriori* Bayes vulnerability $V_1[\pi_s^{a_{sens}|D} \triangleright C^{a_{sens},\mathcal{Q}_{ID}}]$, given by Definitions 6 and 8 in Section 2.3, of the sensitive attribute given the knowledge over the quasi-identifiers, where:

∗ $\pi_s^{a_{sens}|D}$ is the *a priori* probability distribution on the sensitive attribute, according to Equation 3.2, and is defined as:

$$\pi_s^{a_{sens}|D} = \frac{|\{x{\in}D \mid x[a_{sens}]{=}s\}|}{|D|} ;$$

∗ $C^{a_{sens},\mathcal{Q}_{ID}}{:}dom(a_{sens}){\to}\mathbb{D}(dom(\mathcal{Q}_{ID}))$ is the channel that deterministically maps the possible values of the sensitive attribute to values of quasi-identifiers associated with it and is defined, for all $s{\in}dom(a_{sens})$ and $q{\in}dom(\mathcal{Q}_{ID})$, as:

$$C^{a_{sens},\mathcal{Q}_{ID}}_{s,q} = \frac{|\{x{\in}D \mid x[a_{sens}]{=}s, x[\mathcal{Q}_{ID}] = q\}|}{|\{x{\in}D \mid x[a_{sens}]{=}s\}|} .$$

$$
\begin{aligned}
&= \sum_{x \in D} \frac{1}{|D|} \cdot \max_{s \in dom(a_{sens})} \frac{|\{x' \in D \mid x'[a_{sens}] = s, x'[\mathcal{Q}_{ID}] = x[\mathcal{Q}_{ID}]\}|}{|\{x' \in D \mid x'[\mathcal{Q}_{ID}] = x[\mathcal{Q}_{ID}]\}|} \\
&= \frac{1}{|D|} \sum_{q \in dom(\mathcal{Q}_{ID})} \Big(\, |\{x' \in D \mid x'[\mathcal{Q}_{ID}] = q\}| \cdot \\
&\qquad\qquad \max_{s \in dom(a_{sens})} \frac{|\{x' \in D \mid x'[a_{sens}] = s, x'[\mathcal{Q}_{ID}] = q\}|}{|\{x' \in D \mid x'[\mathcal{Q}_{ID}] = q\}|} \Big) \\
&= \frac{1}{|D|} \sum_{q \in dom(\mathcal{Q}_{ID})} \max_{s \in dom(a_{sens})} |\{x \in D \mid x[a_{sens}]{=}s, x[\mathcal{Q}_{ID}]{=}q\}| \ , \qquad (4.12)
\end{aligned}
$$

i.e. for each quasi-identifier value q and each sensitive attribute value s, take the size of the largest available block in the partition of the database induced by $\mathcal{Q}_{ID}$, add them together, and divide by the size of the database.

- **Probabilistic degradation of privacy.** The probabilistic degradation of privacy in a CAS attack is defined as the ratio by which the attack increases the probabilistic success of the adversary:

$$
\begin{aligned}
&\textit{priv-degrad}^{CAS}_{prob}(D, a_{sens}, \{x[\mathcal{Q}_{ID}]\}_{x \in D}) \overset{\text{def}}{=} \\
&\frac{\textit{post-suc}^{CAS}_{prob}(D, a_{sens}, \{x[\mathcal{Q}_{ID}]\}_{x \in D})}{\textit{prior-suc}^{CAS}_{prob}(D, a_{sens})} \ . \qquad (4.13)
\end{aligned}
$$

A detailed numeric example of a CAS attack is presented in Section D.1.2, Example 50.

◁

4.2 Experimental results for collective-target attacks on the School Census of 2018

In this section, we present our quantitative analyses on the privacy risks for individuals whose information are made public on a single database. The attacks performed here were modeled according to the theory developed in Section 4.1. For illustrative individual-target experimental results, in which we target either famous people or our acquaintances selected *a priori*, see Appendix C Section C.1.2.

We begin by detailing our experimental setup, followed by the results for re-identification and then for attribute-inference attacks on INEP's databases.

4.2.1 Experimental setup

As stated in Section 1.3, we have chosen to work with the databases containing information on students and released as microdata, i.e. data at the record level. In this chapter, we consider the database for the School Census of 2018, leaving the analyses for the School Census of 2019 to Appendix E and for the Higher Education Censuses of 2018 and 2019 to Appendix F Section F.1.

A preliminary analysis of the databases containing information on students showed that a student may hold more than one record in a given database, e.g. if a High School student is also enrolled in a Professional Education course. Hence, in order to guarantee that each student holds only one record in each database, according to Assumption **AS1** from Section 4.1.1, we have randomly selected only one record for each data holder of multiple records. This data treatment was based on the unique identification number, the `ID_ALUNO` code, given to each student in the pseudonymization treatment performed by INEP. The `ID_ALUNO` code, which is unique to each student at least in a given database release, easily allowed us to find those students with multiple records and to perform the random selection of just one of them. Furthermore, the treatment has not resulted in a significant decrease in the number of available records. For the School Census of 2018, 92.95% of the records were kept, i.e. 48 176 423 records from a total of 51 829 413.

Even though each database accounts for dozens of attributes, we have chosen just a few for our analyses, given the computational costs, including time and memory usage. The selection criteria was as follows.

- For the quasi-identifying attributes, we have sets restricted to the maximum of ten or eleven attributes, selected according to how easily an adversary could learn them, e.g. the date and city of birth, the city of residency, or the school code. The quasi-identifying attributes chosen for the experiments in this chapter are listed in Table 4.2.
- We have chosen two sensitive attributes selected according to the seriousness of the possible individual privacy breach if revealed, e.g. whether or not the individual has special needs or disabilities. The sensitive attributes chosen for the experiments in this chapter are listed in Table 4.3.

Of course, the selection of those quasi-identifiers and sensitive attributes is arbitrary and can change in significance over time. Nevertheless, our goal is not to ultimately define whether an attribute should be considered as a quasi-identifier or as a sensitive

Variable	Meaning
NU_DIA	Student's day of birth.
NU_MES	Student's month of birth.
NU_ANO	Student's year of birth.
TP_SEXO	Student's gender.
TP_COR_RACA	Student's ethnicity.
TP_NACIONALIDADE	Student's nationality.
CO_PAIS_ORIGEM	Student's country code.
CO_MUNICIPIO_NASC	Student's city of birth code.
CO_MUNICIPIO_END	Student's city of residency code.
CO_ENTIDADE	School code.
TP_DEPENDENCIA	Administrative dependency of the school, i.e. whether public or private.

Table 4.2: Variables from the School Census of 2018 chosen as quasi-identifiers for Collective-target Re-identification Single database (CRS) attacks in Experiment 18, and for Collective-target Attribute-inference Single database (CAS) attacks in Experiment 19.

Variable	Values	Meaning
IN_NECESSIDADE_ESPECIAL	0 (No) 1 (Yes)	Whether the student possesses a disability, global developmental disorder, or autism spectrum disorder, or not.
IN_TRANSPORTE_PUBLICO	-1 (Unavailable) 0 (No) 1 (Yes)	Whether the student uses public school transport, or not.

Table 4.3: Variables from the School Census of 2018 chosen as sensitive attributes for Collective-target Attribute-inference Single database (CAS) attacks in Experiment 19.

attribute. Instead, the results here illustrate possible real-life circumstances and the respective privacy risks for data holders in case of unintended information disclosure.

In the following Sections 4.2.2 and 4.2.3, we present some experimental results for re-identification and attribute-inference attacks, respectively. Both were performed on single databases for collective-targets on the School Census of 2018 released by INEP.

4.2.2 Results of re-identification attacks experiments

In this section, we present our quantitative analyses on the privacy risk of re-identification for individuals whose information are made public on a single database, as modeled in Section 4.1.2. We present one CRS attacks on the School Census of 2018, Experiment 18.

Experiment 18 (Collective-target Re-identification Single database (CRS) attacks on the School Census of 2018)**.** In a CRS attack, the adversary's goal is to re-identify as many individuals as possible, according to Definition 16. Since we want to evaluate the overall re-identification risk, the adversary can gather auxiliary information on the values of a set of quasi-identifying attributes for every individual who holds a record in the database. The following experiment was conducted on the School Census of 2018.

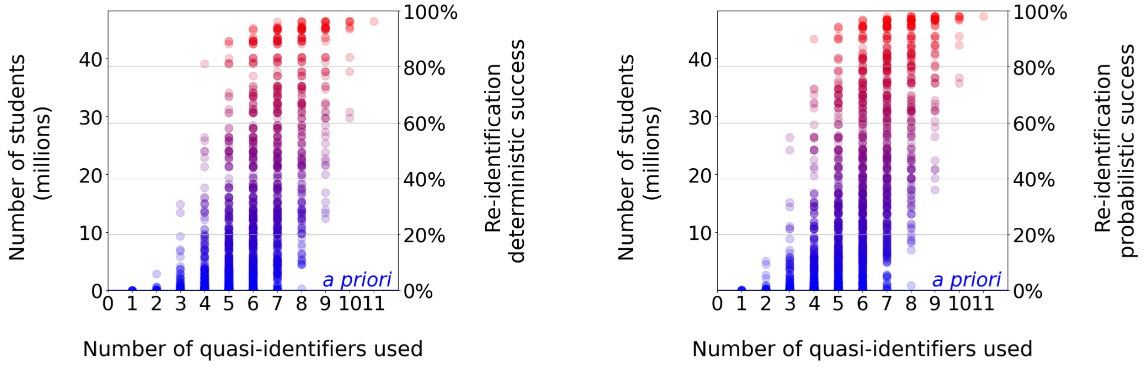
This database was released containing 51 829 413 records, which were reduced to 48 176 423 after the random selection of only one record for each data holder of multiple records, as detailed in Section 4.2.1. For the current experiment, we have selected as quasi-identifiers the eleven attributes listed in Table 4.2. Considering an adversary could gather as auxiliary information any non-empty subset among the $2^{11} - 1 = 2\,047$ possible combinations of quasi-identifying attributes, we have measured both the deterministic and probabilistic degradation of privacy for each combination.

The results are organized as follows.

- Table 4.4 summarizes which combinations of quasi-identifiers result in the highest degradation of privacy according to the subset's size.
- Figure 4.5 shows both the adversary's deterministic and probabilistic measures of success for each combination of quasi-identifying attributes.
- Figure 4.6 shows the worst-case histograms for the distribution of individuals according to the adversary's probabilistic measure of success for some subsets of quasi-identifying attributes with different sizes.

According to those results, we were able to conclude the following.

- **Adversary's deterministic success**. Defined in Equations 4.1 and 4.2, this metric measures the fraction of individuals in the database that can be re-identified with absolute certainty, in a scale from 0% to 100%. According to Table 4.4, the adversary's *a priori* deterministic success equals 0%, i.e. no individual can be re-identified with absolute certainty without the use of auxiliary



(a) Deterministic success.

(b) Probabilistic success.

Figure 4.5: Experiment 18: Adversary's success in Collective-target Re-identification Single database (CRS) attacks on the School Census of 2018. Here, the horizontal "*a priori*" line represents the adversary's *a priori* deterministic or probabilistic success, i.e. before performing the CRS attack. Each dot represents a distinct scenario defined by the selected quasi-identifying attributes among the 2 047 possibilities. The horizontal axis defines the size of the quasi-identifiers subset, while the vertical axis defines the adversary's deterministic or probabilistic success.

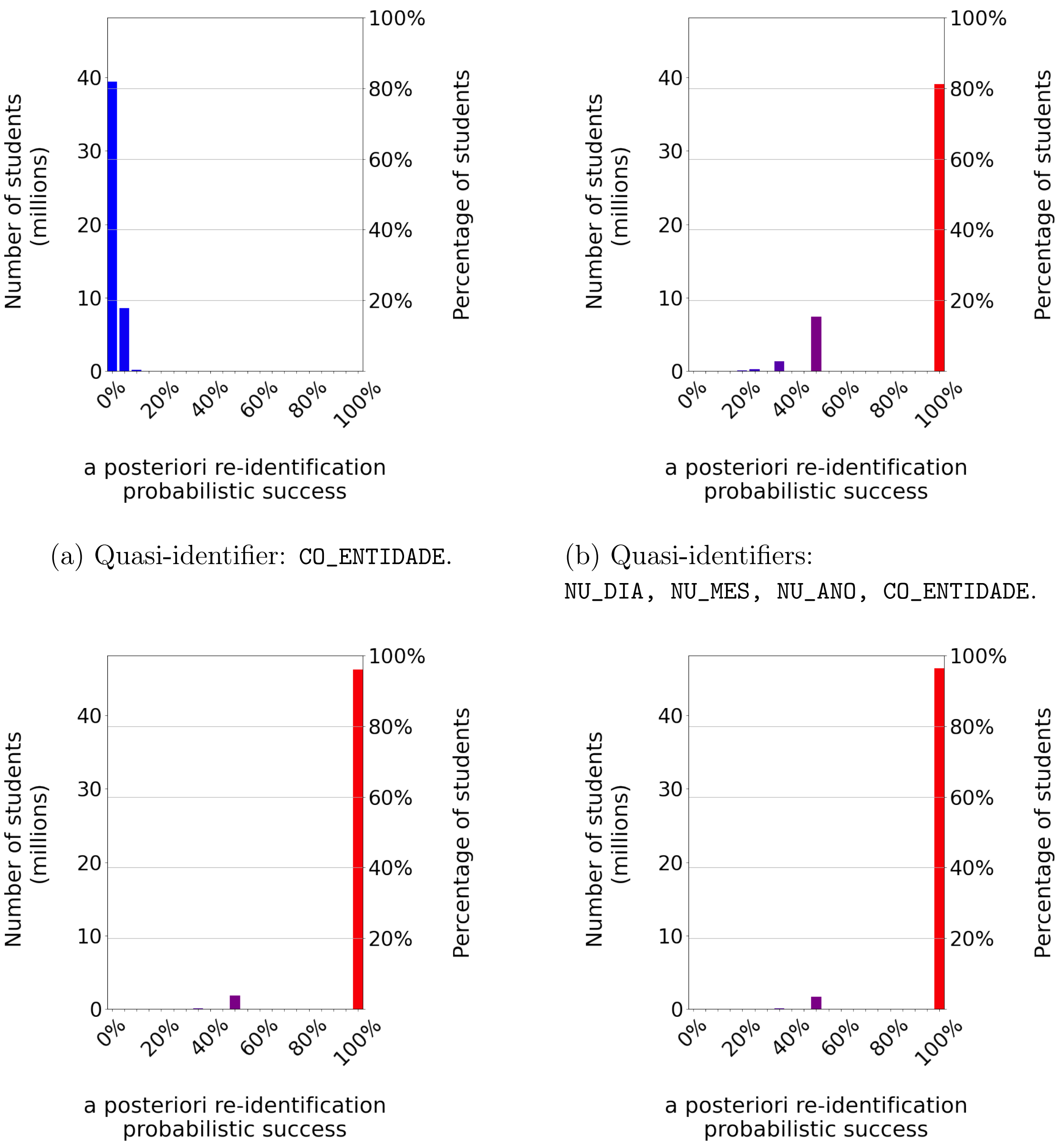
information. However, by using only three quasi-identifying attributes, the adversary can re-identify with absolute certainty up to 30.92% of the individuals in the database, while the use of four quasi-identifiers allows the adversary to re-identify up to 81.13% of the individuals. Finally, by using eight or more of the eleven quasi-identifying attributes available, the risk of re-identification increases to up to 96.34%.

- **Adversary's probabilistic success**. Defined in Equations 4.4 and 4.5, this metric measures the probability of a randomly chosen individual in the database being re-identified, in a scale from 0% to 100%. According to Table 4.4, the adversary's *a priori* probabilistic success is approximately 0.000002%, i.e. the adversary's chance of re-identifying one of the 48 176 423 individuals in the database is almost zero. However, by using only three quasi-identifying attributes, the adversary increases their chance to up to 54.76%, while the use of four quasi-identifiers increases this probability to up to 89.93%. Finally, by using nine or more of the eleven quasi-identifying attributes available, the adversary's chance of re-identifying a randomly chosen individual in the database increases to up to 98.14%.

◁

		Adversary's deterministic success		Adversary's probabilistic success	
		a priori success 0.00%		*a priori* success 0.000002%	
#	**Quasi-identifiers**	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
1	CO_ENTIDADE	0.0002%	0.0002%	0.38%	183 706
2	CO_MUNICIPIO_NASC, CO_ENTIDADE	5.88%	5.88%	9.34%	4 502 009
2	NU_DIA, CO_ENTIDADE	1.39%	1.39%	10.38%	5 000 106
3	NU_DIA, NU_MES, CO_ENTIDADE	30.92%	30.92%	54.76%	26 381 055
4	NU_DIA, NU_MES, NU_ANO, CO_ENTIDADE	81.13%	81.13%	89.93%	43 323 676
5	NU_DIA, NU_MES, NU_ANO, TP_SEXO, CO_ENTIDADE	89.23%	89.23%	94.41%	45 484 648
6	NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_ENTIDADE	93.92%	93.92%	96.88%	46 674 542
7	NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_ENTIDADE	96.05%	96.05%	97.99%	47 206 505
8	NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE	96.34%	96.34%	98.14%	47 279 677
9	NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE	96.34%	96.34%	98.14%	47 279 763
10	NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE	96.34%	96.34%	98.14%	47 279 763
10	NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA	96.34%	96.34%	98.14%	47 279 763
11	NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA	96.34%	96.34%	98.14%	47 279 763

Table 4.4: Experiment 18: Quasi-identifiers with the highest degradation of privacy in Collective-target Re-identification Single database (CRS) attacks on the School Census of 2018. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* successes, while the probabilistic degradation of privacy is their ratio. Table 4.2 lists the English meaning of each quasi-identifying attribute.



(a) Quasi-identifier: `CO_ENTIDADE`.

(b) Quasi-identifiers: `NU_DIA, NU_MES, NU_ANO, CO_ENTIDADE`.

(c) Quasi-identifiers: `NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_ENTIDADE`.

(d) Quasi-identifiers: `NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA`.

Figure 4.6: Experiment 18: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Re-identification Single database (CRS) attacks on the School Census of 2018. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of re-identification.

4.2.3 Results of attribute-inference attacks experiments

In this section, we present our quantitative analyses on the privacy risk of attribute-inference for individuals whose information are made public on a single database, as modeled in Section 4.1.2. We present one CAS attacks on the School Census of 2018, Experiment 19, on sensitive attributes IN_NECESSIDADE_ESPECIAL and IN_TRANSPORTE_PUBLICO.

Experiment 19 (Collective-target Attribute-inference Single database (CAS) attacks on the School Census of 2018)**.** In a CAS attack, the adversary's goal is to infer the values of an attribute considered to be sensitive for as many individuals as possible, according to Definition 17. Since we want to evaluate the overall attribute-inference risk, the adversary can gather auxiliary information on the values of a set of quasi-identifying attributes for every individual who holds a record in the database. The following experiment was conducted on the School Census of 2018.

This database was released containing 51 829 413 records, which were reduced to 48 176 423 after the random selection of only one record for each data holder of multiple records, as detailed in Section 4.2.1. For the current experiment, we have selected as quasi-identifiers the eleven attributes listed in Table 4.2. Considering an adversary could gather as auxiliary information any non-empty subset among the $2^{11} - 1 = 2\,047$ possible combinations, we have measured both the deterministic and probabilistic degradation of privacy for each combination.

Furthermore, we were interested in inferring the value for the attributes IN_NECESSIDADE_ESPECIAL and IN_TRANSPORTE_PUBLICO, both described in Table 4.3 and considered by us to be sensitive.

The results are organized as follows.

- Tables 4.7 and 4.8 summarize which combinations of quasi-identifiers result in the highest degradation of privacy according to the subset's size for the sensitive attributes IN_NECESSIDADE_ESPECIAL and IN_TRANSPORTE_PUBLICO, respectively.
- Figure 4.9 shows both the adversary's deterministic and probabilistic measures of success for each combination of quasi-identifying attributes, and for both sensitive attributes.
- Figures 4.10 and 4.11 show the worst-case histograms for the distribution of individuals according to the adversary's probabilistic measure of success for some subsets of quasi-identifying attributes with different sizes and for the sensitive attributes IN_NECESSIDADE_ESPECIAL and IN_TRANSPORTE_PUBLICO, respectively.

According to those results, we were able to conclude the following.

- **Adversary's deterministic success**. Defined in Equations 4.7 and 4.8, this metric measures the fraction of individuals in the database that can have their values for the sensitive attribute inferred with absolute certainty, in a scale from 0% to 100%.

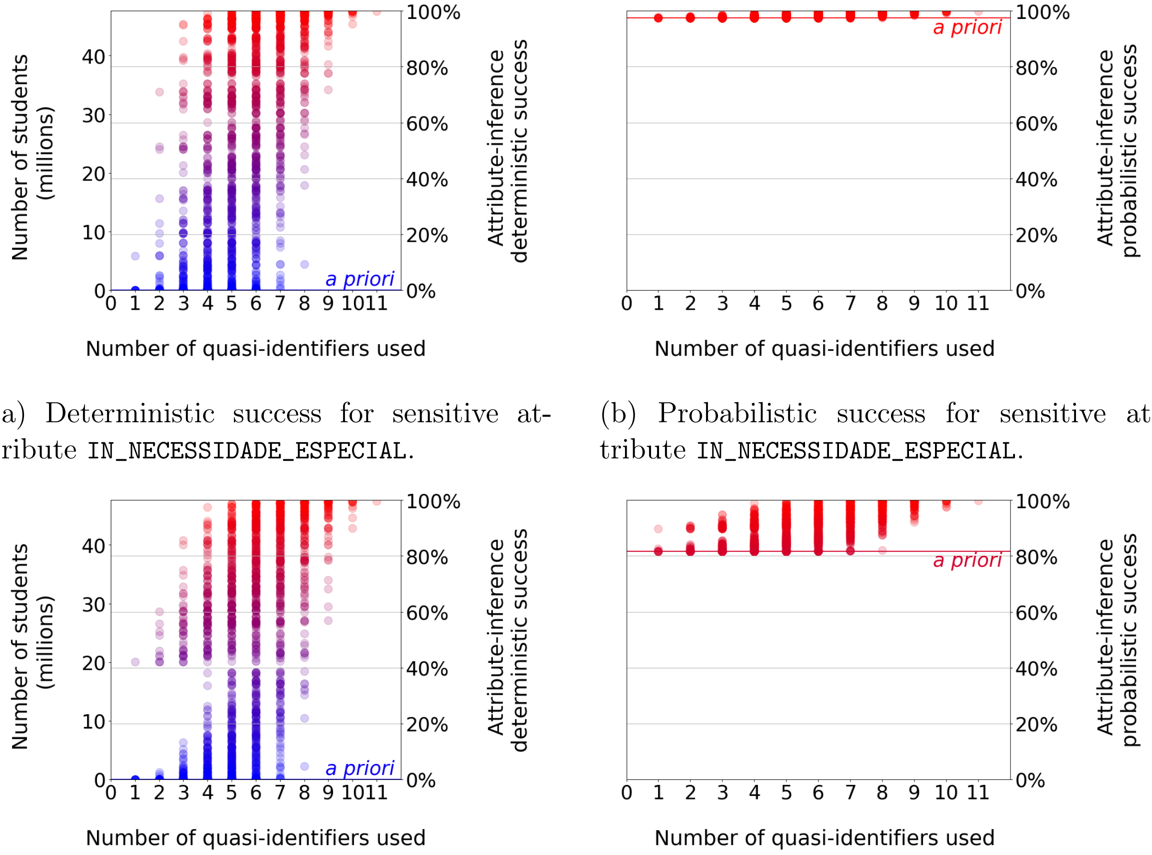
  According to Table 4.7 for the sensitive attribute `IN_NECESSIDADE_ESPECIAL`, the adversary's *a priori* deterministic success equals 0%, i.e. no individual can have their value for the sensitive attribute inferred with absolute certainty without the use of auxiliary information. However, by using only one quasi-identifying attribute, the adversary can infer the values with absolute certainty for up to 12.31% of the individuals in the database, while the use of two quasi-identifiers allows the adversary to infer the values for up to 71.04% of the individuals. Finally, by using four or more of the eleven quasi-identifying attributes available, the risk of attribute-inference increases above 99.31%.

  From Table 4.8 for the sensitive attribute `IN_TRANSPORTE_PUBLICO`, the adversary's *a priori* deterministic success also equals 0%. However, by using only one quasi-identifying attribute, the adversary can infer the values with absolute certainty for up to 42.08% of the individuals in the database, while the use of two quasi-identifiers allows the adversary to infer the values for up to 60.02% of the individuals. Finally, by using six or more of the eleven quasi-identifying attributes available, the risk of attribute-inference increases above 99.30%.

- **Adversary's probabilistic success**. Defined in Equations 4.11 and 4.12, this metric measures the probability of a randomly chosen individual in the database having their value for the sensitive attribute inferred with absolute certainty, in a scale from 0% to 100%.

  According to Table 4.7 for the sensitive attribute `IN_NECESSIDADE_ESPECIAL`, the adversary's *a priori* probabilistic success is already of 97.56%, i.e. the adversary's chance of inferring the value for the sensitive attribute for one of the 48 176 423 individuals in the database is already high even before performing the CAS attack. Furthermore, by using four or more quasi-identifying attributes, the adversary increases their chance above 99.69%.

  From Table 4.8 for the sensitive attribute `IN_TRANSPORTE_PUBLICO`, the adversary's *a priori* probabilistic success is of 81.75%, also high even before performing the CAS attack. Furthermore, by using five or more quasi-identifying attributes, the adversary increases their chance above 99.39%.



(a) Deterministic success for sensitive attribute `IN_NECESSIDADE_ESPECIAL`.

(b) Probabilistic success for sensitive attribute `IN_NECESSIDADE_ESPECIAL`.

(c) Deterministic success for sensitive attribute `IN_TRANSPORTE_PUBLICO`.

(d) Probabilistic success for sensitive attribute `IN_TRANSPORTE_PUBLICO`.

Figure 4.9: Experiment 19: Adversary's success in Collective-target Attribute-inference Single database (CAS) attacks on the School Census of 2018. Here, the horizontal "*a priori*" line represents the adversary's *a priori* deterministic or probabilistic success, i.e. before performing the CAS attack. Each dot represents a distinct scenario defined by the selected quasi-identifying attributes among the 2 047 possibilities. The horizontal axis defines the size of the quasi-identifiers subset, while the vertical axis defines the adversary's deterministic or probabilistic success.

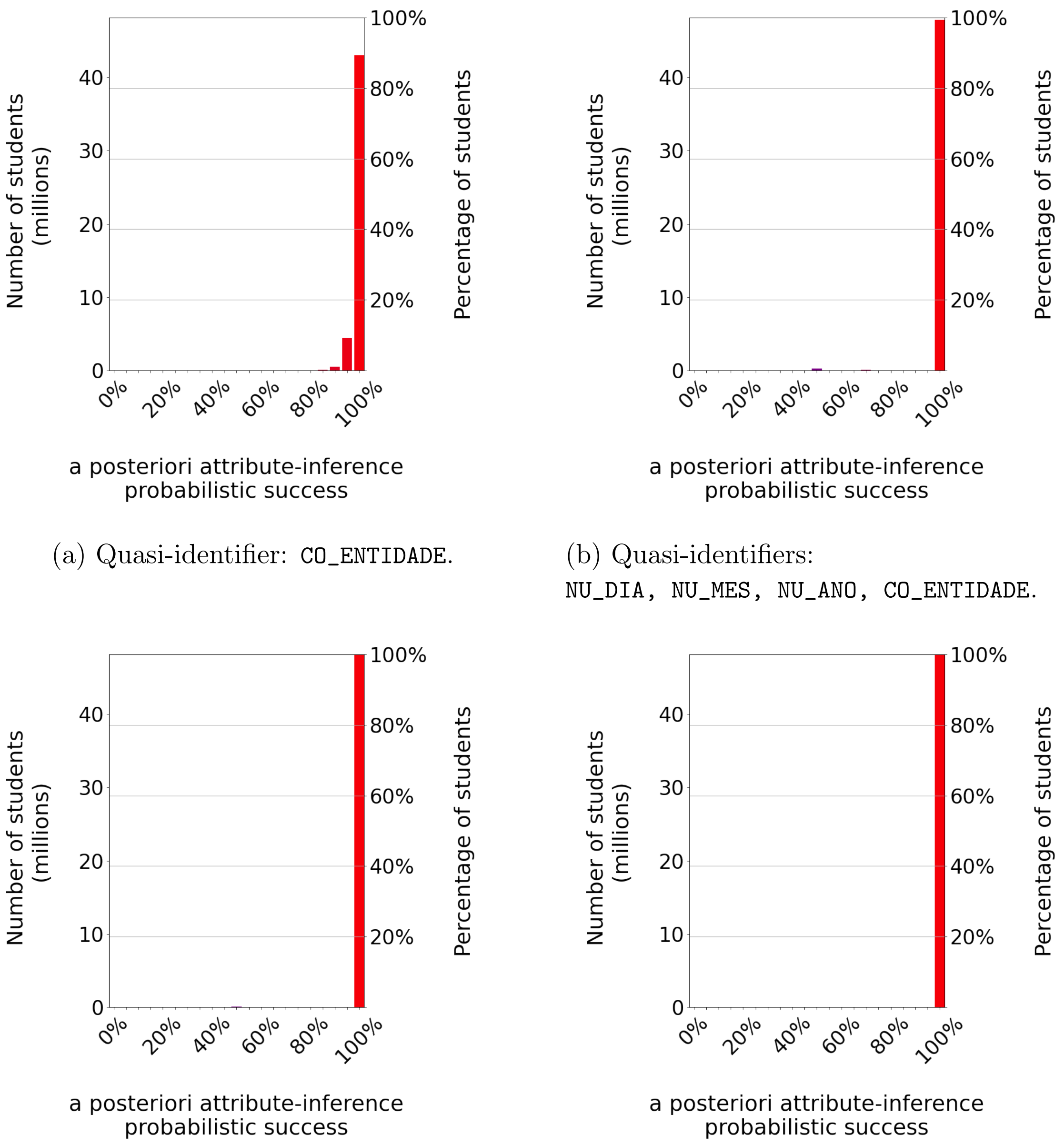
◁

		Adversary's deterministic success		**Adversary's probabilistic success**	
		a priori **success** 0.00%		***a priori*** **success** 97.56%	
#	**Quasi-identifiers**	***a posteriori*** **success**	**Privacy degradation**	***a posteriori*** **success**	**Privacy degradation**
1	`CO_ENTIDADE`	12.31%	12.31%	97.92%	1.0037
2	`NU_DIA, CO_ENTIDADE`	71.04%	71.04%	97.95%	1.0040
2	`CO_MUNICIPIO_NASC, CO_ENTIDADE`	32.80%	32.80%	98.05%	1.0051
3	`NU_DIA, NU_MES, CO_ENTIDADE`	95.35%	95.35%	98.59%	1.0105
3	`NU_DIA, NU_ANO, CO_ENTIDADE`	94.88%	94.88%	98.67%	1.0114
4	`NU_DIA, NU_MES, NU_ANO, CO_ENTIDADE`	99.31%	99.31%	99.69%	1.0218
5	`NU_DIA, NU_MES, NU_ANO, TP_COR_RACA, CO_ENTIDADE`	99.64%	99.64%	99.83%	1.0233
6	`NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_ENTIDADE`	99.81%	99.81%	99.91%	1.0241
7	`NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_ENTIDADE`	99.88%	99.88%	99.94%	1.0245
8	`NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE`	99.89%	99.89%	99.95%	1.0245
9	`NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE`	99.89%	99.89%	99.95%	1.0245
9	`NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE`	99.89%	99.89%	99.95%	1.0245
10	`NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE`	99.89%	99.89%	99.95%	1.0245
10	`NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA`	99.89%	99.89%	99.95%	1.0245
10	`NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA`	99.89%	99.89%	99.95%	1.0245
11	`NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA`	99.89%	99.89%	99.95%	1.0245

Table 4.7: Experiment 19: Quasi-identifiers with the highest degradation of privacy in Collective-target Attribute-inference Single database (CAS) attacks on the School Census of 2018 for the sensitive attribute `IN_NECESSIDADE_ESPECIAL`. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* successes, while the probabilistic degradation of privacy is their ratio. Table 4.2 lists the English meaning of each quasi-identifying attribute.

		Adversary's deterministic success		Adversary's probabilistic success	
		a priori success 0.00%		*a priori* success 81.75%	
#	**Quasi-identifiers**	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
1	CO_ENTIDADE	42.08%	42.08%	89.79%	1.0983
2	NU_DIA, CO_ENTIDADE	60.02%	60.02%	90.66%	1.1090
2	CO_MUNICIPIO_NASC, CO_ENTIDADE	32.80%	32.80%	90.95%	1.1126
3	NU_DIA, NU_MES, CO_ENTIDADE	85.63%	85.63%	94.85%	1.1602
4	NU_DIA, NU_MES, NU_ANO, CO_ENTIDADE	97.42%	97.42%	98.82%	1.2088
5	NU_DIA, NU_MES, NU_ANO, CO_MUNICIPIO_NASC, CO_ENTIDADE	98.68%	98.68%	99.39%	1.2157
6	NU_DIA, NU_MES, NU_ANO, TP_SEXO, CO_MUNICIPIO_NASC, CO_ENTIDADE	99.30%	99.30%	99.67%	1.2191
7	NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_ENTIDADE	99.62%	99.62%	99.82%	1.2210
8	NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE	99.65%	99.65%	99.83%	1.2211
9	NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE	99.65%	99.65%	99.83%	1.2211
10	NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE	99.65%	99.65%	99.83%	1.2211
10	NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA	99.65%	99.65%	99.83%	1.2211
11	NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA	99.65%	99.65%	99.83%	1.2211

Table 4.8: Experiment 19: Quasi-identifiers with the highest degradation of privacy in Collective-target Attribute-inference Single database (CAS) attacks on the School Census of 2018 for the sensitive attribute IN_TRANSPORTE_PUBLICO. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* successes, while the probabilistic degradation of privacy is their ratio. Table 4.2 lists the English meaning of each quasi-identifying attribute.



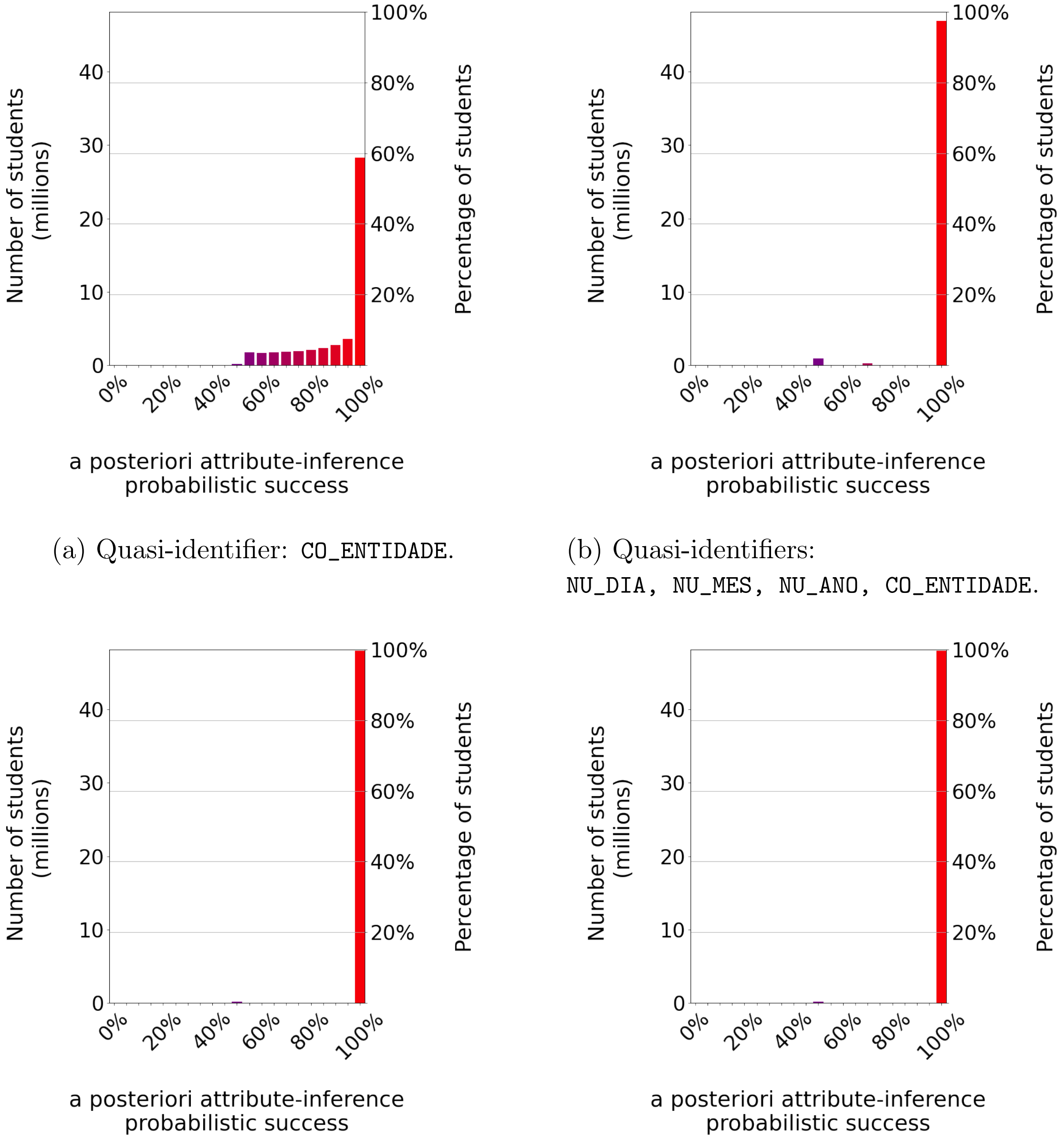
(a) Quasi-identifier: `CO_ENTIDADE`.

(b) Quasi-identifiers: `NU_DIA, NU_MES, NU_ANO, CO_ENTIDADE`.

(c) Quasi-identifiers: `NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_ENTIDADE`.

(d) Quasi-identifiers: `NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA`.

Figure 4.10: Experiment 19: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Attribute-inference Single database (CAS) attacks on the School Census of 2018 for the sensitive attribute `IN_NECESSIDADE_ESPECIAL`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of attribute-inference.



(a) Quasi-identifier: `CO_ENTIDADE`.

(b) Quasi-identifiers: `NU_DIA, NU_MES, NU_ANO, CO_ENTIDADE`.

(c) Quasi-identifiers: `NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_ENTIDADE`.

(d) Quasi-identifiers: `NU_DIA, NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA`.

Figure 4.11: Experiment 19: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Attribute-inference Single database (CAS) attacks on the School Census of 2018 for the sensitive attribute `IN_TRANSPORTE_PUBLICO`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of attribute-inference.

4.3 Takeaways

In this chapter, we have presented our theoretical foundation on attacks against single databases for collective-targets in Section 4.1. We have considered both re-identification and attribute-inference attacks, and both deterministic and probabilistic metrics of success and privacy degradation. Particularly, the probabilistic metrics are applications of the quantitative information flow (QIF) framework introduced in Section 2.3.

We have also presented experimental results for the collective-target attacks on the School Census of 2018 released by INEP in Section 4.2, Experiments 18 and 19, for re-identification and attribute-inference attacks, respectively.

Those results have allowed us to draw the following conclusions.

- The **deterministic re-identification success** of an adversary with knowledge of only three attributes in Experiment 18, i.e. day and month of birth and school code, achieves up to 30.92% of the records, i.e. approximately 14 896 149 students that can be re-identified with absolute certainty. Adding the year of birth to the adversary's knowledge increases their success to 81.13%, i.e. approximately 39 085 531 students. A *plateau* is then reached when adding gender and ethnicity to the adversary's knowledge, which increases their re-identification success to 93.92%, i.e. approximately 45 247 296 students.

  These results are similar to Sweeney's findings in 2000. According to Sweeney [67], 87.10% of the United States population in the Census of 1990 could be uniquely re-identified in the de-identified data by using only a combination of the date of birth, gender, and a 5-digit ZIP code. If we consider that the school code approximately identifies where a given student lives, we can use it as an approximation to that student's ZIP code. Hence, considering the date of birth, gender, and school code, 89.23% of the students can be re-identified in the School Census of 2018.

- The analyses of quasi-identifying attributes sets of different sizes allowed us to further demonstrate that the removal of attributes alone is not sufficient to guarantee individuals' privacy. For instance, even if only the school code attribute was released, approximately 96 students would be susceptible to re-identification attacks, or more than 5.9 million students could have their attribute for disabilities inferred with absolute certainty, in the School Census of 2018. Furthermore, just removing attributes is not a scalable solution since each attribute removed reduces the database's utility by reducing the amount of information available, which may impact the work of data analysts.

- The **probabilistic re-identification success** of an adversary with knowledge of only three attributes in Experiment 18, i.e. day and month of birth and school code, is a 54.76% chance of correctly re-identifying a randomly selected record. Adding the year of birth to the adversary's knowledge increases their chance of success to 89.93%. A *plateau* is then reached when adding gender and ethnicity to the adversary's knowledge, which increases their re-identification success to 96.88%.

- The **deterministic attribute-inference success** of an adversary achieves higher values and increases more rapidly than the respective deterministic re-identification success. This was observed in Experiment 19 for both sensitive attributes `IN_NECESSIDADE_ESPECIAL` and `IN_TRANSPORTE_PUBLICO`, described in Table 4.3. For instance, an adversary could correctly infer the value for both sensitive attributes of more than 99% of the records, i.e. approximately than 47 694 658 students, with either four quasi-identifiers for the sensitive attribute `IN_NECESSIDADE_ESPECIAL`, or six quasi-identifiers for the sensitive attribute `IN_TRANSPORTE_PUBLICO`. In the best scenario for the adversary, they could achieve a 95.35% success, or the re-identification of approximately 45 936 219 students, by knowing only three attributes, i.e. day and month of birth and school code, when trying to infer whether a student possesses a disability or not in Experiment 19.

- The **probabilistic attribute-inference success** of an adversary is initially already high. For the sensitive attribute on students' disabilities, an adversary has an *a priori* chance of correctly inferring the value for a randomly chosen record of 97.56% for the School Census of 2018. This is because of how much skewed the distribution of students is for that attribute. A more skewed distribution implies a higher *a priori* success, as is the case for the attribute on students' disabilities, for which the majority of individuals holds the value 0 (No). Analogously, the opposite holds for the attribute on students' use of public transport, which presents a slightly less skewed distribution and, hence, lower values for an adversary's *a priori* chance, of 81.75% for the School Census of 2018. [7] Therefore, not much increase in an adversary's success is seen as a result of the use of more

[7] We have used average-case metrics for all of our analyses. Instead, a better approach would have been to use worse-case metrics whenever we are dealing with very skewed distributions [4, 5].

> quasi-identifiers for the attribute-inference attacks on both sensitive attributes. Even so, it is remarkable that by using four attributes in Experiment 19, the adversary's chance of correctly inferring the value of the sensitive attribute on students' disabilities for a randomly chosen record reaches values above 99%. [8]

Therefore, as expected from the literature on disclosure control (DC) discussed in Section 2.1.1, INEP's use of de-identification and pseudonymization as the only DC methods for protecting data holders' privacy is clearly insufficient. In fact, almost the entirety of data holders in the database analyzed in our experiments is subject to re-identification or attribute-inference attacks. This is of particular concern when considering an adversary's probabilistic measures of success, which are designed to measure how certain an adversary would be in a given attack. According to our results, that certainty is extremely high for the average data holder in most scenarios. Hence, given the recent enactment of Brazil's LGPD privacy legislation, our work shows unequivocally that mitigating such vulnerabilities is necessary and urgent for INEP.

We have also performed the same collective-target attacks from this chapter on the School Census of 2019 in Appendix E and on the Higher Education Census of 2018 and 2019 in Appendix F Section F.1. Furthermore, for illustrative individual-target experimental results, in which we target either famous people or our acquaintances selected *a priori*, see Appendix C Section C.1.2.

In the following Chapter 5, we formalize similar re-identification and attribute-inference attacks against longitudinal databases and present the respective experimental results for attacks performed on the School Census released by INEP. Longitudinal database scenarios are of particular significance given most DC methods do not account for their specific vulnerabilities. Furthermore, longitudinal database releases are fairly common, particularly in the context of demographic statistics.

[8] For the sensitive attribute on students' use of public transport, an adversary has an *a priori* chance of correctly inferring the value for a randomly chosen record of 81.75% for the School Census of 2018. In this case, however, the adversary's chance of success reaches values above 99% for five attributes in Experiment 19.

Chapter 5

Collective-target attacks on longitudinal databases

We have presented in Chapter 4 our models for attacks against single databases and some experimental results that have demonstrated to what extent the databases considered are vulnerable to the proposed attacks. However, as discussed in Chapter 2, most disclosure control (DC) methods were designed to be applied to single databases, leaving longitudinal databases open to additional vulnerabilities. Furthermore, longitudinal releases of databases are fairly common, particularly in the context of demographic statistics such as national and educational censuses.

Given the serious privacy degradation observed in the experiments on single databases in Chapter 4, we will investigate in the following sections how the constant release of additional databases may further degrade data holders' privacy in INEP's scenarios. Particularly, we expect that small sets of quasi-identifying attributes that were not the most relevant in the single databases scenario to become seriously dangerous for data holders in the longitudinal databases scenario.

In this chapter, we present the theoretical foundation on attacks against longitudinal databases for collective-targets in Section 5.1, followed by the respective experimental results for attacks on the School Census released by INEP in Section 5.2. We conclude this chapter by emphasizing the most relevant contributions in Section 5.3.

5.1 Theoretical foundation

In this section, we present our model for attacks performed on longitudinal databases for collective-targets. We consider our database model introduced in Section 3.2.

5.1.1 Assumptions

We have the following assumptions for all attacks on longitudinal databases:

AL0 There is a *longitudinal collection* $\mathcal{L}_{\mathcal{D}} = \{D_1, D_2, \ldots, D_k\}$, with $k{\geq}1$, of databases of interest, each database D_i on an attributes set $\mathcal{A}_i$, with $1{\leq}i{\leq}k$.

AL1 Each individual of interest holds a record in each database $D_i{\in}\mathcal{L}_{\mathcal{D}}$.

AL2 There is an *attribute of unique identification* a_{id} common to all databases in $\mathcal{L}_{\mathcal{D}}$ and each individual of interest holds a persistent value for this attribute whenever present in a database, i.e. there is $a_{id}{\in}\mathcal{A}_i$, with $1{\leq}i{\leq}k$, such that whenever a given individual holds a record $x{\in}D_i$ and another record $x'{\in}D_j$, with $D_i, D_j{\in}\mathcal{L}_{\mathcal{D}}$, then $x[a_{id}]{=}x'[a_{id}]$. [1] [2]

AL3 There is a set of *quasi-identifiers* $\mathcal{Q}_{ID}{\subseteq}\mathcal{A}_i$, with $1{\leq}i{\leq}k$, common to all databases in $\mathcal{L}_{\mathcal{D}}$, corresponding to attributes whose values for some individuals can be known or learned by the adversary.

AL4 The adversary can aggregate information from the databases in the longitudinal collection $\mathcal{L}_{\mathcal{D}}{=}\{D_1, D_2, \ldots, D_k\}$ by applying *left outer join* operations successively, [3] all applied on the attribute of unique identification a_{id} as follows: At first, the adversary applies a *left outer join* operation on the databases D_1 and

[1] Assumption **AL2** may seem to provide too much power to the adversary, and for sure it cannot be applied to every longitudinal databases release. In our case, the released databases used for the experiments and described in Section 1.3 are real examples of scenarios in which the adversary would be able to rely on Assumption **AL2**. Moreover, we are worried about worst-case scenarios.

[2] Another possibility would be to not rely on an attribute of unique identification, but rather link records across databases in $\mathcal{L}_{\mathcal{D}}$ by using quasi-identifiers. This model was not developed here but is considered as a possible future work in Section 7.1.

[3] The *left outer join* operation is a pairwise database manipulation, e.g. over a *left database* D_l and a *right database* D_r, that returns a new database $D_l ⟕ D_r$ containing all the records from D_l concatenated with the corresponding records from D_r. This pairing is achieved between records from each database that hold the same value for a given set of attributes. If a record in D_l has no correspondent record in D_r, then $D_l ⟕ D_r$ contains only the record from D_l concatenated with null values where should be the values from D_r. In SQL language:

```
select *
from D_l left outer join D_r
where D_l.a_id = D_r.a_id
```

Example 20 shows how a *left outer join* operation can be performed.

D_2, resulting in a new database $D_1 \rightouterjoin D_2$ on the attributes set $\mathcal{A}_1 \uplus (\{\mathcal{A}_2 \backslash \{a_{id}\}\})$. [4] Then, the adversary applies another *left outer join* operation on the databases $D_1 \rightouterjoin D_2$ and D_3, resulting in a new database $(D_1 \rightouterjoin D_2) \rightouterjoin D_3$ on the attributes set $\mathcal{A}_1 \uplus (\{\mathcal{A}_2 \backslash \{a_{id}\}\}) \uplus (\{\mathcal{A}_3 \backslash \{a_{id}\}\})$. This step is then successively repeated until database D_k is reached. Summarizing,

$$\mathit{agreg}(\mathcal{L}_{\mathcal{D}}) \quad \overset{\text{def}}{=} \quad \begin{cases} D_1, & \text{if } k{=}1, \\ \mathit{agreg}(\{D_1, D_2, \ldots, D_{k-1}\}) \rightouterjoin D_k, & \text{if } k{\geq}2 \ , \end{cases}$$

where $\mathit{agreg}(\mathcal{L}_{\mathcal{D}})$ is a database on the attributes set $\mathcal{A}_{\mathcal{L}_{\mathcal{D}}}{=}\mathcal{A}_1 \uplus \left(\biguplus_{i=2}^{k} \mathcal{A}_i \backslash \{a_{id}\} \right)$. For clarity, given a record $x{\in}\mathit{agreg}(\mathcal{L}_{\mathcal{D}})$, we denote by $x[(\mathcal{Q}_{ID}, i)]$ the sub-record of x corresponding to the quasi-identifiers' values $(\mathcal{Q}_{ID}, i)$ of x in database $D_i{\in}\mathcal{L}_{\mathcal{D}}$ and by $x[\mathcal{Q}_{ID}]$ the sub-record corresponding to all the quasi-identifiers' values $\mathcal{Q}_{ID}$ in all databases $D_i{\in}\mathcal{L}_{\mathcal{D}}$, whose domain is $\mathit{dom}(\mathcal{Q}_{ID} \mid \mathcal{L}_{\mathcal{D}}){=}\mathit{dom}(\mathcal{Q}_{ID})^k$.

AL5 The adversary can compute marginal probability distributions on both records and attributes from the aggregated database $\mathit{agreg}(\mathcal{L}_{\mathcal{D}})$ by using functions analogous to those presented in Equations 3.1 and 3.2, respectively. Example 15 in Section 3.2 shows how this could be done by the adversary.

AL6 The adversary wants to learn information on individuals who hold records in the *focal database*, i.e. D_1, and can use the information from remaining, *auxiliary databases*, i.e. $D_2, D_3, \ldots, D_k$, only as auxiliary information. Hence follows:

AL6-A The *a priori* success is relative to the focal database D_1.

AL6-B The *a posteriori* success is relative to the aggregated database $\mathit{agreg}(\mathcal{L}_{\mathcal{D}})$, i.e. the information present in all the databases in the longitudinal collection, including the focal database and the auxiliary databases.

AL7 The adversary knows *a priori* the whole focal database D_1, i.e. they have access to every record and attribute value. Also, the adversary knows *a priori* that every individual of interest holds a record in all the databases in $\mathcal{L}_{\mathcal{D}}$, even though they do not have access to the remaining, auxiliary databases yet. [5] Hence, the

[4] The *disjunct union* of two sets A_i and A_j, with $i{\neq}j$, is a new set $A_i{\uplus}A_j$ in which every element of both A_i and A_j are present and properly labeled according to their original set, i.e.

$$A_i \uplus A_j \overset{\text{def}}{=} (A_i \times \{i\}) \cup (A_j \times \{j\}) \ .$$

For instance, if $A_1{=}\{a, b\}$ and $A_2{=}\{b, c\}$, then $A_1{\uplus}A_2{=}\{(a, 1), (b, 1), (b, 2), (c, 2)\}$.

[5] Assumption **AL7** may seem to provide too much power to the adversary, but again we are worried about worst-case scenarios.

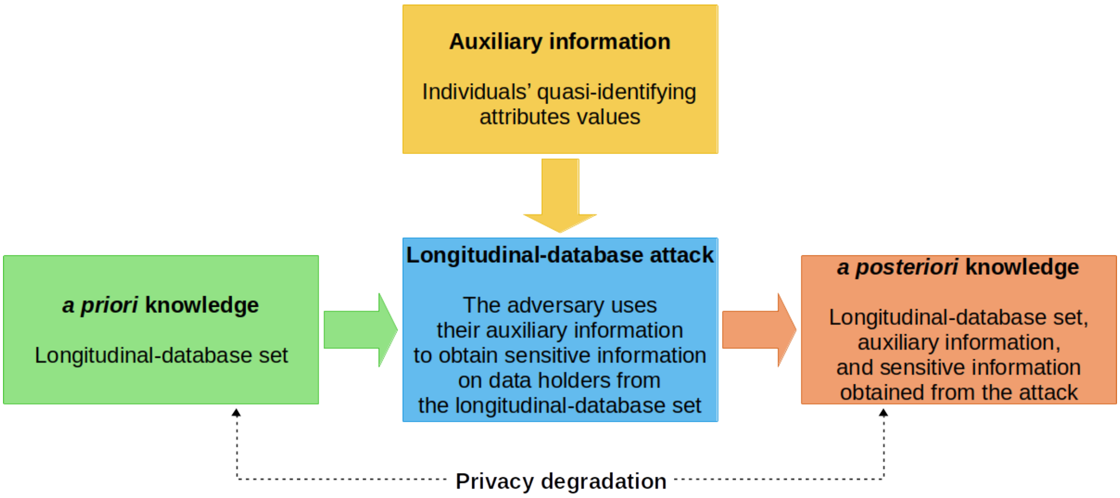

Auxiliary information
Individuals' quasi-identifying attributes values
a priori knowledge
Longitudinal-database set
Longitudinal-database attack
The adversary uses their auxiliary information to obtain sensitive information on data holders from the longitudinal-database set
a posteriori knowledge
Longitudinal-database set, auxiliary information, and sensitive information obtained from the attack
Privacy degradation

Figure 5.1: Schema of longitudinal database attack. The degradation of privacy is a comparative measure between the adversary's *a priori* and *a posteriori* successes and accounts for how much the auxiliary information used as quasi-identifying attributes values can help the adversary on the tasks of re-identification or sensitive attribute-inference.

attack consists in the use of auxiliary information obtained by the adversary from sources other than the longitudinal collection $\mathcal{L}_{\mathcal{D}}$ toghether with being able to perform the left outer join operation on the databases in $\mathcal{L}_{\mathcal{D}}$. The auxiliary information is in turn used by the adversary as quasi-identifier values in order to infer some sensitive information about the targeted-individuals represented in the focal database D_1. Therefore, the degradation of privacy is a comparative measure between the *a priori* and *a posteriori* successes, as schematized in Figure 5.1.

We now present the leading example for longitudinal database attacks.

Example 20 (Leading example for longitudinal database attacks)**.** Consider the longitudinal collection of databases $\mathcal{L}_{\mathcal{D}} = \{D_1, D_2\}$. The focal database, D_1, is defined on the set of attributes $\mathcal{A}_1$=$\{$*id*, *age*, *gender*, *occupation*, *illness*$\}$ and is represented in Table 5.2a, while the auxiliary database, D_2, is defined on the set of attributes $\mathcal{A}_2$=$\{$*id*, *age*, *occupation*$\}$ and is represented in Table 5.2b. Each attribute is specified as follows:

- $dom(id) = \{1, 2, 3, \ldots\}$, consisting of the unique identification number of an individual, constant in all the databases in $\mathcal{L}_{\mathcal{D}}$, given by a positive integer. Hence, given a value for *id* present in more than one database in $\mathcal{L}_{\mathcal{D}}$, the corresponding record is held by the same individual;

id	*age*	*gender*	*occupation*	*illness*
1	25	F	1	no
2	25	F	1	yes
3	25	F	3	yes
4	25	M	2	yes
5	25	M	2	no
6	49	F	3	yes
7	49	F	3	yes
8	49	F	5	no
9	49	M	4	no
10	60	M	4	no

(a) Focal database D_1.

id	*age*	*occupation*
1	26	2
2	26	1
3	26	3
4	26	2
5	26	2
6	50	4
7	50	3
8	50	5
9	50	4
11	19	1

(b) Auxiliary database D_2.

(*id*,1)	(*age*,1)	(*gender*,1)	(*occupation*,1)	(*illness*,1)	(*age*,2)	(*occupation*,2)
1	25	F	1	no	26	2
2	25	F	1	yes	26	1
3	25	F	3	yes	26	3
4	25	M	2	yes	26	2
5	25	M	2	no	26	2
6	49	F	3	yes	50	4
7	49	F	3	yes	50	3
8	49	F	5	no	50	5
9	49	M	4	no	50	4
10	60	M	4	no	—	—

(c) Aggregated database $agreg(\mathcal{L}_{\mathcal{D}}){=}D_1 ⟕ D_2$.

Table 5.2: Longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}{=}\{D_1, D_2\}$ and its aggregation $agreg(\mathcal{L}_{\mathcal{D}})$ from the leading example for longitudinal database attacks, Example 20.

- $dom(age) = \{0, 1, 2, \ldots, 120\}$, consisting of an individual's age, in whole years, given by a positive integer including zero;

- $dom(gender) = \{\texttt{F}, \texttt{M}\}$, denoting an individual's gender, considered here to be either female (F) or male (M) as a simplification for the current example;

- $dom(occupation) = \{\texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}\}$, denoting an individual's occupation category;

- $dom(illness) = \{\texttt{yes}, \texttt{no}\}$, denoting whether an individual is bearer of a given medical condition (yes) or not (no).

According to Assumption **AL4**, the adversary can aggregate the databases in $\mathcal{L}_{\mathcal{D}}$ by performing successive left outer join operations, actually two in this example. The result is a new database $agreg(\mathcal{L}_{\mathcal{D}}){=}D_1 ⟕ D_2$ on the set of attributes $\mathcal{A}_{agreg(\mathcal{L}_{\mathcal{D}})}$=$\{(id, 1), (age, 1), (gender, 1), (occupation, 1), (illness, 1), (age, 2), (occupation, 2)\}$ and represented in Table 5.2c. [6]

As was the case in Example 12 for single database attacks, each record represents an individual from a population of interest and consists of a tuple on the domain spanned by its corresponding set of attributes, i.e. $records(\mathcal{A}_1)$, $records(\mathcal{A}_2)$, or $records(\mathcal{A}_{agreg(\mathcal{L}_{\mathcal{D}})})$, and each domain consists of every combination of the possible values for each attribute.

Finally, according to Assumption **AL5**, the adversary can also compute the marginal probability distributions on both records and attributes from the aggregated database $agreg(\mathcal{L}_{\mathcal{D}})$ by using functions analogous to those presented in Equations 3.1 and 3.2, respectively. Example 15 in Section 3.2 shows how this could be done by the adversary.

◁

Based on the assumptions above, we present in the following section both collective-target re-identification and attribute-inference longitudinal databases attack models.

5.1.2 Attack models

In this section, we formally define both re-identification and attribute-inference attacks on longitudinal databases for collective-targets. Definition 21 accounts for an adversary interested in re-identifying as many individuals as possible, while Definition 22 accounts for an adversary interested in inferring the value of an attribute for as many individuals as possible. Detailed numeric examples of the execution of both attacks are presented in Sections D.2.1 and D.2.2, respectively, and illustrative individual-target attack models are presented in Section C.2.1.

Definition 21 (Collective-target Re-identification Longitudinal database (CRL) attack)**.** In a CRL attack, the adversary can completely access a longitudinal collection of databases, i.e. they have knowledge of every record in each of the databases. The adversary can also gather auxiliary information on the values of a set of quasi-identifiers for every individual who holds a record in the focal database. Particularly, the adversary can use the auxiliary databases from the longitudinal collection as auxiliary information. The adversary's goal is to re-identify as many individuals as possible in the focal database, no matter who they are, i.e. to precisely determine which record in the focal database corresponds to each individual of interest.

[6] Since the record with id attribute equal to 10 is only present in the focal database D_1, the corresponding values for the attributes $(age, 2)$ and $(occupation, 2)$ in $agreg(\mathcal{L}_{\mathcal{D}})$ are null. Similarly, the record with id attribute equal to 11 is only present in the auxiliary database D_2, hence this record is not present in $agreg(\mathcal{L}_{\mathcal{D}})$.

The specification of a CRL attack is analogous to that of a CRS attack, Definition 16. Particularly, the two main differences concern the use of the databases from the longitudinal collection in the definitions of success. For the *a priori* success, the adversary relies only on the information available in the focal database D_1, while for the *a posteriori* success, they rely on the auxiliary information obtained by them from sources other than the longitudinal collection $\mathcal{L}_{\mathcal{D}}$ in addition to the aggregated information from all the databases in $\mathcal{L}_{\mathcal{D}}$.

We now formally define the adversary's knowledge and their inflicted privacy degradation, both deterministic and probabilistic, after performing their attack.

- **Adversary's knowledge.** In a CRL attack, we assume the adversary:

  (i) can completely access a longitudinal collection of (non-empty) databases $\mathcal{L}_{\mathcal{D}}=\{D_1, D_2, \ldots, D_k\}$, according to Assumptions **AL0**–**AL3**, and can aggregate the information from this collection into a new aggregated database $agreg(\mathcal{L}_{\mathcal{D}})$, according to Assumption **AL4**;

  (ii) has every individual x in the focal database D_1 as individuals of interest for re-identification;

  (iii) given a set of quasi-identifiers $\mathcal{Q}_{ID} \subseteq \mathcal{A}_i$, with $1 \leq i \leq k$, that are present in all the databases in $\mathcal{L}_{\mathcal{D}}$, the adversary can gather auxiliary information on the sub-records $\{x[\mathcal{Q}_{ID}]\}_{x \in agreg(\mathcal{L}_{\mathcal{D}})}$ corresponding to the values of the quasi-identifiers for every individual in the new aggregated database $agreg(\mathcal{L}_{\mathcal{D}})$.

- **Deterministic degradation of privacy.** The deterministic degradation of privacy in a CRL attack accounts for the adversary's goal of re-identifying as many individuals as possible in the focal database D_1. Hence, the adversary's success is defined as the ratio of individuals in the focal database that are re-identified with absolute certainty.

  – ***A priori* deterministic success.** Before performing the CRL attack, the adversary can rely only on the focal database D_1, i.e. neither the auxiliary databases nor the quasi-identifiers are used yet. Hence, the adversary can only precisely determine which record in the focal database corresponds to an individual of interest iff the focal database contains only a single record. Therefore, the *a priori* deterministic success in a CRL attack on a longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}$ is defined as:

$$\textit{prior-suc}_{det}^{CRL}(\mathcal{L}_{\mathcal{D}}) \quad \overset{\text{def}}{=} \quad \begin{cases} 1, & \text{if } |D_1|{=}1, \\ 0, & \text{if } |D_1|{\geq}2. \end{cases} \tag{5.1}$$

- ***A posteriori* deterministic success.** By performing a CRL attack, the adversary uses the auxiliary information on the sub-records $\{x[\mathcal{Q}_{ID}]\}_{x\in agreg(\mathcal{L}_{\mathcal{D}})}$ corresponding to the values of the quasi-identifiers for every individual in the aggregated database $agreg(\mathcal{L}_{\mathcal{D}})$ to support the task of re-identification, including the information provided by the auxiliary databases. A CRL attack is then performed as follows: For each individual in the focal database, the adversary uses the corresponding quasi-identifier values to filter the aggregated database $agreg(\mathcal{L}_{\mathcal{D}})$ records, disregarding those that do not match the query. Then, given the remaining database, the adversary can only precisely determine which record corresponds to the individual of interest iff the remaining database contains only a single record. The final deterministic success is determined over the set of all individuals from the focal database and defined as the ratio between the number of re-identified individuals and the total number of individuals in the focal database.

  Therefore, the *a posteriori* deterministic success in a CRL attack on a longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}$, given the adversary's knowledge of the values $\{x[\mathcal{Q}_{ID}]\}_{x\in agreg(\mathcal{L}_{\mathcal{D}})}$ for every individual's quasi-identifiers, is defined as:

$$\textit{post-suc}_{det}^{CRL}(\mathcal{L}_{\mathcal{D}}, \{x[\mathcal{Q}_{ID}]\}_{x\in agreg(\mathcal{L}_{\mathcal{D}})}) \quad \overset{\text{def}}{=} \quad \frac{\left|\{\alpha{\in} agreg(\mathcal{L}_{\mathcal{D}})/{\sim_{\mathcal{Q}_{ID}}} \mid |\alpha|{=}1\}\right|}{|agreg(\mathcal{L}_{\mathcal{D}})|}\,, \tag{5.2}$$

  where $agreg(\mathcal{L}_{\mathcal{D}})/{\sim_{\mathcal{Q}_{ID}}}$ is the partition on set $\mathcal{Q}_{ID}$ of all individuals that share the same quasi-identifiers in the aggregated database $agreg(\mathcal{L}_{\mathcal{D}})$ and α is a variable corresponding to each block within the partition.

- **Deterministic degradation of privacy.** The deterministic degradation of privacy in a CRL attack is defined as the difference between the adversary's *a posteriori* and *a priori* successes, i.e. by how much the attack increases the adversary's success:

$$\begin{aligned} & \mathit{priv\text{-}degrad}_{det}^{CRL}(\mathcal{L}_{\mathcal{D}}, \{x[\mathcal{Q}_{ID}]\}_{x \in \mathit{agreg}(\mathcal{L}_{\mathcal{D}})}) \overset{\text{def}}{=} \\ & \mathit{post\text{-}suc}_{det}^{CRL}(\mathcal{L}_{\mathcal{D}}, \{x[\mathcal{Q}_{ID}]\}_{x \in \mathit{agreg}(\mathcal{L}_{\mathcal{D}})}) - \mathit{prior\text{-}suc}_{det}^{CRL}(\mathcal{L}_{\mathcal{D}}) \ . \end{aligned} \tag{5.3}$$

- **Probabilistic degradation of privacy.** The adversary's probabilistic success in a CRL attack does not rely on precisely determining which record in the focal database D_1 corresponds to each individual of interest. Rather, we compute the probability of correctly re-identifying a randomly selected target in the focal database, i.e. the greater this probability, the more successful the adversary is and greater the expected risk of re-identification for each individual in the focal database.

  - ***A priori* probabilistic success.** Before performing the CRL attack, the adversary can rely only on the focal database D_1, i.e. neither the auxiliary databases nor the quasi-identifiers are used yet. Hence, given a random target chosen by the adversary, their best course of action is to randomly select a record from the focal database, according to the maximum entropy principle. Since each individual only holds one record in the focal database, the probability of the adversary being successful is the inverse of the focal database size. Therefore, the *a priori* probabilistic success in a CRL attack on a longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}$ is defined as: [7]

$$\mathit{prior\text{-}suc}_{prob}^{CRL}(\mathcal{L}_{\mathcal{D}}) \overset{\text{def}}{=} \frac{1}{|D_1|} \ . \tag{5.4}$$

  - ***A posteriori* probabilistic success.** By performing a CRL attack, the adversary uses the auxiliary information on the sub-records $\{x[\mathcal{Q}_{ID}]\}_{x \in \mathit{agreg}(\mathcal{L}_{\mathcal{D}})}$ corresponding to the values of quasi-identifiers for every individual in the aggregated database $\mathit{agreg}(\mathcal{L}_{\mathcal{D}})$ to support the task of re-identification, including the information provided by the auxiliary databases. Given a randomly selected target in the focal database, the adversary uses the corresponding quasi-identifier values to filter the aggregated database $\mathit{agreg}(\mathcal{L}_{\mathcal{D}})$ records, disregarding those that do not match the query. From there, their best course of action is to randomly select a record from the filtered database, according to the maximum entropy principle. Since each individual only

[7] This definition corresponds to the *a priori* Bayes vulnerability $V_1[\pi_x^{D_1}]$, given by Definitions 6 and 8 in Section 2.3, of the focal database's records on the *a priori* probability distribution $\pi_x^{D_1}$.

holds one record in each database, the probability of the adversary being successful is the inverse of the filtered database size.

Therefore, the *a posteriori* probabilistic success in a CRL attack on a longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}$, given the adversary's knowledge of the values $\{x[\mathcal{Q}_{ID}]\}_{x\in agreg(\mathcal{L}_{\mathcal{D}})}$ for every individual's quasi-identifiers, is defined as the expected value of the probability of success taken from the probability distribution for selecting each individual from the focal database as the target: [8]

$$post\text{-}suc^{CRL}_{prob}(\mathcal{L}_{\mathcal{D}}, \{x[\mathcal{Q}_{ID}]\}_{x\in agreg(\mathcal{L}_{\mathcal{D}})}) \overset{\text{def}}{=} \sum_{x\in agreg(\mathcal{L}_{\mathcal{D}})} \pi^{D_1}_x \cdot \frac{1}{|\{x'{\in} agreg(\mathcal{L}_{\mathcal{D}}) \mid x'[\mathcal{Q}_{ID}]{=}x[\mathcal{Q}_{ID}]\}|} = \frac{\left|agreg(\mathcal{L}_{\mathcal{D}})/{\sim_{\mathcal{Q}_{ID}}}\right|}{|agreg(\mathcal{L}_{\mathcal{D}})|}\,, \tag{5.5}$$

i.e. the ratio between the number of blocks in the partition on set $\mathcal{Q}_{ID}$ of the aggregated database $agreg(\mathcal{L}_{\mathcal{D}})$ and the aggregated database size.

– **Probabilistic degradation of privacy.** The probabilistic degradation of privacy in a CRL attack is defined as the ratio by which the attack increases the probabilistic success of the adversary:

$$priv\text{-}degrad^{CRL}_{prob}(\mathcal{L}_{\mathcal{D}}, \{x[\mathcal{Q}_{ID}]\}_{x\in agreg(\mathcal{L}_{\mathcal{D}})}) \overset{\text{def}}{=} \frac{post\text{-}suc^{CRL}_{prob}(\mathcal{L}_{\mathcal{D}}, \{x[\mathcal{Q}_{ID}]\}_{x\in agreg(\mathcal{L}_{\mathcal{D}})})}{prior\text{-}suc^{CRL}_{prob}(\mathcal{L}_{\mathcal{D}})} = |D_1|\frac{\left|agreg(\mathcal{L}_{\mathcal{D}})/{\sim_{\mathcal{Q}_{ID}}}\right|}{|agreg(\mathcal{L}_{\mathcal{D}})|}\,, \tag{5.6}$$

i.e. the focal database size multiplied by the ratio between the number of blocks in the partition on set $\mathcal{Q}_{ID}$ of the aggregated database $agreg(\mathcal{L}_{\mathcal{D}})$ and the aggregated database size.

A detailed numeric example of a CRL attack is presented in Section D.2.1, Example 52.

[8] This definition corresponds to the *a posteriori* Bayes vulnerability $V_1[\pi^{D_1}_x \triangleright C^{\mathcal{Q}_{ID}}]$, given by Definitions 6 and 8 in Section 2.3, of the focal database's records given the knowledge of the quasi-identifiers values, where:

- $*$ $\pi^{D_1}_x$ is the *a priori* probability distribution on the focal database's records;
- $*$ $C^{\mathcal{Q}_{ID}} : D_1 \to \mathbb{D}(dom(\mathcal{Q}_{ID})^k)$, where k accounts for the number of databases in the longitudinal collection, is the channel that deterministically maps each record in the focal database to the respective values of quasi-identifiers and is defined, for all $x{\in}D_1$ and $q{\in}dom(\mathcal{Q}_{ID})^k$, as:

$$C^{\mathcal{Q}_{ID}}_{x,q} = \begin{cases} 1, & \text{if } x[\mathcal{Q}_{ID}] = q, \\ 0, & \text{otherwise.} \end{cases}$$

◁

We now present the attribute-inference attack on longitudinal databases for collective-targets.

Definition 22 (Collective-target Attribute-inference Longitudinal database (CAL) attack)**.** In a CAL attack, the adversary can completely access a longitudinal collection of databases, i.e. they have knowledge of every record in each of the databases. The adversary can also gather auxiliary information on the values of a set of quasi-identifiers for every individual who holds a record in the focal database. Particularly, the adversary can use the auxiliary databases from the longitudinal collection as auxiliary information. The adversary's goal is to infer the values of an attribute considered to be sensitive for as many individuals as possible in the focal database, no matter who they are.

The specification of a CAL attack is analogous to that of a CAS attack, Definition 17. Particularly, the two main differences concern the use of the databases from the longitudinal collection in the definitions of success. For the *a priori* success, the adversary relies only on the information available in the focal database D_1, while for the *a posteriori* success, they rely on the auxiliary information obtained by them from sources other than the longitudinal collection $\mathcal{L}_{\mathcal{D}}$ in addition to the aggregated information from all the databases in $\mathcal{L}_{\mathcal{D}}$.

We now formally define the adversary's knowledge and their inflicted privacy degradation, both deterministic and probabilistic, after performing their attack.

- **Adversary's knowledge.** In an CAL attack, we assume the adversary:
  - (i) can completely access a longitudinal collection of (non-empty) databases $\mathcal{L}_{\mathcal{D}}=\{D_1, D_2, \ldots, D_k\}$, according to Assumptions **AL0**–**AL3**, and can aggregate the information from this collection into a new aggregated database $agreg(\mathcal{L}_{\mathcal{D}})$, according to Assumption **AL4**;
  - (ii) has every individual x in the focal database D_1 as individuals of interest for attribute-inference;
  - (iii) given a set of quasi-identifiers $\mathcal{Q}_{ID}\subseteq\mathcal{A}_i$, with $1\leq i\leq k$, that are present in all the databases in $\mathcal{L}_{\mathcal{D}}$, the adversary can gather auxiliary information on the sub-records $\{x[\mathcal{Q}_{ID}]\}_{x\in agreg(\mathcal{L}_{\mathcal{D}})}$ corresponding to the values of the quasi-identifiers for every individual in the new aggregated database $agreg(\mathcal{L}_{\mathcal{D}})$;

(iv) given an attribute $a_{sens}{\in}\mathcal{A}_1$ in the focal base D_1 such that $a_{sens}{\notin}\mathcal{Q}_{ID}$, considered to be sensitive, the adversary's goal is to infer the attribute's value $x[a_{sens}]$ for as many individuals as possible.

- **Deterministic degradation of privacy.** The deterministic degradation of privacy in a CAL attack accounts for the adversary's goal of precisely determining the value of the sensitive attribute for as many individuals as possible in the focal database D_1. Hence, the adversary's success is defined as the ratio of individuals in the focal database whose value for the sensitive attribute can be inferred with absolute certainty.

  – ***A priori* deterministic success.** Before performing the CAL attack, the adversary can rely only on the focal database D_1, i.e. neither the auxiliary databases nor the quasi-identifiers are used yet. Hence, the adversary can only precisely determine the value of the sensitive attribute for any individual of interest iff all the records in the focal database have the same value for the sensitive attribute. Therefore, the *a priori* deterministic success in a CAL attack on a longitudinal collection of databases $\mathcal{L}_\mathcal{D}$ and with respect to the sensitive attribute a_{sens} is defined as:

$$\textit{prior-suc}^{CAL}_{det}(\mathcal{L}_\mathcal{D}, a_{sens}) \overset{\text{def}}{=} \begin{cases} 1, & \text{if for all } x, x'{\in}D_1, \\ & x[a_{sens}]{=}x'[a_{sens}], \\ 0, & \text{otherwise.} \end{cases} \tag{5.7}$$

  – ***A posteriori* deterministic success.** By performing a CAL attack, the adversary uses the auxiliary information on the sub-records $\{x[\mathcal{Q}_{ID}]\}_{x\in agreg(\mathcal{L}_\mathcal{D})}$ corresponding to the values of quasi-identifiers for every individual in the aggregated database *agreg*($\mathcal{L}_\mathcal{D}$) to support the task of inferring the value of the sensitive attribute, including the information provided by the auxiliary databases. A CAL attack is then performed as follows: For each individual in the focal database, the adversary uses the corresponding quasi-identifier values to filter the aggregated database *agreg*($\mathcal{L}_\mathcal{D}$) records, disregarding those that do not match the query. Then, given the remaining database, the adversary can only precisely determine the value of the sensitive attribute for the individual of interest iff all the records in the remaining database have the same value for the sensitive attribute. The final deterministic success is determined over the set of all individuals from the focal database and defined as the ratio between the number of individuals whose value for the sensitive attribute could be precisely determined and the total number of individuals in the focal database.

Therefore, the *a posteriori* deterministic success in a CAL attack on a longitudinal collection of databases $\mathcal{L}_\mathcal{D}$ and with respect to the sensitive attribute a_{sens}, given the adversary's knowledge of the values $\{x[\mathcal{Q}_{ID}]\}_{x \in agreg(\mathcal{L}_\mathcal{D})}$ for every individual's quasi-identifiers, is defined as:

$$\begin{aligned} &\textit{post-suc}^{CAL}_{det}(\mathcal{L}_\mathcal{D}, a_{sens}, \{x[\mathcal{Q}_{ID}]\}_{x \in agreg(\mathcal{L}_\mathcal{D})}) \overset{\text{def}}{=} \\ &\frac{\sum_{\alpha \in agreg(\mathcal{L}_\mathcal{D})/\sim_{\mathcal{Q}_{ID}}} |\alpha| \cdot \textit{unique-sens}(\alpha, a_{sens})}{|agreg(\mathcal{L}_\mathcal{D})|} \ , \end{aligned} \tag{5.8}$$

where $agreg(\mathcal{L}_\mathcal{D})/_{\sim_{\mathcal{Q}_{ID}}}$ is the partition on set $\mathcal{Q}_{ID}$ of all individuals that share the same quasi-identifiers in the aggregated database $agreg(\mathcal{L}_\mathcal{D})$ and α is a variable corresponding to each block within the partition. Also, the function $\textit{unique-sens}(\alpha, a_{sens})$ is defined as:

$$\textit{unique-sens}(\alpha, a_{sens}) \overset{\text{def}}{=} \begin{cases} 1, & \text{if for all } x, x' \in \alpha, \\ & x[(a_{sens}, 1)] = x'[(a_{sens}, 1)], \\ 0, & \text{otherwise.} \end{cases} \tag{5.9}$$

- **Deterministic degradation of privacy.** The deterministic degradation of privacy in a CAL attack is defined as the difference between the adversary's *a posteriori* and *a priori* successes, i.e. by how much the attack increases the adversary's success:

$$\begin{aligned} &\textit{priv-degrad}^{CAL}_{det}(\mathcal{L}_\mathcal{D}, a_{sens}, \{x[\mathcal{Q}_{ID}]\}_{x \in agreg(\mathcal{L}_\mathcal{D})}) \overset{\text{def}}{=} \\ &\textit{post-suc}^{CAL}_{det}(\mathcal{L}_\mathcal{D}, a_{sens}, \{x[\mathcal{Q}_{ID}]\}_{x \in agreg(\mathcal{L}_\mathcal{D})}) - \textit{prior-suc}^{CAL}_{det}(\mathcal{L}_\mathcal{D}, a_{sens}) \ . \end{aligned} \tag{5.10}$$

- **Probabilistic degradation of privacy.** The adversary's probabilistic success in a CAL attack does not rely on precisely determining the value of the sensitive attribute for each individual of interest in the focal database D_1. Rather, we compute the probability of correctly determining the value of the sensitive attribute for a randomly selected target in the focal database, i.e. the greater this probability, the more successful the adversary is and greater the expected risk of sensitive attribute-inference for each individual in the focal database.

  - ***A priori* probabilistic success.** Before performing the CAL attack, the adversary can rely only on the focal database D_1, i.e. neither the auxiliary databases nor the quasi-identifiers are used yet. Hence, given a random target chosen by the adversary, their best course of action is to randomly select a possible value for the sensitive attribute, according to the maximum entropy principle. Since each individual only holds one record in the focal database, the most probable value would be the most frequent one in the focal database. Therefore, the *a priori* probabilistic success in a CAL attack on a longitudinal collection of databases $\mathcal{L}_\mathcal{D}$ and with respect to the sensitive attribute a_{sens}, given the probability distribution of values for the sensitive attribute $\pi_s^{a_{sens}|D_1}$, according to Equation 3.2, is defined as: [9]

$$\begin{aligned} \textit{prior-suc}_{prob}^{CAL}(\mathcal{L}_\mathcal{D}, a_{sens}) &\overset{\text{def}}{=} \max_{s \in dom(a_{sens})} \pi_s^{a_{sens}|D_1} \\ &= \max_{s \in dom(a_{sens})} \frac{|\{x{\in}D_1 \mid x[a_{sens}]{=}s\}|}{|D_1|} \ . \end{aligned} \tag{5.11}$$

– ***A posteriori* probabilistic success.** By performing a CAL attack, the adversary uses the auxiliary information on the sub-records $\{x[\mathcal{Q}_{ID}]\}_{x \in agreg(\mathcal{L}_\mathcal{D})}$ corresponding to the values of quasi-identifiers for every individual in the aggregated database $agreg(\mathcal{L}_\mathcal{D})$ to support the task of inferring the sensitive attribute value, including the information provided by the auxiliary databases. Given a randomly selected target in the focal database, the adversary uses the corresponding quasi-identifier values to filter the aggregated database $agreg(\mathcal{L}_\mathcal{D})$ records, disregarding those that do not match the query. From there, their best course of action is to randomly select a possible value for the sensitive attribute among those available in the remaining database, according to the maximum entropy principle. Since each individual only holds one record in each database, the most probable value would be the most frequent one in the remaining database.

Therefore, the *a posteriori* probabilistic success in a CAL attack on a longitudinal collection of databases $\mathcal{L}_\mathcal{D}$ and with respect to the sensitive attribute a_{sens}, given the adversary's knowledge of the values $\{x[\mathcal{Q}_{ID}]\}_{x \in agreg(\mathcal{L}_\mathcal{D})}$ for every individual's quasi-identifiers, is defined as the expected value of the probability of success taken from the probability distribution for selecting

[9] This definition corresponds to the *a priori* Bayes vulnerability $V_1[\pi_s^{a_{sens}|D_1}]$, given by Definitions 6 and 8 in Section 2.3, of the sensitive attribute in the focal database on the *a priori* probability distribution $\pi_s^{a_{sens}|D_1}$.

each individual as the target: [10]

$$\textit{post-suc}^{CAL}_{prob}(\mathcal{L}_{\mathcal{D}}, a_{sens}, \{x[\mathcal{Q}_{ID}]\}_{x\in agreg(\mathcal{L}_{\mathcal{D}})}) \stackrel{\text{def}}{=} \frac{\sum_{q\in dom(\mathcal{Q}_{ID}|\mathcal{L}_{\mathcal{D}})} \max_{s\in dom(a_{sens})} |\{x\in agreg(\mathcal{L}_{\mathcal{D}}) \mid x[(a_{sens},1)]{=}s, x[\mathcal{Q}_{ID}]{=}q\}|}{|agreg(\mathcal{L}_{\mathcal{D}})|}, \tag{5.12}$$

i.e. for each quasi-identifier value q and each sensitive attribute value s, take the size of the largest available block in the partition on set $\mathcal{Q}_{ID}$ of the aggregated database $agreg(\mathcal{L}_{\mathcal{D}})$, add them together, and divide by the size of the aggregated database.

- **Probabilistic degradation of privacy.** The probabilistic degradation of privacy in a CAL attack is defined as the ratio by which the attack increases the probabilistic success of the adversary:

$$\textit{priv-degrad}^{CAL}_{prob}(\mathcal{L}_{\mathcal{D}}, a_{sens}, \{x[\mathcal{Q}_{ID}]\}_{x\in agreg(\mathcal{L}_{\mathcal{D}})}) \stackrel{\text{def}}{=} \frac{\textit{post-suc}^{CAL}_{prob}(\mathcal{L}_{\mathcal{D}}, a_{sens}, \{x[\mathcal{Q}_{ID}]\}_{x\in agreg(\mathcal{L}_{\mathcal{D}})})}{\textit{prior-suc}^{CAL}_{prob}(\mathcal{L}_{\mathcal{D}}, a_{sens})}. \tag{5.13}$$

A detailed numeric example of a CAL attack is presented in Section D.2.2, Example 54.

$\triangleleft$

[10] This definition corresponds to the *a posteriori* Bayes vulnerability $V_1[\pi_s^{a_{sens}|D_1} \triangleright C^{a_{sens},\mathcal{Q}_{ID}}]$, given by Definitions 6 and 8 in Section 2.3, of the sensitive attribute in the focal database given the knowledge over the quasi-identifiers, where:

- $\pi_s^{a_{sens}|D_1}$ is the *a priori* probability distribution on the sensitive attribute in the focal database, according to Equation 3.2, and is defined as:

$$\pi_s^{a_{sens}|D_1} = \frac{|\{x\in D_1 \mid x[a_{sens}]{=}s\}|}{|D_1|};$$

- $C^{a_{sens},\mathcal{Q}_{ID}}{:}dom(a_{sens}){\to}\mathbb{D}(dom(\mathcal{Q}_{ID})^k)$, where k accounts for the number of databases in the longitudinal collection, is the channel that deterministically maps the possible values of the sensitive attribute to values of quasi-identifiers associated with it and is defined, for all $s{\in}dom(a_{sens})$ and $q{\in}dom(\mathcal{Q}_{ID})^k$, as:

$$C^{a_{sens},\mathcal{Q}_{ID}}_{s,q} = \frac{|\{x\in\mathcal{L}_{\mathcal{D}} \mid x[a_{sens}]{=}s, x[\mathcal{Q}_{ID}] = q\}|}{|\{x\in\mathcal{L}_{\mathcal{D}} \mid x[a_{sens}]{=}s\}|}.$$

5.2 Experimental results for collective-target attacks on the School Censuses

In this section, we present our quantitative analyses on the privacy risks for individuals whose information are made public on longitudinal databases. The attacks performed here were modeled according to the theory developed in Section 5.1. For illustrative individual-target experimental results, in which we target either famous people or our acquaintances selected *a priori*, see Appendix C Section C.2.2.

We begin by detailing our experimental setup, followed by the results for re-identification and then for attribute-inference attacks on INEP's databases.

5.2.1 Experimental setup

As stated in Section 1.3, we have chosen to work with the databases containing information on students and released as microdata, i.e. data at the record level. In this chapter, we consider databases from the School Censuses, leaving the analyses for the Higher Education Census to Appendix F Section F.1.

A preliminary analysis of the databases containing information on students showed that a student may hold more than one record in a given database, e.g. if a High School student is also enrolled in a Professional Education course. Hence, in order to guarantee that each student holds only one record in each database, according to Assumption **AL1** from Section 5.1.1, we have randomly selected only one record for each data holder of multiple records. This data treatment was based on the unique identification number, the `ID_ALUNO` code, given to each student in the pseudonymization treatment performed by INEP. The `ID_ALUNO` code, which is unique to each student at least in a given database release, easily allowed us to find those students with multiple records and to perform the random selection of just one of them. [11]

Even though each database accounts for dozens of attributes, we have chosen just a few for our analyses given the computational costs, including time and memory usage. The selection criteria was as follows.

[11] Until the Census of 2017, each student would receive a unique `ID_ALUNO` code, which would stay constant for every annual Census. Such an identification number allows an adversary to easily follow a given student through different years, which increases the amount of information leaked. From the Census of 2018 onward, INEP changed how the `ID_ALUNO` code is assigned to each student, continuing to be unique for a student in a given database release, but changing among different releases. Therefore, from the Census of 2018 onward that code could not be used to follow a given student through different years anymore.

Variable	Meaning
FK_COD_MUNICIPIO_END	Student's city of residency code.
PK_COD_ENTIDADE	School code.
FK_COD_ETAPA_ENSINO	Student's educational stage code.

Table 5.3: Variables from the School Census of 2014 chosen as quasi-identifiers for Collective-target Re-identification Longitudinal databases (CRL) attacks in Experiment 23 and for Collective-target Attribute-inference Longitudinal databases (CAL) attacks in Experiment 24.

Variable	Values	Meaning
ID_POSSUI_NEC_ESPECIAL	0 (No) 1 (Yes)	Whether the student possesses a disability or global developmental disorder, or not.
ID_N_T_E_P	0 (No) 1 (Yes)	Whether the student uses public school transport, or not.

Table 5.4: Variables from the School Census of 2014 chosen as sensitive attributes for Collective-target Attribute-inference Longitudinal databases (CAL) attacks in Experiment 24.

- We have chosen three quasi-identifying attributes selected according to how easily an adversary could learn them and to their variability through the years, i.e. whether or not they could change and hence be captured in longitudinal databases attacks. The quasi-identifying attributes chosen for the experiments in this chapter are listed in Table F.19.

- We have chosen two sensitive attributes selected according to the seriousness of the possible individual privacy breach if revealed, e.g. whether or not the individual has special needs or disabilities. The sensitive attributes chosen for the experiments in this chapter are listed in Table 5.4.

Of course, the selection of those quasi-identifiers and sensitive attributes is arbitrary and can change in significance over time. Nevertheless, our goal is not to ultimately define whether an attribute should be considered as a quasi-identifier or as a sensitive attribute. Instead, the results here illustrate possible real-life circumstances and the respective privacy risks for the data holders in case of unintended information disclosure.

In the following Sections 5.2.2 and 5.2.3, we present some experimental results for re-identification and attribute-inference attacks, respectively. Both were performed on longitudinal databases for collective-targets and on the School Censuses released by INEP.

5.2.2 Results of re-identification attack experiments

In this section, we present our quantitative analyses on the privacy risk of re-identification for individuals whose information are made public on longitudinal databases, as modeled in Section 5.1.2. We present one CRL attack on the School Censuses from 2014 to 2017, Experiment 23.

Experiment 23 (Collective-target Re-identification Longitudinal databases (CRL) attack on the School Censuses)**.** In a CRL attack, the adversary's goal is to re-identify as many individuals as possible, according to Definition 21. Since we want to evaluate the overall re-identification risk, the adversary can gather auxiliary information on the values of a set of quasi-identifying attributes for every individual who holds a record in the focal database. The following experiment was conducted on the School Censuses from 2014 to 2017 and had the database of 2014 as the focal one.

The focal database was released containing 56 064 675 records, which were reduced to 49 491 319 after the random selection of only one record for each data holder of multiple records, as detailed in Section 5.2.1. For the current experiment, we have selected as quasi-identifiers the attributes `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO`, as specified in Table 5.3. We have measured both the deterministic and probabilistic degradation of privacy for the set composed of those three attributes.

The results are organized as follows.

- Table 5.5 summarizes the adversary's measures of success and privacy degradation as more databases are aggregated to the focal database.
- Figure 5.6 shows the adversary's deterministic and probabilistic measures of success according to Table 5.5.
- Figure 5.7 shows the histograms for the distribution of individuals according to the adversary's probabilistic measure of success as more databases are aggregated to the focal database.

According to those results, we were able to conclude the following.

- **Adversary's deterministic success**. Defined in Equations 5.1 and 5.2, this metric measures the fraction of individuals in the focal database that can be re-identified with absolute certainty, in a scale from 0% to 100%. According to Table 5.5, the adversary's *a priori* deterministic success equals 0%, i.e. no individual can be re-identified with absolute certainty in the focal database without the use of auxiliary information. However, by using only the attributes `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO` as quasi-identifiers and only the database of 2014, the adversary can re-identify with absolute certainty 1.44% of the individuals in the focal database, i.e. approximately 712 675 students.

  As the longitudinal collection of databases available to the adversary increases, allowing the use of more School Census databases as auxiliary information, so increases the fraction of individuals in the focal database that can be re-identified with absolute certainty. Particularly, the adversary's *a posteriori* deterministic success increases to 12.88% by using only the database of 2015 as auxiliary information, to 25.26% by using the databases of 2015 and 2016, and to 36.31% by using the databases from 2015 to 2017.

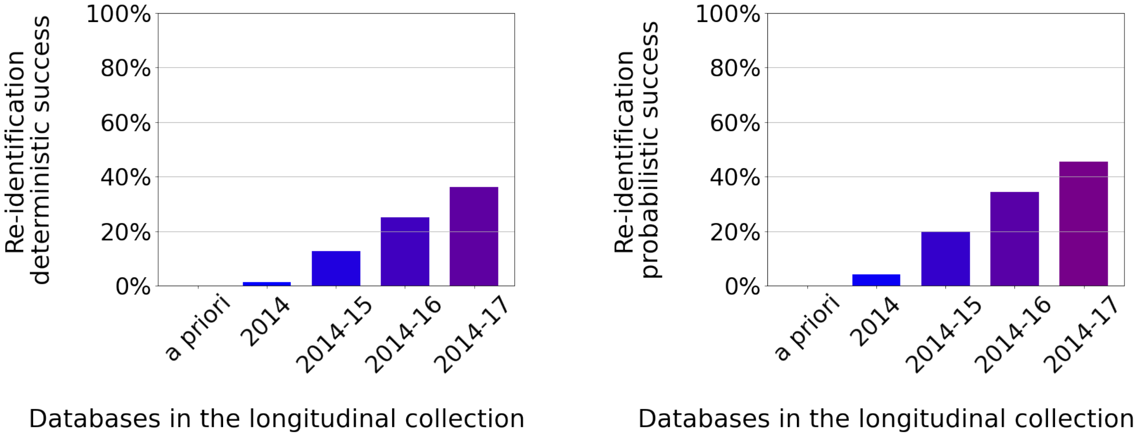
- **Adversary's probabilistic success**. Defined in Equations 5.4 and 5.5, this metric measures the probability of a randomly chosen individual in the focal database being re-identified, in a scale from 0% to 100%. According to Table 5.5, the adversary's *a priori* probabilistic success is approximately 0.000002%, i.e. the adversary's chance of re-identifying one of the 49 491 319 individuals in the focal database is almost zero. However, by using only the attributes `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO` as quasi-identifiers and only the database of 2014, the adversary increases their chance to 4.24%.

  As the longitudinal collection of databases available to the adversary increases, so increases the adversary's chance of re-identifying a randomly chosen individual in the focal database. Particularly, the adversary's *a posteriori* probabilistic success increases to 20.08% by using only the database of 2015 as auxiliary information, to 34.37% by using the databases of 2015 and 2016, and to 45.60% by using the databases from 2015 to 2017.

◁

	Adversary's deterministic success		**Adversary's probabilistic success**	
	***a priori* success** 0.00%		***a priori* success** 0.000002%	
Databases in the longitudinal aggregation	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
`2014`	1.44%	1.44%	4.24%	2 103 225
`2014 to 2015`	12.88%	12.88%	20.08%	9 937 236
`2014 to 2016`	25.26%	25.26%	34.37%	17 008 936
`2014 to 2017`	36.31%	36.31%	45.60%	22 566 084

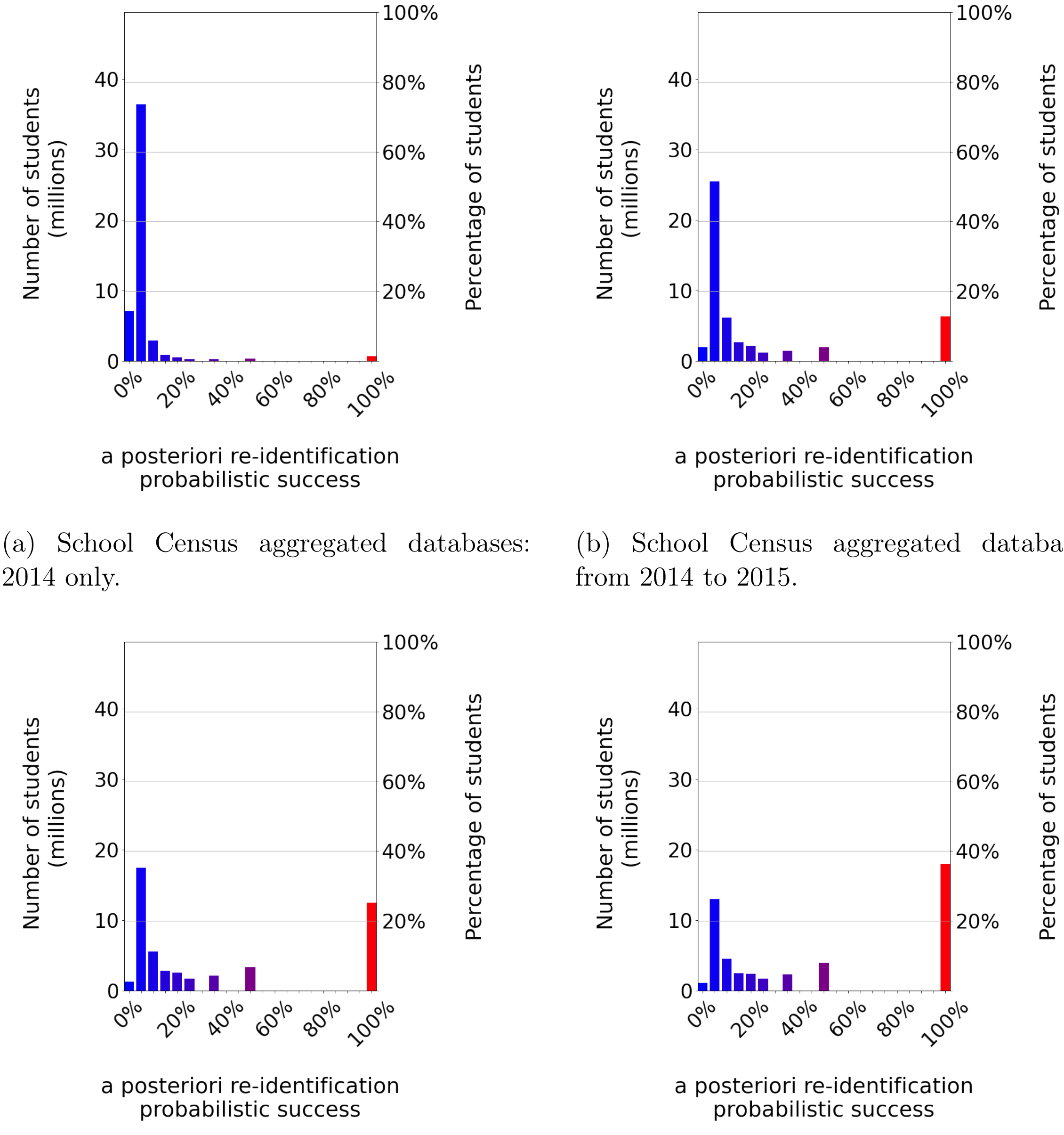
Table 5.5: Experiment 23: Privacy degradation in Collective-target Re-identification Longitudinal databases (CRL) attacks on the School Censuses from 2014 to 2017 using the quasi-identifiers `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO`. Here, the focal database is that for the School Census of 2014. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* success, while the probabilistic degradation of privacy is their ratio.



(a) Deterministic success.

(b) Probabilistic success.

Figure 5.6: Experiment 23: Adversary's success in Collective-target Re-identification Longitudinal databases (CRL) attacks on the School Censuses from 2014 to 2017 using the quasi-identifiers `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO`. Here, each bar represents a different longitudinal aggregation of databases, always having the School Census of 2014 as the focal one, except for the bar with label *a priori*, which indicates the adversary's *a priori* success relying only on the focal database. The height of each bar represents the adversary's deterministic or probabilistic success.



(a) School Census aggregated databases: 2014 only.

(b) School Census aggregated databases: from 2014 to 2015.

(c) School Census aggregated databases: from 2014 to 2016.

(d) School Census aggregated databases: from 2014 to 2017.

Figure 5.7: Experiment 23: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Re-identification Longitudinal databases (CRL) attacks on the School Censuses from 2014 to 2017 using the quasi-identifiers `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of re-identification.

5.2.3 Results of attribute-inference attack experiments

In this section, we present our quantitative analyses on the privacy risk of attribute-inference for individuals whose information are made public on longitudinal databases, as modeled in Section 5.1.2. We present one CAL attack on the School Censuses from 2014 to 2017, Experiment 24, on sensitive attributes `ID_POSSUI_NEC_ESPECIAL` and `ID_N_T_E_P`.

Experiment 24 (Collective-target Attribute-inference Longitudinal databases (CAL) attack on the School Censuses)**.** In a CAL attack, the adversary's goal is to infer the values of an attribute considered to be sensitive for as many individuals as possible, according to Definition 22. Since we want to evaluate the overall attribute-inference risk, the adversary can gather auxiliary information on the values of a set of quasi-identifying attributes for every individual who holds a record in the focal database. The following experiment was conducted on the School Censuses from 2014 to 2017 and had the database of 2014 as the focal one.

The focal database was released containing 56 064 675 records, which were reduced to 49 491 319 after the random selection of only one record for each data holder of multiple records, as detailed in Section 5.2.1. For the current experiment, we have selected as quasi-identifiers the attributes `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO`, as specified in Table 5.3. We have measured both the deterministic and probabilistic degradation of privacy for the set composed of those three attributes.

Furthermore, we were interested in inferring the value for the attributes `ID_POSSUI_NEC_ESPECIAL` and `ID_N_T_E_P`, both described in Table 5.4 and considered by us to be sensitive.

The results are organized as follows.

- Tables 5.8a and 5.8b summarize the adversary's measures of success and privacy degradation as more databases are aggregated to the focal database for the sensitive attributes `ID_POSSUI_NEC_ESPECIAL` and `ID_N_T_E_P`, respectively.

- Figure 5.9 shows both the adversary's deterministic and probabilistic measures of success according to Table 5.8, for both sensitive attributes.

- Figures 5.10 and 5.11 show the histograms for the distribution of individuals according to the adversary's probabilistic measure of success as more databases are aggregated to the focal database for the sensitive attributes `ID_POSSUI_NEC_ESPECIAL` and `ID_N_T_E_P`, respectively.

According to those results, we were able to conclude the following.

- **Adversary's deterministic success**. Defined in Equations 5.7 and 5.8, this metric measures the fraction of individuals in the focal database that can have their values for the sensitive attribute inferred with absolute certainty, in a scale from 0% to 100%.

  According to Table 5.8a for the sensitive attribute `ID_POSSUI_NEC_ESPECIAL`, the adversary's *a priori* deterministic success equals 0%, i.e. no individual can have their values for the sensitive attribute inferred with absolute certainty in the focal database without the use of auxiliary information. However, by using only the attributes `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO` as quasi-identifiers and only the database of 2014, the adversary can infer the values with absolute certainty for up to 57.17% of the individuals in the focal database, i.e. approximately 28 294 187 students.

  As the longitudinal collection of databases available to the adversary increases, allowing the use of more School Census databases as auxiliary information, so increases the fraction of individuals in the focal database that can have their values for the sensitive attribute `ID_POSSUI_NEC_ESPECIAL` inferred with absolute certainty. Particularly, the adversary's *a posteriori* deterministic success increases to 79.21% by using only the database of 2015 as auxiliary information, to 87.59% by using the databases of 2015 and 2016, and to 91.28% by using the databases from 2015 to 2017.

  From Table 5.8b for the sensitive attribute `ID_N_T_E_P`, the adversary's *a priori* deterministic success also equals 0%. However, by using only the attributes `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO` as quasi-identifiers and only the database of 2014, the adversary can infer the values with absolute certainty for up to 58.07% of the individuals in the focal database, i.e. approximately 28 739 608 students.

  As the longitudinal collection of databases available to the adversary increases, so increases the fraction of individuals in the focal database that can have their values for the sensitive attribute `ID_N_T_E_P` inferred with absolute certainty. Particularly, the adversary's *a posteriori* deterministic success increases to 68.60% by using only the database of 2015 as auxiliary information, to 75.32% by using the databases of 2015 and 2016, and to 79.92% by using the databases from 2015 to 2017.

- **Adversary's probabilistic success**. Defined in Equations 5.11 and 5.12, this metric measures the probability of a randomly chosen individual in the focal database having their value for the sensitive attribute inferred with absolute certainty, in a scale from 0% to 100%.

  According to Table 5.8a for the sensitive attribute `ID_POSSUI_NEC_ESPECIAL`, the adversary's *a priori* probabilistic success is already of 98.21%, i.e. the adversary's chance of inferring the value for the sensitive attribute for one of the 49 491 319 individuals in the focal database is already high even before performing the CAL attack. Furthermore, by using only the attributes `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO` as quasi-identifiers and only the database of 2014, the adversary increases their chance to up to 98.71%.

  As the longitudinal collection of databases available to the adversary increases, allowing the use of more School Census databases as auxiliary information, so increases the adversary's chance of inferring the value for the sensitive attribute `ID_POSSUI_NEC_ESPECIAL`. Particularly, the adversary's *a posteriori* probabilistic success increases to 99.03% by using only the database of 2015 as auxiliary information, to 99.30% by using the databases of 2015 and 2016, and to 99.49% by using the databases from 2015 to 2017.

  From Table 5.8b for the sensitive attribute `ID_N_T_E_P`, the adversary's *a priori* probabilistic success is of 82.50%. Furthermore, by using only the attributes `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO` as quasi-identifiers and only the database of 2014, the adversary increases their chance to up to 91.64%.

  As the longitudinal collection of databases available to the adversary increases, so increases the adversary's chance of inferring the value for the sensitive attribute `ID_N_T_E_P`. Particularly, the adversary's *a posteriori* probabilistic success increases to 93.03% by using only the database of 2015 as auxiliary information, to 94.17% by using the databases of 2015 and 2016, and to 95.07% by using the databases from 2015 to 2017.

◁

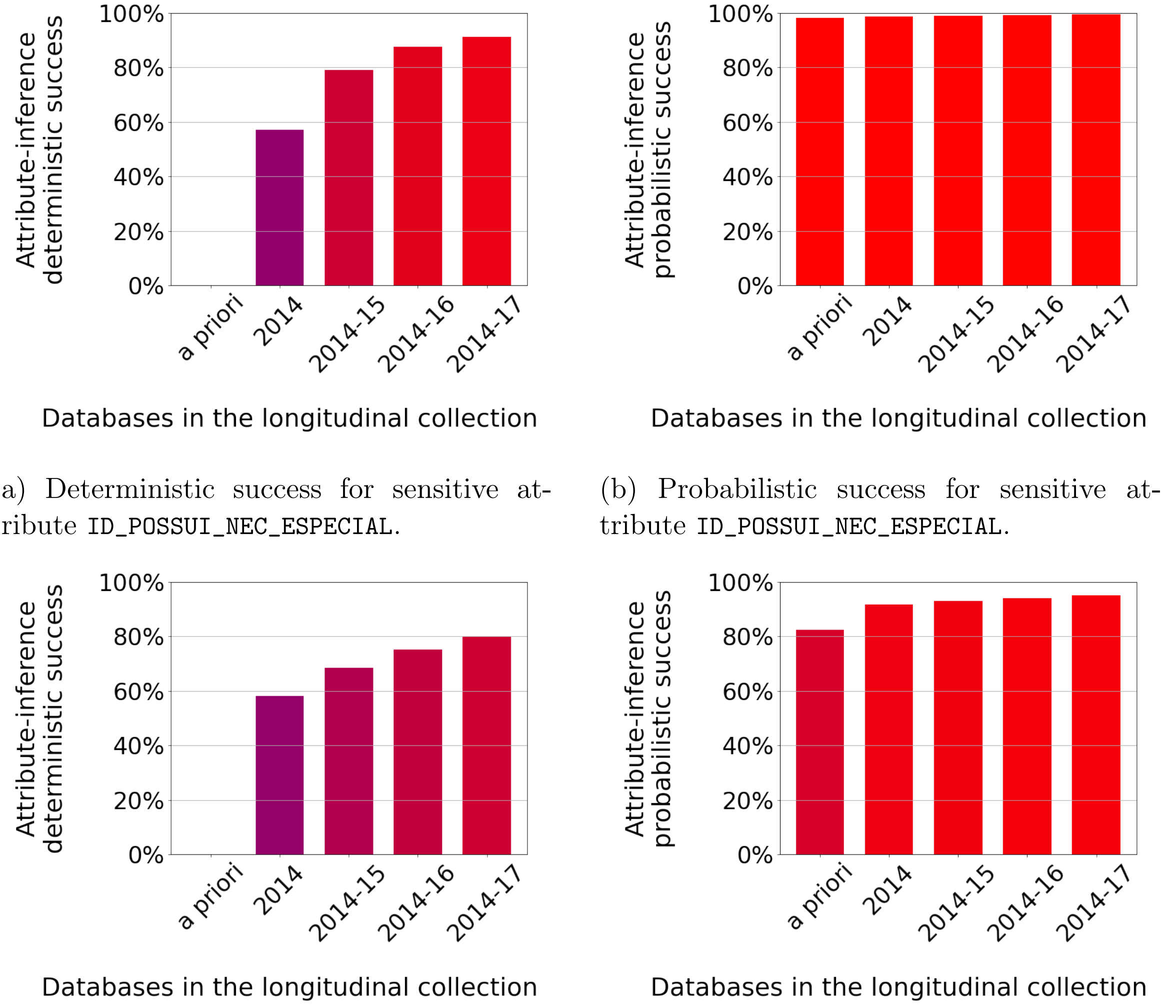
	Adversary's deterministic success		**Adversary's probabilistic success**	
	***a priori* success** 0.00%		***a priori* success** 98.21%	
Databases in the longitudinal aggregation	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
`2014`	57.17%	57.17%	98.71%	1.0051
`2014 to 2015`	79.21%	79.21%	99.03%	1.0083
`2014 to 2016`	87.59%	87.59%	99.30%	1.0111
`2014 to 2017`	91.28%	91.28%	99.49%	1.0130

(a) Sensitive attribute: `ID_POSSUI_NEC_ESPECIAL`.

	Adversary's deterministic success		**Adversary's probabilistic success**	
	***a priori* success** 0.00%		***a priori* success** 82.50%	
Databases in the longitudinal aggregation	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
`2014`	58.07%	58.07%	91.64%	1.1109
`2014 to 2015`	68.60%	68.60%	93.03%	1.1277
`2014 to 2016`	75.32%	75.32%	94.17%	1.1414
`2014 to 2017`	79.92%	79.92%	95.07%	1.1524

(b) Sensitive attribute: `ID_N_T_E_P`.

Table 5.8: Experiment 24: Privacy degradation in Collective-target Attribute-inference Longitudinal databases (CAL) attacks on the School Censuses from 2014 to 2017 using the quasi-identifiers `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO`. Here, the focal database is that for the School Census of 2014. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* success, while the probabilistic degradation of privacy is their ratio.



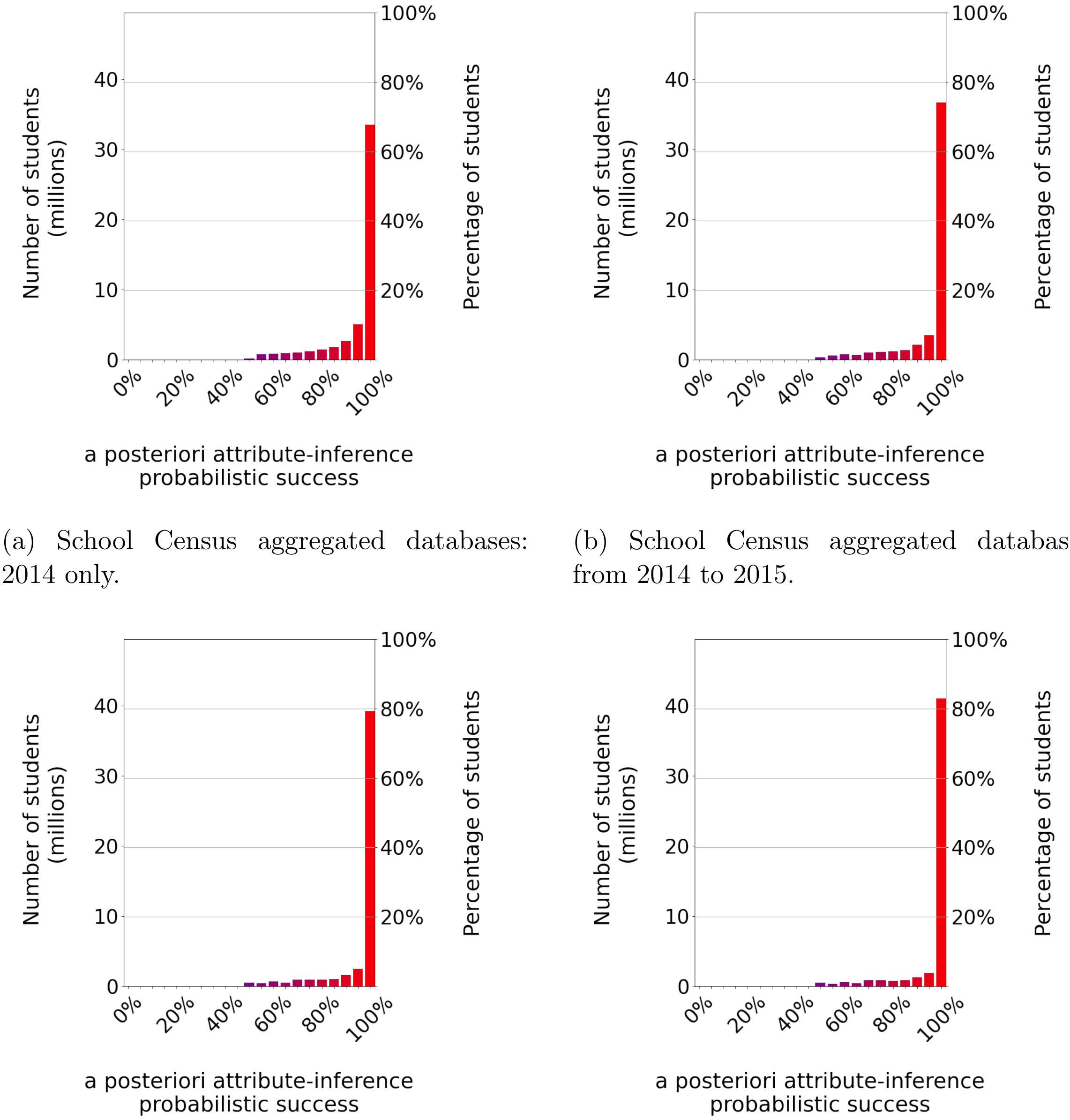
(a) Deterministic success for sensitive attribute `ID_POSSUI_NEC_ESPECIAL`.

(b) Probabilistic success for sensitive attribute `ID_POSSUI_NEC_ESPECIAL`.

(c) Deterministic success for sensitive attribute `ID_N_T_E_P`.

(d) Probabilistic success for sensitive attribute `ID_N_T_E_P`.

Figure 5.9: Experiment 24: Adversary's success in Collective-target Attribute-inference Longitudinal databases (CAL) attacks on the School Censuses from 2014 to 2017 using the quasi-identifiers `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO`. Here, each bar represents a different longitudinal aggregation of databases, always having the School Census of 2014 as the focal one, except for the bar with label *a priori*, which indicates the adversary's *a priori* success relying only on the focal database. The height of each bar represents the adversary's deterministic or probabilistic success.



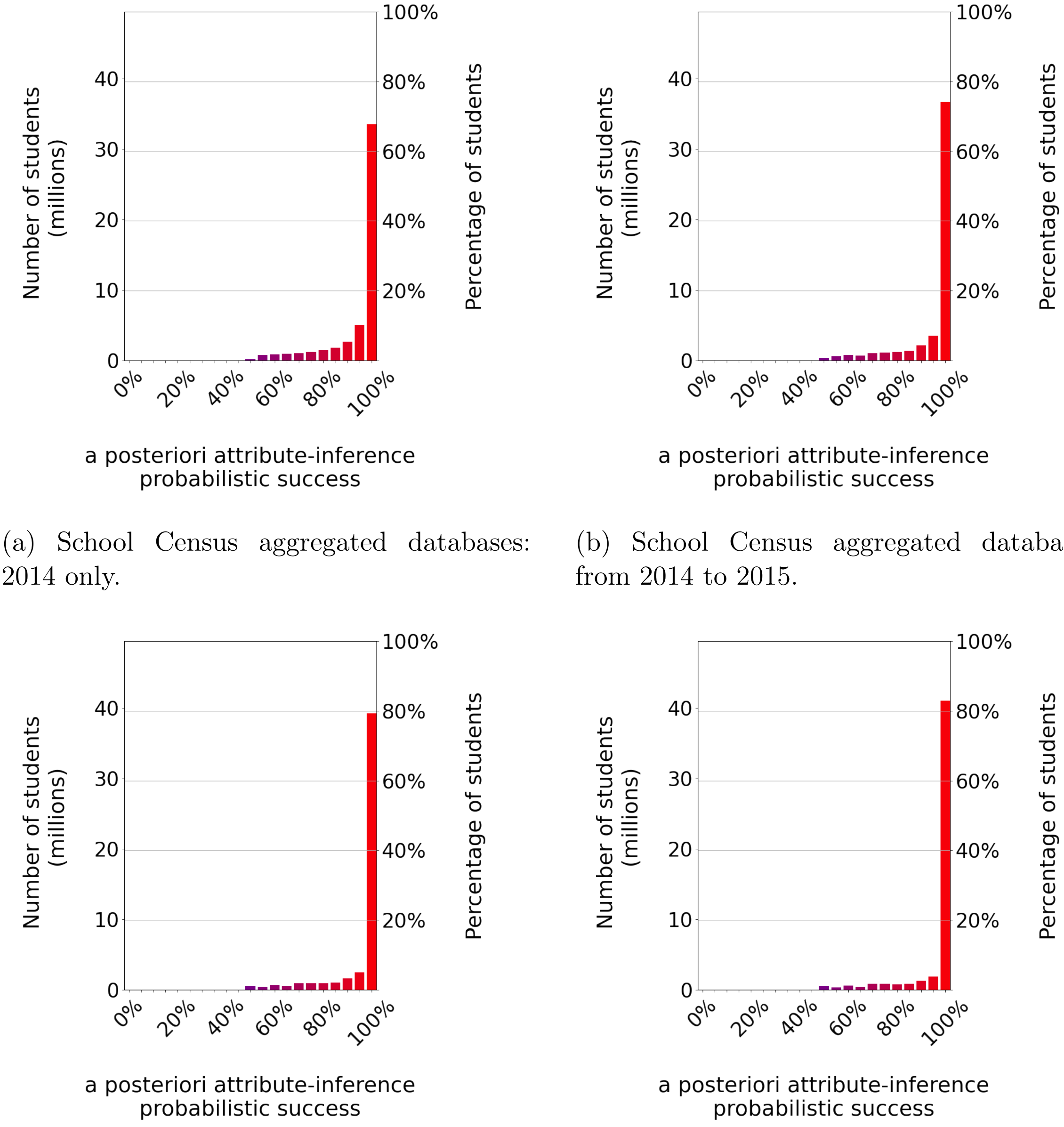
(a) School Census aggregated databases: 2014 only.

(b) School Census aggregated databases: from 2014 to 2015.

(c) School Census aggregated databases: from 2014 to 2016.

(d) School Census aggregated databases: from 2014 to 2017.

Figure 5.10: Experiment 24: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Attribute-inference Longitudinal databases (CAL) attacks on the School Censuses from 2014 to 2017 using the quasi-identifiers `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO` for the sensitive attribute `ID_POSSUI_NEC_ESPECIAL`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of attribute-inference.



(a) School Census aggregated databases: 2014 only.

(b) School Census aggregated databases: from 2014 to 2015.

(c) School Census aggregated databases: from 2014 to 2016.

(d) School Census aggregated databases: from 2014 to 2017.

Figure 5.11: Experiment 24: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Attribute-inference Longitudinal databases (CAL) attacks on the School Censuses from 2014 to 2017 using the quasi-identifiers `FK_COD_MUNICIPIO_END`, `PK_COD_ENTIDADE`, and `FK_COD_ETAPA_ENSINO` for the sensitive attribute `ID_N_T_E_P`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of attribute-inference.

5.3 Takeaways

In this chapter, we have introduced the theoretical foundation on attacks against longitudinal databases for collective-targets in Section 5.1. We have considered both re-identification and attribute-inference attacks, and both deterministic and probabilistic metrics of success and privacy degradation. Particularly, the probabilistic metrics are applications of the quantitative information flow (QIF) framework introduced in Section 2.3.

We have also presented experimental results for the collective-target attacks on the School Census released by INEP in Section 5.2. Both Experiments 23 and 24 were performed on the School Census from 2014 to 2017.

Those results have allowed us to draw the following conclusions.

- The **deterministic re-identification success** of an adversary with knowledge of only three attributes whose values may change through the years, i.e. city of residency, school, and educational stage, is initially very low. By considering only the focal database for the School Census of 2014, the adversary can re-identify with absolute certainty only 1.44% of the records, i.e. approximately 712 674 students. But as we increase the longitudinal collection of databases, the adversary's success increases to up to 36.31%, i.e. approximately 17 970 297 students, when using the databases for School Census of 2015, 2016, and 2017 as auxiliary information.

- The **probabilistic re-identification success** of an adversary with knowledge of only three attributes whose values may change through the years, i.e. city of residency, school, and educational stage, is also initially very low. By considering only the focal database for the School Census of 2014, the adversary's chance of correctly re-identifying a randomly selected record is of only 4.24%. But as we increase the longitudinal collection of databases, the adversary's chance increases to up to 45.60% when using the databases for School Census of 2015, 2016, and 2017 as auxiliary information.

- The **deterministic attribute-inference success** of an adversary achieves higher values than the respective deterministic re-identification success. For both sensitive attributes chosen as described in Table 5.4 and by considering only the focal database for the School Census of 2014, an adversary could correctly infer the value for either of the sensitive attributes for around 58% of the records, i.e. approximately 28 704 965 students. But as we increase the longitudinal collection of databases by adding the School Census of 2015, 2016, and 2017 as auxiliary information, the adversary could increase their success to a remarkable 97.28% of the records, i.e. approximately 48 145 155 students, when trying to infer whether a student possesses a disability or not. Similarly, the adversary could increase their success to 79.92% of the records, i.e. approximately 39 553 462 students, when trying to infer whether a student uses public school transport or not.

- The **probabilistic attribute-inference success** of an adversary is initially already high. For the sensitive attribute on students' disabilities, an adversary has an *a priori* chance of correctly inferring the value for a randomly chosen record in the focal database for the School Census of 2014 of 98.21%. This is because of how much skewed the distribution of students is for that attribute. A more skewed distribution implies a higher *a priori* success, as is the case for the attribute on students' disabilities, for which the majority of individuals holds the value 0 (No). Analogously, the opposite holds for the attribute on students' use of public transport, which presents a slightly less skewed distribution and, hence, lower values for an adversary's *a priori* chance, of 82.50%. [12] Therefore, not much increase in an adversary's success is seen as a result of increasing the longitudinal collection of databases. Even so, the adversary's success achieves a remarkable 99.49% chance for the sensitive attribute on students' disabilities by adding the School Census of 2015, 2016, and 2017 as auxiliary information. [13]

The experiments presented in this chapter were performed with only a small set of three quasi-identifying attributes, all of them seemingly innocuous when it comes to their perceived danger for individuals' privacy. Even though, the results have allowed us to further demonstrate that the removal of attributes alone is not sufficient to guarantee individuals' privacy, particularly in INEP's scenario of annual releases for their educational censuses, which allows adversaries to perform longitudinal databases attacks.

Therefore, the results reported for attacks on longitudinal databases demonstrate the high degree of privacy degradation for individuals even when considering only seemly innocuous information, such as city of residency, school, and educational stage. Those

[12] We have used average-case metrics for all of our analyses. Instead, a better approach would have been to use worse-case metrics whenever we are dealing with very skewed distributions [4, 5].

[13] For the sensitive attribute on students' use of public transport, an adversary has an *a priori* chance of correctly inferring the value for a randomly chosen record in the focal database of 82.50%. In this case, the adversary's chance reaches 95.07% by adding the databases for the School Census of 2015, 2016, and 2017 as auxiliary information.

results are expected from the literature on disclosure control (DC) discussed in Section 2.2, given the extremely low privacy guarantees provided by data de-identification, which is subject to both linkage [65] and composition [24] attacks.

In fact, most of the data holders in the databases analyzed in our experiments are subject to attribute-inference attacks. [14] This is of particular concern when considering an adversary's probabilistic measures of success, which are designed to measure how certain an adversary would be in a given attack. According to our results, that certainty is extremely high for the average data holder in every attribute-inference scenario. Hence, given the recent enactment of Brazil's LGPD privacy legislation, mitigating such vulnerabilities is necessary and urgent for INEP, particularly because of their annual release for their educational censuses.

We have also performed the same collective-target attacks from this chapter on the Higher Education Census released by INEP. The results for those experiments can be found in Appendix F Section F.2. Furthermore, for illustrative individual-target experimental results, in which we target either famous people or our acquaintances selected *a priori*, see Appendix C Section C.2.2.

[14] Even though the results of re-identification attacks were lower than those reported for attribute-inference attacks, almost 18 million students could be re-identified with seemly innocuous information in the worst scenario.

Chapter 6

Privacy and utility analyses in differential privacy

We have presented in Chapters 4 and 5 our models for attacks against single and longitudinal databases, respectively, along with extensive experimental results that have demonstrated INEP's databases are vulnerable to severe and generalized privacy risks. However, given the recent enactment of Brazil's LGPD privacy legislation, mitigating such vulnerabilities is necessary and urgent, but this always comes with a cost in utility, which in turn impacts data analysts. Hence the importance of knowing the balance between privacy and utility when deploying disclosure control (DC) methods.

We have also discussed in Chapter 2 some DC methods and presented examples from the literature in which imprecise definitions of adversaries and their context could lead to unexpected vulnerabilities in publicly released databases. Furthermore, syntactic methods are proven to be insufficient in providing privacy guarantees to data holders, which leads us to focus on semantic methods.

Given the importance of providing accurate formalization to guarantee what vulnerabilities a given database publication is subject to, we provide here privacy and utility analyses for two different implementations of differential privacy, a state-of-the-art DC method described in Chapter 2. Due to the challenge of extending differential privacy to longitudinal databases, we consider only scenarios for single databases.

In this chapter, we introduce the theoretical foundation to perform privacy and utility analyses when implementing both oblivious and local differential privacy mechanisms in Section 6.1. We also present the corresponding experimental results for analyses carried on a small sample from the School Census of 2019 in Section 6.2. We conclude this chapter by emphasizing the most relevant contributions in Section 6.3.

6.1 Theoretical foundation

In this section, we present our QIF model for analyzing both privacy and utility in two distinct implementations of differential privacy, one based on the oblivious mechanism and the other based on the local mechanism, both described in Section 2.2.3. For both mechanisms, we are interested in how privacy loss and utility metrics vary as we change the amount of noise introduced by each mechanism, particularly in the context of correlated-databases. Some remarks, propositions, and proofs on this theoretical model are provided in Appendix G.

As shown by Alvim et al. [6], the existence of correlations in databases exposes them to inference attacks on sensitive attributes. Hence the impossibility of having optimal noise-adding mechanisms with the best possible utility or privacy in all scenarios, including those based on the current state-of-the-art solution, differential privacy.

Considering our database model introduced in Section 3.2, we now present the leading example for privacy and utility analyses that will help us illustrate new concepts throughout this chapter.

Example 25 (Leading example for privacy and utility analyses)**.** Consider the following scenario in which the set of attributes is $\mathcal{A}$=$\{id, income, gender, country\}$, specified as follows:

- $dom(id) = \{1, 2, 3, \ldots\}$, consisting of the unique identification number of an individual given by a positive integer;
- $dom(income) = \{\texttt{low}, \texttt{medium}, \texttt{high}\}$, consisting of an individual's income bracket;
- $dom(gender) = \{\texttt{F}, \texttt{M}\}$, denoting an individual's gender, considered here to be either female ($\texttt{F}$) or male ($\texttt{M}$) as a simplification for the current example;
- $dom(country) = \{\texttt{AUS}, \texttt{BRA}\}$, denoting if an individual's country of residency is Australia ($\texttt{AUS}$) or Brazil ($\texttt{BRA}$).

Table 6.1 represents a database D consisting of eight records, each of which is a tuple representing an individual from a population of interest, e.g. subjects in a census or in a medical study. For instance, consider the subset of attributes $\mathcal{A}' = \{income, country\}$ and take individual x to be the one represented by the record $\langle 3, \texttt{medium}, \texttt{F}, \texttt{AUS} \rangle$. Then, this individual's sub-record with respect to $\mathcal{A}'$ is $x[\mathcal{A}'] = \langle \texttt{medium}, \texttt{AUS} \rangle$.

$\triangleleft$

id	*income*	*gender*	*country*
1	`low`	`F`	`BRA`
2	`low`	`M`	`BRA`
3	`medium`	`F`	`AUS`
4	`medium`	`M`	`AUS`
5	`medium`	`F`	`AUS`
6	`medium`	`M`	`AUS`
7	`high`	`F`	`BRA`
8	`high`	`M`	`BRA`

Table 6.1: Database D from the leading example for privacy and utility analyses, Example 25.

In the following section, we present our adversary model and their context, and define both *privacy loss* and *utility* in terms of quantitative information flow measures.

6.1.1 Adversary model in an attack context

We consider two possibly distinct agents. First, a *data analyst* who is interested in obtaining useful statistical information from the database via a query mechanism that returns the, possibly randomized, result for a query performed on it. Second, an *adversary* who is trying to infer sensitive information about the database's records and who also has access to the query mechanism output. Based on the two agents' goals, we model the *privacy loss* as the increase in the adversary's success in inferring sensitive information, and the *utility* as the data analyst's posterior success in inferring useful information, both after accessing the output of the query mechanism.

We now formally define an attack context and its components, followed by the definition of a joint probability distribution induced by a given context.

Definition 26 (Attack context)**.** An *attack context* is a tuple:

$$\Gamma = \langle D, \mathcal{A}, a_s, a_u, \texttt{count}_q, \pi^\star, M_{\texttt{count}_q} \rangle,$$

where:

- D is a database on the set of attributes $\mathcal{A}$ and the adversary knows that each individual appears only once in the database.

- $a_s \in \mathcal{A}$ is a *sensitive attribute* whose values are to be kept private for every record.

- $a_u \in \mathcal{A}$ is a *useful attribute* whose values are of interest to a data analyst who indirectly accesses them via a counting query.

- $\texttt{count}_q : \mathcal{D} \to \mathbb{N}$ is a counting query mapping databases to the number of records in the database satisfying a predicate $q : \mathit{dom}(a_u) \to \texttt{Bool}$, which maps each possible value for the record's useful attribute to a Boolean value: [1]

$$\texttt{count}_q(D) \overset{\text{def}}{=} \sum_{\substack{x \in \mathit{records}(\mathcal{A}),\\ q(x[a_u])}} \mathit{mult}_D(x) \ ,$$

where the result ranges over $[0, 1, \ldots |D|]$ and is called the *real count* for counting query $\texttt{count}_q$ on database D.

- $\pi^\star \in \mathbb{D}(\mathit{records}(\mathcal{A}))$ is an *a priori probability distribution* on the values of a new record to be added to the database D. The adversary can compute $\pi^\star$ by deriving the probability of each record $x \in \mathit{records}(\mathcal{A})$ from their frequency in D as:

$$\pi^\star_x \overset{\text{def}}{=} \frac{\mathit{mult}_D(x)}{|D|} \ . \tag{6.1}$$

From the probability distribution $\pi^\star$, the adversary can derive a marginal probability distribution $\pi^{\star\mathcal{A}'} : \mathbb{D}(\mathit{dom}(\mathcal{A}'))$ on the values of any subset of attributes $\mathcal{A}' \subseteq \mathcal{A}$ for the new added record as:

$$\pi^{\star\mathcal{A}'|D}_a \quad = \quad \sum_{\substack{x \in \mathit{records}(\mathcal{A})\\ x[\mathcal{A}']=a}} \pi^D_x \ . \tag{6.2}$$

- $M_{\texttt{count}_q} : \mathcal{D} \to \mathbb{D}(\mathbb{N})$ is a *query mechanism*, i.e. a function mapping databases to a probability distribution on natural numbers. Its output is made available to both the data analyst and any other interested observer, including the adversary, as a possibly noisy version of the real count $\texttt{count}_q(D)$. [2] The output of $M_{\texttt{count}_q}(D)$ is called the *reported count* for counting query $\texttt{count}_q$ on database D.

$\triangleleft$

[1] For instance, consider the database D on the set of attributes $\mathcal{A}$ from Example 25. Let q be a predicate on the useful column *income* defined as $q(i) \overset{\text{def}}{=} (i = \texttt{medium})$, for $i \in \mathit{dom}(\mathit{income})$. Then $\texttt{count}_q(D) = 4$.

[2] Note that the mechanism may introduce noise to preserve privacy. We will soon cover two approaches for implementing such a mechanism: the *oblivious* and the *local* models.

Definition 27 (Joint probability distribution induced by the context)**.** We denote by $p^{\Gamma} : \mathbb{D}(\mathit{records}(\mathcal{A}), \mathit{dom}(a_s), \mathbb{N}, \mathbb{N})$ the *joint probability distribution* induced by context Γ, which depends on the coin tosses of $\pi^{\star}$ and $M_{\texttt{count}_q}$, and such that:

$$p^{\Gamma}(x^{\star}{=}x, x^{\star}[a_s]{=}s, \texttt{count}_q(D{\cup}x^{\star}){=}u, M_{\texttt{count}_q}(D{\cup}x^{\star}){=}u')$$

is the probability that in Γ:

- the new added record $x^{\star}$ assumes value $x \in \mathit{records}(\mathcal{A})$;
- the sensitive value $x^{\star}[a_s]$ of the new added record $x^{\star}$ assumes value $s \in \mathit{dom}(a_s)$;
- the real count of query $\texttt{count}_q$ performed on the extended database $D{\cup}x^{\star}$ assumes value $u \in \mathbb{N}$;
- the reported count of query $\texttt{count}_q$ produced by the mechanism $M_{\texttt{count}_q}$, with respect to the extended database $D{\cup}x^{\star}$, assumes value $u' \in \mathbb{N}$.

$\lhd$

Also, we assume the following assumptions on the adversary and the data analyst.

- The adversary:

  A1 Has full knowledge of the context Γ, including every record in D. Hence, the adversary is capable of computing an *a priori* probability distribution according to Equation 6.1.

  A2 Is unsure about the value of one extra record $x^{\star} \in \mathit{records}(\mathcal{A})$ that will be added to the database, and has the goal to learn the value $x^{\star}[a_s]$ for the sensitive attribute a_s of this new record. With a slight abuse of notation, we denote by $D{\cup}x^{\star}$ the extended database $D{\cup}\{\!\{x^{\star}\}\!\}$ consisting in the union of all records originally in D with the new record $x^{\star}$.

  A3 Believes the value $x^{\star}[a_s]$ for the sensitive attribute a_s of the new added record $x^{\star}$ follows the probability distribution $\pi^{\star a_s^{\star}} \in \mathbb{D}(\mathit{dom}(a_s))$, derived from $\pi^{\star}$ according to Equation 6.2.

- The data analyst:

  A4 Has full knowledge of $\mathcal{A}$, a_u, $\texttt{count}_q$, $\pi^{\texttt{count}_q(D{\cup}x^{\star})}$, and $M_{\texttt{count}_q}$, but not necessarily of the database D.

A5 Has the goal to obtain the real count $\texttt{count}_q(D{\cup}x^\star)$ for query $\texttt{count}_q$ on the extended database $D{\cup}x^\star$. However, the only way they can interact with the database $D{\cup}x^\star$ is via the query mechanism $M_{\texttt{count}_q}$, which outputs a possibly randomized reported count $M_{\texttt{count}_q}(D{\cup}x^\star)$ which is an approximation of the real count.

A6 Believes the real count $\texttt{count}_q(D{\cup}x^\star)$ for the query $\texttt{count}_q$ performed on the extended database $D{\cup}x^\star$ follows a probability distribution compatible with $\pi^\star$ and D.

Although both the data analyst and the adversary can observe the reported count that is output by the query mechanism, their goals do not necessarily coincide. Whereas the data analyst wants to infer the real count $\texttt{count}_q(D{\cup}x^\star)$ of the query $\texttt{count}_q$ on the extended database $D{\cup}x^\star$, the adversary wants only to infer the sensitive value $x^\star[a_s]$ of the new added record $x^\star$.

We now define both *privacy loss* and *utility* in terms of quantitative information flow measures. Since both definitions depend on the probability distribution p^Γ, which itself depends on the coin tosses of the query mechanism $M_{\texttt{count}_q}$ used, we will also define how such mechanism operates, which is done in the next Section 6.1.2.

Definition 28 (Privacy loss in an attack context)**.** Consider the attack context $\Gamma = \langle D, \mathcal{A}, a_s, a_u, \texttt{count}_q, \pi^\star, M_{\texttt{count}_q}\rangle$ and its corresponding joint distribution $p^\Gamma{:}\mathbb{D}(\textit{records}(\mathcal{A}), \textit{dom}(a_s), \mathbb{N}, \mathbb{N})$ on the new added record $x^\star$, the sensitive value $x^\star[a_s]$ for the new added record, the real count of query $\texttt{count}_q$ performed on the extended database $D{\cup}x^\star$, and the reported count output by query mechanism $M_{\texttt{count}_q}$ on the extended database $D{\cup}x^\star$.

Then, the *privacy loss* of Γ is defined as the *multiplicative Bayes leakage* of the secret value $x^\star[a_s]$ for the added individual given the reported count $M_{\texttt{count}_q}(D{\cup}x^\star)$ on the extended database $D{\cup}x^\star$:

$$\textit{privacy-loss}(\Gamma) \overset{\text{def}}{=} \frac{\textit{post-vul}(\Gamma)}{\textit{prior-vul}(\Gamma)} \ , \tag{6.3}$$

where $\textit{post-vul}(\Gamma)$ is the *a posteriori Bayes vulnerability* of the secret value $x^\star[a_s]$ given the reported count $M_{\texttt{count}_q}(D{\cup}x^\star)$, defined as:

$$\textit{post-vul}(\Gamma) \overset{\text{def}}{=} \sum_{u'\in\mathbb{N}} \max_{s\in \textit{dom}(a_s)} p^\Gamma(x^\star[a_s]{=}s, M_{\texttt{count}_q}(D{\cup}x^\star){=}u') \ , \tag{6.4}$$

and *prior-vul*(Γ) is the *a priori Bayes vulnerability* of the secret value $x^\star[a_s]$, defined as:

$$\textit{prior-vul}(\Gamma) \overset{\text{def}}{=} \max_{s \in \textit{dom}(a_s)} p^\Gamma(x^\star[a_s]{=}s) \ . \tag{6.5}$$

◁

Definition 29 (Utility in an attack context)**.** Consider the attack context $\Gamma = \langle D, \mathcal{A}, a_s, a_u, \texttt{count}_q, \pi^\star, M_{\texttt{count}_q} \rangle$ and its corresponding joint distribution $p^\Gamma{:}\mathbb{D}(\textit{records}(\mathcal{A}), \textit{dom}(a_s), \mathbb{N}, \mathbb{N})$ on the new added record $x^\star$, the sensitive value $x^\star[a_s]$ for the new added record, the real count of query $\texttt{count}_q$ performed on the extended database $D{\cup}x^\star$, and the reported count output by query mechanism $M_{\texttt{count}_q}$ on the extended database $D{\cup}x^\star$.

Then, the *utility* of Γ is defined as the *a posteriori Bayes vulnerability* of the real count $\texttt{count}_q(D{\cup}x^\star)$ given the reported count $M_{\texttt{count}_q}(D{\cup}x^\star)$ on the extended database $D{\cup}x^\star$:

$$\textit{utility}(\Gamma) \overset{\text{def}}{=} \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} p^\Gamma(\texttt{count}_q(D{\cup}x^\star){=}u, M_{\texttt{count}_q}(D{\cup}x^\star){=}u') \ . \tag{6.6}$$

◁

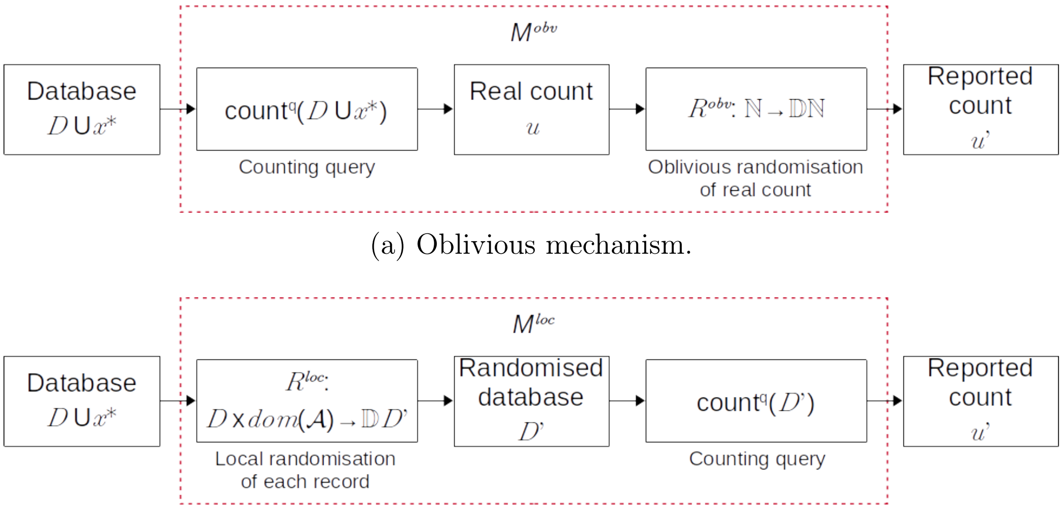
In the following section, we present our adversary model given the chosen mechanism $M_{\texttt{count}_q}$ is differential privacy, and in order to do se, we also define both the oblivious and local models for implementing a differential privacy mechanism.

6.1.2 Adversary model for differential privacy mechanisms

In this section, we define both the oblivious and local models for implementing a differential privacy mechanism, which in turn will be used as the mechanism $M_{\texttt{count}_q}$ in an attack context $\Gamma = \langle D, \mathcal{A}, a_s, a_u, \texttt{count}_q, \pi^\star, M_{\texttt{count}_q} \rangle$. The respective mechanisms are schematized in Figure 6.2.

6.1.2.1 Oblivious differential privacy

As discussed in Section 2.2.3.1, the *oblivious differential privacy* framework considers that the data curator is responsible for controlling and performing the queries on the database. Furthermore, the reported answer for each query possibly differs from the real answer given that the data curator perturbs it according to a chosen differentially-



(a) Oblivious mechanism.

(b) Locally-private mechanism.

Figure 6.2: Schemes for the different types of differential privacy mechanisms considered.

private mechanism and value of ϵ. In this case, the answer reported by the data curator is ϵ-differentially private.

Definition 30 (Oblivious mechanism)**.** Consider the attack context on the extended database $D{\cup}x^\star$, $\Gamma^{obv} = \langle D, \mathcal{A}, a_s, a_u, \texttt{count}_q, \pi^\star, M^{obv}_{\texttt{count}_q}\rangle$, and the *oblivious randomization function* $R^{obv} : \mathbb{N} \to \mathbb{D}(\mathbb{N})$ such that $R^{obv}(u' \mid u)$ is the probability of real count u being mapped to reported count u'.

The query mechanism $M^{obv}_{\texttt{count}_q} : \mathcal{D} \to \mathbb{D}(\mathbb{N})$ is *oblivious* iff the joint probability distribution $p^{\Gamma^{obv}} : \mathbb{D}(\mathit{records}(\mathcal{A}) \times \mathit{dom}(a_s) \times \mathbb{N} \times \mathbb{N})$ it induces on the new added record $x^\star$, the sensitive value $x^\star[a_s]$ for the new added record, the real count of query $\texttt{count}_q$ performed on the extended database $D{\cup}x^\star$, and the reported count output by query mechanism $M_{\texttt{count}_q}$ on the extended database $D{\cup}x^\star$ satisfies, for every record $x \in \mathit{records}(\mathcal{A})$, sensitive value $s \in \mathit{dom}(a_s)$, real count $u \in \mathbb{N}$, and reported count $u' \in \mathbb{N}$: [3]

$$p^{\Gamma^{obv}}(x^\star{=}x, x^\star[a_s]{=}s, \texttt{count}_q(D{\cup}x^\star){=}u, M^{obv}_{\texttt{count}_q}(D{\cup}x^\star){=}u') = \\ \pi^\star_x \cdot \delta_{x[a_s]}(s) \cdot \delta_{\texttt{count}_q(D\cup x)}(u) \cdot R^{obv}(u' \mid u) \ . \tag{6.7}$$

See Remark 63 in Appendix G for a detailed derivation the formula above.

[3] Here, δ denotes the *Dirac's delta function*, defined as:

$$\delta_a(b) \overset{\text{def}}{=} \begin{cases} 1, & \text{if } a = b \ , \\ 0, & \text{otherwise} \ . \end{cases}$$

The *privacy loss* of such oblivious mechanism is obtained by substituting Equation 6.7 in Definition 28, yielding:

$$\textit{privacy-loss}(\Gamma^{obv}) = \frac{\textit{post-vul}(\Gamma^{obv})}{\textit{prior-vul}(\Gamma^{obv})} \ , \tag{6.8}$$

where

$$\begin{aligned}\textit{post-vul}(\Gamma^{obv}) &= \sum_{u' \in \mathbb{N}} \max_{s \in \textit{dom}(a_s)} p^{\Gamma^{obv}}(x^{\star}[a_s]{=}s, M^{obv}_{\texttt{count}_q}(D{\cup}x^{\star}){=}u') \\ &= \sum_{u' \in \mathbb{N}} \max_{s \in \textit{dom}(a_s)} \sum_{\substack{x \in \textit{records}(\mathcal{A}),\\ x[a_s]=s}} \pi^{\star}_x \cdot R^{obv}(u' \mid \texttt{count}_q(D{\cup}x)) \ ,\end{aligned}$$

and

$$\textit{prior-vul}(\Gamma^{obv}) = \max_{s \in \textit{dom}(a_s)} \sum_{\substack{x \in \textit{records}(\mathcal{A}),\\ x[a_s]=s}} \pi^{\star}_x \ .$$

See Proposition 64 in Appendix G for a detailed derivation of the formula above.

The *utility* of such oblivious mechanism is obtained by substituting Equation 6.7 in Definition 29, yielding:

$$\begin{aligned}\textit{utility}(\Gamma^{obv}) &= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} p^{\Gamma^{obv}}(\texttt{count}_q(D{\cup}x^{\star}){=}u, M^{obv}_{\texttt{count}_q}(D{\cup}x^{\star}){=}u') \\ &= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} \sum_{\substack{x \in \textit{records}(\mathcal{A}),\\ \texttt{count}_q(D \cup x)=u}} \pi^{\star}_x \cdot R^{obv}(u' \mid u) \ . \end{aligned} \tag{6.9}$$

See Proposition 65 in Appendix G for a detailed derivation of the formula above. ◁

6.1.2.2 Local differential privacy

As discussed in Section 2.2.3.2, the *local differential privacy* framework considers that the chosen mechanism perturbs the data at the record level, making the microdata itself ϵ-differentially private. This allows for uncontrolled queries on the perturbed database or even the publication of the whole set of microdata for widespread use.

Definition 31 (Local mechanism)**.** Consider the attack context on the extended database $D{\cup}x^{\star}$, $\Gamma^{loc} = \langle D, \mathcal{A}, a_s, a_u, \texttt{count}_q, \pi^{\star}, M^{loc}_{\texttt{count}_q} \rangle$, and the *local randomization function* $R^{loc} : \textit{dom}(a_u) \rightarrow \mathbb{D}(\textit{dom}(a_u))$ such that $R^{loc}(w' \mid w)$ is the probability of useful value w for a record being mapped to another useful value w'.

The query mechanism $M^{loc}_{\texttt{count}_q} : \mathcal{D}(\mathcal{A}) \to \mathbb{D}(\mathbb{N})$ is *local* iff the joint probability distribution $p^{\Gamma^{loc}} : \mathbb{D}(\mathit{records}(\mathcal{A}) \times \mathit{dom}(a_s) \times \mathbb{N} \times \mathbb{N})$ it induces on the new added record $x^\star$, the sensitive value $x^\star[a_s]$ for the new added record, the real count of query $\texttt{count}_q$ performed on the extended database $D{\cup}x^\star$, and the reported count output by query mechanism $M_{\texttt{count}_q}$ on the extended database $D{\cup}x^\star$ satisfies, for every record $x \in \mathit{records}(\mathcal{A})$, sensitive value $s \in \mathit{dom}(a_s)$, real count $u \in \mathbb{N}$, and reported count $u' \in \mathbb{N}$:

$$\begin{aligned} &p^{\Gamma^{loc}}(x^\star{=}x, x^\star[a_s]{=}s, \texttt{count}_q(D{\cup}x^\star){=}u, M^{loc}_{\texttt{count}_q}(D{\cup}x^\star){=}u') = \\ &\pi^\star_x \cdot \delta_{x[a_s]}(s) \cdot \delta_{\texttt{count}_q(D\cup x)}(u) \cdot p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(D{\cup}x){=}u') \ , \end{aligned} \tag{6.10}$$

where $p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(D{\cup}x){=}u')$ is the conditional probability of the local mechanism returning reported answer u' when the input database is $D{\cup}x$, defined recursively for every database $d \in \mathcal{D}(\mathcal{A})$, record $y \in \mathit{records}(\mathcal{A})$, and reported count $c \in \mathbb{N}$ as:

$$p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(d{\cup}y){=}c) =$$
$$\begin{cases} 0, & \text{if } c < 0 \text{ or } c > |d{\cup}y|; \\ 1, & \text{if } d{\cup}y = \emptyset \text{ and } c = 0; \\ \sum_{\substack{w\in \mathit{dom}(a_u), \\ q(w)=\texttt{false}}} R^{loc}(w \mid y[a_u]), & \text{if } d{\cup}y = \{\!\{y\}\!\} \text{ and } c = 0; \\ \sum_{\substack{w\in \mathit{dom}(a_u), \\ q(w)=\texttt{true}}} R^{loc}(w \mid y[a_u]), & \text{if } d{\cup}y = \{\!\{y\}\!\} \text{ and } c = 1; \\ (p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(\{\!\{y\}\!\}){=}1) \cdot p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(d){=}c{-}1) + & \\ p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(\{\!\{y\}\!\}){=}0) \cdot p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(d){=}c)), & \text{if } 0 \le c \le |d{\cup}y| \text{ and} \\ & \emptyset \neq d{\cup}y \neq \{\!\{y\}\!\}. \end{cases} \tag{6.11}$$

See Remark 63 in Appendix G for a detailed derivation the formula above.

The *privacy loss* of such local mechanism is obtained by substituting Equation 6.10 in Definition 28, yielding:

$$\mathit{privacy\text{-}loss}(\Gamma^{loc}) = \frac{\mathit{post\text{-}vul}(\Gamma^{loc})}{\mathit{prior\text{-}vul}(\Gamma^{loc})} \ , \tag{6.12}$$

where

$$\mathit{post\text{-}vul}(\Gamma^{loc}) = \sum_{u'\in\mathbb{N}} \max_{s\in \mathit{dom}(a_s)} p^{\Gamma^{loc}}(x^\star[a_s]{=}s, M^{loc}_{\texttt{count}_q}(D{\cup}x^\star){=}u')$$

$$= \sum_{u' \in \mathbb{N}} \max_{s \in dom(a_s)} \sum_{\substack{x \in records(\mathcal{A}),\\ x[a_s]=s}} \pi^\star_x \cdot p^{\Gamma^{loc}}(M^{loc}_{\mathtt{count}_q}(D \cup x) = u') \ ,$$

and

$$\textit{prior-vul}(\Gamma^{loc}) = \max_{s \in dom(a_s)} \sum_{\substack{x \in records(\mathcal{A}),\\ x[a_s]=s}} \pi^\star_x \ .$$

See Proposition 67 in Appendix G for a detailed derivation of the formula above.

The *utility* of such local mechanism is obtained by substituting Equation 6.10 in Definition 29, yielding:

$$\begin{aligned} \textit{utility}(\Gamma^{loc}) &= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} p^{\Gamma}(\mathtt{count}_q(D \cup x^\star) = u, M^{loc}_{\mathtt{count}_q}(D \cup x^\star) = u') \\ &= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} \sum_{\substack{x \in records(\mathcal{A}),\\ \mathtt{count}_q(D \cup x) = u}} \pi^\star_x \cdot p^{\Gamma^{loc}}(M^{loc}_{\mathtt{count}_q}(D \cup x) = u') \ . \end{aligned} \tag{6.13}$$

See Proposition 68 in Appendix G for a detailed derivation of the formula above. ◁

6.1.3 Truncated geometric mechanism for differential privacy

We now define the truncated geometric mechanism, first proposed by Ghosh et al. [28], which will be used for both the oblivious and local differential privacy models in Example 33 and in the experiments presented in Section 6.2.

Definition 32 (Truncated geometric mechanism)**.** Given $\epsilon > 0$, let $\alpha = e^{-\epsilon}$. The *truncated geometric mechanism* with domain $\{0, 1, 2, \ldots, m\}$ and co-domain $\{0, 1, 2, \ldots, n\}$ is defined as:

$$G(j \mid i) = \begin{cases} \frac{1}{1+\alpha} \cdot \alpha^i, & \text{if } j = 0, \\ \frac{1-\alpha}{1+\alpha} \cdot \alpha^{|i-j|}, & \text{if } 0 < j < n, \\ \frac{1}{1+\alpha} \cdot \alpha^{|i-n|}, & \text{if } j = n, \end{cases} \tag{6.14}$$

for every $0 \leq i \leq m$ and $0 \leq j \leq n$. Each value $G(j \mid i)$ represents the probability that integer i is remapped to integer j.

For instance, Table 6.3 represents the matrix for a truncated geometric mechanism with $\alpha = 1/2$ and $n = 5$. For each input value ranging from 0 to 5, represented in the lines, there is a probability for the output value, in the columns. Each line sums to 1.

IN \ OUT	0	1	2	3	4	5
0	$^{2}/_{3}$	$^{1}/_{6}$	$^{1}/_{12}$	$^{1}/_{24}$	$^{1}/_{48}$	$^{1}/_{48}$
1	$^{1}/_{3}$	$^{1}/_{3}$	$^{1}/_{6}$	$^{1}/_{12}$	$^{1}/_{24}$	$^{1}/_{24}$
2	$^{1}/_{6}$	$^{1}/_{6}$	$^{1}/_{3}$	$^{1}/_{6}$	$^{1}/_{12}$	$^{1}/_{12}$
3	$^{1}/_{12}$	$^{1}/_{12}$	$^{1}/_{6}$	$^{1}/_{3}$	$^{1}/_{6}$	$^{1}/_{6}$
4	$^{1}/_{24}$	$^{1}/_{24}$	$^{1}/_{12}$	$^{1}/_{6}$	$^{1}/_{3}$	$^{1}/_{3}$
5	$^{1}/_{48}$	$^{1}/_{48}$	$^{1}/_{24}$	$^{1}/_{12}$	$^{1}/_{6}$	$^{2}/_{3}$

Table 6.3: Truncated geometric mechanism with $\alpha = {}^{1}/_{2}$ and $n = 5$.

The truncated geometric mechanism is specialized to the oblivious and local models as follows:

1. The oblivious randomization function $R^{obv} : \mathbb{N} \to \mathbb{D}(\mathbb{N})$ on database $D \cup x$ has domain and co-domain both equal to $\{0, 1, \ldots, |D \cup x|\}$, since the result of the counting query, either real or reported, must be a non-negative integer smaller than or equal to the size of the database, i.e. $|D \cup x|$. [4]

2. The local randomization function $R^{loc} : dom(a_u) \to \mathbb{D}(dom(a_u))$ has its domain $dom(a_u)$ converted to integers representing the distance between values of the domain, and the local mechanism is applied with both domain and co-domain equal to $\{0, 1, \ldots, |dom(a_u)| - 1\}$.

$\triangleleft$

6.1.4 Example of a privacy and utility analyses

We now present an example of how the proposed analyses of privacy and utility in terms of the privacy loss and utility metrics would be conducted for both the local and oblivious models based on the database presented in Example 25.

Example 33 (Execution of privacy and utility analyses for both the local and oblivious differential privacy mechanisms)**.** Consider again the scenario described in Example 25, the adversary model from Section 6.1.1, the differential privacy mechanisms from Section 6.1.2, and the truncated geometric mechanism from Definition 32.

Based on the attributes specified in Table 6.1, we have selected various pairs (a_u, a_s) of useful and sensitive attributes, respectively, with different degrees of correlation between them. Particularly, we have considered the following scenarios.

[4] We assume the adversary knows the database size, hence trying to report an answer outside of this range would be pointless.

Scenario A: highly-correlated (a_u, a_s)**.** In this case, we selected:

- a_s : *income*,
- a_u : *income*,
- Count query $\texttt{count}_q$: "Count x from D if $x[income] ==$ `medium`".

Scenario B: lowly-correlated (a_u, a_s)**.** In this case, we selected:

- a_s : *income*,
- a_u : *gender*,
- Count query $\texttt{count}_q$: "Count x from D if $x[gender] ==$ `F`".

Scenario C: sensitive-inferrable-from-useful (a_u, a_s)**.** In this case, we selected: [5]

- a_s : *country*,
- a_u : *income*,
- Count query $\texttt{count}_q$: "Count x from D if $x[income] ==$ `medium`".

Scenario D: useful-inferrable-from-sensitive (a_u, a_s)**.** In this case, we selected: [6]

- a_s : *income*,
- a_u : *country*,
- Count query $\texttt{count}_q$: "Count x from D if $x[country] ==$ `BRA`".

Also, we have selected the arithmetic loss function as a utility metric, i.e. for every $i, j \in \mathbb{N}$, the arithmetic loss function is given by:

$$\ell(i, j) = |i - j| \ .$$

The results for those experiments can be seen in Table 6.4.

In order to protect the data holders' privacy, we want the privacy loss metric to be as close to 1 as possible. Since the privacy loss in an attack context is defined as the multiplicative Bayes leakage, according to Definition 28, i.e. the ratio between the *a posteriori* and *a priori* Bayes vulnerabilities, a result equal to one means the information flow in a given scenario does not increase the adversary's *a priori* knowledge.

[5] Since country is a deterministic function of income, knowledge of income implies complete knowledge of country, but knowledge of country does not give as much information on income.

[6] Just the opposite of Scenario D.

However, a result greater than one means the adversary has successfully gathered new information and has increased their *a priori* knowledge. This result can go as high as the inverse of the adversary's *a priori* knowledge, which means that the posterior Bayes vulnerability is maximum and so is the data holders' privacy degradation.

Finally, from the data analyst's perspective, we also want the utility metric to be as close to one as possible, but with a different reasoning. Since the utility in an attack context is defined as the *a posteriori* Bayes vulnerability, according to Definition 29, a result close to one means that the true value of the query performed can be inferred from the reported value with high certainty by the data analyst. However, a result lower than one means the reported value is expected to be less accurate. This result can go as low as the proportion of records in the database for the most probable value of the useful attribute.

◁

6.2 Experimental results

In this section, we present our quantitative analyses for both privacy loss and utility metrics in both the oblivious and local implementations of differential privacy. The experiments performed here were modeled according to the theoretical model developed in Section 6.1.

We begin by detailing our experimental setup followed by the results for privacy and utility analyses on INEP's databases.

6.2.1 Experimental setup

As stated in Section 1.3, we have chosen to work with the databases containing information on students and released as microdata, i.e. data at the record level. In this chapter, we have selected just a small sample from the School Census of 2019 due to the high computational costs of analysis.

A preliminary analysis of the databases containing information on students showed that a student may hold more than one record in a given database, e.g. if a High School student is also enrolled in a Professional Education course. Hence, because differential privacy implicitly assumes that each individual holds only one record in the database, we have randomly selected only one record for each data holder of multiple records. This data treatment was based on the unique identification number, the `ID_ALUNO` code, given to each student in the pseudonymization treatment performed by INEP.

	Oblivious context: Γ^{obv}			Local context: Γ^{loc}		
	$\epsilon = \ln 1.5$	$\epsilon = \ln 3$	$\epsilon = \ln 10$	$\epsilon = \ln 1.5$	$\epsilon = \ln 3$	$\epsilon = \ln 10$
privacy-loss(Γ)	1.0000	1.1250	1.3636	1.0000	1.0000	1.0308
utility(Γ)	0.6000	0.7500	0.9091	0.5124	0.5492	0.6561

(a) Sensitive attribute a_s: *income*. Useful attribute a_u: *income*.
Count query $\texttt{count}_q$: "Count x from D if x[*income*] == `medium`".

	Oblivious context: Γ^{obv}			Local context: Γ^{loc}		
	$\epsilon = \ln 1.5$	$\epsilon = \ln 3$	$\epsilon = \ln 10$	$\epsilon = \ln 1.5$	$\epsilon = \ln 3$	$\epsilon = \ln 10$
privacy-loss(Γ)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
utility(Γ)	0.6000	0.7500	0.9091	0.5280	0.5812	0.7221

(b) Sensitive attribute a_s: *income*. Useful attribute a_u: *gender*.
Count query $\texttt{count}_q$: "Count x from D if x[*gender*] == `F`".

	Oblivious context: Γ^{obv}			Local context: Γ^{loc}		
	$\epsilon = \ln 1.5$	$\epsilon = \ln 3$	$\epsilon = \ln 10$	$\epsilon = \ln 1.5$	$\epsilon = \ln 3$	$\epsilon = \ln 10$
privacy-loss(Γ)	1.2000	1.5000	1.8182	1.0249	1.0984	1.3122
utility(Γ)	0.6000	0.7500	0.9091	0.5124	0.5492	0.6561

(c) Sensitive attribute a_s: *country*. Useful attribute a_u: *income*.
Count query $\texttt{count}_q$: "Count x from D if x[*income*] == `medium`".

	Oblivious context: Γ^{obv}			Local context: Γ^{loc}		
	$\epsilon = \ln 1.5$	$\epsilon = \ln 3$	$\epsilon = \ln 10$	$\epsilon = \ln 1.5$	$\epsilon = \ln 3$	$\epsilon = \ln 10$
privacy-loss(Γ)	1.0000	1.1250	1.3636	1.0000	1.0021	1.0831
utility(Γ)	0.6000	0.7500	0.9091	0.5280	0.5812	0.7221

(d) Sensitive attribute a_s: *income*. Useful attribute a_u: *country*.
Count query $\texttt{count}_q$: "Count x from D if x[*country*] == `BRA`".

Table 6.4: Privacy loss and utility results for Example 33.

The `ID_ALUNO` code, which is unique to each student at least in a given database release, easily allowed us to find those students with multiple records and to perform the random selection of just one of them.

For all the following experiments, we have the sample consisting of all students in federal institutions in Belo Horizonte, Brazil, i.e. attributes `TP_DEPENDENCIA` equals 1 and `CO_MUNICIPIO` equals `3 106 200`. This sample accounts for 4 676 unique records with the guarantee that no student holds more than one record in the original database.

We are interested here in just two attributes: age and ethnicity, i.e. `NU_IDADE` and `TP_COR_RACA`, respectively. For age, we have 3 278 students ranging from 6 to 18 years old and the remaining 1 398 from 19 to 82 years old, 1 150 of them with up to 40 years old. Particularly, the largest age groups are for 17, 16, and 18 years old students, with 801, 799, and 641 individuals, respectively. For ethnicity, 2 306 students have not responded, 1 132 have declared to be white, 200 to be black, 1 015 to be mixed, 21 to be asian, and 2 to be indigenous.

6.2.2 Results for privacy and utility analyses

In Experiment 34, we present a comparison between Scenarios A and B, as described in Example 33, both having the attribute `TP_COR_RACA` as the sensitive one. But while the query for Scenario A is on the attribute `TP_COR_RACA`, for Scenario B it is on the attribute `NU_IDADE`. Scenarios C and D were left as future work in Section 7.1.

Experiment 34 (Privacy and utility analyses for Scenarios A and B for students in federal institutions on the School Census of 2019 in Belo Horizonte, Brazil, for the sensitive attribute `TP_COR_RACA`)**.** In this experiment, we consider both Scenarios A and B, i.e. for a highly-correlated (a_u, a_s) pair of attributes and for another lowly-correlated one, respectively. For both scenarios, the sensitive attribute which interests the adversary is `TP_COR_RACA`, but for Scenario A, the query is `count * where TP_COR_RACA == 3`, i.e. those students who declared to be mixed, while for Scenario B, the query is `count * where NU_IDADE > 18`, i.e. those students with age greater than 18 years old.

The experiment results for values of ϵ ranging from $\ln 3$ to $\ln 10^5$ can be seen in Table 6.5.

Here we can observe two well-known results from the literature. First, for both scenarios and for each value of ϵ, the local mechanism is always more private, but less useful, than the oblivious mechanism, i.e. both values for the privacy loss and utility metrics for the local mechanism are always lower than the corresponding values for the oblivious mechanism. Second, given that our database for this experiment is small, with only 4 676 records, it was expected that the local mechanism would not provide much utility even for relatively high values of ϵ.

Furthermore, as explained in Example 33, we want both the privacy loss and utility metrics to be as close to one as possible. For privacy loss, this would mean that the adversary has not gathered much information from the attack. For utility, this would mean that the data analyst can have more confidence in the reported value for the query since it is more likely to be equal to the real value.

From the results presented for Scenario A in Table 6.5, it is clear that one cannot expect to have both the privacy loss and utility metrics with their optimal values if the

	Scenario A				**Scenario B**			
	Oblivious mechanism		**Local mechanism**		**Oblivious mechanism**		**Local mechanism**	
ϵ	**Priv. loss**	**Util.**	**Priv. loss**	**Util.**	**Priv. loss**	**Util.**	**Priv. loss**	**Util.**
theoretical minimum	1.00000	0.49316	1.00000	0.49316	1.00000	0.17130	1.00000	0.17130
$\ln \mathbf{3}$	1.08012	0.78293	1.00000	0.78293	1.00000	0.75000	1.00000	0.70103
$\ln \mathbf{5}$	1.20013	0.83333	1.00000	0.78293	1.00000	0.83333	1.00000	0.70103
$\ln \mathbf{10}$	1.30923	0.90909	1.00000	0.78293	1.00000	0.90909	1.00000	0.70103
$\ln \mathbf{30}$	1.39370	0.96774	1.00000	0.78293	1.00000	0.96774	1.00000	0.70103
$\ln \mathbf{60}$	1.41655	0.98361	1.00000	0.78293	1.00000	0.98361	1.00000	0.70103
$\ln \mathbf{100}$	1.42590	0.99010	1.00002	0.78293	1.00000	0.99010	1.00000	0.70106
$\ln \mathbf{200}$	1.43299	0.99502	1.00092	0.78295	1.00000	0.99502	1.00000	0.70190
$\ln \mathbf{500}$	1.43728	0.99800	1.01287	0.78463	1.00000	0.99800	1.00000	0.71218
$\ln \mathbf{10^3}$	1.43872	0.99900	1.05148	0.79899	1.00000	0.99900	1.00000	0.73904
$\ln \mathbf{10^5}$	1.44014	0.99999	1.42916	0.99395	1.00000	0.99999	1.00000	0.99447
theoretical maximum	2.02774	1.00000	2.02774	1.00000	2.02774	1.00000	2.02774	1.00000

Table 6.5: Experiment 34: Privacy loss and utility on Scenarios A (with highly correlated counting query and secret) and B (with practically independent counting query and secret) for students in federal institutions in Belo Horizonte, Brazil, on the School Census of 2019. Scenario A accounts for attribute `TP_COR_RACA` as the secret and for the query `count * where TP_COR_RACA == 3`, while Scenario B accounts for attribute `TP_COR_RACA` as the secret and for the query `count * where NU_IDADE > 18`.

useful and sensitive attributes are highly-correlated. However, for Scenario B, it would be possible to achieve closer to optimal results for both the privacy loss and utility metrics, particularly if using an oblivious mechanism. Hence the necessity of carefully analyzing the privacy and utility balance whenever implementing differential privacy.

$\triangleleft$

6.3 Takeaways

In this chapter, we have presented the theoretical foundation to perform privacy and utility analyses for both global, or oblivious, and local differential privacy mechanisms in Section 6.1, particularly in the context of possibly correlated-databases [6]. The models developed here are based on the quantitative information flow (QIF) framework introduced in Section 2.3.

We have also presented the corresponding experimental results for analyses performed on a small sample from the School Census of 2019 in Section 6.2. Experiment 34 accounted for Scenarios A and B, as described in Example 33.

Those results have allowed us to draw the following conclusions.

- For **highly-correlated** useful and sensitive attributes, data curators should not expect to have both the privacy loss and utility metrics with their optimal values, no matter the differential privacy mechanism used.
- For **lowly-correlated** useful and sensitive attributes, data curators may achieve closer to optimal results for both the privacy loss and utility metrics, particularly if using an oblivious mechanism.
- For both **highly-correlated** and **lowly-correlated** useful and sensitive attributes pairs, the local mechanism is always more private, but less useful, than the oblivious mechanism.

Therefore, as expected from the literature on differential privacy, the existence of correlations in databases exposes them to inference attacks on sensitive attributes [6, 50]. The importance of this result cannot be overstated, since it is not unusual to find claims in the literature that differential privacy is capable of protecting data holders in any circumstances.

Finally, the results presented in this chapter for the analyses of privacy and utility in differential privacy mechanisms are evidence of the difficulties INEP currently faces to continue with the release of microdata for educational censuses. Given the recent enactment of Brazil's LGPD privacy legislation, mitigating the vulnerabilities presented in Chapters 4 and 5 is necessary and urgent for INEP. However, syntactic methods are insufficient to mitigate individuals' privacy risks, while semantic methods such as differential privacy are challenging to calibrate.

Chapter 7

Conclusions

The work developed in this thesis has directly contributed to informing INEP, the most important educational data curator in Brazil, on how to properly analyze the vulnerabilities present in their current databases released in the form of microdata. Given the recent enactment of Brazil's LGPD privacy legislation, mitigating such vulnerabilities is necessary and urgent for INEP. Particularly, our work was fundamental to Reports 1 and 2, and contributed to Report 5, of the Decentralized Execution Term 8750 signed between INEP and the Federal University of Minas Gerais, whose goal was to improve INEP's public release of microdata in the context of the new privacy legislation. Moreover, our work provides the largest, most thorough study of actual privacy threats in official government data releases in Brazil, comprising more than fifty million individuals, or around 25% of the country's population.

In addition, we have developed a model for analyzing the implementation of both oblivious and local differential privacy mechanisms in terms of privacy loss and utility, particularly in the context of possibly correlated-databases. Differential privacy is the golden standard in the area of Disclosure Control (DC) and is currently being adopted by the United States Census Bureau for their decennial census. The analyses we have performed are particularly important given that possible correlations between attributes in a database may expose sensitive attributes to unexpected inference attacks.

Furthermore, our contributions include a new proposal for the categorization of attacks classification. The literature on DC provides some commonly used privacy and risk models to categorize the possible attacks against databases, as discussed in Section 2.4. However, the resulting categorization does not account for all possible scenarios and presents some overlapping definitions. Our proposed categorization, detailed in Section 3.1, untangles the possible category dimensions and solves the overlapping problems. Based on this new classification, we have developed four different models of collective-target attacks against databases, which have allowed us to widely explore the vulnerabilities present in the databases released by INEP as microdata. [1]

In terms of our research questions, defined in Section 1.4, we were able to provide extensive answers that we summarize here.

RQ1 The experiments performed on single databases in Section 4 have demonstrated that INEP's use of de-identification and pseudonymization as the only DC methods for protecting data holders' privacy is clearly insufficient. For instance, an adversary with knowledge of only three quasi-identifying attributes, i.e. day and month of birth and school code, could re-identify with absolute certainty up to 30.92% of the records on the School Census of 2018, which is equivalent to approximately 14 896 149 students. Furthermore, if the adversary could increase their knowledge with only the attribute for year of birth, they would be able to re-identify with absolute certainty up to 81.13% of the records, or approximately 39 085 531 students.

Those results only get worse for the data holders' privacy if we consider the adversary's probabilistic measures of success, which are designed to measure how certain an adversary would be in a given attack. According to our results, that certainty is extremely high for the average data holder in most scenarios. For instance, with knowledge of the same four quasi-identifying attributes as before, i.e. date of birth and school code, the adversary's success in correctly re-identifying a randomly selected record from the School Census of 2018 is of 89.93%. Even worse, an adversary with the goal of inferring the values for the sensitive attribute on students' disabilities on the same database and using the same quasi-identifying attributes would have a staggering 99.69% success.

Additionally, one of the main characteristics of the statistical studies released by INEP that we have considered is their annual frequency. As discussed in Chapter 2, most DC methods were designed to be applied to single databases, leaving longitudinal databases such as the School Census released by INEP open to additional vulnerabilities. Particularly, it is known from the literature that de-identification is vulnerable to both linkage [65] and composition [24] attacks. This leads us to our second research question.

RQ2 We were able to demonstrate with the experiments performed on longitudinal databases in Section 5 that even when considering only seemly innocuous information,

[1] We have also developed another four different models of individual-target attacks, which are presented in Sections C.1 and C.2, for single databases and longitudinal databases, respectively.

such as city of residency, school, and educational stage, an adversary would be able to severely degrade the data holders' privacy. For instance, with knowledge of only those three quasi-identifying attributes and access to three auxiliary databases, up to 36.31% of the students on the School Census of 2014 could be re-identified with absolute certainty, which is equivalent to approximately 17 970 297 students.

Again, those results would get worse for the data holders' privacy if we consider the adversary's probabilistic measures of success and, particularly, an adversary with the goal of inferring the values for the sensitive attribute on students' disabilities. For instance, an adversary with knowledge of only those three quasi-identifying attributes, i.e. city of residency, school, and educational stage, and access to the same three auxiliary databases would have a staggering 99.49% success.

Given the recent enactment of Brazil's LGPD privacy legislation, mitigating such vulnerabilities is necessary and urgent for INEP. Therefore, we have also proposed a third research question considering the knowledge from the literature that syntactic methods are inefficient to mitigate individuals privacy risks, hence the need to investigate a semantic method such as differential privacy. Particularly, we were interested on how to mitigate data holders' risks while maintaining data utility for analyst.

RQ3 We have developed a model for analyzing the privacy and utility balance in differential privacy implementations for both global and local mechanisms in Chapter 6. Differential privacy is a state-of-the-art DC method described in Chapter 2 that sometimes has its privacy guarantees overstated in the literature [50]. Our results demonstrate that differential privacy is highly dependent on the mechanism and parameter for noise addition used, but also on correlations existent in the database itself. Furthermore, our results demonstrate some challenges of properly setting a differential privacy implementation with optimal privacy and utility.

7.1 Future work

As we have presented our work, we have also pointed out some possible future work. We summarize those ideas here, in no particular order.

1. We have performed all the longitudinal databases attacks by using a unique identification number for each individual in every database in the longitudinal collection. This approach was well-suited for the databases we have considered here, given most of them provide such a unique identification for individuals. It

would be interesting to also perform those composition attacks based on quasi-identifying attributes, since INEP no longer publishes a unique identification for individuals since 2018.

2. We have performed all our database attacks on the unmodified microdata released by INEP. It would be interesting to model similar attacks on anonymized databases to measure how vulnerable data holders would still be given different anonymization methods.

3. We have selected a specific sample for our privacy and utility analyses in differential privacy, as discussed in Section 6.2. It would be interesting to randomly select a series of samples and statistically analyze the results.

4. We have proposed four scenarios for the possible correlations between useful and sensitive attributes in Example 33 for privacy and utility analyses in differential privacy, as discussed in Section 6. But so far we have performed only the experiments for Scenarios A and B, for highly and lowly-correlated useful and sensitive attributes, respectively. It would be interesting to also explore the experimental results for Scenarios C and D, for sensitive-inferrable-from-useful and useful-inferrable-from-sensitive cases, respectively.

Appendix A

Overview of privacy legislation

In this appendix, we present an overview of privacy legislation around the world. This topic was briefly discussed in Section 1.2.

On January 29, 2014, the General Assembly of the United Nations (UN) adopted Resolution 68/261, which establishes a set of fundamental principles for official statistics [74, 75]. The principles defined in this Resolution came from work developed by the Conference of European Statisticians in 1992 and adopted for the first time by the UN Statistical Commission in April 1994 [73].

Among the UN General Assembly motivations for adopting those principles is the necessity of official statistical systems to keep public trust. According to the Resolution's preamble, this trust comes from respecting the fundamental values and principles contained therein. In turn, among the established principles, the following stands out:

Principle 6. Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.

Prior to the principles' adoption by the UN General Assembly, when they had only been adopted by the entity's Statistics Committee, they served as a basis for the European Statistical Office (EUROSTAT) to draw up the European Statistics Code of Practice in 2001, revised in 2017 [21]. Subsequently, this document was used for the development of the Quality Assurance Framework of the European Statistical System, adopted in 2005 and revised in 2019 [22].

Similarly, the Statistical Conference of the Americas of the United Nations Economic Commission for Latin America and the Caribbean (ECLAC) developed, between 2009 and 2011, the Code of Good Practice in Statistics for Latin America and the Caribbean [17], approved in November 2011 during the Sixth Meeting of the ECLAC. The development of this Code had the collaboration of EUROSTAT and fourteen member countries of ECLAC. Later, this Code served as a basis for both the Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística*, or IBGE, in Portuguese) [46] and the National Institute of Educational Studies and Research (*Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*, or INEP, in Portuguese) [48], the two major statistical data publishers in Brazil, to elaborate their respective documents on good practices for statistics.

However, those Resolutions or Codes only establish principles and good practices on which governmental statistical entities should rely, but in practice they do not have the force of Law. Hence, we now briefly discuss the legislation adopted by some governments in order to guarantee individuals' privacy, including Australia, Brazil, the European Union, and the United States.

A.1 In Australia

The Australian Bureau of Statistics (ABS) was created by Law in 1975 [31] and has a mandate for collecting statistical data established by the Census and Statistics Act of 1905 [30], which includes the realization of the National Census. The 1905 Law requires the ABS to publish statistical information while guaranteeing the confidentiality of the collected information.

Furthermore, the Privacy Act of 1988 [32], revised in November 2015, establishes the national regulation for privacy and treatment of personal information, particularly with the definition of the Australian Privacy Principles. Among the established principles, we highlight the guaranteed use of "anonymization", i.e. the disassociation of individuals from their respective records, or "pseudonymization", i.e. the attribution of an individual code to each record in a database to replace the name or other direct identifiers, in addition to transparency by government entities in the collection and treatment of personal information, as well as the adoption of rules for the use and disclosure of personal data.

A.2 In Brazil

The Constitution of the Federative Republic of Brazil from 1988 [35], in its Article 5, guarantees to all Brazilians and foreigners residing in the country that:

X the privacy, private life, honour and image of persons are inviolable, and the right to compensation for property or moral damages resulting from their violation is ensured;

XXXIII all persons have the right to receive, from the public agencies, information of private interest to such persons, or of collective or general interest, which shall be provided within the period established by law, subject to liability, except for the information whose secrecy is essential to the security of society and of the State;

Hence, item X of Article 5 sets the constitutional, individual right to privacy, whereas item XXXIII of the same Article sets the constitutional right, individual and collective, to transparency by the State. However, there is no definition on how to balance those two principles or on what would be the legal limits of each one of them.

Therefore, to regulate the rights to privacy and transparency determined in those entrenched clauses of the 1988 Constitution, Law 12 527 of 2011 was sanctioned to regulate access to information and Law 13 709 of 2018 was sanctioned to regulate the protection of personal data.

Law 12 527 of 2011 [39], known as the Access to Information Law (*Lei de Acesso à Informação*, or LAI, in Portuguese), regulates item XXXIII of Article 5 of the Federal Constitution. Articles 6 and 8 of LAI determine to the public authority the duty to guarantee broad access to information, particularly that considered to be of collective or general interest, which must be made available via the Internet regardless of requirement. In addition, Article 7 guarantees the right of access to other information not immediately available via the Internet. However, according to Article 22, access to information may be denied in whole or in part if the information is considered to be confidential, including cases of judicial or industrial secrecy.

Particularly for the treatment of personal information, Article 31 establishes that it must be done in a transparent manner and with respect to individual freedoms and guarantees, according to item X of Article 5 of the Federal Constitution. However, provisions on the handling of personal information are open to subsequent regulation.

Law 13 709 of 2018 [40], known as the General Personal Data Protection Law (*Lei Geral de Proteção de Dados Pessoais*, or LGPD, in Portuguese), as per its Article 1, aims to protect the fundamental rights of freedom and privacy. Article 7 determines in which cases the processing of personal data is allowed, and Article 11 does the same specifically for "sensitive personal data". [1] In both cases, processing is permitted with

[1] From the LGPD Article 5 [40]:

the consent of the data subject or with the consent of the legal guardian for sensitive data.

One important definition comes in Article 12 of the LGPD for “anonymous data”, [2] which is not to be considered personal data, except when the anonymization process can be reversed with reasonable efforts. Therefore, objective factors such as the cost and time needed to reverse the anonymization process should be considered given the available technologies and disregarding the use of third party means. But again, the proper definition of what would be considered a reasonable effort, or which anonymization methods should be used, were left to subsequent regulation.

In Brazil, data on individuals are published by three main federal institutes: the Brazilian Institute of Geography and Statistics, the National Institute of Educational Studies and Research, and the Comptroller General of the Union.

The Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística*, or IBGE, in Portuguese) is the main statistical institute in the country and responsible for the production and analysis of the Decennial Demographic Censuses and various other statistical products in the social, demographic, agricultural and economic spheres. Its mandate comes from Law 5 534 of 1968 [33] and from Decree 73 177 of 1973 [34]. According to IBGE’s mandate, every natural or legal person, under public or private law, is obliged to provide the information requested by the institute. However, this information is protected by confidentiality and can be used, exclusively, for statistical purposes [47]. Furthermore, based on the Code of Good Practice in Statistics for Latin America and the Caribbean [17], the institute published in 2013 the IBGE Statistics Code of Good Practice (*Código de Boas Práticas das Estatísticas do IBGE*, in Portuguese) [46].

The National Institute of Educational Studies and Research (*Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*, or INEP, in Portuguese) is responsible for the development and maintenance of educational statistics systems and assessment projects, as well as the dissemination of this information, according to Law 9 448 of 1997 [36] and Decree 6 317 of 2007 [37]. As per the Decree, both Basic and Higher

II Sensitive personal data: personal data on racial or ethnic origin, religious belief, political opinion, union membership or affiliation to organizations of a religious, philosophical, or political nature, data relating to health or sexual life, genetic or biometric data, when linked to a natural person.

[2] From the LGPD Article 5 [40]:

III Anonymous data: data relating to an unidentifiable holder, considering the use of reasonable technical means available at the time of processing.

Education establishments, whether public or private, are required to provide the information requested by INEP, while also ensuring the confidentiality of personal data collected and prohibiting its use for other purposes.

Finally, the Comptroller General of the Union (*Controladoria-Geral da União*, or CGU, in Portuguese) is responsible for the maintenance of the Transparency Portal (*Portal da Transparência*, in Portugueses), launched as a web-page in 2004 and remodeled in June 2018. Based on the LAI [39] and the Complementary Law 131 of 2009 [38], the Portal is a channel through which citizens can monitor the use of federal resources collected with taxes in the provision of public services to the population, in addition to informing themselves about other matters related to the Federal Public Administration. This includes data on expenses with public servants' remuneration and resources made available and withdrawn by beneficiaries of social programs.

A.3 In the European Union

The European Parliament and the Council of the European Union (EU) enacted in October 1995 the Directive 95/46/EC, also known as *Data Protection Directive*, which regulated the processing and the free movement of personal data within the EU to protect fundamental rights and freedoms of individuals [71]. This Directive was then superseded by the Regulation (EU) 2016/679, implemented in May 2018 and known as *The General Data Protection Regulation* (GDPR) [72]. Among other reasons, the GDPR enactment was necessary to harmonize the application of data protection across the EU, given different implementations of the Directive 95/46/EC by Member States, but also as a response to increasing changes in the collection and use of personal data, as highlighted in the GDPR's preamble:

> (6) Rapid technological developments and globalisation have brought new challenges for the protection of personal data. The scale of the collection and sharing of personal data has increased significantly. Technology allows both private companies and public authorities to make use of personal data on an unprecedented scale in order to pursue their activities. Natural persons increasingly make personal information available publicly and globally. Technology has transformed both the economy and social life, and should further facilitate the free flow of personal data within the Union and the transfer to third countries and international organisations, while ensuring a high level of the protection of personal data.

The GDPR deals with the privacy and protection of personal data of all individuals in the EU and the European Economic Area (EEA). This includes entities that trade with the EU, regardless of their country of origin, as well as exports of personal data to outside those regions. Violations of the GDPR may result in fines to businesses of up to 20 million Euros or up to 4% of the annual worldwide turnover of the preceding financial year in case of an enterprise, whichever is greater.

A.4 In the United States

The United States (US) created its first Privacy Act in 1974, through an amendment to the US Code Title 5 that added Section 552a [42]. Under the new Section, federal agencies that collect, maintain, use, or disseminate any record of *personally identifiable information* (PII) must ensure that those actions take place legally, and that adequate safeguards are provided to prevent their misuse. Particularly in the use of those records for statistical research, the transfer of information must take place in a way that ensures those records are not individually identifiable, according to section 552a(b)(5) [45].

In 2002, the *Confidential Information Protection and Statistical Efficiency Act* (CIPSEA) [43] was approved. This Act established new safeguards to PII provided to federal agencies for statistical purposes under the promise of confidentiality. Thereafter, any intentional disclosure of such information for non-statistical purposes without the consent of the "data holder" became a federal crime.

Several federal agencies in the US collect and disclose PII for statistical purposes, including the National Center for Education Statistics (NCES) and the Census Bureau (USCB). However, since each agency is subordinate to a specific Department of the federal government, and hence subject to the rules established by it, they have implemented different methods to protect PII. Furthermore, in some cases there is also specific legislation that governs the mandate of certain agencies, such as the USCB Constitutional obligation to perform the Decennial Census, regulated by the US Code Title 13 [41]. According to Section 9 of Title 13, all information obtained based on this Title must be used exclusively for the statistical purposes for which it was collected, and no individual or establishment in particular can have its identity revealed from the information provided.

The methods in use by fourteen federal agencies up to 2005 were reported in the Statistical Policy Working Paper 22 [44]. Those methods, known in the literature as "syntactic methods" and discussed here in Section 2.2.2, were proven during the decade of 2000 to be inherently insecure and subject to unintended disclosure of PII [13, 24].

Furthermore, this was the case independently of the data being aggregated or not, i.e. released as tabular data or as “microdata”. Consequently, the USCB started the most important change in data disclosure control practices in the US until then by adopting formal, provably secure privacy methods, known as “semantic methods”, from the 2020 Census onward [26, 76, 77].

Appendix B

Development of the concept of vulnerability

In this appendix, we present an overview on why to consider min-entropy instead of Shannon entropy for some scenarios and how this relates to the concept of "vulnerability". As discussed in Chapter 2.3, semantic disclosure control (DC) methods consider the uninformative principle, which requires that the adversary's *a posteriori* knowledge on any data holder should not be much larger than their *a priori* knowledge.

One way to quantitatively measure how much information an adversary has gathered by interacting with a database is by analyzing the quantitative flow of information in the process. Until 2009, the consensus metrics for such measurements were either the *Shannon entropy* [1] or the *mutual information.* [2]

As was demonstrated by Geoffrey Smith in 2009, not always those metrics have meaningful operational interpretation with respect to information security [66]. Particularly,

[1] The Shannon entropy $H(X)$ of a random variable X with possible values in $\mathcal{X}$ is:

$$H(X) = \sum_{x \in \mathcal{X}} P[X = x] \log \frac{1}{P[X = x]},$$

i.e. the expected number of bits required to transmit X optimally, measured in bits. Informally, the Shannon entropy can be interpreted as the uncertainty about X.

[2] The mutual information $I(X;Y)$ of two jointly distributed random variables X and Y is:

$$I(X;Y) = H(X) - H(X|Y),$$

where $H(X)$ is the Shannon entropy of X and $H(X|Y)$ is the conditional Shannon entropy of X given Y:

$$H(X|Y) = \sum_{y \in \mathcal{Y}} P[Y = y] H(X|Y = y),$$

Smith considered in his work a program that receives an input IN and produces an output OUT accessible by an adversary, from where he formulated that the information leaked by the program should be equal to the adversary's initial uncertainty minus the adversary's remaining uncertainty. By considering the then consensus definitions, his formulation can be translated to:

$$I(\texttt{IN};\texttt{OUT}) = H(\texttt{IN}) - H(\texttt{IN}|\texttt{OUT}).$$

Given this scenario, Smith showed that the only meaningful interpretation of the Shannon entropy for the information leaked, i.e. $H(\texttt{OUT})$, is the distinction between zero and nonzero values. $H(\texttt{OUT})$ equals zero iff the deterministic program satisfies the *non-interference principle*, i.e. the output is independent from the input and nothing can be learned from the former. For nonzero values, two different results for $H(\texttt{OUT})$ cannot be directly compared since the Shannon entropy can be arbitrarily large.

Furthermore, Smith showed that the conditional entropy $H(\texttt{IN}|\texttt{OUT})$, which corresponds to the adversary's remaining uncertainty, produces results incompatible with the programs considered in the examples for his proposed scenario [66].

As an alternative, Smith formulated a new definition for the adversary's remaining uncertainty that would already consider the desired security guarantees, named "vulnerability" and defined as follows.

Definition 35 (Vulnerability [66])**.** Given a random variable X with space of possible values $\mathcal{X}$, the vulnerability of X, denoted $V(X)$, is given by: [3]

$$V(X) = \max_{x \in \mathcal{X}} P[X = x].$$

◁

The proposed vulnerability, a worst-case metric, measures the probability that an adversary could correctly guess the value of X in one try. Based on this definition of vulnerability, Smith defined the min-entropy $H_\infty(X)$ as the new measure for the adversary's initial uncertainty, the conditional min-entropy $H_\infty(X|Y)$ as the new measure

measured in bits and where

$$H(X|Y = y) = \sum_{x \in \mathcal{X}} P[X = x|Y = y] \log \frac{1}{P[X = x|Y = y]}.$$

Informally, the conditional entropy can be interpreted as the uncertainty about X given Y, while the mutual information can be interpreted as the amount of information shared between X and Y.

[3] Nowadays, this is known as the *Bayes vulnerability* $V_1(\pi)$, induced by the gain function g_{id}.

for the adversary's remaining uncertainty, and their difference $H_\infty(X) - H_\infty(X|Y)$ as the new measure for the information leaked. Both min-entropy and conditional min-entropy are defined as follows.

Definition 36 (Min-entropy [66])**.** The min-entropy $H_\infty(X)$ of a X, denoted $H_\infty(X)$, is given by:

$$H_\infty(X) = \log \frac{1}{V(X)}.$$

$\triangleleft$

Definition 37 (Conditional vulnerability [66])**.** Given (jointly distributed) random variables X and Y, the *conditional vulnerability* $V(X|Y)$ is

$$V(X|Y) = \sum_{y \in \mathcal{Y}} P[Y = y] V(X|Y = y)$$

where

$$V(X|Y = y) = \max_{x \in \mathcal{X}} P[X = x|Y = y].$$

$\triangleleft$

Definition 38 (Conditional min-entropy [66])**.** The conditional min-entropy $H_\infty(X|Y)$ is:

$$H_\infty(X|Y) = \log \frac{1}{V(X|Y)}.$$

$\triangleleft$

Differently from the then consensus definitions, the new ones allowed Smith to achieve meaningful operational interpretations with respect to information security for the computed values in the considered scenario. Furthermore, his work was foundational for the development of the theory of *Quantitative Information Flow* (QIF) [5], introduced in Section 2.3.

Appendix C

Individual-target attacks

In this appendix, we present the theoretical foundation on attacks against single and longitudinal databases for individual-targets and for both re-identification and attribute-inference attacks. Additionally, we provide the respective experimental results for attacks on the Educational Censuses released by INEP. Individual-target models are illustrative to their collective-target counterparts in the sense that they clearly show the privacy risks that a targeted-individual can be subjected to given the knowledge of values for some trivial quasi-identifying attributes.

C.1 Attacks on single databases

In this section, we present the theoretical foundation on attacks against single databases for individual-targets in Section C.1.1, followed by the respective experimental results for attacks on the Educational Censuses released by INEP in Section C.1.2.

C.1.1 Theoretical foundation

For all attacks on single databases, we follow the assumptions introduced in Section 4.1.1. In this section, we formally define both re-identification and attribute-inference attacks on single databases for individual-targets. Definition 39 accounts for an adversary interested in re-identifying only one individual, while Definition 40 accounts for an adversary interested in inferring the value of an attribute for the targeted-individual. Detailed examples of both attacks are presented in Appendix D Sections D.1.1 and D.1.2, respectively.

Definition 39 (Individual-target Re-identification Single database (IRS) attack)**.** In an IRS attack, the adversary can completely access a single database, i.e. they have knowledge of every record in the database. The adversary can also gather auxiliary information on the values of a set of quasi-identifiers for a predetermined, specific individual of interest. The adversary's goal is to re-identify the targeted-individual, i.e. to precisely determine which record in the database corresponds to the individual of interest.

- **Adversary's knowledge.** In an IRS attack, we assume the adversary:

  (i) can completely access a (non-empty) database D on a set of attributes $\mathcal{A}$, according to Assumptions **AS0**–**AS2**;

  (ii) has a predetermined, specific individual of interest $x^{\star}{\in}D$ as the re-identification target;

  (iii) given a set of quasi-identifiers $\mathcal{Q}_{ID}{\subseteq}\mathcal{A}$, the adversary can gather auxiliary information on the individual of interest $x^{\star}$.

- **Deterministic degradation of privacy.** The adversary is considered to be deterministically successful in an IRS attack iff capable of precisely determining which record in the database corresponds to the individual of interest.

  - ***A priori* deterministic success.** Before performing the IRS attack, the adversary can rely only on the database itself, i.e. the quasi-identifiers are not used yet. Hence, the adversary can only precisely determine which record in the database corresponds to the individual of interest iff the database contains only a single record. Therefore, the *a priori* deterministic success in an IRS attack on a database D is defined as

$$\textit{prior-suc}^{\textit{IRS}}_{\textit{det}}(D) \quad \overset{\text{def}}{=} \quad \begin{cases} \texttt{true}, & \text{if } |D|{=}1, \\ \texttt{false}, & \text{if } |D|{\geq}2. \end{cases} \tag{C.1}$$

  - ***A posteriori* deterministic success.** By performing an IRS attack, the adversary uses the auxiliary information as values for the quasi-identifiers of the individual of interest to support the task of re-identification. An IRS attack is then performed as follows: At first, the adversary filters the database according to the values for the quasi-identifiers of the individual of interest, disregarding those records that do not match the query. Then, given the remaining database, the adversary can only precisely determine which record corresponds to the individual of interest iff the remaining database contains only a single record.

Therefore the *a posteriori* deterministic success in an IRS attack on a database D, given the adversary's knowledge of the values for the target's quasi-identifiers $x^\star[\mathcal{Q}_{ID}]$, is defined as:

$$\textit{post-suc}^{\textit{IRS}}_{\textit{det}}(D, x^\star[\mathcal{Q}_{ID}]) \quad \stackrel{\text{def}}{=} \quad \begin{cases} \texttt{true}, & \text{if there is exactly one} \\ & \text{record } x{\in}D \text{ such that} \\ & x[\mathcal{Q}_{ID}]{=}x^\star[\mathcal{Q}_{ID}], \\ \texttt{false}, & \text{otherwise.} \end{cases} \tag{C.2}$$

- **Deterministic degradation of privacy.** The adversary is considered to be deterministically successful in an IRS attack iff **not** capable of precisely re-identifying the individual of interest **before** the attack, but otherwise capable of doing so **after** the attack is performed:

$$\begin{aligned} \textit{priv-degrad}^{\textit{IRS}}_{\textit{det}}(D, x^\star[\mathcal{Q}_{ID}]) \quad \stackrel{\text{def}}{=} \quad & \textit{prior-suc}^{\textit{IRS}}_{\textit{det}}(D){=}\texttt{false} \quad \text{and} \\ & \textit{post-suc}^{\textit{IRS}}_{\textit{det}}(D, x^\star[\mathcal{Q}_{ID}]){=}\texttt{true}. \end{aligned} \tag{C.3}$$

- **Probabilistic degradation of privacy.** The adversary's probabilistic success in an IRS attack does not rely on precisely determining which record in the database corresponds to the individual of interest. Rather, we compute the probability of correctly re-identifying the target in the database, i.e. the greater this probability, the more successful the adversary is.

  - ***A priori* probabilistic success.** Before performing the IRS attack, the adversary can rely only on the database itself, i.e. the quasi-identifiers are not used yet. Hence, the best course of action for the adversary is to randomly select a record from the database, according to the maximum entropy principle. Since each individual only holds one record in the database, the probability of the adversary being successful is the inverse of the database size. Therefore, the *a priori* probabilistic success in an IRS attack on a database D is defined as:

$$\textit{prior-suc}^{\textit{IRS}}_{\textit{prob}}(D) \quad \stackrel{\text{def}}{=} \quad \frac{1}{|D|}\ . \tag{C.4}$$

  - ***A posteriori* probabilistic success.** By performing an IRS attack, the adversary uses the auxiliary information as values for the quasi-identifiers of the individual of interest to support the task of re-identification. An IRS attack is then performed as follows: At first, the adversary filters the database according to the values for the quasi-identifiers of the individual of interest, disregarding records that do not match the query. Then, given the remaining database, the best course of action for the adversary is to randomly select a record, according to the maximum entropy principle. Since each individual only holds one record in the database, the probability of the adversary being successful is the inverse of the remaining database size.

Therefore, the *a posteriori* probabilistic success in an IRS attack on a database D, given the adversary's knowledge of the values for the target's quasi-identifiers $x^\star[\mathcal{Q}_{ID}]$, is defined as:

$$\textit{post-suc}_{\textit{prob}}^{\textit{IRS}}(D, x^\star[\mathcal{Q}_{ID}]) \overset{\text{def}}{=} \frac{1}{|\textit{block}(x^\star[\mathcal{Q}_{ID}])|}\ , \tag{C.5}$$

where $\textit{block}(x^\star[\mathcal{Q}_{ID}])$ is the block defined by the partition on set $\mathcal{Q}_{ID}$ in which the individual of interest $x^\star$ belongs.

– **Probabilistic degradation of privacy.** The probabilistic degradation of privacy in an IRS attack is defined as the ratio by which the attack increases the probabilistic success of the adversary:

$$\begin{aligned}\textit{priv-degrad}_{\textit{prob}}^{\textit{IRS}}(D, x^\star[\mathcal{Q}_{ID}]) &\overset{\text{def}}{=} \frac{\textit{post-suc}_{\textit{prob}}^{\textit{IRS}}(D, x^\star[\mathcal{Q}_{ID}])}{\textit{prior-suc}_{\textit{prob}}^{\textit{IRS}}(D)}\\ &= \frac{|D|}{|\textit{block}(x^\star[\mathcal{Q}_{ID}])|}\\ &= \frac{|D|}{|\{x{\in}D \mid x[\mathcal{Q}_{ID}] = x^\star[\mathcal{Q}_{ID}]\}|}\ . \end{aligned} \tag{C.6}$$

A detailed numeric example of an IRS attack is presented in Section D.1.1, Example 47.

◁

We now present the attribute-inference attack on single databases for individual-targets.

Definition 40 (Individual-target Attribute-inference Single database (IAS) attack)**.** In an IAS attack, the adversary can completely access a single database, i.e. they have knowledge of every record in the database. The adversary can also gather auxiliary information on the values of a set of quasi-identifiers for a predetermined, specific individual of interest. The adversary's goal is to infer the value of an attribute considered to be sensitive for the targeted-individual regardless of being able to re-identify the individual of interest.

- **Adversary's knowledge.** In an IAS attack, we assume the adversary:
  - (i) can completely access a (non-empty) database D on a set of attributes $\mathcal{A}$;
  - (ii) has a predetermined, specific individual of interest $x^{\star}{\in}D$ as the attribute-inference target;
  - (iii) given a set of quasi-identifiers $\mathcal{Q}_{ID}{\subseteq}\mathcal{A}$, the adversary can gather auxiliary information on the individual of interest $x^{\star}$;
  - (iv) given an attribute $a_{sens}{\in}\mathcal{A}$ such that $a_{sens}{\notin}\mathcal{Q}_{ID}$, considered to be sensitive, the adversary's goal is to infer the attribute's value $x^{\star}[a_{sens}]$ for the targeted-individual $x^{\star}$.
- **Deterministic degradation of privacy.** The adversary is considered to be deterministically successful in an IAS attack iff capable of precisely determining the value of the sensitive attribute for the individual of interest.
  - ***A priori* deterministic success.** Before performing the IAS attack, the adversary can rely only on the database itself, i.e. the quasi-identifiers are not used yet. Hence, the adversary can only precisely determine the value of the sensitive attribute for the individual of interest iff all the records in the database have the same value for the sensitive attribute. Therefore the *a priori* deterministic success in an IAS attack on a database D and with respect to the sensitive attribute a_{sens} is defined as:

$$\textit{prior-suc}^{IAS}_{det}(D, a_{sens}) \quad \overset{\text{def}}{=} \quad \begin{cases} \texttt{true}, & \text{if for all } x, x'{\in}D, \\ & x[a_{sens}]{=}x'[a_{sens}], \\ \texttt{false}, & \text{otherwise.} \end{cases} \tag{C.7}$$

  - ***A posteriori* deterministic success.** By performing an IAS attack, the adversary uses the auxiliary information as values for the quasi-identifiers of the individual of interest to support the task of inferring the value of the sensitive attribute. An IAS attack is then performed as follows: At first, the adversary filters the database according to the values for the quasi-identifiers of the individual of interest, disregarding those records that do not match the query. Then, given the remaining database, the adversary can only precisely determine the value of the sensitive attribute for the individual of interest iff all the records in the remaining database have the same value for the sensitive attribute.

Therefore, the *a posteriori* deterministic success in an IAS attack on a database D, with respect to the sensitive attribute a_{sens}, given the adversary's knowledge of the values for the target's quasi-identifiers $x^{\star}[\mathcal{Q}_{ID}]$, is defined as:

$$\begin{aligned} &\textit{post-suc}^{IAS}_{det}(D, a_{sens}, x^{\star}[\mathcal{Q}_{ID}]) \overset{\text{def}}{=} \\ &\begin{cases} \texttt{true}, & \text{if for all } x, x'{\in}D \text{ such that} \\ & x[\mathcal{Q}_{ID}]{=}x'[\mathcal{Q}_{ID}]{=}x^{\star}[\mathcal{Q}_{ID}], \\ & \text{we have } x[a_{sens}]{=}x'[a_{sens}], \\ \texttt{false}, & \text{otherwise.} \end{cases} \end{aligned} \tag{C.8}$$

– **Deterministic degradation of privacy.** The adversary is considered to be deterministically successful in an IAS attack iff **not** capable of precisely determining the value of the sensitive attribute for the individual of interest **before** the attack, but otherwise capable of doing so **after** the attack is performed:

$$\begin{aligned} &\textit{priv-degrad}^{IAS}_{det}(D, a_{sens}, x^{\star}[\mathcal{Q}_{ID}]) \overset{\text{def}}{=} \\ &(\textit{prior-suc}^{IAS}_{det}(D, a_{sens}){=}\texttt{false}) \quad \text{and} \\ &(\textit{post-suc}^{IAS}_{det}(D, a_{sens}, x^{\star}[\mathcal{Q}_{ID}]){=}\texttt{true})\ . \end{aligned} \tag{C.9}$$

- **Probabilistic degradation of privacy.** The adversary's probabilistic success in an IAS attack does not rely on precisely determining the value of the sensitive attribute for the individual of interest. Rather, we compute the probability of correctly inferring the target's value for the sensitive attribute in the database, i.e. the greater this probability, the more successful the adversary is.

  – ***A priori* probabilistic success.** Before performing the IAS attack, the adversary can rely only on the database itself, i.e. the quasi-identifiers are not used yet. Hence, the best course of action for the adversary is to randomly select a possible value for the sensitive attribute, according to the maximum entropy principle. Since each individual only holds one record in the database, the most probable value would be the most frequent one in the database. Therefore, the *a priori* probabilistic success in an IAS attack on a database D and with respect to the sensitive attribute a_{sens}, given the probability distribution of values for the sensitive attribute $\pi^{a_{sens}|D}$, according to Equation 3.2, is defined as:

$$\begin{aligned} \textit{prior-suc}^{IAS}_{prob}(D, a_{sens}) &\overset{\text{def}}{=} \max_{s \in dom(a_{sens})} \pi_s^{a_{sens}|D} \\ &= \max_{s \in dom(a_{sens})} \frac{|\{x{\in}D \mid x[a_{sens}]{=}s\}|}{|D|} . \end{aligned} \tag{C.10}$$

- ***A posteriori* probabilistic success.** By performing an IAS attack, the adversary uses the auxiliary information as values for the quasi-identifiers of the individual of interest to support the task of inferring the value of the sensitive attribute. An IAS attack is then performed as follows: At first, the adversary filters the database according to the values for the quasi-identifiers of the individual of interest, disregarding records that do not match the query. Then, given the remaining database, the best course of action for the adversary is to randomly select a possible value for the sensitive attribute among those available in the remaining database, according to the maximum entropy principle. Since each individual only holds one record in the database, the most probable value would be the most frequent one in the remaining database.

  Therefore, the *a posteriori* probabilistic success in an IAS attack on a database D with respect to the sensitive attribute a_{sens}, given the probability distribution of values for the sensitive attribute $\pi^{a_{sens}|D}$, according to Equation 3.2, is defined as:

$$\textit{post-suc}^{IAS}_{prob}(D, a_{sens}, x^\star[\mathcal{Q}_{ID}]) \overset{\text{def}}{=}$$

$$\max_{s \in dom(a_{sens})} \frac{|\{x{\in}D \mid x[\mathcal{Q}_{ID}] = x^\star[\mathcal{Q}_{ID}] \text{ and } x[a_{sens}] = s\}|}{|\{x{\in}D \mid x[\mathcal{Q}_{ID}] = x^\star[\mathcal{Q}_{ID}]\}|} = \tag{C.11}$$

$$\max_{s \in dom(a_{sens})} \frac{|\{x{\in}\textit{block}(x^\star[\mathcal{Q}_{ID}]) \mid x[a_{sens}]{=}s\}|}{|\textit{block}(x^\star[\mathcal{Q}_{ID}])|} , \tag{C.12}$$

  where $\textit{block}(x^\star[\mathcal{Q}_{ID}]){=}\{x{\in}D \mid x[\mathcal{Q}_{ID}]{=}x^\star[\mathcal{Q}_{ID}]\}$ is the block defined by the partition on set $\mathcal{Q}_{ID}$ of all individuals that share the same quasi-identifiers with the targeted-individual $x^\star$.

- **Probabilistic degradation of privacy.** The probabilistic degradation of privacy in an IAS attack is defined as the ratio by which the attack increases

the probabilistic success of the adversary:

$$priv\text{-}degrad^{IAS}_{prob}(D, a_{sens}, x^{\star}[\mathcal{Q}_{ID}]) \quad \overset{\text{def}}{=} \quad \frac{post\text{-}suc^{IAS}_{prob}(D, a_{sens}, x^{\star}[\mathcal{Q}_{ID}])}{prior\text{-}suc^{IAS}_{prob}(D, a_{sens})} \ . \tag{C.13}$$

A detailed numeric example of an IAS attack is presented in Section D.1.1, Example 49.

◁

In the following section, we present some experimental results for both re-identification and attribute-inference attacks on single databases for individual-targets on the Educational Censuses released by INEP.

C.1.2 Experimental results for attacks on the INEP databases

In this section, we present an illustrative, quantitative analysis on the privacy risks of re-identification and attribute-inference for individual, predetermined-targets whose information is made public on a single database, as modeled in the previous section. First, we present an Individual-target Re-identification Single database (IRS) attack against a well-known Brazilian political figure, Experiment 41. Then, we present an Individual-target Attribute-inference Single database (IAS) attack against a person known to the author, Experiment 42.

Experiment 41 (Individual-target Re-identification Single database (IRS) attack on the Higher Education Censuses of 2014 and 2018)**.** In an IRS attack, the adversary's goal is to re-identify the targeted-individual, according to Definition 39. Here, we have selected a well-known Brazilian political figure as our target and all the information used as quasi-identifiers were widely available on the Internet. Those attributes include the date and city of birth, in which Higher Education Institutions this individual was enrolled at a given year, and the respective undergraduate course.

- **Probabilistic degradation of privacy**. In a probabilistic scenario, the adversary is not required to precisely determine which record in the database corresponds to the individual of interest in order to be successful. Rather, we compute the adversary's probability of correctly re-identifying the target in the database.

  The Higher Education Census of 2014 database was released containing 10 793 936 records, which were reduced to 9 773 350 after the random selection of only one record for each data holder of multiple records.

Before executing the IRS attack, the adversary cannot rely on the quasi-identifiers in the attempt of target re-identification. Hence, the best course of action for them is to randomly select a record from the database, according to the maximum entropy principle. Therefore, the adversary's *a priori* probabilistic success is, according to Equation C.4:

$$\frac{1}{9\,773\,350} = 1.0232 \cdot 10^{-7} \approx 0.00001\%.$$

However, by preforming the IRS attack, the adversary can rely on the values of the quasi-identifiers to assist the attack. Here, we have used only the date of birth and the Higher Education Institution in which the target was enrolled in 2014, i.e. January 28, 1996, and *Universidade Federal do ABC*, code `4925`, respectively. Both information are widely available on the Internet.

After filtering the database using this auxiliary information, only two records corresponded to the same values of the quasi-identifiers. Since no more information is available at this point, the best course of action for the adversary is to randomly select a record from the database, according to the maximum entropy principle. Therefore, the adversary's *a posteriori* probabilistic success is, according to Equation C.5:

$$\frac{1}{2} = 0.50 = 50\%.$$

Finally, the probabilistic degradation of privacy for the targeted-individual is, according to Equation C.6:

$$\frac{0.50}{1.0232 \cdot 10^{-7}} = 4\,886\,675.$$

Therefore, for the selected political figure and the knowledge of their date of birth and Higher Education Institution in 2014, an adversary increases their chance of re-identifying the targeted-individual by a factor of 4 886 675. [1]

- **Deterministic degradation of privacy**. In a deterministic scenario, the adversary is successful iff they are capable of precisely determining which record in the database corresponds to the individual of interest.

[1] We note that the targeted-individual could be uniquely re-identified by adding only the undergraduate course in which they were enrolled in 2014, i.e. Economic Sciences (*Ciências Econômicas*, in Portuguese), code `123345`, also widely available on the Internet.

For the deterministic IRS attack, the adversary's target is again the same Brazilian political figure, but now on the Higher Education Census of 2018. This database was released containing 12 043 994 records, which were reduced to 10 811 601 after the random selection of only one record for each data holder of multiple records.

Since the database has more than one record, the target cannot be re-identified with absolute certainty *a priori*. According to Equation C.1, the *a priori* deterministic success is equal to `false`.

However, by preforming the IRS attack, the adversary can rely on the values of the quasi-identifiers to assist the attack. Once again, we have used only the date of birth and the Higher Education Institution in which the target was enrolled in 2018, i.e. January 28, 1996, and *Instituto de Direito Público de São Paulo*, code `17672`, respectively. Both information are widely available on the Internet.

After filtering the database using this auxiliary information, only one record corresponded to the same values of the quasi-identifiers. Therefore, the adversary's *a posteriori* deterministic success is, according to Equation C.2, equal to `true`.

Finally, the deterministic degradation of privacy for the targeted-individual is, according to Equation C.3, equal to `true`. Therefore, for the selected political figure and the knowledge of their date of birth and Higher Education Institution in 2018, an adversary can uniquely re-identify the target.

◁

We now present an IAS attack against a person known to the author.

Experiment 42 (Individual-target Attribute-inference Single database (IAS) attack on the School Census of 2019)**.** In an IAS attack, the adversary's goal is to infer the value of an attribute considered to be sensitive for the targeted-individual regardless of being able to re-identify them, according to Definition 40. Here, we have selected a person known to the author as our target and all the information used as quasi-identifiers were easily obtained. Those attributes include the date and city of birth, gender, city of residency, and in which School this individual was enrolled at a given year. Also, we were interested in inferring the value for the attribute `IN_TRANSPORTE_PUBLICO`, described in Table E.2 and considered by us to be sensitive.

- **Probabilistic degradation of privacy**. In a probabilistic scenario, the adversary is not required to precisely determine which value the sensitive attribute takes for the individual of interest in order to be successful. Rather, we compute the adversary's probability of correctly inferring that value.

The School Census of 2019 database was released containing 51 166 723 records, which were reduced to 47 640 822 after the random selection of only one record for each data holder of multiple records.

Before executing the IAS attack, the adversary cannot rely on the quasi-identifiers in the attempt of attribute-inference. Hence, the best course of action for the adversary is to randomly select a value among the possible values for the sensitive attribute in the database, according to the maximum entropy principle. Therefore, the adversary's *a priori* probabilistic success is, according to Equation C.10:

$$\frac{39\,093\,415}{47\,640\,822} = 82.06\%,$$

since the attribute `IN_TRANSPORTE_PUBLICO` can hold three distinct values and the most common among them (`0`, i.e. the student does not use public school transport) occurs for 39 093 415 of the 47 640 822 records in the database.

However, by preforming the IAS attack, the adversary can rely on the values of the quasi-identifiers to assist the attack. Here, we have used only the gender, month, year, and city of birth, i.e. male, November 2001, and *Belo Horizonte*, code `3106200`, respectively. All the information is known to people acquainted to the target.

After filtering the database using this auxiliary information, we have found 1 589 records corresponding to the same values of the quasi-identifiers. Since no more information is available at this point, the best course of action for the adversary is to randomly select a value among the possible values for the sensitive attribute in the filtered database, according to the maximum entropy principle. Therefore, the adversary's *a posteriori* probabilistic success is, according to Equation C.11:

$$\frac{1\,482}{1\,589} = 93.27\%,$$

since the most common value for the attribute `IN_TRANSPORTE_PUBLICO` in the filtered database (again equal to `0`) occurs for 1 482 of the 1 589 records.

Finally, the probabilistic degradation of privacy for the targeted-individual is, according to Equation C.13:

$$\frac{93.27\%}{82.06\%} = 1.1366.$$

Therefore, for the selected target and the knowledge of their gender, month, year, and city of birth, an adversary increases their chance of inferring the value for the attribute `IN_TRANSPORTE_PUBLICO` by a factor of 1.1366, i.e. an increase of 13.66%.

- **Deterministic degradation of privacy**. In a deterministic scenario, the adversary is successful iff they are capable of precisely determining which value the sensitive attribute takes for the individual of interest. For the deterministic IAS attack, we keep both the adversary's target and database from the probabilistic scenario.

  Since the attribute `IN_TRANSPORTE_PUBLICO` can hold three distinct values and all of them occur for at least one record in the database, the target's value for the sensitive attribute could not be inferred with absolute certainty *a priori*. According to Equation C.7, the *a priori* deterministic success is equal to `false`.

  However, by preforming the IAS attack, the adversary can rely on the values of the quasi-identifiers to assist the attack. Once again, we have used the gender, month, year, and city of birth, i.e. male November 2001, and *Belo Horizonte*, code `3106200`, respectively. Furthermore, we have added the target's city of residency, i.e. *Contagem*, code `3118601`. Again, all the information is known to people acquainted to the target.

  After filtering the database using this auxiliary information, we have found 26 records corresponding to the same values of the quasi-identifiers. Since all those records held the same value for the attribute `IN_TRANSPORTE_PUBLICO` (again equal to `0`), the target's value for the sensitive attribute could be inferred with absolute certainty *a posteriori*. Therefore, the adversary's *a posteriori* deterministic success is, according to Equation C.8, equal to `true`.

  Finally, the deterministic degradation of privacy for the targeted-individual is, according to Equation C.9, equal to `true`. Therefore, for the selected target and the knowledge of their gender, month and year of birth, and cities of birth and of residency, an adversary can precisely infer the value for the target's sensitive attribute. [2]

◁

[2] We note that it was not necessary for the adversary to deterministically re-identify the targeted-individual in the database to precisely infer their value for the sensitive attribute. In fact, after executing this IAS attack, the target was still indistinguishable among the 26 records found to match the auxiliary information used as quasi-identifiers.

The results presented above for IRS and IAS attacks show us the knowledge of values for some trivial quasi-identifying attributes are enough to cause a serious degradation of privacy for the selected targets. Naturally, one could argue that each of our targets is just one individual among millions in the databases we have used for those attacks. Unfortunately, as we presented in Section 4.2 for the collective-targets attacks on the School Census of 2018, [3] our selected targets are not exceptions, but rather typical cases of privacy degradation in the databases here analyzed. We also present in Appendix F Section F.1 the equivalent results for collective-targets attacks on the Higher Education Censuses of 2018 and 2019.

C.2 Attacks on longitudinal databases

In this section, we present the theoretical foundation on attacks against longitudinal databases for individual-targets in Section C.2.1, followed by the respective experimental results for attacks on the Educational Censuses released by INEP in Section C.2.2.

C.2.1 Theoretical foundation

For all attacks on single databases, we follow the assumptions introduced in Section 5.1.1. In this section, we formally define both re-identification and attribute-inference attacks on longitudinal databases for individual-targets. Definition 43 accounts for an adversary interested in re-identifying only one individual, while Definition 44 accounts for an adversary interested in inferring the value of an attribute for the targeted-individual. Detailed examples of both attacks are presented in Appendix D Sections D.2.1 and D.2.2, respectively.

Definition 43 (Individual-target Re-identification Longitudinal database (IRL) attack)**.** In an IRL attack, the adversary can completely access a longitudinal collection of databases, i.e. they have knowledge of every record in each of the databases. The adversary can also gather auxiliary information on the values of a set of quasi-identifiers for a predetermined, specific individual of interest from the focal database. Particularly, the adversary can use the auxiliary databases from the longitudinal collection as auxiliary information. The adversary's goal is to re-identify the targeted-individual, i.e. to precisely determine which record in the focal database corresponds to the individual of interest.

[3] The results for the collective-targets attacks on the School Census of 2019 are presented in Appendix E.

The specification of an IRL attack is analogous to that of an IRS attack, Definition 39. Particularly, the two main differences concern the use of the databases from the longitudinal collection in the definitions of success. For the *a priori* success, the adversary relies only on the information available in the focal database D_1, while for the *a posteriori* success, they rely on the auxiliary information obtained by them from sources other than the longitudinal collection $\mathcal{L}_{\mathcal{D}}$ in addition to the aggregated information from all the databases in $\mathcal{L}_{\mathcal{D}}$.

- **Adversary's knowledge.** In an IRL attack, we assume the adversary:

  (i) can completely access a longitudinal collection of (non-empty) databases $\mathcal{L}_{\mathcal{D}}{=}\{D_1, D_2, \ldots, D_k\}$, according to Assumptions **AL0**–**AL3**, and can aggregate the information from this collection into a new aggregated database $agreg(\mathcal{L}_{\mathcal{D}})$ according to Assumption **AL4**;

  (ii) has a predetermined, specific individual of interest $x^{\star}{\in}D_1$ as the re-identification target;

  (iii) given a set of quasi-identifiers $\mathcal{Q}_{ID}{\subseteq}\mathcal{A}_i$, with $1{\leq}i{\leq}k$, that are present in all the databases in $\mathcal{L}_{\mathcal{D}}$, the adversary knows the sub-record $x^{\star}[\mathcal{Q}_{ID}]$ corresponding to the values of the quasi-identifiers of the individual of interest $x^{\star}$, on whom the adversary can gather auxiliary information.

- **Deterministic degradation of privacy.** The adversary is considered to be deterministically successful in an IRL attack iff capable of precisely determining which record in the focal database D_1 corresponds to the individual of interest.

  – ***A priori* deterministic success.** Before performing the IRL attack, the adversary can rely only on the focal database D_1, i.e. neither the auxiliary databases nor the quasi-identifiers are used yet. Hence, the adversary can only precisely determine which record in the focal database corresponds to the individual of interest iff the focal database contains only a single record. Therefore, the *a priori* deterministic success in an IRL attack on a longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}$ is defined as:

$$prior\text{-}suc^{IRL}_{det}(\mathcal{L}_{\mathcal{D}}) \quad \stackrel{\text{def}}{=} \quad \begin{cases} \texttt{true}, & \text{if } |D_1|{=}1, \\ \texttt{false}, & \text{if } |D_1|{\geq}2. \end{cases} \tag{C.14}$$

  – ***A posteriori* deterministic success.** By performing an IRL attack, the adversary uses the auxiliary information as values for the quasi-identifiers of the individual of interest to support the task of re-identification, including the information provided by the auxiliary databases. An IRL attack is then performed as follows: At first, the adversary filters the longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}$ according to the values for the quasi-identifiers of the individual of interest, disregarding those records that do not match the query. Then, given the remaining database, the adversary can only precisely determine which record corresponds to the individual of interest iff the remaining database contains only a single record.

Therefore, the *a posteriori* deterministic success in an IRL attack on a longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}$, given the adversary's knowledge of the values for the target's quasi-identifiers $x^{\star}[\mathcal{Q}_{ID}]$, is defined as:

$$\textit{post-suc}^{\textit{IRL}}_{\textit{det}}(\mathcal{L}_{\mathcal{D}}, x^{\star}[\mathcal{Q}_{ID}]) \overset{\text{def}}{=} \begin{cases} \texttt{true}, & \text{if there is exactly one} \\ & \text{record } x{\in}\textit{agreg}(\mathcal{L}_{\mathcal{D}}) \\ & \text{such that } x[\mathcal{Q}_{ID}]{=}x^{\star}[\mathcal{Q}_{ID}], \\ \texttt{false}, & \text{otherwise.} \end{cases} \tag{C.15}$$

- **Deterministic degradation of privacy.** The adversary is considered to be deterministically successful in an IRL attack iff **not** capable of precisely re-identifying the individual of interest **before** the attack, but otherwise capable of doing so **after** the attack is performed:

$$\textit{priv-degrad}^{\textit{IRL}}_{\textit{det}}(\mathcal{L}_{\mathcal{D}}, x^{\star}[\mathcal{Q}_{ID}]) \overset{\text{def}}{=} \textit{prior-suc}^{\textit{IRL}}_{\textit{det}}(\mathcal{L}_{\mathcal{D}}){=}\texttt{false} \quad \text{and} \tag{C.16}$$

$$\textit{post-suc}^{\textit{IRL}}_{\textit{det}}(\mathcal{L}_{\mathcal{D}}, x^{\star}[\mathcal{Q}_{ID}]){=}\texttt{true}. \tag{C.17}$$

- **Probabilistic degradation of privacy.** The adversary's probabilistic success in an IRL attack does not rely on precisely determining which record in the focal database D_1 corresponds to the individual of interest. Rather, we compute the probability of correctly re-identifying the target in the focal database, i.e. the greater this probability, the more successful the adversary is.

  - ***A priori* probabilistic success.** Before performing the IRL attack, the adversary can rely only on the focal database D_1, i.e. neither the auxiliary databases nor the quasi-identifiers are used yet. Hence, the best course of action for the adversary is to randomly select a record from the focal database, according to the maximum entropy principle. Since each individual only holds one record in each database, the probability of the adversary being successful is the inverse of the focal database size. Therefore, the *a priori* probabilistic success in an IRL attack on a collection of longitudinal databases $\mathcal{L}_{\mathcal{D}}$ is defined as:

$$\textit{prior-suc}^{\textit{IRL}}_{\textit{prob}}(\mathcal{L}_{\mathcal{D}}) \quad \stackrel{\text{def}}{=} \quad \frac{1}{|D_1|} \, . \tag{C.18}$$

- ***A posteriori* probabilistic success.** By performing an IRL attack, the adversary uses the auxiliary information as values for the quasi-identifiers of the individual of interest to support the task of re-identification, including the information provided by the auxiliary databases. An IRL attack is then performed as follows: At first, the adversary filters the collection of longitudinal databases $\mathcal{L}_{\mathcal{D}}$ according to the values for the quasi-identifiers of the individual of interest, disregarding those records that do not match the query. Then, given the remaining database, the best course of action for the adversary is to randomly select a record, according to the maximum entropy principle. Since each individual only holds one record in each database, the probability of the adversary being successful is the inverse of the remaining database size.

  Therefore, the *a posteriori* probabilistic success in an IRL attack on a collection of longitudinal databases $\mathcal{L}_{\mathcal{D}}$, given the adversary's knowledge of the values for the target's quasi-identifiers $x^{\star}[\mathcal{Q}_{ID}]$, is defined as:

$$\textit{post-suc}^{\textit{IRL}}_{\textit{prob}}(\mathcal{L}_{\mathcal{D}}, x^{\star}[\mathcal{Q}_{ID}]) \quad \stackrel{\text{def}}{=} \quad \frac{1}{|\{x{\in}\textit{agreg}(\mathcal{L}_{\mathcal{D}}) \mid x[\mathcal{Q}_{ID}]{=}x^{\star}[\mathcal{Q}_{ID}]\}|} \, . \tag{C.19}$$

- **Probabilistic degradation of privacy.** The probabilistic degradation of privacy in an IRL attack is defined as the ratio by which the attack increases the probabilistic success of the adversary:

$$\textit{priv-degrad}^{\textit{IRL}}_{\textit{prob}}(\mathcal{L}_{\mathcal{D}}, x^{\star}[\mathcal{Q}_{ID}]) \quad \stackrel{\text{def}}{=} \quad \frac{\textit{post-suc}^{\textit{IRL}}_{\textit{prob}}(\mathcal{L}_{\mathcal{D}}, x^{\star}[\mathcal{Q}_{ID}])}{\textit{prior-suc}^{\textit{IRL}}_{\textit{prob}}(\mathcal{L}_{\mathcal{D}})} \tag{C.20}$$

$$= \quad \frac{|D|}{|\textit{block}(x^{\star}[\mathcal{Q}_{ID}])|} \tag{C.21}$$

$$= \frac{|D|}{|\{x{\in}D \mid x[\mathcal{Q}_{ID}] = x^{\star}[\mathcal{Q}_{ID}]\}|} \ . \qquad \text{(C.22)}$$

A detailed numeric example of an IRL attack is presented in Section D.2.1, Example 51.
◁

We now present the attribute-inference attack on longitudinal databases for individual-targets.

Definition 44 (Individual-target Attribute-inference Longitudinal database (IAL) attack)**.** In an IAL attack, the adversary can completely access a longitudinal collection of databases, i.e. has knowledge of every record in each of the databases. The adversary can also gather auxiliary information on the values of a set of quasi-identifiers for a predetermined, specific individual of interest from the focal database. Particularly, the adversary can use the auxiliary databases from the longitudinal collection as auxiliary information. The adversary's goal is to infer the value of an attribute considered to be sensitive for the targeted-individual regardless of being able to re-identify the individual of interest.

The specification of the IAL attack is analogous to that of an IAS attack, Definition 40. The two main differences concern the use of the databases from the longitudinal collection in the definitions of success. For the *a priori* success, we rely only on the information available in the focal database, while for the *a posteriori* success, we rely on the auxiliary information obtained by the adversary from sources other than the longitudinal collection $\mathcal{L}_{\mathcal{D}}$ in addition to the aggregated information from all the databases in $\mathcal{L}_{\mathcal{D}}$.

- **Adversary's knowledge.** In an IAL attack, we assume the adversary:
  (i) can completely access a longitudinal collection of (non-empty) databases $\mathcal{L}_{\mathcal{D}}{=}\{D_1, D_2, \ldots, D_k\}$ according to Assumptions **AL0**–**AL3**, and can aggregate the information from this collection into a new aggregated database $agreg(\mathcal{L}_{\mathcal{D}})$ according to Assumption **AL4**;
  (ii) has a predetermined, specific individual of interest $x^{\star}{\in}D_1$ as the attribute-inference target;
  (iii) given a set of quasi-identifiers $\mathcal{Q}_{ID}{\subseteq}\mathcal{A}_i$, with $1{\leq}i{\leq}k$, that are present in all the databases in $\mathcal{L}_{\mathcal{D}}$, the adversary knows the sub-record $x^{\star}[\mathcal{Q}_{ID}]$ corresponding to the values of the quasi-identifiers of the individual of interest $x^{\star}$, on whom the adversary can gather auxiliary information.

(iv) given an attribute $a_{sens} \in \mathcal{A}_1$ in the focal base D_1 such that $a_{sens} \notin \mathcal{Q}_{ID}$, considered to be sensitive, the adversary's goal is to infer the attribute's value $x^\star[a_{sens}]$ for the targeted-individual $x^\star$.

- **Deterministic degradation of privacy.** The adversary is considered to be deterministically successful in an IAL attack iff capable of precisely determining the value of the sensitive attribute for the individual of interest in the focal database D_1.

  – ***A priori* deterministic success.** Before performing the IAL attack, the adversary can rely only on the focal database D_1, i.e. neither the auxiliary databases nor the quasi-identifiers are used yet. Hence, the adversary can only precisely determine the value of the sensitive attribute for the individual of interest in the focal database iff all the records in the focal database have the same value for the sensitive attribute. Therefore, the *a priori* deterministic success in an IAL attack on a longitudinal collection of databases $\mathcal{L}_\mathcal{D}$ and with respect to the sensitive attribute a_{sens} is defined as:

$$\textit{prior-suc}_{det}^{IAL}(\mathcal{L}_\mathcal{D}, a_{sens}) \overset{\text{def}}{=} \begin{cases} \texttt{true}, & \text{if for all } x, x' \in D_1, \\ & \quad x[a_{sens}] = x'[a_{sens}], \\ \texttt{false}, & \text{otherwise.} \end{cases} \tag{C.23}$$

  – ***A posteriori* deterministic success.** By performing an IAL attack, the adversary uses the auxiliary information as values for the quasi-identifiers of the individual of interest to support the task of inferring the value of the sensitive attribute, including the information provided by the auxiliary databases. An IAL attack is then performed as follows: At first, the adversary filters the longitudinal collection of databases $\mathcal{L}_\mathcal{D}$ according to the values for the quasi-identifiers of the individual of interest, disregarding those records that do not match the query. Then, given the remaining database, the adversary can only precisely determine the value of the sensitive attribute for the individual of interest iff all the records in the remaining database have the same value for the sensitive attribute.

  Therefore, the *a posteriori* deterministic success in an IAL attack on a longitudinal collection of databases $\mathcal{L}_\mathcal{D}$ and with respect to the sensitive attribute a_{sens}, given the adversary's knowledge of the values for the target's quasi-identifiers $x^{\star}[\mathcal{Q}_{ID}]$, is defined as:

$$\begin{aligned}&\textit{post-suc}^{IAL}_{det}(\mathcal{L}_{\mathcal{D}}, a_{sens}, x^{\star}[\mathcal{Q}_{ID}]) \overset{\text{def}}{=} \\ &\begin{cases}\texttt{true}, & \text{if for all } x, x'{\in}\textit{agreg}(\mathcal{L}_{\mathcal{D}}) \text{ such that} \\ & x[\mathcal{Q}_{ID}]{=}x'[\mathcal{Q}_{ID}]{=}x^{\star}[\mathcal{Q}_{ID}], \text{ we have} \\ & x[(a_{sens},1)]{=}x'[(a_{sens},1)], \\ \texttt{false}, & \text{otherwise.}\end{cases}\end{aligned} \tag{C.24}$$

- **Deterministic degradation of privacy.** The adversary is considered to be deterministically successful in an IAL attack iff **not** capable of precisely determining the value of the sensitive attribute for the individual of interest **before** the attack, but otherwise capable of doing so **after** the attack is performed:

$$\begin{aligned}&\textit{priv-degrad}^{IAL}_{det}(\mathcal{L}_{\mathcal{D}}, a_{sens}, x^{\star}[\mathcal{Q}_{ID}]) \overset{\text{def}}{=} \\ &(\textit{prior-suc}^{IAL}_{det}(\mathcal{L}_{\mathcal{D}}, a_{sens}){=}\texttt{false}) \quad \text{and} \\ &(\textit{post-suc}^{IAL}_{det}(\mathcal{L}_{\mathcal{D}}, a_{sens}, x^{\star}[\mathcal{Q}_{ID}]){=}\texttt{true})\ .\end{aligned} \tag{C.25}$$

- **Probabilistic degradation of privacy.** The adversary's probabilistic success in an IAL attack does not rely on precisely determining the value of the sensitive attribute for the individual of interest in the focal database D_1. Rather, we compute the probability of correctly inferring the target's value for the sensitive attribute in the database, i.e. the greater this probability, the more successful the adversary is.

  - ***A priori* probabilistic success.** Before performing the IAL attack, the adversary can rely only on the focal database D_1, i.e. neither the auxiliary databases nor the quasi-identifiers are used yet. Hence, the best course of action for the adversary is to randomly select a possible value for the sensitive attribute, according to the maximum entropy principle. Since each individual only holds one record in each database, the most probable value would be the most frequent one in the focal database. Therefore, the *a priori* probabilistic success in an IAL attack on a collection of longitudinal databases $\mathcal{L}_{\mathcal{D}}$ and with respect to the sensitive attribute a_{sens}, given the probability distribution of values for the sensitive attribute $\pi^{a_{sens}|D_1}$, according to Equation 3.2, is defined as:

$$\begin{aligned} \textit{prior-suc}^{IAL}_{prob}(\mathcal{L}_{\mathcal{D}}, a_{sens}) &\overset{\text{def}}{=} \max_{s \in dom(a_{sens})} \pi_s^{a_{sens}|D_1} \\ &= \max_{s \in dom(a_{sens})} \frac{|\{x{\in}D_1 \mid x[a_{sens}]{=}s\}|}{|D_1|} \ . \end{aligned} \tag{C.26}$$

- ***A posteriori* probabilistic success.** By performing an IAL attack, the adversary uses the auxiliary information as values for the quasi-identifiers of the individual of interest to support the task of inferring the value of the sensitive attribute, including the information provided by the auxiliary databases. An IAL attack is then performed as follows: At first, the adversary filters the collection of longitudinal databases $\mathcal{L}_{\mathcal{D}}$ according to the values for the quasi-identifiers of the individual of interest, disregarding those records that do not match the query. Then, given the remaining database, the best course of action for the adversary is to randomly select a possible value for the sensitive attribute among those available in the remaining database, according to the maximum entropy principle. Since each individual only holds one record in each database, the most probable value would be the most frequent one in the remaining database.

  Therefore, the *a posteriori* probabilistic success in an IAL attack on a longitudinal databases $\mathcal{L}_{\mathcal{D}}$ and with respect to the sensitive attribute a_{sens}, given the probability distribution of values for the sensitive attribute $\pi^{a_{sens}|D_1}$, according to Equation 3.2, and the adversary's knowledge of the values for the target's quasi-identifiers $x^{\star}[\mathcal{Q}_{ID}]$, is defined as:

$$\begin{aligned} &\textit{post-suc}^{IAL}_{prob}(\mathcal{L}_{\mathcal{D}}, a_{sens}, x^{\star}[\mathcal{Q}_{ID}]) \overset{\text{def}}{=} \\ &\max_{s \in dom(a_{sens})} \frac{|\{x{\in}\textit{block}(x^{\star}[\mathcal{Q}_{ID}]) \mid x[(a_{sens},1)]{=}s\}|}{|\textit{block}(x^{\star}[\mathcal{Q}_{ID}])|} \ , \end{aligned} \tag{C.27}$$

  where $\textit{block}(x^{\star}[\mathcal{Q}_{ID}]){=}\{x{\in}\textit{agreg}(\mathcal{L}_{\mathcal{D}}) \mid x[\mathcal{Q}_{ID}]{=}x^{\star}[\mathcal{Q}_{ID}]\}$ is the block defined by the partition on set $\mathcal{Q}_{ID}$ of all individuals in the aggregated database $\textit{agreg}(\mathcal{L}_{\mathcal{D}})$ that share the same quasi-identifiers with the targeted-individual $x^{\star}$.

- **Probabilistic degradation of privacy.** The probabilistic degradation of privacy in an IAL attack is defined as the ratio by which the attack increases

the probabilistic success of the adversary:

$$\textit{priv-degrad}^{IAL}_{prob}(\mathcal{L}_{\mathcal{D}}, a_{sens}, x^{\star}[\mathcal{Q}_{ID}]) \quad \stackrel{\text{def}}{=} \quad \frac{\textit{post-suc}^{IAL}_{prob}(\mathcal{L}_{\mathcal{D}}, a_{sens}, x^{\star}[\mathcal{Q}_{ID}])}{\textit{prior-suc}^{IAL}_{prob}(\mathcal{L}_{\mathcal{D}}, a_{sens})} \, . \tag{C.28}$$

A detailed numeric example of an IAL attack is presented in Appendix D.2.2, Example 53. ◁

In the following section, we present some experimental results for both re-identification and attribute-inference attacks on longitudinal databases for individual-targets on the Educational Censuses released by INEP.

C.2.2 Experimental results for attacks on the INEP databases

In this section, we present an illustrative, quantitative analysis on the privacy risks of re-identification and attribute-inference for individual, predetermined-targets whose information is made public on longitudinal databases, as modeled in the previous section. First, we present a deterministic Individual-target Re-identification Longitudinal databases (IRL) attack against a well-known Brazilian political figure and a probabilistic IRL attack against the author, Experiment 45. Then, we present a deterministic Individual-target Attribute-inference Longitudinal databases (IAL) attack against a person known to the author and a probabilistic IAL attack against the author, Experiment 46.

Experiment 45 (Individual-target Re-identification Longitudinal databases (IRL) attack on the Higher Education Censuses)**.** In an IRL attack, the adversary's goal is to re-identify the targeted-individual, according to Definition 43. Particularly, the adversary can use the auxiliary databases from the longitudinal collection as additional auxiliary information.

First, for the deterministic IRL attack, we have selected a well-known Brazilian political figure as our target and all the information used as quasi-identifiers are widely available on the Internet. Then, for the probabilistic IRL attack, we have selected this thesis author and all the information used as quasi-identifiers are also easily obtainable by third-parties.

We have performed both attacks using the database released for the Higher Education Census of 2014 as the focal one. This database was released containing 10 793 936 records, which were reduced to 9 773 350 after the random selection of only one record for each data holder of multiple records. Also, we have access to the databases released for the Higher Education Census from 2009 to 2019, all containing microdata, i.e. data at the record level.

- **Deterministic degradation of privacy**. In a deterministic scenario, the adversary is successful iff they are capable of precisely determining which record in the focal database corresponds to the individual of interest.

  For the deterministic IRL attack, the adversary's target is a well-known Brazilian political figure, who we try to re-identify in the focal database for the Higher Education Census of 2014. Here, we also use the database for the Higher Education Census of 2016 as auxiliary information.

  Since the focal database has more than one record, the target cannot be re-identified with absolute certainty *a priori*. According to Equation C.14, the *a priori* deterministic success is equal to `false`.

  However, by preforming the IRL attack, the adversary can rely on the values of the quasi-identifiers, not only on the focal database, but also on the auxiliary one, to assist the attack. Here, we have used only the Higher Education Institution in which the target was enrolled in 2014 and in 2016, i.e. *Universidade Federal do ABC*, code `4925`, and *Instituto de Direito Público de São Paulo*, code `17672`, respectively. Both information are widely available on the Internet.

  After filtering the aggregated database using this auxiliary information, only one record corresponded to the same values of the quasi-identifiers. Therefore, the adversary's *a posteriori* deterministic success is, according to Equation C.15, equal to `true`.

  Finally, the deterministic degradation of privacy for the targeted-individual is, according to Equation C.16, equal to `true`. Therefore, for the selected political figure and the knowledge of their Higher Education Institution in 2014 and in 2016, an adversary can uniquely re-identify the target.

- **Probabilistic degradation of privacy**. In a probabilistic scenario, the adversary is not required to precisely determine which record in the focal database corresponds to the individual of interest in order to be successful. Rather, we compute the adversary's probability of correctly re-identifying the target in that database.

  For the probabilistic IRL attack, the adversary's target is this thesis author, who we try to re-identify in the focal database for the Higher Education Census of

2014. Here, we also use the database for the Higher Education Census of 2011 as auxiliary information.

Before executing the IRL attack, the adversary cannot rely on the quasi-identifiers, nor on the auxiliary database, in the attempt of target re-identification. Hence, the best course of action for the adversary is to randomly select a record from the database, according to the maximum entropy principle. Therefore, the adversary's *a priori* probabilistic success is, according to Equation C.18:

$$\frac{1}{9\,773\,350} = 1.0232 \cdot 10^{-7} \approx 0.00001\%.$$

However, by preforming the IRL attack, the adversary can rely on the values of the quasi-identifiers, not only on the focal database, but also on the auxiliary one, to assist the attack. Here, we have only used the Higher Education Institution course in which the author was enrolled in 2014 and in 2011, i.e. Physics (variable `NO_CURSO` equal to `FISICA`) and Medicine (variable `NO_CURSO` equal to `MEDICINA`), respectively, where the variable `NO_CURSO` indicates the Higher Education Institution course name in Portuguese.

After filtering the aggregated database using this auxiliary information, only seven records corresponded to the same values of the quasi-identifiers. Since no more information is available at this point, the best course of action for the adversary is to randomly select a record from the database, according to the maximum entropy principle. Therefore, the adversary's *a posteriori* probabilistic success is, according to Equation C.19:

$$\frac{1}{7} = 14.29\%.$$

Finally, the probabilistic degradation of privacy for the targeted-individual is, according to Equation C.20:

$$\frac{1/7}{1/9\,773\,350} \approx 1\,396\,192.$$

Therefore, for this thesis author and the knowledge of his Higher Education Institution course in 2014 and 2011, an adversary increases their chance of re-identifying the targeted-individual by a factor of 1 396 192. [4]

[4] We note that the targeted-individual could be uniquely re-identified by specifying only one of the

◁

We now present a deterministic IAL against a person known to the author and a probabilistic IAL attack against the author.

Experiment 46 (Individual-target Attribute-inference Longitudinal databases (IAL) attack on the School Censuses)**.** In an IAL attack, the adversary's goal is to infer the value of an attribute considered to be sensitive for the targeted-individual regardless of being able to re-identify them, according to Definition 44. Particularly, the adversary can use the auxiliary databases from the longitudinal collection as additional auxiliary information.

First, for the deterministic IAL attack, we have selected a person known to the author as our target and all the information used as quasi-identifiers are easily obtained. This attack was performed using the database released for the School Census of 2016 as the focal one, which was released containing 52 356 383 records, but reduced to 48 561 221 after the random selection of only one record for each data holder of multiple records. Also, we have access to the databases released for the School Census from 2007 to 2019, all containing microdata, i.e. data at the record level. Furthermore, we were interested in inferring the value for the attribute `IN_NECESSIDADE_ESPECIAL`, i.e. whether or not the student possesses a disability or global developmental disorder.

Then, for the probabilistic IAL attack, we have selected this thesis author as our target and all the information used as quasi-identifiers are also easily obtainable by third-parties. This attack was performed using the database released for the Higher Education Census of 2014 as the focal one, which was released containing 10 793 936 records, but reduced to 9 773 350 after the random selection of only one record for each data holder of multiple records. Also, we have access to the databases released for the Higher Education Census from 2009 to 2019. Furthermore, we were interested in inferring the value for the attribute `CO_MUNICIPIO_NASCIMENTO`, i.e. the student's city of birth code.

- **Deterministic degradation of privacy**. In a deterministic scenario, the adversary is successful iff they are capable of precisely determining which value the sensitive attribute takes for the individual of interest.

  Since the attribute `IN_NECESSIDADE_ESPECIAL` can hold two distinct values and both of them occur for at least one record in the focal database, i.e. the School

Higher Education Institutions in which he was enrolled in 2014 or 2011.

Census of 2016, the target's value for the sensitive attribute could not be inferred with absolute certainty *a priori*. According to Equation C.23, the *a priori* deterministic success is equal to `false`.

However, by preforming the IAL attack, the adversary can rely on the values of the quasi-identifiers, not only on the focal database, but also on the auxiliary one, to assist the attack. Here, we have used only the School in which the target was enrolled in 2016 and in 2017, i.e. *Colégio São Judas Tadeu*, code `31014150`, and *IEC - Unidade CENTEC*, code `31014001`. All the information is known to people acquainted to the target.

After filtering the aggregated database using this auxiliary information, we have found 14 records corresponding to the same values of the quasi-identifiers. Since all those records hold the same value for the attribute `IN_NECESSIDADE_ESPECIAL` (0, i.e. the student does not possess a disability or global developmental disorder), the target's value for the sensitive attribute could be inferred with absolute certainty *a posteriori*. Therefore, the adversary's *a posteriori* deterministic success is, according to Equation C.24, equal to `true`.

Finally, the deterministic degradation of privacy for the targeted-individual is, according to Equation C.25, equal to `true`. Therefore, for the selected target and the knowledge of their School in 2016 and 2017, an adversary can precisely infer the value for the target's sensitive attribute. [5]

- **Probabilistic degradation of privacy**. In a probabilistic scenario, the adversary is not required to precisely determine which value the sensitive attribute takes for the individual of interest in order to be successful. Rather, we compute the adversary's probability of correctly inferring that value.

  Before executing the IAL attack, the adversary cannot rely on the quasi-identifiers, nor on the auxiliary database, in the attempt of attribute-inference. Hence, the best course of action for the adversary is to randomly select a value among the possible values for the sensitive attribute in the focal database, i.e. the Higher Education Census of 2014, according to the maximum entropy principle. Therefore, the adversary's *a priori* probabilistic success is, according to

[5] We note that it was not necessary for the adversary to deterministically re-identify the targeted-individual in the focal database to precisely infer their value for the sensitive attribute. In fact, after executing this IAL attack, the target was still indistinguishable among the 14 records found to match the auxiliary information used as quasi-identifiers.

Equation C.26:

$$\frac{556\,398}{9\,773\,350} = 5.69\%,$$

since the attribute `CO_MUNICIPIO_NASCIMENTO` can hold 5 561 distinct values in the focal database and the most common among them (code `3550308`, i.e. the city of *São Paulo*) occurs for 556 398 of the 9 773 350 records in the database.

However, by preforming the IAL attack, the adversary can rely on the values of the quasi-identifiers, not only on the focal database, but also on the auxiliary one, to assist the attack. Here, we have used only the Higher Education Institution course in which the author was enrolled in 2014 and in 2011, i.e. Physics (variable `NO_CURSO` equal to `FISICA`) and Medicine (variable `NO_CURSO` equal to `MEDICINA`), respectively, where the variable `NO_CURSO` indicates the Higher Education Institution course name in Portuguese.

After filtering the aggregated database using this auxiliary information, only seven records corresponded to the same values of the quasi-identifiers. Since no more information is available at this point, the best course of action for the adversary is to randomly select a value among the possible values for the sensitive attribute in the filtered database, according to the maximum entropy principle. Therefore, the adversary's *a posteriori* probabilistic success is, according to Equation C.27:

$$\frac{2}{7} = 28.57\%,$$

since the most common value for the attribute `CO_MUNICIPIO_NASCIMENTO` in the filtered database (code `3304557`, i.e. the city of *Rio de Janeiro*) occurs for 2 of the 7 records.

Finally, the probabilistic degradation of privacy for the targeted-individual is, according to Equation C.28:

$$\frac{0.2857}{0.0569} = 5.0211.$$

Therefore, for this thesis author and the knowledge of his Higher Education Institution course in 2014 and 2011, an adversary increases their chance of inferring the value for the attribute `CO_MUNICIPIO_NASCIMENTO` by a factor of 5.0211. [6]

[6] Once again, we note that it was not necessary for the adversary to deterministically re-identify

◁

Similarly to IRS and IAS attacks presented in Section C.1, for IRL and IAL attacks we also notice that the knowledge of values for some trivial quasi-identifying attributes are enough to cause a serious degradation of privacy for the selected targets. Naturally, one could argue that each of our targets is just one individual among millions in the databases we have used for those attacks. Unfortunately, as we presented in Section 5.2 for the collective-targets attacks on the School Census of 2018, [7] our selected targets are not exceptions, but rather typical cases of privacy degradation in the databases here analyzed. We also present in Appendix F Section F.2 the equivalent results for collective-targets attacks on the Higher Education Censuses.

the targeted-individual in the focal database to precisely infer their value for the sensitive attribute. In fact, after executing this IAL attack, the target was still indistinguishable among the 7 records found to match the auxiliary information used as quasi-identifiers.

[7] The results for the collective-targets attacks on the School Census of 2019 are presented in Appendix E.

Appendix D

Examples of attacks

In this appendix, we present detailed numeric examples of the attacks defined in this thesis for both individual and collective-targets, single and longitudinal databases, and for re-identification and attribute-inference attacks.

D.1 Attacks on single databases

In this section, we present detailed examples of each attack performed on single databases, as discussed in Chapter 4 and Appendix C.1. The leading example upon which we perform the attacks was presented in Section 3.2, Example 12.

In Section D.1.1, we show examples of both individual and collective re-identification attacks, and in Section D.1.2, we show examples of both individual and collective attribute-inference attacks.

D.1.1 Examples of re-identification attacks

In this section, we present both examples of re-identification attacks on single databases. We begin with Example 47, which accounts for an adversary interested in re-identifying only predetermined targeted-individuals, according to Definition 39. Then, we present Example 48, which accounts for the general case, i.e. an adversary interested in re-identifying as many individuals as possible, according to Definition 16.

Example 47 (Execution of an Individual-target Re-identification Single database (IRS) attack)**.** Consider the database D on the set of attributes $\mathcal{A}$ from Example 12 and the assumptions made in Section 4.1.1. In this example, the adversary wants to separately re-identify the records of two predetermined targeted-individuals: Alex and

Bia. Furthermore, the adversary can obtain the following auxiliary information to be used as quasi-identifiers:

- Alex is a 60 years old man, i.e. his quasi-identifiers are $\mathcal{Q}'_{ID} = \{gender, age\}$ and the adversary knows the values $x^{\star}_{Alex}[\{gender, age\}] = \langle \mathtt{M}, 60 \rangle$;
- Bia is a woman whose *occupation* category is given by `1`, i.e. her quasi-identifiers are $\mathcal{Q}''_{ID} = \{gender, occupation\}$, and the adversary knows the values $x^{\star}_{Bia}[\{gender, occupation\}] = \langle \mathtt{F}, \mathtt{1} \rangle$.

We now compute the metrics from Definition 39.

- **Deterministic degradation of privacy.** Since the database D has more than one record, neither Alex nor Bia can be re-identified with absolute certainty *a priori*, i.e. without the use of auxiliary information. Formally, Equation C.1 for the *a priori* success returns a `false` value for both individuals.

  However, by performing the IRS attack against Alex, the adversary can rely on the sub-record $x^{\star}_{Alex}[\{gender, age\}] = \langle \mathtt{M}, 60 \rangle$ as auxiliary information, corresponding to the first target's quasi-identifiers. Since there is only one record $x \in D$ such that $x[\{gender, age\}] = x^{\star}_{Alex}[\{gender, age\}]$, the adversary can precisely determine that the record with *id* 9 is held by Alex. Hence, the adversary can re-identify the target with absolute certainty and achieves *a posteriori*, deterministic success. Formally, Equation C.2 for the *a posteriori* success returns a `true` value for the IRS attack against Alex.

  Therefore, the deterministic re-identification attack against Alex succeeds, i.e. Equation C.3 for the deterministic degradation of privacy returns a `true` value.

  Similarly, by performing the IRS attack against Bia, the adversary can rely on the sub-record $x^{\star}_{Bia}[\{gender, occupation\}] = \langle \mathtt{F}, \mathtt{1} \rangle$ as auxiliary information, corresponding to the second target's quasi-identifiers. In this case, however, the adversary cannot re-identify Bia with absolute certainty *a posteriori*, since there are two records $x \in D$ such that $x[\{gender, occupation\}] = x^{\star}_{Bia}[\{gender, occupation\}]$, the records with *id* 1 and 2. Formally, Equation C.2 for the *a posteriori* success returns a `false` value for the IRS attack against Bia.

  Thus, the deterministic re-identification attack against Bia does not succeed, i.e. Equation C.3 for the deterministic degradation of privacy returns a `false` value.

- **Probabilistic degradation of privacy.** Although Bia did not suffer a deterministic degradation of privacy in the adversary's IRS attack based on the sub-record $x^{\star}_{Bia}[\{gender, occupation\}] = \langle \mathtt{F}, \mathtt{1} \rangle$ as auxiliary information, the adversary did increase their knowledge on Bia. We can quantify this gain of information probabilistically as follows.

  Whenever the adversary has more than one record to choose from, their best course of action is to randomly select one of them, according to the maximum entropy principle. Since there are ten records in the database D, the adversary has a 10% chance of randomly selecting Bia's record in the database before performing the attack, i.e. without relying on any auxiliary information. Formally, Equation C.4 for the *a priori*, probabilistic success equals:

$$\textit{prior-suc}^{IRS}_{prob}(D) = \frac{1}{|D|} = \frac{1}{10} = 0.1 \ .$$

  However, by performing the IRS attack against Bia, the adversary can rely on the sub-record $x^{\star}_{Bia}[\{gender, occupation\}] = \langle \mathtt{F}, \mathtt{1} \rangle$ as auxiliary information, which decreases the number of possible records from ten to just two of them. Hence, the adversary has a 50% chance of randomly selecting Bia's record in the database D after performing the attack. Formally, Equation C.5 for the *a posteriori*, probabilistic success equals:

$$\textit{post-suc}^{IRS}_{prob}(D, x^{\star}_{Bia}[\{gender, occupation\}]) = \\ \frac{1}{|\{x{\in}D \mid x[\{gender, occupation\}]{=}x^{\star}_{Bia}[\{gender, occupation\}]\}|} = 0.5 \ .$$

  Therefore, the probabilistic degradation of privacy in a re-identification attack against Bia, according to Equation C.6, is the ratio ${}^{50\%}/_{10\%} = 5$, i.e. the adversary's chance of re-identifying Bia increases by a factor of five after performing the IRS attack.

◁

We now present the collective-target example of re-identification attacks on single databases. This allows us to quantify the overall privacy degradation in a given database, i.e. how susceptible to privacy degradation are those individuals who hold a record in the database.

Example 48 (Execution of a Collective-target Re-identification Single database (CRS) attack)**.** Consider the database D on the set of attributes $\mathcal{A}$ from Example 12 and the assumptions made in Section 4.1.1. In this example, the adversary wants to re-identify as many records as possible from those present in the database. Furthermore, the adversary can obtain auxiliary information on the values of the attribute in $\mathcal{Q}'_{ID} = \{age\}$ for every individual, which are used as quasi-identifiers.

We now compute the metrics from Definition 16.

- **Deterministic degradation of privacy.** Since the database D has more than one record, no individual can be re-identified with absolute certainty *a priori*, i.e. without the use of auxiliary information, and the fraction of individuals that can be re-identified by the adversary *a priori* is of 0%. Formally, Equation 4.1 for the *a priori* success equals:

  $$\textit{prior-suc}^{CRS}_{det}(D) = 0 \ .$$

  However, by performing the CRS attack, the adversary can rely on the values of the attribute in $\mathcal{Q}'_{ID} = \{age\}$ for every individual in the database D. Since there is only one record with a unique value for the attribute *age*, i.e. the record with *id* 10 whose *age* is 60, only this record can be re-identified with absolute certainty by the adversary *a posteriori*. Hence, the adversary can re-identify 10% of the records in the database given the knowledge of the attribute *age*'s value for every individual. Formally, Equation 4.2 for the *a posteriori*, deterministic success equals:

  $$\begin{aligned} \textit{post-suc}^{CRS}_{det}(D, \{x[\mathcal{Q}'_{ID}]\}_{x \in D}) &= \frac{\left|\{\alpha{\in}D/_{\sim_{\mathcal{Q}'_{ID}}} \mid |\alpha|{=}1\}\right|}{|D|} \\ &= \frac{|\{\langle age{=}60\rangle\}|}{|D|} \\ &= \frac{1}{10} = 0.1 \ . \end{aligned}$$

  Therefore, the deterministic degradation of privacy in a collective re-identification attack, according to Equation 4.3, is:

  $$\textit{priv-degrad}^{CRS}_{det}(D, \{x[\mathcal{Q}'_{ID}]\}_{x \in D}) = \\ \textit{post-suc}^{CRS}_{det}(D, \{x[\mathcal{Q}'_{ID}]\}_{x \in D}) - \textit{prior-suc}^{CRS}_{det}(D) = 0.1 - 0 = 0.1 \ ,$$

i.e. the adversary's absolute chance of deterministically re-identifying an individual in the database D increases by 10% after performing the CRS attack.

- **Probabilistic degradation of privacy.** Whenever the adversary has more than one record to choose from, their best course of action is to randomly select one of them, according to the maximum entropy principle. Since there are ten records in the database D, the adversary has a 10% chance of randomly re-identifying a record in the database before performing the attack, i.e. without relying on any auxiliary information. Formally, Equation 4.4 for the *a priori*, probabilistic success equals:

$$\textit{prior-suc}^{\textit{CRS}}_{\textit{prob}}(D) = \frac{1}{|D|} = \frac{1}{10} = 0.1 \ .$$

However, by performing the CRS attack, the adversary can rely on the values of the attribute in $\mathcal{Q}'_{ID} = \{\textit{age}\}$ for every individual in the database D. In this case, the database can be partitioned in three blocks, each containing all the records that share the same value for the attribute *age*, i.e. one block with five records whose *age* equal 25, one block with four records whose *age* equal 49, and one block with a single record whose *age* is 60.

Based on how many records there are in each block, the adversary has a distinct probability of success, i.e. 20%, 25%, and 100%, respectively. Similarly, a randomly selected record has a distinct chance of belonging to each one of those blocks, i.e. 50%, 40%, or 10%, respectively. Hence, the adversary's success can be computed as $50\% \cdot 20\% + 40\% \cdot 25\% + 10\% \cdot 100\% = 30\%$, i.e. the adversary has a 30% chance of re-identifying a randomly selected record in the database D after performing the attack. Formally, Equation 4.5 for the *a posteriori*, probabilistic success equals:

$$\begin{aligned}\textit{post-suc}^{\textit{CRS}}_{\textit{prob}}(D, \{x[\mathcal{Q}'_{ID}]\}_{x\in D}) &= \frac{\left|D/\!\sim_{\mathcal{Q}'_{ID}}\right|}{|D|} \\ &= \frac{|\{\langle \textit{age}{=}25\rangle, \langle \textit{age}{=}49\rangle, \langle \textit{age}{=}60\rangle\}|}{|D|} \\ &= \frac{3}{10} = 0.3 \ .\end{aligned}$$

Therefore, the probabilistic degradation of privacy in a collective re-identification attack, according to Equation 4.6, is:

$$\textit{priv-degrad}^{CRS}_{prob}(D, \{x[\mathcal{Q}'_{ID}]\}_{x \in D}) = \frac{\textit{post-suc}^{CRS}_{prob}(D, \{x[\mathcal{Q}'_{ID}]\}_{x \in D})}{\textit{prior-suc}^{CRS}_{prob}(D)} = \frac{0.3}{0.1} = 3 \ ,$$

i.e. the adversary's chance of re-identifying a randomly selected record in the database D increases by a factor of three after performing the CRS attack, from 10% to 30%.

$\triangleleft$

D.1.2 Examples of attribute-inference attacks

In this section, we present both examples of attribute-inference attacks on single databases. We begin with Example 49, which accounts for an adversary interested in inferring the value of an attribute for only predetermined targeted-individuals, according to Definition 40. Then, we present Example 50, which accounts for the general case, i.e. an adversary interested in inferring the value of an attribute for as many individuals as possible, according to Definition 17.

Example 49 (Execution of an Individual-target Attribute-inference Single database (IAS) attack)**.** Consider the database D on the set of attributes $\mathcal{A}$ from Example 12 and the assumptions made in Section 4.1.1. In this example, the adversary wants to separately infer the value of the attribute *illness*, considered here to be sensitive, for two predetermined targeted-individuals: Caio and Dora. Furthermore, the adversary can obtain the following auxiliary information to be used as quasi-identifiers:

- Caio is a man whose *occupation* category is given by 4, i.e. his quasi-identifiers are $\mathcal{Q}'_{ID} = \{\textit{gender}, \textit{occupation}\}$, and the adversary knows the values $x^{\star}_{Caio}[\{\textit{gender}, \textit{occupation}\}] = \langle \texttt{M}, 4 \rangle$;
- Dora is a 49 years old woman, i.e. her quasi-identifiers are $\mathcal{Q}''_{ID} = \{\textit{gender}, \textit{age}\}$, and the adversary knows the values $x^{\star}_{Dora}[\{\textit{gender}, \textit{age}\}] = \langle \texttt{F}, 49 \rangle$.

We now compute the metrics from Definition 40.

- **Deterministic degradation of privacy.** The attribute *illness* can assume two distinct values in the database D: five records hold value `yes` and the other five hold value `no`. Hence, neither Caio nor Dora can have the value of their sensitive attribute inferred by the attacker with absolute certainty *a priori*, i.e. without the use of auxiliary information. Formally, Equation C.7 for the *a priori* success returns a `false` value for both individuals.

However, by performing the IAS attack against Caio, the adversary can rely on the sub-record $x^{\star}_{Caio}[\{gender, occupation\}] = \langle \texttt{M}, 4 \rangle$ as auxiliary information, corresponding to the first target's quasi-identifiers. Even though two records hold the same values as Caio for the attributes *gender* and *occupation*, i.e. the records with *id* 9 and 10, it is still possible for the adversary to infer the value of the sensitive attribute with absolute certainty. Because both records hold the same value for the attribute *illness*, i.e. `no`, the adversary does achieve *a posteriori*, deterministic success. Formally, Equation C.8 for the *a posteriori* success returns a `true` value for the IAS attack against Caio.

Therefore, the deterministic attribute-inference attack against Caio succeeds, i.e. Equation C.9 for the deterministic degradation of privacy returns a `true` value.

Similarly, by performing the IAS attack against Dora, the adversary can rely on the sub-record $x^{\star}_{Dora}[\{gender, age\}] = \langle \texttt{F}, 49 \rangle$ as auxiliary information, corresponding to the second target's quasi-identifiers. In this case, however, the adversary cannot infer the value of the sensitive attribute with absolute certainty *a posteriori*, since not all the records that hold the same values for the attributes *gender* and *age*, i.e. the records with *id* 6, 7, and 8, hold the same value for the attribute *illness*. While the records with *id* 6 and 7 hold the value `yes`, the record with *id* 8 holds the value `no`. Formally, Equation C.8 for the *a posteriori* success returns a `false` value for the IAS attack against Dora.

Thus, the deterministic attribute-inference attack against Dora does not succeed, i.e. Equation C.9 for the deterministic degradation of privacy returns a `false` value.

- **Probabilistic degradation of privacy.** Although Dora did not suffer a deterministic degradation of privacy in the adversary's IAS attack based on the sub-record $x^{\star}_{Dora}[\{gender, age\}] = \langle \texttt{F}, 49 \rangle$ as auxiliary information, the adversary did increase their knowledge on Dora. We can quantify this gain of information probabilistically as follows.

  Since half of the records in the database D holds value `yes` for the sensitive attribute *illness*, while the other half holds value `no`, by applying the maximum entropy principle stated in Section 2.1, the adversary's best course of action is to randomly select one of the values. Hence, the adversary has a 50% chance of correctly inferring Dora's value for the sensitive attribute before performing the attack, i.e. without relying on any auxiliary information. Formally, Equation C.10 for the *a priori*, probabilistic success equals:

$$\begin{aligned}
\textit{prior-suc}^{\textit{IAS}}_{\textit{prob}}(D, \textit{illness}) &= \max_{s \in \textit{dom}(\textit{illness})} \frac{|\{x{\in}D \mid x[\textit{illness}]{=}s\}|}{|D|} \\
&= \frac{|\{x{\in}D \mid x[\textit{illness}]{=}\texttt{yes}\}|}{|D|} \\
&= \frac{|\{x{\in}D \mid x[\textit{illness}]{=}\texttt{no}\}|}{|D|} \\
&= \frac{5}{10} = 0.50 \ .
\end{aligned}$$

However, by performing the IAS attack against Dora, the adversary can rely on the sub-record $x^{\star}_{\textit{Dora}}[\{\textit{gender}, \textit{age}\}] = \langle \texttt{F}, 49\rangle$ as auxiliary information, which decreases the number of records from ten to just three of them. Since two of the records hold the value `yes` for the attribute *illness*, i.e. those with *id* 6 and 7, while one of them holds the value `no`, i.e. the record with *id* 8, the adversary has a 67% chance of correctly inferring Dora's sensitive attribute value after performing the attack. Formally, Equation C.11 for the *a posteriori*, probabilistic success equals:

$$\begin{aligned}
&\textit{post-suc}^{\textit{IAS}}_{\textit{prob}}(D, \textit{illness}, x^{\star}_{\textit{Dora}}[\{\textit{gender}, \textit{age}\}]) = \\
\max_{s \in \textit{dom}(\textit{illness})} &\frac{|\{x{\in}\textit{block}(x^{\star}_{\textit{Dora}}[\{\textit{gender}, \textit{age}\}]) \mid x[\textit{illness}]{=}s\}|}{|\textit{block}(x^{\star}_{\textit{Dora}}[\{\textit{gender}, \textit{age}\}])|} = \\
&\frac{|\{x{\in}\textit{block}(x^{\star}_{\textit{Dora}}[\{\textit{gender}, \textit{age}\}]) \mid x[\textit{illness}]{=}\texttt{yes}\}|}{|\textit{block}(x^{\star}_{\textit{Dora}}[\{\textit{gender}, \textit{age}\}])|} = \frac{2}{3} = 0.67 \ ,
\end{aligned}$$

where $\textit{block}(x^{\star}_{\textit{Dora}}[\{\textit{gender}, \textit{age}\}]){=}\{x{\in}D \mid x[\{\textit{gender}, \textit{age}\}]{=}x^{\star}_{\textit{Dora}}[\{\textit{gender}, \textit{age}\}]\}$ is the block of all individuals in the database D that hold the same quasi-identifier values as the targeted-individual, $x^{\star}_{\textit{Dora}}$.

Therefore, the probabilistic degradation of privacy in an attribute-inference attack against Dora, according to Equation C.13, is the ratio $^{67\%}/_{50\%} = 1.34$, i.e. the adversary's chance of correctly inferring the value of Dora's sensitive attribute increases by approximately 34% after performing the IAS attack.

◁

We now present the collective-target example of attribute-inference attacks on single databases. This allows us to quantify the overall privacy degradation in a given database, i.e. how susceptible to privacy degradation are those individuals who hold a record in the database.

Example 50 (Execution of a Collective-target Attribute-inference Single database (CAS) attack)**.** Consider the database D on the set of attributes $\mathcal{A}$ from Example 12 and the assumptions made in Section 4.1.1. In this example, the adversary wants to infer the value of the attribute *illness*, considered to be sensitive, for as many records as possible from those present in the database. Furthermore, consider the adversary can obtain auxiliary information on the values of the attributes in $\mathcal{Q}'_{ID} = \{age\}$ for every individual, which are used as quasi-identifiers.

We now compute the metrics from Definition 17.

- **Deterministic degradation of privacy.** The attribute *illness* can assume two distinct values in the database D: five records hold value `yes` and the other five hold value `no`. Hence, no individual can have the value of their sensitive attribute inferred by the attacker with absolute certainty *a priori*, i.e. without the use of auxiliary information. Formally, Equation 4.7 for the *a priori* success equals:

  $$\textit{prior-suc}^{CAS}_{det}(D, \textit{illness}) = 0 \ .$$

  However, by performing the CAS attack, the adversary can rely on the values of the attribute in $\mathcal{Q}'_{ID} = \{age\}$ for every individual in the database D. In this case, the database can be partitioned in three blocks, each containing all the records that share the same value for the attribute *age*, i.e. one block with five records whose *age* equal 25, one block with four records whose *age* equal 49, and one block with a single record whose *age* is 60. But for only one of those three blocks all records hold the same value for the attribute *illness*, i.e. the block with a single record, with *id* 10 and whose value for *illness* equals `no`. Hence, the adversary can correctly infer the value of the sensitive attribute for only one in ten records, i.e. the adversary has a 10% chance of success. Formally, Equation 4.8 for the *a posteriori*, deterministic success equals:

  $$\begin{aligned} &\textit{post-suc}^{CAS}_{det}(D, \textit{illness}, \{x[\mathcal{Q}'_{ID}]\}_{x \in D}) = \\ &\frac{\sum_{\alpha \in D/\sim_{\mathcal{Q}'_{ID}}} |\alpha| \cdot \textit{unique-sens}(\alpha, \textit{illness})}{|D|} = \\ &\qquad \frac{5 \cdot 0 + 4 \cdot 0 + 1 \cdot 1}{10} = \frac{1}{10} = 0.1 \ , \end{aligned}$$

where function *unique-sens* is defined by Equation 4.9. Here, $\textit{unique-sens}(\alpha, \textit{illness}) = 1$ only for the block α whose value for the attribute *age* equals 60 since only in this block all records hold the same value for the sensitive attribute *illness*.

Therefore, the deterministic degradation of privacy in a collective attribute-inference attack, according to Equation 4.10, is:

$$\begin{aligned} \textit{priv-degrad}^{CAS}_{det}(D, \textit{illness}, \{x[\mathcal{Q}'_{ID}]\}_{x\in D}) &= \\ \textit{post-suc}^{CAS}_{det}(D, \textit{illness}, \{x[\mathcal{Q}'_{ID}]\}_{x\in D}) - \textit{prior-suc}^{CAS}_{det}(D, \textit{illness}) &= \\ &= 0.1 - 0 = 0.1 \ , \end{aligned}$$

i.e. the adversary's chance of deterministically inferring the value of the sensitive attribute *illness* for an individual in the database D increases by 10% after performing the CAS attack, from 0% to 10%.

- **Probabilistic degradation of privacy.** Since half of the records in the database D holds value `yes` for the sensitive attribute *illness*, while the other half holds value `no`, by applying the maximum entropy principle, the adversary's best course of action is to randomly select one of the values. Hence, the adversary has a 50% chance of correctly inferring the value for the sensitive attribute of a randomly selected record before performing the attack, i.e. without relying on any auxiliary information. Formally, Equation 4.11 for the *a priori*, probabilistic success equals:

$$\begin{aligned} \textit{prior-suc}^{CAS}_{prob}(D, \textit{illness}) &= \max_{s\in dom(\textit{illness})} \frac{|\{x{\in}D \mid x[\textit{illness}]{=}s\}|}{|D|} \\ &= \frac{|\{x{\in}D \mid x[\textit{illness}]{=}\texttt{yes}\}|}{|D|} \\ &= \frac{|\{x{\in}D \mid x[\textit{illness}]{=}\texttt{no}\}|}{|D|} \\ &= \frac{5}{10} = 0.5 \ . \end{aligned}$$

However, by performing the CAS attack, the adversary can rely on the values of the attribute in $\mathcal{Q}'_{ID} = \{\textit{age}\}$ for every individual in the database D. In this case, the database can be partitioned in three blocks, each containing all the records that share the same value for the attribute *age*, i.e. one block with five records whose *age* equal 25, one block with four records whose *age* equal 49, and one block with a single record whose *age* is 60.

Based on how many records within a block hold the same value for the attribute *illness*, the adversary has a distinct probability of success, i.e. 60% for the first block in which three in five records hold the same value for the sensitive attribute, 50% for the second, and 100% for the third block. Similarly, a randomly selected record has a distinct chance of belonging to each one of them, i.e. 50% chance of belonging to the block of *age* 25, 40% chance for the block of *age* 49, and 10% chance for the block of age 60. Hence, the adversary's success can be computed as $50\% \cdot 60\% + 40\% \cdot 50\% + 10\% \cdot 100\% = 60\%$, i.e. the adversary has a 60% chance of correctly inferring the value of the sensitive attribute for a randomly selected record after performing the CAS attack. Formally, Equation 4.12 for the *a posteriori*, probabilistic success equals:

$$\begin{aligned} \textit{post-suc}^{CAS}_{prob}(D, \textit{illness}, \{x[\mathcal{Q}'_{ID}]\}_{x\in D}) &= \\ \frac{\sum_{q\in dom(\mathcal{Q}'_{ID})} \max_{s\in dom(illness)} |\{x{\in}D \mid x[\textit{illness}]{=}s, x[\mathcal{Q}'_{ID}]{=}q\}|}{|D|} &= \\ &= \frac{3+2+1}{10} \\ &= \frac{6}{10} = 0.6 \ . \end{aligned}$$

Therefore, the probabilistic degradation of privacy in a collective attribute-inference attack, according to Equation 4.13, is:

$$\begin{aligned} &\textit{priv-degrad}^{CAS}_{prob}(D, \textit{illness}, \{x[\mathcal{Q}'_{ID}]\}_{x\in D}) = \\ &\frac{\textit{post-suc}^{CAS}_{prob}(D, \textit{illness}, \{x[\mathcal{Q}'_{ID}]\}_{x\in D})}{\textit{prior-suc}^{CAS}_{prob}(D, \textit{illness})} = \frac{0.6}{0.5} = 1.2 \ , \end{aligned}$$

i.e. the adversary's chance of correctly inferring the value for the sensitive attribute *illness* for a randomly selected record in the database D increases by 20% after performing the CAS attack, from 50% to 60%.

◁

D.2 Attacks on longitudinal databases

In this section, we present detailed examples of each attack performed on longitudinal databases, as discussed in Chapter 5 and Appendix C.2. The leading example upon which we perform the attacks was presented in Section 5.1.1, Example 20.

In Section D.2.1, we show examples of both individual and collective re-identification attacks, and in Section D.2.2, we show examples of both individual and collective attribute-inference attacks.

D.2.1 Examples of re-identification attacks

In this section, we present both examples of re-identification attacks on longitudinal databases. We begin with Example 51, which accounts for an adversary interested in re-identifying only predetermined targeted-individuals, according to Definition 43. Then, we present Example 52, which accounts for the general case, i.e. an adversary interested in re-identifying as many individuals as possible, according to Definition 21.

Example 51 (Execution of an Individual-target Re-identification Longitudinal database (IRL) attack)**.** Consider the longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}{=}\{D_1, D_2\}$ from Example 20, its aggregation $agreg(\mathcal{L}_{\mathcal{D}}){=}D_1{\Join}D_2$ on the set of attributes $\mathcal{A}_{agreg(\mathcal{L}_{\mathcal{D}})}$, as shown in Table 5.2c, and the assumptions made in Section 5.1.1. In this example, the adversary wants to separately re-identify the records of two predetermined targeted-individuals in the focal database D_1: André and Bia (the same Bia from Example 47). Furthermore, the adversary can obtain the following auxiliary information to be used as quasi-identifiers:

- André is a 25 years old man in database D_1 and a 26 years old man in database D_2, i.e. his quasi-identifiers are $\mathcal{Q}'_{ID} = \{age, gender\}$ in both databases, and the adversary knows the values $x^{\star}_{André}[\{(age,1),(gender,1),(age,2)\}] = \langle 25, \texttt{M}, 26\rangle$;
- Bia is a woman whose *occupation* category is given by `1` in both databases D_1 and D_2, i.e. her quasi-identifiers are $\mathcal{Q}''_{ID} = \{gender, occupation\}$ in both databases, and the adversary knows the values $x^{\star}_{Bia}[\{(gender,1),(occupation,1),(occupation,2)\}] = \langle \texttt{F}, \texttt{1}, \texttt{1}\rangle$.

We now compute the metrics from Definition 43.

- **Deterministic degradation of privacy.** Since the focal database D_1 has more than one record, neither André nor Bia can be re-identified with absolute certainty *a priori*, i.e. without the use of auxiliary information. Formally, Equation C.14 for the *a priori* success returns a `false` value both individuals.

  However, by performing the IRL attack against Bia in the focal database, the adversary can rely on the sub-record $x^{\star}_{Bia}[\{(gender,1),(occupation,1),$

$(occupation, 2)\}] = \langle \mathtt{F}, \mathtt{1}, \mathtt{1} \rangle$ as auxiliary information, corresponding to the first target's quasi-identifiers in the aggregation $agreg(\mathcal{L}_{\mathcal{D}})$. Since there is only one record $x{\in}agreg(\mathcal{L}_{\mathcal{D}})$ such that $x[\{(gender, 1), (occupation, 1), (occupation, 2)\}] = x^{\star}_{Bia}[\{(gender, 1), (occupation, 1), (occupation, 2)\}]$, the adversary can precisely determine that the record with *id* 2 is held by Bia. Hence, differently from the outcome in Example 47, the adversary can now re-identify Bia with absolute certainty and achieves *a posteriori*, deterministic success. Formally, Equation C.15 for the *a posteriori* success returns a `true` value for the IRL attack against Bia.

Therefore, the deterministic re-identification attack against Bia succeeds, i.e. Equation C.16 for the deterministic degradation of privacy returns a `true` value. Interestingly, Bia's re-identification was only possible due to an attribute change in the record with *id* equal to 1, not due to some change in the record held by Bia.

Similarly, by performing the IRL attack against André, the adversary can rely on the sub-record $x^{\star}_{André}[\{(age, 1), (gender, 1), (age, 2)\}] = \langle 25, \mathtt{M}, 26 \rangle$ as auxiliary information, corresponding to the second target's quasi-identifiers. In this case, however, the adversary cannot re-identify André with absolute certainty *a posteriori*, since there are two records $x{\in}agreg(\mathcal{L}_{\mathcal{D}})$ such that $x[\{(age, 1), (age, 2), (gender, 1)\}] = x^{\star}_{André}[\{(age, 1), (age, 2), (gender, 1)\}]$, the records with *id* 4 and 5. Formally, Equation C.15 for the *a posteriori* success returns a `false` value for the IRL attack against André.

Thus, the deterministic re-identification attack against André does not succeed, i.e. Equation C.16 for the deterministic degradation of privacy returns a `false` value.

- **Probabilistic degradation of privacy.** Although André did not suffer a deterministic degradation of privacy in the adversary's IRL attack based on the sub-record $x^{\star}_{André}[\{(age, 1), (gender, 1), (age, 2)\}] = \langle 25, \mathtt{M}, 26 \rangle$ auxiliary information, the adversary did increase his knowledge on André. We can quantify this gain of information probabilistically as follows.

  Whenever the adversary has more than one record to choose from, their best course of action is to randomly select one of them, according to the maximum entropy principle. Since there are ten records in the database D_1, the adversary has a 10% chance of randomly selecting André's record in the focal database before performing the attack, i.e. without relying on any auxiliary information.

Formally, Equation C.18 for the *a priori*, probabilistic success equals:

$$\textit{prior-suc}^{IRL}_{prob}(\mathcal{L}_{\mathcal{D}}) = \frac{1}{|D_1|} = \frac{1}{10} = 0.1 \ .$$

However, by performing the IRL attack against André, the adversary can rely on the sub-record $x^{\star}_{André}[\{(age, 1), (gender, 1), (age, 2)\}] = \langle 25, \texttt{M}, 26\rangle$ as auxiliary information, which decreases the number of possible records to just two of them. Hence, the adversary has a 50% chance of randomly selecting André's record in the focal database D_1 after performing the attack. Formally, Equation C.19 for the *a posteriori*, probabilistic success equals:

$$\textit{post-suc}^{IRL}_{prob}(\mathcal{L}_{\mathcal{D}}, x^{\star}_{André}[\mathcal{Q}'_{ID}]) = \\ \frac{1}{|\{x{\in}agreg(\mathcal{L}_{\mathcal{D}}) \mid x[\mathcal{Q}'_{ID}]{=}x^{\star}_{André}[\mathcal{Q}'_{ID}]\}|} = \frac{1}{2} = 0.5 \ ,$$

where $[\mathcal{Q}'_{ID}] = [\{(age, 1), (gender, 1), (age, 2)\}]$.

Therefore, the probabilistic degradation of privacy in a re-identification attack against André, according to Equation C.20, is the ratio $^{50\%}/_{10\%} = 5$, i.e. the adversary's chance of re-identifying André increases by a factor of five after performing the attack.

◁

We now present the collective-target example of re-identification attacks on longitudinal databases. This allows us to quantify the overall privacy degradation in a focal database given the longitudinal collection of databases it pertains to, i.e. how susceptible to privacy degradation are those individuals who hold a record in the focal database if the adversary has access to other databases in the longitudinal collection.

Example 52 (Execution of a Collective-target Re-identification Longitudinal database (CRL) attack)**.** Consider the longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}{=}\{D_1, D_2\}$ from Example 20, its aggregation $agreg(\mathcal{L}_{\mathcal{D}}){=}D_1 \bowtie D_2$ on the set of attributes $\mathcal{A}_{agreg(\mathcal{L}_{\mathcal{D}})}$, as shown in Table 5.2c, and the assumptions made in Section 5.1.1. In this example, the adversary wants to re-identify as many records as possible from those present in the focal database D_1. Furthermore, consider the adversary can obtain auxiliary information on the values of the attributes in $\mathcal{Q}'_{ID}{=}\{gender, occupation\}$ for every individual, which are used as quasi-identifiers.

We now compute the metrics from Definition 21.

- **Deterministic degradation of privacy.** Since the focal database D_1 has more than one record, no individual can be re-identified with absolute certainty *a priori*, i.e.without the use of auxiliary information, and the fraction of individuals that can be re-identified by the adversary *a priori* is of 0%. Formally, Equation 5.1 for the *a priori* success equals:

$$\textit{prior-suc}_{det}^{CRL}(\mathcal{L}_{\mathcal{D}}) = 0 \ .$$

However, by performing the CRL attack, the adversary can rely on the values of attributes in $\mathcal{Q}'_{ID}$=$\{\textit{gender}, \textit{occupation}\}$ for every individual and for every database in the longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}$. In this case, there are six records in $\textit{agreg}(\mathcal{L}_{\mathcal{D}})$ with unique values for the attributes $\{(\textit{gender}, 1), (\textit{occupation}, 1), (\textit{occupation}, 2)\}$, i.e. the records with *id* 1, 2, 6, 8, 9, and 10 can be re-identified with absolute certainty by the adversary *a posteriori*. Hence, the adversary can re-identify 60% of the records in the focal database D_1 given the knowledge of every value for the attributes $\{(\textit{gender}, 1), (\textit{occupation}, 1), (\textit{occupation}, 2)\}$ for every individual. Formally, Equation C.15 for the *a posteriori*, deterministic success equals:

$$\begin{aligned}\textit{post-suc}_{det}^{CRL}(\mathcal{L}_{\mathcal{D}}, \{x[\mathcal{Q}'_{ID}]\}_{x\in \textit{agreg}(\mathcal{L}_{\mathcal{D}})}) &= \frac{\left|\{\alpha{\in}\textit{agreg}(\mathcal{L}_{\mathcal{D}})/{\sim_{\mathcal{Q}'_{ID}}} \mid |\alpha|{=}1\}\right|}{|\textit{agreg}(\mathcal{L}_{\mathcal{D}})|} \\ &= \frac{6}{10} = 0.6 \ .\end{aligned}$$

Therefore, the deterministic degradation of privacy in a collective re-identification attack, according to Equation 5.3, is:

$$\begin{aligned}\textit{priv-degrad}_{det}^{CRL}(\mathcal{L}_{\mathcal{D}}, \{x[\mathcal{Q}_{ID}]\}_{x\in \textit{agreg}(\mathcal{L}_{\mathcal{D}})}) &= \\ \textit{post-suc}_{det}^{CRL}(\mathcal{L}_{\mathcal{D}}, \{x[\mathcal{Q}_{ID}]\}_{x\in \textit{agreg}(\mathcal{L}_{\mathcal{D}})}) - \textit{prior-suc}_{det}^{CRL}(\mathcal{L}_{\mathcal{D}}) &= \\ &= 0.6 - 0 = 0.6 \ ,\end{aligned}$$

i.e. the adversary's absolute chance of deterministically re-identifying an individual in the focal database D_1 increases by 60% after performing the CRL attack. [1]

[1] For comparison, consider the adversary was restricted to a single database scenario, i.e. the adversary can access only the database D_1. In this case, there is only one record in D_1 with unique values for the attributes *gender* and *occupation*, i.e. the record with *id* 8 is the only one that can be re-identified with absolute certainty by the adversary *a posteriori*. Therefore, both the adversary's *a posteriori*, deterministic success and degradation of privacy would be equal to 1/10 = 10%, far less than the degradation of privacy achieved in a longitudinal databases scenario.

- **Probabilistic degradation of privacy.** Whenever the adversary has more than one record to choose from, their best course of action is to randomly select one of them, according to the maximum entropy principle. Since there are ten records in the focal database D_1, the adversary has a 10% chance of randomly re-identifying a record in the focal database before performing the attack, i.e. without relying on any auxiliary information. Formally, Equation 5.4 for the *a priori*, probabilistic success equals:

  $$\textit{prior-suc}^{CRL}_{\textit{prob}}(\mathcal{L}_{\mathcal{D}}) = \frac{1}{|D_1|} = \frac{1}{10} = 0.1 \ .$$

  However, by performing the CRL attack, the adversary can rely on the values of attributes in $\mathcal{Q}'_{ID} = \{\textit{gender}, \textit{occupation}\}$ for every individual and for every database in the longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}$. In this case, the database $\textit{agreg}(\mathcal{L}_{\mathcal{D}})$ can be partitioned in eight blocks, each containing all the records that share the same value for the attributes $\{(\textit{gender}, 1), (\textit{occupation}, 1), (\textit{occupation}, 2)\}$, i.e. one block with two records whose values are $\langle \texttt{F}, \texttt{3}, \texttt{3}\rangle$, one block with two records whose values are $\langle \texttt{M}, \texttt{2}, \texttt{2}\rangle$, and six blocks with a single record in each corresponding to the remaining records whose values are unique for this combination of quasi-identifiers.

  Based on how many records there are in each block, the adversary has a distinct probability of success, i.e. 50% for each block with two records, and 100% for each block with only one record. Similarly, a randomly selected record in the focal database D_1 has a distinct chance of belonging to each one of those blocks, i.e. 40% for a block with two records, and 60% for a block with only one record. Hence, the adversary's success can be computed as $40\% \cdot 50\% + 60\% \cdot 100\% = 80\%$, i.e. the adversary has a 80% chance of re-identifying a randomly selected record in the focal database D_1 after performing the attack. Formally, Equation 5.5 for the *a posteriori*, probabilistic success equals:

  $$\textit{post-suc}^{CRL}_{\textit{prob}}(\mathcal{L}_{\mathcal{D}}, \{x[\mathcal{Q}'_{ID}]\}_{x \in \textit{agreg}(\mathcal{L}_{\mathcal{D}})}) = \frac{\left| \textit{agreg}(\mathcal{L}_{\mathcal{D}})/_{\sim_{\mathcal{Q}'_{ID}}} \right|}{|\textit{agreg}(\mathcal{L}_{\mathcal{D}})|} = \frac{8}{10} = 0.8 \ .$$

  Therefore, the probabilistic degradation of privacy in a collective re-identification attack, according to Equation 5.6, is:

  $$\textit{priv-degrad}^{CRL}_{\textit{prob}}(\mathcal{L}_{\mathcal{D}}, \{x[\mathcal{Q}_{ID}]\}_{x \in \textit{agreg}(\mathcal{L}_{\mathcal{D}})}) =$$

$$\frac{\mathit{post\text{-}suc}^{CRL}_{prob}(\mathcal{L}_{\mathcal{D}}, \{x[\mathcal{Q}'_{ID}]\}_{x \in agreg(\mathcal{L}_{\mathcal{D}})})}{\mathit{prior\text{-}suc}^{CRL}_{prob}(\mathcal{L}_{\mathcal{D}})} = \frac{0.80}{0.10} = 8 \ ,$$

i.e. the adversary's chance of re-identifying a randomly selected record in the focal database D_1 increases by a factor of eight after performing the CRL attack, from 10% up to 80%. [2]

$\triangleleft$

D.2.2 Examples of attribute-inference attacks

In this section, we present both examples of attribute-inference attacks on longitudinal databases. We begin with Example 53, which accounts for an adversary interested in inferring the value of an attribute for only predetermined targeted-individuals, according to Definition 44. Then, we present Example 54, which accounts for the general case, i.e. an adversary interested in inferring the value of an attribute for as many individuals as possible, according to Definition 22.

Example 53 (Execution of an Individual-target Attribute-inference Longitudinal database (IAL) attack)**.** Consider the longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}$=$\{D_1, D_2\}$ from Example 20, its aggregation $agreg(\mathcal{L}_{\mathcal{D}})$=$D_1$⟗$D_2$ on the set of attributes $\mathcal{A}_{agreg(\mathcal{L}_{\mathcal{D}})}$, as shown in Table 5.2c, and the assumptions made in Section 5.1.1. In this example, the adversary wants to separately infer the value of the attribute *illness*, considered to be sensitive, for two predetermined targeted-individuals: Eva and Fábio. Furthermore, the adversary can obtain the following auxiliary information to be used as quasi-identifiers:

- Eva is a woman whose *occupation* category is given by `3` in both databases D_1 and D_2, i.e. her quasi-identifiers are $\mathcal{Q}'_{ID} = \{gender, occupation\}$ in both databases, and the adversary knows the values $x^{\star}_{Eva}[\{(gender, 1), (occupation, 1), (occupation, 2)\}] = \langle \mathtt{F}, \mathtt{3}, \mathtt{4} \rangle$;
- Fábio is a 25 years old man in database D_1 and a 26 years old man in database D_2, i.e. his quasi-identifiers are $\mathcal{Q}''_{ID} = \{age, gender\}$ in both databases, and the adversary knows the values $x^{\star}_{Fábio}[\{(age, 1), (gender, 1), (age, 2)\}] = \langle 25, \mathtt{M}, 26 \rangle$.

[2] For comparison, consider the adversary was restricted to a single database scenario, i.e. the adversary can access only the database D_1. In this case, the database D_1 can be partitioned in five blocks by using the attributes *gender* and *occupation* as quasi-identifiers. Therefore, the adversary's *a posteriori*, probabilistic success would be equal to $60\% \cdot 50\% + 30\% \cdot 33\% + 10\% \cdot 100\% \approx 50\%$ and the degradation of privacy would be approximately equal to five, less than the degradation of privacy achieved in a longitudinal databases scenario.

We now compute the metrics from Definition 44.

- **Deterministic degradation of privacy.** The attribute *illness* can assume two distinct values in the focal database D_1: five records hold value `yes` and the other five hold value `no`. Hence, neither Eva nor Fábio can have the value of their sensitive attribute inferred by the attacker with absolute certainty *a priori*, i.e. without the use of auxiliary information. Formally, Equation C.23 for the *a priori* success returns a `false` value for both individuals.

  However, by performing the IAL attack against Eva in the focal database, the adversary can rely on the sub-record $x^{\star}_{Eva}[\{(gender, 1), (occupation, 1), (occupation, 2)\}] = \langle \mathtt{F}, \mathtt{3}, \mathtt{4} \rangle$ as auxiliary information, corresponding to the first target's quasi-identifiers in the aggregation $agreg(\mathcal{L}_{\mathcal{D}})$. Even though two records in $agreg(\mathcal{L}_{\mathcal{D}})$ hold the same values as Eva for the attributes $\{(gender, 1), (occupation, 1), (occupation, 2)\}$, i.e. the records with *id* 3 and 7, it is still possible for the adversary to infer the value of the sensitive attribute with absolute certainty. Because both records hold the same value for the attribute *illness*, i.e. `yes`, the adversary does achieve *a posteriori*, deterministic success. Formally, Equation C.24 for the *a posteriori* success returns a `true` value for the IAL attack against Eva.

  Therefore, the deterministic attribute-inference attack against Eva succeeds, i.e. Equation C.25 for the deterministic degradation of privacy returns a `true` value.

  Similarly, by performing the IAL attack against Fábio, the adversary can rely on the sub-record $x^{\star}_{Fábio}[\{(age, 1), (gender, 1), (age, 2)\}] = \langle 25, \mathtt{M}, 26 \rangle$ as auxiliary information, corresponding to the second target's quasi-identifiers. In this case, the adversary cannot infer the value of the sensitive attribute with absolute certainty *a posteriori*, since not all the records $x \in agreg(\mathcal{L}_{\mathcal{D}})$ that hold the same values for attributes $\{(gender, 1), (occupation, 1), (occupation, 2)\}$, i.e. the records with *id* 4 and 5, hold the same value for the attribute *illness*. While the record with *id* 4 holds the value `yes`, the record with *id* 5 holds the value `no`. Formally, Equation C.24 for the *a posteriori* success returns a `false` value for the IAL attack against Fábio.

  Thus, the deterministic attribute-inference attack against Fábio does not succeed, i.e. Equation C.25 for the deterministic degradation of privacy returns a `false` value.

- **Probabilistic degradation of privacy.** Although Fábio did not suffer a deterministic degradation of privacy in the adversary's IAL attack based on the sub-record $x^{\star}_{\textit{Fábio}}[\{(\textit{age},1),(\textit{gender},1),(\textit{age},2)\}] = \langle 25,\texttt{M},26\rangle$ as auxiliary information, the adversary did increase his knowledge on Fábio. We can quantify this gain of information probabilistically as follows.

Since half of the records in the focal database D_1 holds value `yes` for the sensitive attribute *illness*, while the other half holds value `no`, by applying the maximum entropy principle, the adversary's best course of action is to randomly select one of the values. Hence, the adversary has a 50% chance of correctly inferring Fábio's value for the sensitive attribute before performing the attack, i.e. without relying on any auxiliary information. Formally, Equation C.26 for the *a priori*, probabilistic success equals:

$$\begin{aligned}\textit{prior-suc}^{\textit{IAL}}_{\textit{prob}}(\mathcal{L}_{\mathcal{D}},\textit{illness}) &= \max_{s\in \textit{dom}(\textit{illness})} \frac{|\{x{\in}D_1 \mid x[\textit{illness}]{=}s\}|}{|D_1|}\\ &= \frac{|\{x{\in}D_1 \mid x[\textit{illness}]{=}\texttt{yes}\}|}{|D_1|}\\ &= \frac{|\{x{\in}D_1 \mid x[\textit{illness}]{=}\texttt{no}\}|}{|D_1|}\\ &= \frac{5}{10} = 0.5\ .\end{aligned}$$

However, by performing the IAL attack against Fábio, the adversary can rely on the sub-record $x^{\star}_{\textit{Fábio}}[\{(\textit{age},1),(\textit{gender},1),(\textit{age},2)\}] = \langle 25,\texttt{M},26\rangle$ as auxiliary information, which decreases the number of records from ten to just two of them. Since one of the records hold the value `yes` for the attribute *illness*, i.e. that with *id* 4, while the other holds the value `no`, i.e. that with *id* 5, the adversary has a 50% chance of correctly inferring Fábio's sensitive attribute value in the focal database D_1 after performing the attack. Formally, Equation C.27 for the *a posteriori*, probabilistic success equals:

$$\begin{aligned}&\textit{post-suc}^{\textit{IAL}}_{\textit{prob}}(\mathcal{L}_{\mathcal{D}},\textit{illness},x^{\star}_{\textit{Fábio}}[\mathcal{Q}''_{\textit{ID}}]) =\\ &\max_{s\in \textit{dom}(\textit{illness})} \frac{|\{x{\in}\textit{block}(x^{\star}_{\textit{Fábio}}[\mathcal{Q}''_{\textit{ID}}]) \mid x[\textit{illness}]{=}s\}|}{|\textit{block}(x^{\star}_{\textit{Fábio}}[\mathcal{Q}''_{\textit{ID}}])|} =\\ &\frac{|\{x{\in}\textit{block}(x^{\star}_{\textit{Fábio}}[\mathcal{Q}''_{\textit{ID}}]) \mid x[\textit{illness}]{=}\texttt{yes}\}|}{|\textit{block}(x^{\star}_{\textit{Fábio}}[\mathcal{Q}''_{\textit{ID}}])|} = \frac{1}{2} = 0.5\ ,\end{aligned}$$

where $\textit{block}(x^{\star}_{\textit{Fábio}}[\mathcal{Q}''_{\textit{ID}}]){=}\{x{\in}\textit{agreg}(\mathcal{L}_{\mathcal{D}}) \mid x[\mathcal{Q}''_{\textit{ID}}]{=}x^{\star}_{\textit{Fábio}}[\mathcal{Q}''_{\textit{ID}}]\}$ is the block of all individuals in the aggregated database $\textit{agreg}(\mathcal{L}_{\mathcal{D}})$ that hold the same quasi-identifier values as the targeted-individual, $x^{\star}_{\textit{Fábio}}$.

Therefore, the probabilistic degradation of privacy in an attribute-inference attack against Fábio, according to Equation C.28, is the ratio $^{50\%}/_{50\%} = 1$, i.e. the adversary's chance of correctly inferring the value of Fábio's sensitive attribute does not change after performing the IAL attack.

◁

We now present the collective-target example of attribute-inference attacks on longitudinal databases. This allows us to quantify the overall privacy degradation in a focal database given the longitudinal collection of databases it pertains to, i.e. how susceptible to privacy degradation are those individuals who hold a record in the focal database if the adversary has access to other databases in the longitudinal collection.

Example 54 (Execution of a Collective-target Attribute-inference Longitudinal database (CAL) attack)**.** Consider the longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}{=}\{D_1, D_2\}$ from Example 20, its aggregation $\mathit{agreg}(\mathcal{L}_{\mathcal{D}}){=}D_1{\bowtie}D_2$ on the set of attributes $\mathcal{A}_{\mathit{agreg}(\mathcal{L}_{\mathcal{D}})}$, as shown in Table 5.2c, and the assumptions made in Section 5.1.1. In this example, the adversary wants to infer the value of the attribute *illness*, considered to be sensitive, for as many records as possible from those present in the focal database D_1. Furthermore, consider the adversary can obtain auxiliary information on the values of attributes in $\mathcal{Q}'_{ID}{=}\{\mathit{gender}, \mathit{occupation}\}$ for every individual, which are used as quasi-identifiers.

We now compute the metrics from Definition 22.

- **Deterministic degradation of privacy.** The attribute *illness* can assume two distinct values in the focal database D_1: five records hold value `yes` and the other five hold value `no`. Hence, no individual can have the value of their sensitive attribute inferred by the attacker with absolute certainty *a priori*, i.e. without the use of auxiliary information. Formally, Equation 5.7 for the *a priori* success equals:

  $$\mathit{prior\text{-}suc}^{CAL}_{det}(\mathcal{L}_{\mathcal{D}}, \mathit{illness}) = 0 \ .$$

  However, by performing the CAL attack, the adversary can rely on the values of the attributes in $\mathcal{Q}'_{ID}{=}\{\mathit{gender}, \mathit{occupation}\}$ for every individual in the longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}$. In this case, there are six records in $\mathit{agreg}(\mathcal{L}_{\mathcal{D}})$ with unique values for the attributes $\{(\mathit{gender}, 1), (\mathit{occupation}, 1), (\mathit{occupation}, 2)\}$, i.e. the records with *id* 1, 2, 6, 8, 9, and 10 can have their values for the sensitive attribute *illness* immediately inferred. Also, even though the records with *id* 3 and

7 share the same value for the quasi-identifier attributes, they also share the same value for the sensitive attribute, which can then be immediately inferred. Since the adversary can infer the values of the sensitive attribute *illness* for eight of the records in the focal database D_1, from a total of ten records in the database, the adversary has an 80% chance of success. Formally, Equation 5.8 for the *a posteriori*, deterministic success equals:

$$\begin{aligned}&\mathit{post\text{-}suc}^{CAL}_{det}(\mathcal{L}_{\mathcal{D}}, \mathit{illness}, \{x[\mathcal{Q}'_{ID}]\}_{x \in \mathit{agreg}(\mathcal{L}_{\mathcal{D}})}) = \\ &\frac{\sum_{\alpha \in \mathit{agreg}(\mathcal{L}_{\mathcal{D}})/\sim_{\mathcal{Q}'_{ID}}} |\alpha| \cdot \mathit{unique\text{-}sens}(\alpha, \mathit{illness})}{|\mathit{agreg}(\mathcal{L}_{\mathcal{D}})|} = \\ &\frac{6 \cdot (1 \cdot 1) + 1 \cdot (2 \cdot 1) + 1 \cdot (2 \cdot 0)}{10} = \frac{8}{10} = 0.8 \ ,\end{aligned}$$

where the function *unique-sens* is defined by Equation 5.9. Here, $\mathit{unique\text{-}sens}(\alpha, \mathit{illness}) = 1$ for all blocks α except for the one containing the records with *id* 4 and 5.

Therefore, the deterministic degradation of privacy in a collective attribute-inference attack, according to Equation 5.10, is:

$$\begin{aligned}&\mathit{priv\text{-}degrad}^{CAL}_{det}(\mathcal{L}_{\mathcal{D}}, \mathit{illness}, \{x[\mathcal{Q}'_{ID}]\}_{x \in \mathit{agreg}(\mathcal{L}_{\mathcal{D}})}) = \\ &\mathit{post\text{-}suc}^{CAL}_{det}(\mathcal{L}_{\mathcal{D}}, \mathit{illness}, \{x[\mathcal{Q}'_{ID}]\}_{x \in \mathit{agreg}(\mathcal{L}_{\mathcal{D}})}) - \mathit{prior\text{-}suc}^{CAL}_{det}(\mathcal{L}_{\mathcal{D}}, \mathit{illness}) = \\ &0.8 - 0 = 0.8 \ ,\end{aligned}$$

i.e. the adversary's absolute chance of deterministically inferring the value of the sensitive attribute *illness* for an individual in the focal database D_1 increases by 80% after performing the attack. [3]

- **Probabilistic degradation of privacy.** Since half of the records in the focal database D_1 holds value `yes` for the sensitive attribute *illness*, while the other half holds value `no`, by applying the maximum entropy principle, the adversary's best course of action is to randomly select one of the values. Hence, the adversary has a 50% chance of correctly inferring the value for the sensitive attribute of a randomly selected record before performing the attack, i.e. without relying on

[3] For comparison, consider the adversary was restricted to a single database scenario, i.e. the adversary can access only the database D_1. In this case, there would be three blocks such that all records share the same value for the sensitive attribute *illness*, i.e. one block for the records with *id* 3, 6, and 7, one block for the record with *id* 8, and one block for the records with *id* 9 and 10. Therefore, both the adversary's *a posteriori*, deterministic success and degradation of privacy would be equal to $6/10 = 60\%$, less than the degradation of privacy achieved in a longitudinal database scenario.

any auxiliary information. Formally, Equation 5.11 for the *a priori*, probabilistic success equals:

$$\begin{aligned}
\mathit{prior\text{-}suc}^{CAL}_{prob}(\mathcal{L}_{\mathcal{D}}, \mathit{illness}) &= \max_{s\in \mathit{dom}(\mathit{illness})} \frac{|\{x{\in}D_1 \mid x[\mathit{illness}]{=}s\}|}{|D_1|} \\
&= \frac{|\{x{\in}D_1 \mid x[\mathit{illness}]{=}\texttt{yes}\}|}{|D_1|} \\
&= \frac{|\{x{\in}D_1 \mid x[\mathit{illness}]{=}\texttt{no}\}|}{|D_1|} \\
&= \frac{5}{10} = 0.5 \ .
\end{aligned}$$

However, by performing the CAL attack, the adversary can rely on the values of the attributes in $\mathcal{Q}'_{ID} = \{\mathit{gender}, \mathit{occupation}\}$ for every individual in the longitudinal collection of databases $\mathcal{L}_{\mathcal{D}}$. In this case, the database $\mathit{agreg}(\mathcal{L}_{\mathcal{D}})$ can be partitioned in eight blocks, each containing all the records that share the same value for the attributes $\{(\mathit{gender}, 1), (\mathit{occupation}, 1), (\mathit{occupation}, 2)\}$, i.e. one block with two records whose values are $\langle \texttt{F}, \texttt{3}, \texttt{3}\rangle$, one block with two records whose values are $\langle \texttt{M}, \texttt{2}, \texttt{2}\rangle$, and six blocks with a single record in each corresponding to the remaining records whose values are unique for this combination of quasi-identifiers.

Based on how many records within a block hold the same value for the attribute *illness*, the adversary has a distinct probability of success, i.e. 50% for the block with two records whose values are $\langle \texttt{M}, \texttt{2}, \texttt{2}\rangle$, and 100% for the block with two records whose values are $\langle \texttt{F}, \texttt{3}, \texttt{3}\rangle$ as well as for each of the six blocks with a single record. Similarly, a randomly selected record in the focal database D_1 has a distinct chance of belonging to each one of those blocks, i.e. 20% for the block with two records whose values are $\langle \texttt{M}, \texttt{2}, \texttt{2}\rangle$, 20% for the block with two records whose values are $\langle \texttt{F}, \texttt{3}, \texttt{3}\rangle$, and 10% for each of the six blocks with only one record. Hence, the adversary's success can be computed as $20\% \cdot 50\% + 20\% \cdot 100\% + 6 \cdot 10\% \cdot 100\% = 90\%$, i.e. the adversary has a 90% chance of correctly inferring the value of the sensitive attribute for a randomly selected record in the focal database after performing the CAL attack. Formally, Equation 5.12 for the *a posteriori*, probabilistic success equals:

$$\mathit{post\text{-}suc}^{CAL}_{prob}(\mathcal{L}_{\mathcal{D}}, \mathit{illness}, \{x[\mathcal{Q}'_{ID}]\}_{x\in \mathit{agreg}(\mathcal{L}_{\mathcal{D}})}) =$$

$$\frac{\displaystyle\sum_{q\in \mathit{dom}(\mathcal{Q}'_{ID}|\mathcal{L}_{\mathcal{D}})} \max_{s\in \mathit{dom}(\mathit{illness})} |\{x{\in}\mathit{agreg}(\mathcal{L}_{\mathcal{D}}) \mid x[(\mathit{illness}, 1)]{=}s, x[\mathcal{Q}'_{ID}]{=}q\}|}{|\mathit{agreg}(\mathcal{L}_{\mathcal{D}})|} =$$

$$\frac{6 \cdot 1 + 1 \cdot 2 + 1 \cdot 1}{10} = \frac{9}{10} = 0.9 \ ,$$

where $\mathcal{Q}'_{ID}$=$\{$*gender*, *occupation*$\}$.

Therefore, the probabilistic degradation of privacy in a collective attribute-inference attack, according to Equation 5.13, is:

$$\begin{aligned} &\textit{priv-degrad}^{CAL}_{prob}(\mathcal{L}_{\mathcal{D}}, \textit{illness}, \{x[\mathcal{Q}'_{ID}]\}_{x \in agreg(\mathcal{L}_{\mathcal{D}})}) = \\ &\frac{\textit{post-suc}^{CAL}_{prob}(\mathcal{L}_{\mathcal{D}}, \textit{illness}, \{x[\mathcal{Q}'_{ID}]\}_{x \in agreg(\mathcal{L}_{\mathcal{D}})})}{\textit{prior-suc}^{CAL}_{prob}(\mathcal{L}_{\mathcal{D}}, \textit{illness})} = \frac{0.9}{0.5} = 1.8 \ , \end{aligned}$$

i.e. the adversary's chance of correctly inferring the value for the sensitive attribute *illness* for a randomly selected record in the focal database D_1 increases by 80% after performing the CAL attack from 50% to 90%. [4]

◁

[4] For comparison, consider the adversary was restricted to a single database scenario, i.e. the adversary can access only the database D_1. In this case, the database can be partitioned in five blocks by using the attributes *gender* and *occupation* as quasi-identifiers. Therefore, the adversary's *a posteriori*, probabilistic success would be equal to $10\% \cdot 100\% + 2 \cdot 20\% \cdot 50\% + 20\% \cdot 100\% + 30\% \cdot 100\% = 80\%$ and the degradation of privacy would be an increase of 60% in the adversary's chance, less than the degradation of privacy achieved in a longitudinal databases scenario.

Appendix E

Experimental results for collective-target attacks on the School Census of 2019

In this appendix, we present additional quantitative analyses on the privacy risks for individuals whose information are made public on a single database. The attacks performed here were modeled according to the theory developed in Section 4.1. But instead of the database for the School Census of 2018, here we consider that for the School Census of 2019, which features a change adopted by INEP as a tentative measure to decrease the privacy risk to which students were subject, i.e. the removal of the variable `NU_DIA` for the student's day of birth. We verify this change decreases the maximum privacy risk to which students are subject, but is still insufficient to guarantee individuals' privacy.

We begin by detailing our experimental setup, followed by the results for re-identification and then for attribute-inference attacks on INEP's databases.

E.1 Experimental setup

As stated in Section 1.3, we have chosen to work with the databases containing information on students and released as microdata, i.e. data at the record level. In this appendix, we consider the database for the School Census of 2019. The analyses for the School Census of 2018 is in Chapter 4 and for the Higher Education Censuses of 2018 and 2019 is in Appendix F.

A preliminary analysis of the databases containing information on students showed that a student may hold more than one record in a given database, e.g. if a High School student is also enrolled in a Professional Education course. Hence, in order to guarantee that each student holds only one record in each database, according to Assumption **AS1** from Section 4.1.1, we have randomly selected only one record for each data holder of multiple records. This data treatment was based on the unique identification number, the `ID_ALUNO` code, given to each student in the pseudonymization treatment performed by INEP. The `ID_ALUNO` code, which is unique to each student at least in a given database release, easily allowed us to find those students with multiple records and to perform the random selection of just one of them. Furthermore, the treatment has not resulted in a significant decrease in the number of available records. For the School Census of 2019, 93.11% of the records were kept, i.e. 47 640 822 records from a total of 51 166 723.

Even though each database accounts for dozens of attributes, we have chosen just a few for our analyses, given the computational costs, including time and memory usage. The selection criteria was as follows.

- For the quasi-identifying attributes, we have sets restricted to the maximum of ten or eleven attributes, selected according to how easily an adversary could learn them, e.g. the date and city of birth, the city of residency, or the school code. The quasi-identifying attributes chosen for the experiments in this chapter are listed in Table E.1.
- We have chosen two sensitive attributes selected according to the seriousness of the possible individual privacy breach if revealed, e.g. whether or not the individual has special needs or disabilities. The sensitive attributes chosen for the experiments in this chapter are listed in Table E.2.

Of course, the selection of those quasi-identifiers and sensitive attributes is arbitrary and can change in significance over time. Nevertheless, our goal is not to ultimately define whether an attribute should be considered as a quasi-identifier or as a sensitive attribute. Instead, the results here illustrate possible real-life circumstances and the respective privacy risks for data holders in case of unintended information disclosure.

In the following Sections 4.2.2 and 4.2.3, we present some experimental results for re-identification and attribute-inference attacks, respectively. Both were performed on single databases for collective-targets on the School Census of 2019 released by INEP.

Variable	Meaning
NU_MES	Student's month of birth.
NU_ANO	Student's year of birth.
TP_SEXO	Student's gender.
TP_COR_RACA	Student's ethnicity.
TP_NACIONALIDADE	Student's nationality.
CO_PAIS_ORIGEM	Student's country code.
CO_MUNICIPIO_NASC	Student's city of birth code.
CO_MUNICIPIO_END	Student's city of residency code.
CO_ENTIDADE	School code.
TP_DEPENDENCIA	Administrative dependency of the school, i.e. whether public or private.

Table E.1: Variables from the School Census of 2019 chosen as quasi-identifiers for Collective-target Re-identification Single database (CRS) attacks in Experiment 55, and for Collective-target Attribute-inference Single database (CAS) attacks in Experiment 56. Even though the variable NU_DIA for the student's day of birth was listed as available in the 2019 database's variables dictionary, it was not present in the actual released database.

Variable	Values	Meaning
IN_NECESSIDADE_ESPECIAL	0 (No) 1 (Yes)	Whether the student possesses a disability, global developmental disorder, or autism spectrum disorder, or not.
IN_TRANSPORTE_PUBLICO	-1 (Unavailable) 0 (No) 1 (Yes)	Whether the student uses public school transport, or not.

Table E.2: Variables from the School Census of 2019 chosen as sensitive attributes for Collective-target Attribute-inference Single database (CAS) attacks in Experiment 56.

E.2 Results of re-identification attacks experiments

In this section, we present our quantitative analyses on the privacy risk of re-identification for individuals whose information are made public on a single database, as modeled in Section 4.1.2. We present one CRS attacks on the School Census of 2019, Experiment 55. Note that the quasi-identifying attribute `NU_DIA` was not released by INEP for this database.

Experiment 55 (Collective-target Re-identification Single database (CRS) attacks on the School Census of 2019)**.** In a CRS attack, the adversary's goal is to re-identify as many individuals as possible, according to Definition 16. Since we want to evaluate the overall re-identification risk, the adversary can gather auxiliary information on the values of a set of quasi-identifying attributes for every individual who holds a record in the database. The following experiment was conducted on the School Census of 2019.

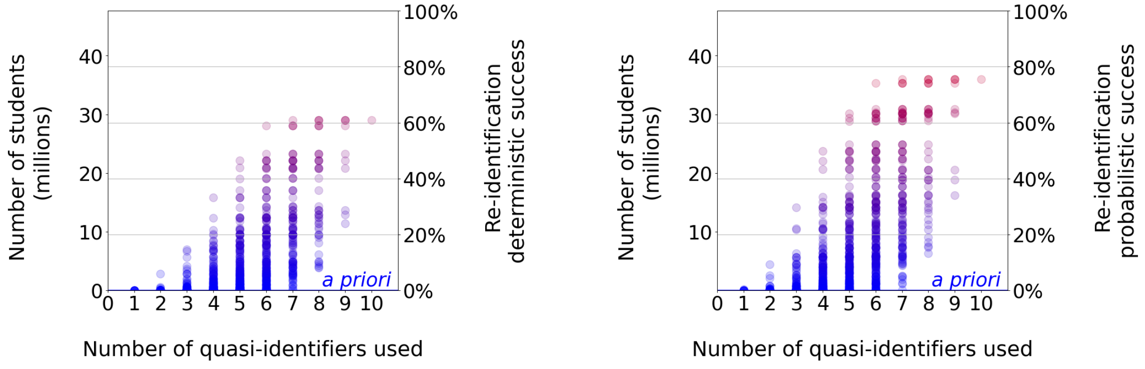
This database was released containing 51 166 723 records, which were reduced to 47 640 822 after the random selection of only one record for each data holder of multiple records, as detailed in Section 4.2.1. For the current experiment, we have selected as quasi-identifiers the ten attributes listed in Table E.1. Considering an adversary could gather as auxiliary information any non-empty subset among the $2^{10} - 1 = 1\,023$ possible combinations of quasi-identifying attributes, we have measured both the deterministic and probabilistic degradation of privacy for each combination.

The results are organized as follows.

- Table E.3 summarizes which combinations of quasi-identifiers result in the highest degradation of privacy according to the subset's size.
- Figure E.4 shows both the adversary's deterministic and probabilistic measures of success for each combination of quasi-identifying attributes.
- Figure E.5 shows the worst-case histograms for the distribution of individuals according to the adversary's probabilistic measure of success for some subsets of quasi-identifying attributes with different sizes.

According to those results, we were able to conclude the following.

- **Adversary's deterministic success**. Defined in Equations 4.1 and 4.2, this metric measures the fraction of individuals in the database that can be re-identified with absolute certainty, in a scale from 0% to 100%. According to



(a) Deterministic success.

(b) Probabilistic success.

Figure E.4: Experiment 55: Adversary's success in Collective-target Re-identification Single database (CRS) attacks on the School Census of 2019. Here, the horizontal "*a priori*" line represents the adversary's *a priori* deterministic or probabilistic success, i.e. before performing the CRS attack. Each dot represents a distinct scenario defined by the selected quasi-identifying attributes among the 1 023 possibilities. The horizontal axis defines the size of the quasi-identifiers subset, while the vertical axis defines the adversary's deterministic or probabilistic success.

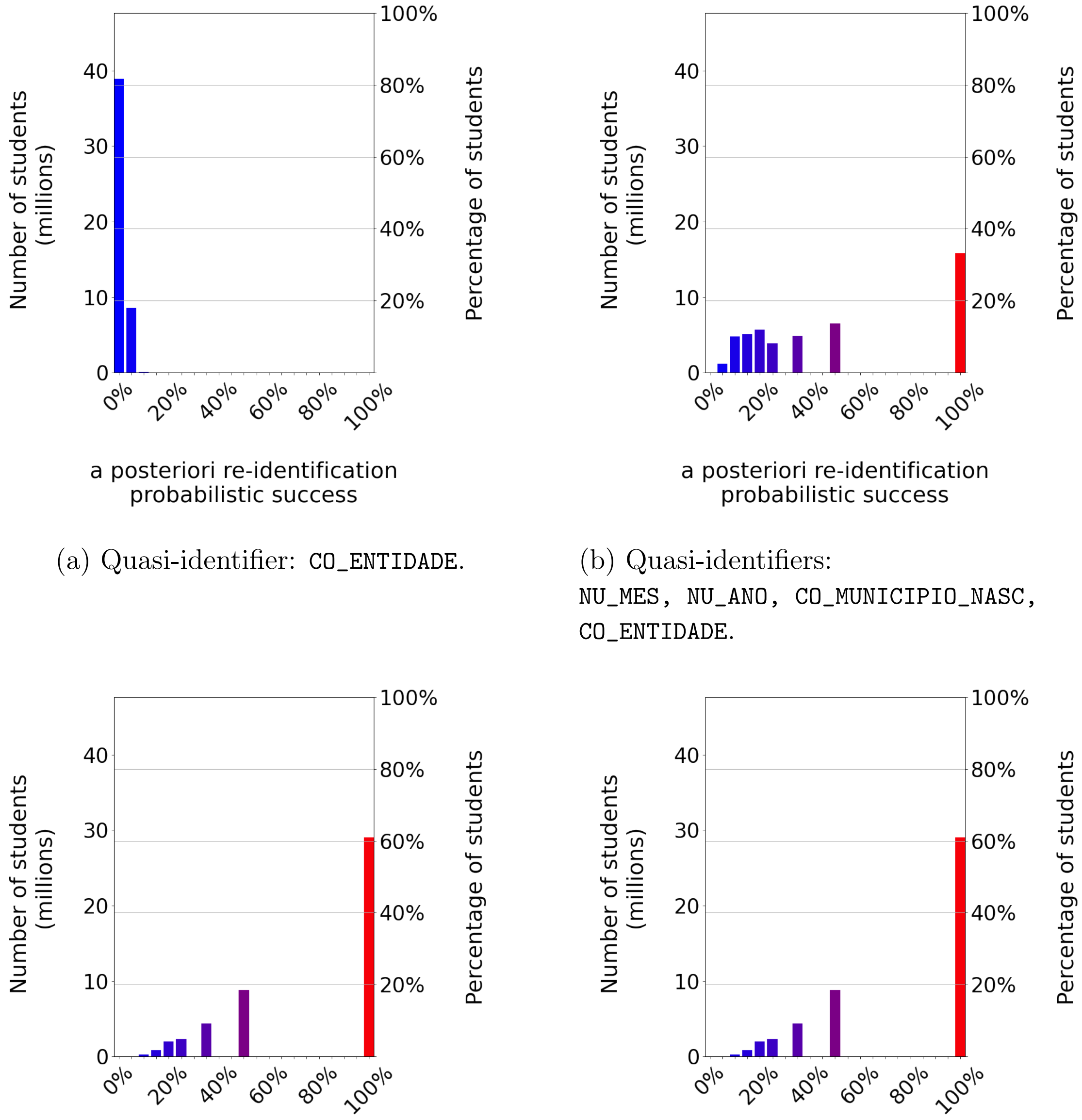
Table E.3, the adversary's *a priori* deterministic success equals 0%, i.e. no individual can be re-identified with absolute certainty without the use of auxiliary information. However, by using only three quasi-identifying attributes, the adversary can re-identified with absolute certainty up to 14.54% of the individuals in the database, while the use of four quasi-identifiers allows the adversary to re-identify up to 33.12% of the individuals. Finally, by using seven or more of the ten quasi-identifying attributes available, the risk of re-identification increases to up to 60.90%.

- **Adversary's probabilistic success**. Defined in Equations 4.4 and 4.5, this metric measures the probability of a randomly chosen individual in the database being re-identified, in a scale from 0% to 100%. According to Table E.3, the adversary's *a priori* probabilistic success is approximately 0.000002%, i.e. the adversary's chance of re-identifying one of the 47 640 822 individuals in the database is almost zero. However, by using only three quasi-identifying attributes, the adversary increases their chance to up to 29.64%, while the use of four quasi-identifiers increases this probability to up to 49.86%. Finally, by using seven or more of the ten quasi-identifying attributes available, the adversary's chance of re-identifying a randomly chosen individual in the database increases to up to 75.51%.

◁

		Adversary's deterministic success		**Adversary's probabilistic success**	
		***a priori* success** 0.00%		***a priori* success** 0.000002%	
#	**Quasi-identifiers**	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
1	`CO_ENTIDADE`	0.0002%	0.0002%	0.38%	182 434
2	`CO_MUNICIPIO_NASC, CO_ENTIDADE`	5.92%	5.92%	9.38%	4 468 395
3	`NU_MES, NU_ANO, CO_ENTIDADE`	11.80%	11.80%	29.64%	14 119 984
3	`NU_ANO, CO_MUNICIPIO_NASC, CO_ENTIDADE`	14.54%	14.54%	21.73%	10 352 649
4	`NU_MES, NU_ANO, CO_MUNICIPIO_NASC, CO_ENTIDADE`	33.12%	33.12%	49.86%	23 752 768
5	`NU_MES, NU_ANO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_ENTIDADE`	46.46%	46.46%	63.20%	30 107 590
6	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_ENTIDADE`	58.94%	58.94%	74.18%	35 338 492
7	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE`	60.90%	60.90%	75.51%	35 972 591
8	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE`	60.90%	60.90%	75.51%	35 973 629
9	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE`	60.90%	60.90%	75.51%	35 973 629
9	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA`	60.90%	60.90%	75.51%	35 973 629
10	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA`	60.90%	60.90%	75.51%	35 973 629

Table E.3: Experiment 55: Quasi-identifiers with the highest degradation of privacy in Collective-target Re-identification Single database (CRS) attacks on the School Census of 2019. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* successes, while the probabilistic degradation of privacy is their ratio. Table E.1 lists the English meaning of each quasi-identifying attribute.



(a) Quasi-identifier: `CO_ENTIDADE`.

(b) Quasi-identifiers: `NU_MES`, `NU_ANO`, `CO_MUNICIPIO_NASC`, `CO_ENTIDADE`.

(c) Quasi-identifiers: `NU_MES`, `NU_ANO`, `TP_SEXO`, `TP_COR_RACA`, `CO_MUNICIPIO_NASC`, `CO_MUNICIPIO_END`, `CO_ENTIDADE`.

(d) Quasi-identifiers: `NU_MES`, `NU_ANO`, `TP_SEXO`, `TP_COR_RACA`, `TP_NACIONALIDADE`, `CO_PAIS_ORIGEM`, `CO_MUNICIPIO_NASC`, `CO_MUNICIPIO_END`, `CO_ENTIDADE`, `TP_DEPENDENCIA`.

Figure E.5: Experiment 55: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Re-identification Single database (CRS) attacks on the School Census of 2019. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of re-identification.

E.3 Results of attribute-inference attacks experiments

In this section, we present our quantitative analyses on the privacy risk of attribute-inference for individuals whose information are made public on a single database, as modeled in Section 4.1.2. We present one CAS attacks on the School Census of 2019, Experiment 56, on sensitive attributes `IN_NECESSIDADE_ESPECIAL` and `IN_TRANSPORTE_PUBLICO`. Note that the quasi-identifying attribute `NU_DIA` was not released by INEP for this database.

Experiment 56 (Collective-target Attribute-inference Single database (CAS) attacks on the School Census of 2019)**.** In a CAS attack, the adversary's goal is to infer the values of an attribute considered to be sensitive for as many individuals as possible, according to Definition 17. Since we want to evaluate the overall attribute-inference risk, the adversary can gather auxiliary information on the values of a set of quasi-identifying attributes for every individual who holds a record in the database. The following experiment was conducted on the School Census of 2019.

This database was released containing 51 166 723 records, which were reduced to 47 640 822 after the random selection of only one record for each data holder of multiple records, as detailed in Section 4.2.1. For the current experiment, we have selected as quasi-identifiers the ten attributes listed in Table E.1. Considering an adversary could gather as auxiliary information any non-empty subset among the $2^{10} - 1 = 1\,023$ possible combinations, we have measured both the deterministic and probabilistic degradation of privacy for each combination.

Furthermore, we were interested in inferring the value for the attributes `IN_NECESSIDADE_ESPECIAL` and `IN_TRANSPORTE_PUBLICO`, both described in Table E.2 and considered by us to be sensitive.

The results are organized as follows.

- Tables E.6 and E.7 summarize which combinations of quasi-identifiers result in the highest degradation of privacy according to the subset's size for the sensitive attributes `IN_NECESSIDADE_ESPECIAL` and `IN_TRANSPORTE_PUBLICO`, respectively.

- Figure E.8 shows both the adversary's deterministic and probabilistic measures of success for each combination of quasi-identifying attributes, and for both sensitive attributes.

- Figures E.9 and E.10 show the worst-case histograms for the distribution of individuals according to the adversary's probabilistic measure of success for some subsets of quasi-identifying attributes with different sizes and for the sensitive attributes `IN_NECESSIDADE_ESPECIAL` and `IN_TRANSPORTE_PUBLICO`, respectively.

According to those results, we were able to conclude the following.

- **Adversary's deterministic success**. Defined in Equations 4.7 and 4.8, this metric measures the fraction of individuals in the database that can have their values for the sensitive attribute inferred with absolute certainty, in a scale from 0% to 100%.

  According to Table E.6 for the sensitive attribute `IN_NECESSIDADE_ESPECIAL`, the adversary's *a priori* deterministic success equals 0%, i.e. no individual can have their value for the sensitive attribute inferred with absolute certainty without the use of auxiliary information. However, by using only one quasi-identifying attribute, the adversary can infer the values with absolute certainty for up to 11.29% of the individuals in the database, while the use of two quasi-identifiers allows the adversary to infer the values for up to 49.52% of the individuals. Finally, by using four or more of the ten quasi-identifying attributes available, the risk of attribute-inference increases above 93.28%.

  From Table E.7 for the sensitive attribute `IN_TRANSPORTE_PUBLICO`, the adversary's *a priori* deterministic success also equals 0%. However, by using only one quasi-identifying attribute, the adversary can infer the values with absolute certainty for up to 41.48% of the individuals in the database, while the use of two quasi-identifiers allows the adversary to infer the values for up to 55.44% of the individuals. Finally, by using six or more of the ten quasi-identifying attributes available, the risk of attribute-inference increases above 93.05%.

- **Adversary's probabilistic success**. Defined in Equations 4.11 and 4.12, this metric measures the probability of a randomly chosen individual in the database having their value for the sensitive attribute inferred with absolute certainty, in a scale from 0% to 100%.

  According to Table E.6 for the sensitive attribute `IN_NECESSIDADE_ESPECIAL`, the adversary's *a priori* probabilistic success is already of 97.38%, i.e. the adversary's chance of inferring the value for the sensitive attribute for one of the 47 640 822 individuals in the database is already high even before performing the

CAS attack. Furthermore, by using seven or more quasi-identifying attributes, the adversary increases their chance to up to 99.26%.

From Table E.7 for the sensitive attribute `IN_TRANSPORTE_PUBLICO`, the adversary's *a priori* probabilistic success is of 82.06%, also high even before performing the CAS attack. Furthermore, by using only two quasi-identifying attributes, the adversary increases their chance to up to 90.82%, while the use of four quasi-identifiers increases the adversary's chance to up to 95.23%. Finally, by using seven or more of the ten quasi-identifying attributes available, the risk of attribute-inference increases to up to 97.53%.

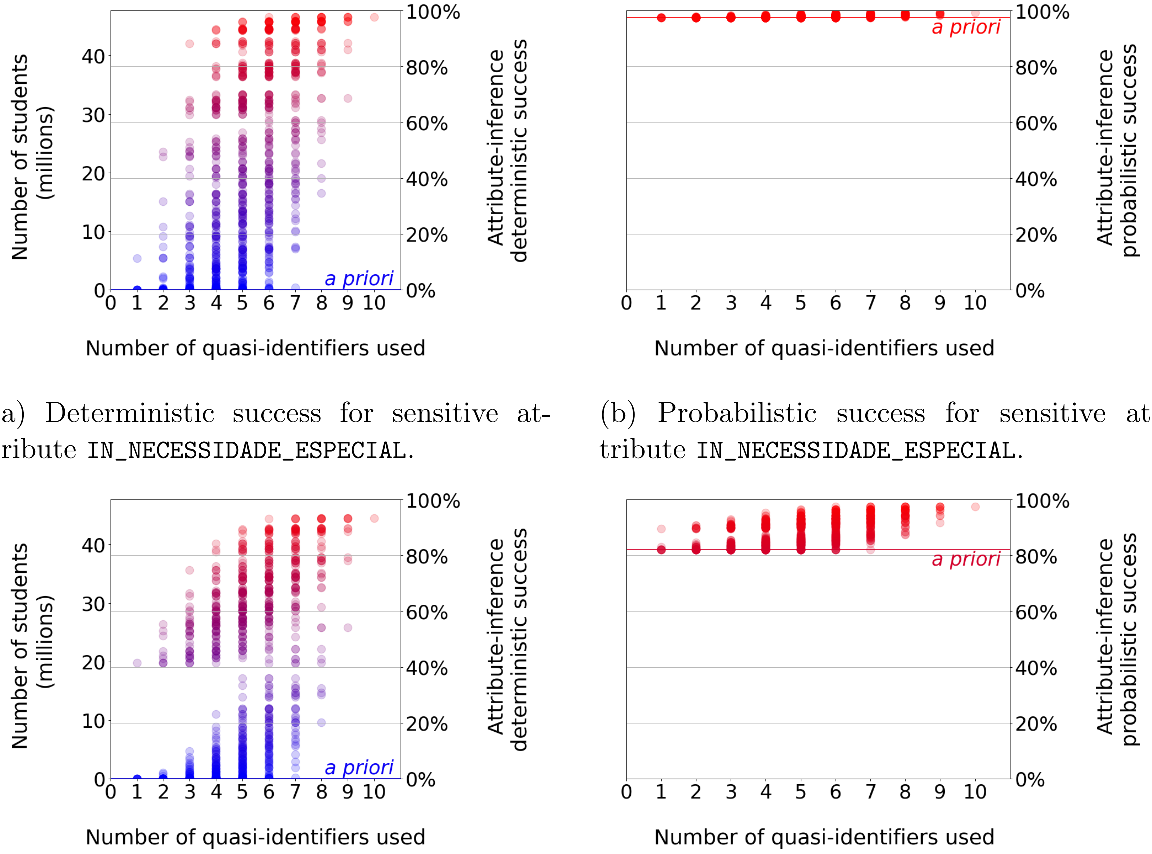
◁

		Adversary's deterministic success		Adversary's probabilistic success	
		a priori success 0.00%		*a priori* success 97.38%	
#	**Quasi-identifiers**	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
1	`CO_ENTIDADE`	11.29%	11.29%	97.74%	1.0037
2	`CO_MUNICIPIO_NASC, CO_ENTIDADE`	31.54%	31.54%	97.88%	1.0052
2	`NU_MES, CO_ENTIDADE`	49.52%	49.52%	97.74%	1.0038
3	`NU_ANO, CO_MUNICIPIO_NASC, CO_ENTIDADE`	65.29%	65.29%	98.19%	1.0083
3	`NU_MES, NU_ANO, CO_ENTIDADE`	88.17%	88.17%	98.18%	1.0082
4	`NU_MES, NU_ANO, CO_MUNICIPIO_NASC, CO_ENTIDADE`	92.88%	92.88%	98.66%	1.0131
4	`NU_MES, NU_ANO, TP_COR_RACA, CO_ENTIDADE`	93.28%	93.28%	98.53%	1.0118
5	`NU_MES, NU_ANO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_ENTIDADE`	95.91%	95.91%	98.97%	1.0163
5	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_ENTIDADE`	96.08%	96.08%	98.84%	1.015
6	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_ENTIDADE`	97.63%	97.63%	99.22%	1.0189
7	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE`	97.77%	97.77%	99.26%	1.0193
8	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE`	97.77%	97.77%	99.26%	1.0193
9	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE`	97.77%	97.77%	99.26%	1.0193
9	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA`	97.77%	97.77%	99.26%	1.0193
10	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA`	97.77%	97.77%	99.26%	1.0193

Table E.6: Experiment 56: Quasi-identifiers with the highest degradation of privacy in Collective-target Attribute-inference Single database (CAS) attacks on the School Census of 2019 for the sensitive attribute `IN_NECESSIDADE_ESPECIAL`. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* successes, while the probabilistic degradation of privacy is their ratio. Table E.1 lists the English meaning of each quasi-identifying attribute.

		Adversary's deterministic success		**Adversary's probabilistic success**	
		***a priori* success** 0.00%		***a priori* success** 82.06%	
#	**Quasi-identifiers**	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
1	`CO_ENTIDADE`	41.48%	41.48%	89.65%	1.0926
2	`CO_MUNICIPIO_NASC, CO_ENTIDADE`	51.25%	51.25%	90.82%	1.1067
2	`NU_ANO, CO_ENTIDADE`	55.44%	55.44%	90.71%	1.1054
3	`NU_MES, NU_ANO, CO_ENTIDADE`	74.66%	74.66%	92.85%	1.1315
4	`NU_MES, NU_ANO, CO_MUNICIPIO_NASC, CO_ENTIDADE`	84.29%	84.29%	95.23%	1.1605
5	`NU_MES, NU_ANO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_ENTIDADE`	89.34%	89.34%	96.43%	1.1751
6	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_ENTIDADE`	93.05%	93.05%	97.41%	1.1871
7	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE`	93.38%	93.38%	97.53%	1.1885
8	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE`	93.38%	93.38%	97.53%	1.1885
9	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE`	93.38%	93.38%	97.53%	1.1885
9	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA`	93.38%	93.38%	97.53%	1.1885
10	`NU_MES, NU_ANO, TP_SEXO, TP_COR_RACA, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_MUNICIPIO_NASC, CO_MUNICIPIO_END, CO_ENTIDADE, TP_DEPENDENCIA`	93.38%	93.38%	97.53%	1.1885

Table E.7: Experiment 56: Quasi-identifiers with the highest degradation of privacy in Collective-target Attribute-inference Single database (CAS) attacks on the School Census of 2019 for the sensitive attribute `IN_TRANSPORTE_PUBLICO`. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* successes, while the probabilistic degradation of privacy is their ratio. Table E.1 lists the English meaning of each quasi-identifying attribute.



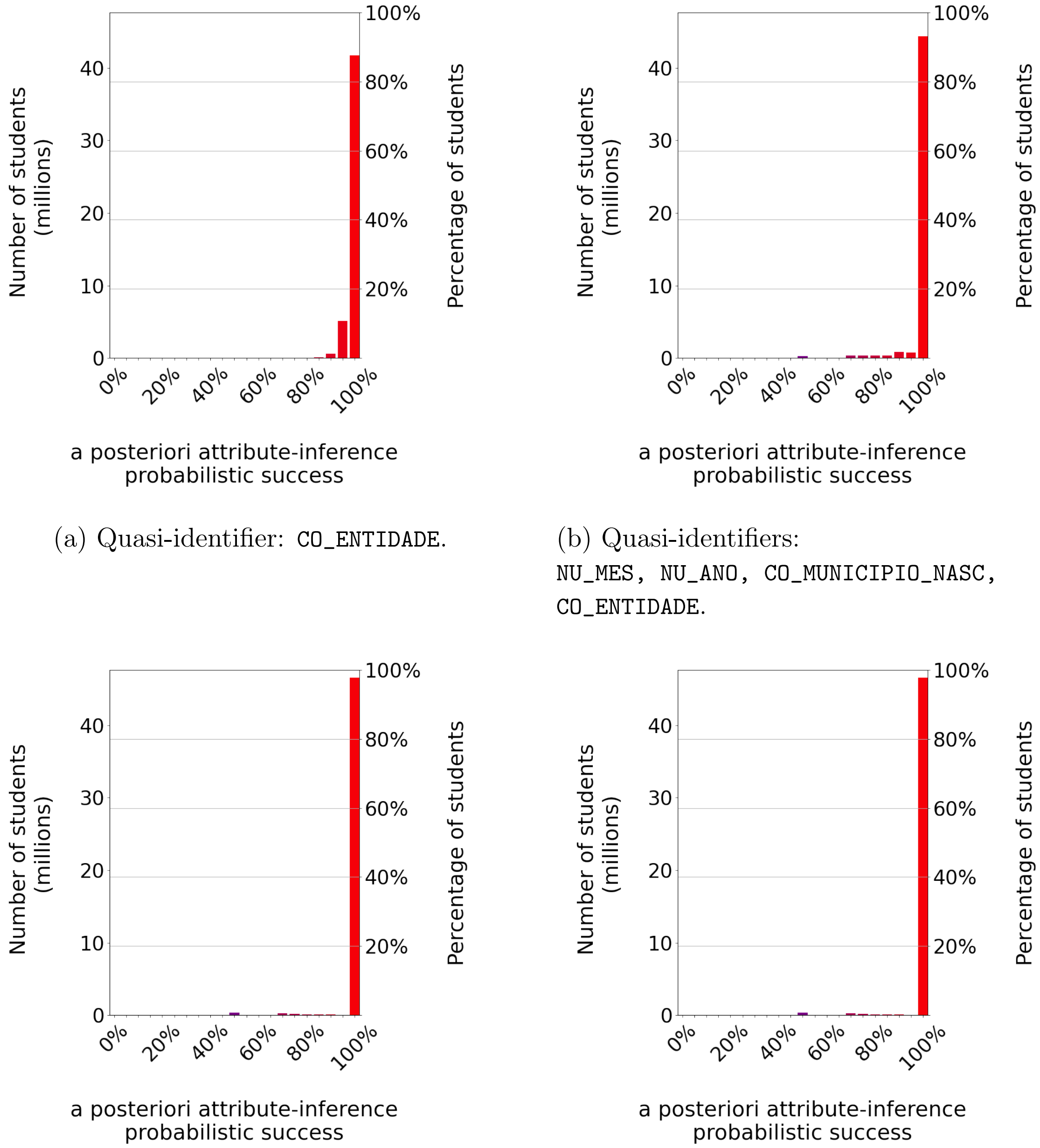
(a) Deterministic success for sensitive attribute `IN_NECESSIDADE_ESPECIAL`.

(b) Probabilistic success for sensitive attribute `IN_NECESSIDADE_ESPECIAL`.

(c) Deterministic success for sensitive attribute `IN_TRANSPORTE_PUBLICO`.

(d) Probabilistic success for sensitive attribute `IN_TRANSPORTE_PUBLICO`.

Figure E.8: Experiment 56: Adversary's success in Collective-target Attribute-inference Single database (CAS) attacks on the School Census of 2019. Here, the horizontal "*a priori*" line represents the adversary's *a priori* deterministic or probabilistic success, i.e. before performing the CAS attack. Each dot represents a distinct scenario defined by the selected quasi-identifying attributes among the 1 023 possibilities. The horizontal axis defines the size of the quasi-identifiers subset, while the vertical axis defines the adversary's deterministic or probabilistic success.



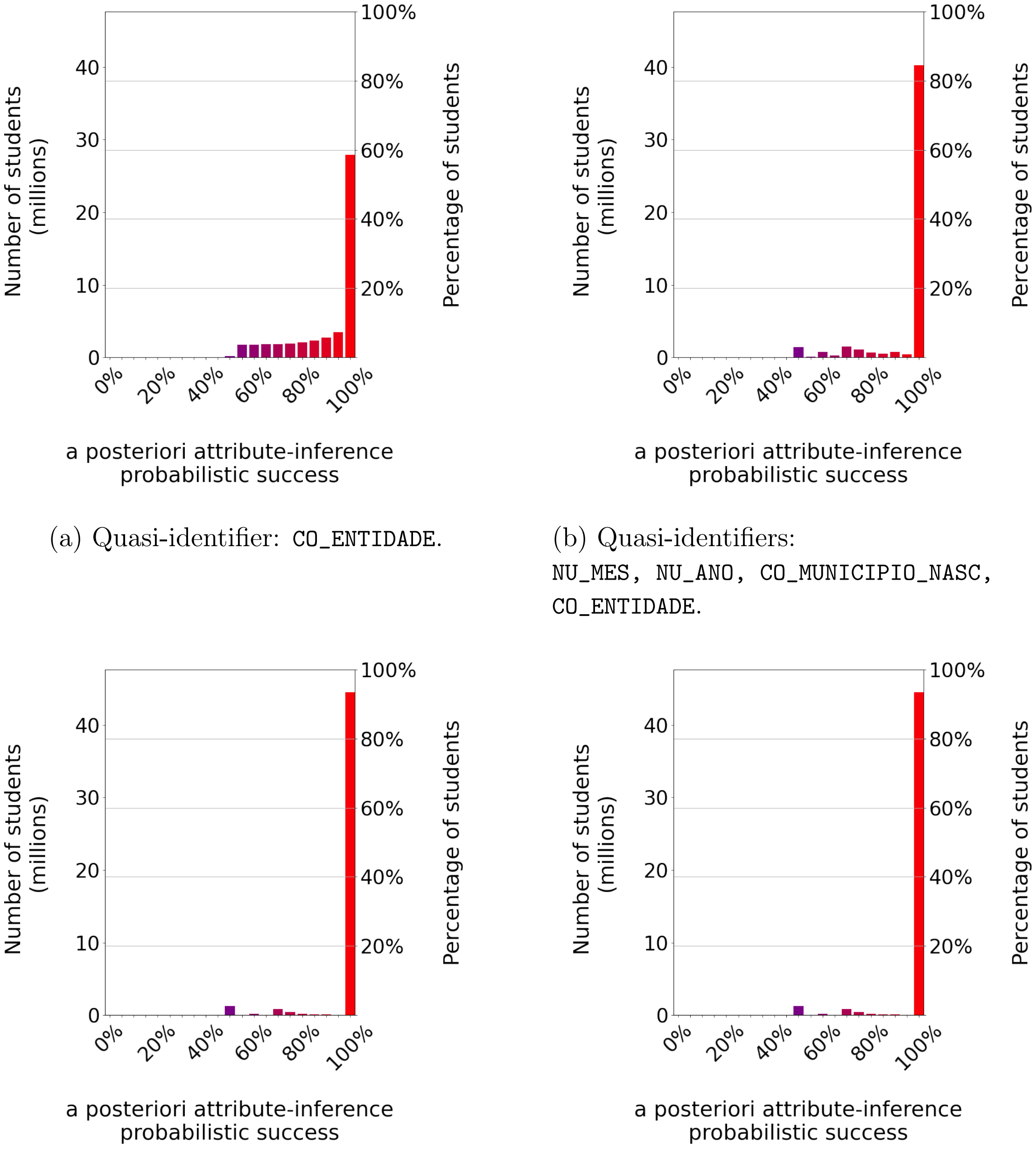
(a) Quasi-identifier: `CO_ENTIDADE`.

(b) Quasi-identifiers: `NU_MES`, `NU_ANO`, `CO_MUNICIPIO_NASC`, `CO_ENTIDADE`.

(c) Quasi-identifiers: `NU_MES`, `NU_ANO`, `TP_SEXO`, `TP_COR_RACA`, `CO_MUNICIPIO_NASC`, `CO_MUNICIPIO_END`, `CO_ENTIDADE`.

(d) Quasi-identifiers: `NU_MES`, `NU_ANO`, `TP_SEXO`, `TP_COR_RACA`, `TP_NACIONALIDADE`, `CO_PAIS_ORIGEM`, `CO_MUNICIPIO_NASC`, `CO_MUNICIPIO_END`, `CO_ENTIDADE`, `TP_DEPENDENCIA`.

Figure E.9: Experiment 56: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Attribute-inference Single database (CAS) attacks on the School Census of 2019 for the sensitive attribute `IN_NECESSIDADE_ESPECIAL`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of attribute-inference.



(a) Quasi-identifier: `CO_ENTIDADE`.

(b) Quasi-identifiers: `NU_MES`, `NU_ANO`, `CO_MUNICIPIO_NASC`, `CO_ENTIDADE`.

(c) Quasi-identifiers: `NU_MES`, `NU_ANO`, `TP_SEXO`, `TP_COR_RACA`, `CO_MUNICIPIO_NASC`, `CO_MUNICIPIO_END`, `CO_ENTIDADE`.

(d) Quasi-identifiers: `NU_MES`, `NU_ANO`, `TP_SEXO`, `TP_COR_RACA`, `TP_NACIONALIDADE`, `CO_PAIS_ORIGEM`, `CO_MUNICIPIO_NASC`, `CO_MUNICIPIO_END`, `CO_ENTIDADE`, `TP_DEPENDENCIA`.

Figure E.10: Experiment 56: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Attribute-inference Single database (CAS) attacks on the School Census of 2019 for the sensitive attribute `IN_TRANSPORTE_PUBLICO`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of attribute-inference.

E.4 Takeaways

In this appendix, we have presented experimental results for the collective-target attacks on the School Census of 2019 released by INEP, Experiments 55 and 56, for re-identification and attribute-inference attacks, respectively.

Those results have allowed us to draw the following conclusions.

- The **deterministic re-identification success** of an adversary with knowledge of only three attributes in Experiment 55, i.e. year and city of birth and school code, achieves up to 14.54% of the records, i.e. approximately 6 926 975 students that can be re-identified with absolute certainty. Adding the month of birth to the adversary's knowledge increases their success to 33.12%, i.e. approximately 15 778 640 students. A *plateau* is then reached when adding gender and ethnicity to the adversary's knowledge, which increases their re-identification success to 58.94%, i.e. approximately 28 079 500 students.

- The main difference between the two databases for the School Census of 2018 and 2019 is the removal of the day of birth attribute from the latter. As can be seen by comparing Figures 4.5 and E.4, the absence of this attribute greatly reduces the adversary's re-identification successes, with both a slower increase rate and a considerably lower *plateau* as the number of quasi-identifying attributes used increases. [1]

- Differently from Sweeney's findings in 2000 for the United States population, the removal of the day of birth attribute did not reduce so drastically the results for the deterministic re-identification success in Experiment 55. According to Sweeney [67], only 3.7% of the United States population in the Census of 1990 could be uniquely re-identified in the de-identified data considering the month and year of birth, gender, and a 5-digit ZIP code. Considering again the school code as an approximation of the ZIP code, 21.51% of the students in the School Census of 2019 can be re-identified by using the month and year of birth, gender, and school code. [2] [3]

[1] A similar behavior can be seen by comparing Figures 4.9 and E.8 for the adversary's attribute-inference successes.

[2] For Experiment 18 on the School Census of 2018, 21.76% of the students can be re-identified by using the month and year of birth, gender, and school code.

[3] The results mentioned here for the quasi-identifiers month and year of birth, gender, and school code are not available in Tables 4.4 and E.3 since they do not have the highest values among all the results for four attributes.

- The **probabilistic re-identification success** of an adversary with knowledge of only three attributes, i.e. month and year of birth and school code, is a 29.64% chance of correctly re-identifying a randomly selected record. Adding the city of birth to the adversary's knowledge increases their chance of success to 49.86%. A *plateau* is then reached when adding gender and ethnicity to the adversary's knowledge, which increases their re-identification success to 74.18%.

- The **deterministic attribute-inference success** of an adversary achieves higher values and increases more rapidly than the respective deterministic re-identification success. This was observed in Experiment 56 for both sensitive attributes `IN_NECESSIDADE_ESPECIAL` and `IN_TRANSPORTE_PUBLICO`, described in Table E.2. Even in the absence of the day of birth attribute in the School Census of 2019, Experiment 56, an adversary could correctly infer the value for both sensitive attributes of more than 90% of the records, or approximately 42 876 739 students, with either four quasi-identifiers for the sensitive attribute `IN_NECESSIDADE_ESPECIAL`, or six quasi-identifiers for the sensitive attribute `IN_TRANSPORTE_PUBLICO`.

- The **probabilistic attribute-inference success** of an adversary is initially already high. For the sensitive attribute on students' disabilities, an adversary has an *a priori* chance of correctly inferring the value for a randomly chosen record of 97.38% for the School Census of 2019. This is because of how much skewed the distribution of students is for that attribute. A more skewed distribution implies a higher *a priori* success, as is the case for the attribute on students' disabilities, for which the majority of individuals holds the value 0 (No). Analogously, the opposite holds for the attribute on students' use of public transport, which presents a slightly less skewed distribution and, hence, lower values for an adversary's *a priori* chance, of 82.06% for the School Census of 2019. [4] Therefore, not much increase in an adversary's success is seen as a result of the use of more quasi-identifiers for the attribute-inference attacks on both sensitive attributes. Even so, it is remarkable that by using six attributes in Experiment 56, the adversary's chance of correctly inferring the value of the sensitive attribute on students' disabilities for a randomly chosen record reaches values above 99%. [5]

[4] We have used average-case metrics for all of our analyses. Instead, a better approach would have been to use worse-case metrics whenever we are dealing with very skewed distributions [4, 5].

[5] For the sensitive attribute on students' use of public transport, an adversary has an *a priori* chance of correctly inferring the value for a randomly chosen record of 82.06% for the School Census of 2019. In this case, however, the adversary's chance of success reaches a *plateau* at 97.41% for six attributes in Experiment 56.

As expected, the removal by INEP of the variable NU_DIA for the student's day of birth as a measure to decrease the privacy risk to which students were subject was insufficient. In fact, the analyses of quasi-identifying attributes sets of different sizes allowed us to further demonstrate that the removal of attributes alone is insufficient to guarantee individuals' privacy. Furthermore, each attribute removed reduces the database's utility by reducing the amount of information available, which may impact the work of data analysts. Therefore, given the high privacy risks reported in our results and the recent enactment of Brazil's LGPD privacy legislation, mitigating such vulnerabilities is necessary and urgent for INEP.

Appendix F

Experimental results for collective-target attacks on the Higher Education Censuses

In this appendix, we present additional quantitative analyses on the privacy risks for collective-target attacks on single and longitudinal databases. But instead of the School Censuses, here we consider the Higher Education Censuses, also released by INEP. The databases for the Higher Education Censuses feature a sensitive attribute on the student use of financing that presents a far less skewed distribution of students. We verify this difference reduces the *a priori* probabilistic attribute-inference success for that attribute, and hence increases the privacy degradation attained by the adversary after performing their attack.

In Section F.1, we present the single database attacks on the Higher Education Censuses of 2018 and 2019. Those attacks were introduced in Chapter 4, in which the theoretical foundation was presented in Section 4.1. The equivalent results for attacks on the School Censuses of 2018 and 2019 were presented in Section 4.2 and Appendix E, respectively.

In Section F.2, we present the longitudinal database attacks on the Higher Education Censuses from 2014 to 2017. Those attacks were introduced in Chapter 5, in which the theoretical foundation was presented in Section 5.1. The equivalent results for attacks on the School Censuses from 2014 to 2017 were presented in Section 5.2.

F.1 Attacks on single databases

In this section, we present additional quantitative analyses on the privacy risks for individuals whose information are made public on a single database. The attacks performed here were modeled according to the theory developed in Section 4.1. For illustrative individual-target experimental results, see Section C.1.2.

F.1.1 Experimental setup

As stated in Section 1.3, we have chosen to work with the databases containing information on students and released as microdata, i.e. data at the record level. In this appendix, we consider the database for the Higher Education Census of 2018 and 2019. Also, in order to guarantee that each student holds only one record in each database, according to Assumption **AS1** from Section 4.1.1, we have randomly selected only one record for each data holder of multiple records.

Even though each database accounts for dozens of attributes, we have chosen just a few for our analysis, given the computational costs, including time and memory usage. The selection criteria was as follows.

- For the quasi-identifying attributes, we have sets restricted to the maximum of eleven attributes, selected according to how easily an adversary could learn them, e.g. the date and city of birth or the Higher Education Institution code. All the quasi-identifying attributes chosen for the experiments in this chapter are listed in Table F.1.

- We have chosen two sensitive attributes selected according to the possible individual privacy breach if revealed, e.g. whether or not the individual has special needs or disabilities. The sensitive attributes chosen for the experiments in this chapter are listed in Table F.2.

Of course, the selection of those quasi-identifiers and sensitive attributes is arbitrary and can change in significance over time. Nevertheless, the results here provided illustrate possible real-life circumstances and the respective privacy risks for the data holders.

F.1.2 Results of re-identification attack experiments

In this section, we present our quantitative analyses on the privacy risk of re-identification for individuals whose information are made public on a single database,

Variable	Meaning
NU_DIA_NASCIMENTO	Student's day of birth.
NU_MES_NASCIMENTO	Student's month of birth.
NU_ANO_NASCIMENTO	Student's year of birth.
TP_SEXO	Student's gender.
TP_COR_RACA	Student's ethnicity.
TP_NACIONALIDADE	Student's nationality.
CO_PAIS_ORIGEM	Student's country code.
CO_MUNICIPIO_NASCIMENTO	Student's city of birth code.
CO_IES	Higher Education Institution code.
CO_CURSO	Higher Education Institution course code to which the student is enrolled.
TP_ESCOLA_CONCLUSAO_ENS_MEDIO	Administrative dependency of the school in which the student completed High School, i.e. whether public or private.

Table F.1: Variables from the Higher Education Censuses of 2018 and 2019 chosen as quasi-identifiers for Collective-target Re-identification Single database (CRS) attacks in Experiments 57 and 58, and for Collective-target Attribute-inference Single database (CAS) attacks in Experiments 59 and 60.

Variable	Values	Meaning
IN_DEFICIENCIA	0 (No) 1 (Yes) 9 (No answer)	Whether or not the student possesses a disability or global developmental disorder.
IN_FINANCIAMENTO_ESTUDANTIL	-1 (Unavailable) 0 (No) 1 (Yes)	Whether or not the student uses student financing.

Table F.2: Variables from the Higher Education Censuses of 2018 and 2019 chosen as sensitive attributes for Collective-target Attribute-inference Single database (CAS) attacks in Experiments 59 and 60.

as modeled in Section 4.1.2. We present two CRS attacks, one on the Higher Education Census of 2018, Experiment 57, and the other on the Higher Education Census of 2019, Experiment 58.

Experiment 57 (Collective-target Re-identification Single database (CRS) attacks on the Higher Education Census of 2018)**.** In a CRS attack, the adversary's goal is to re-identify as many individuals as possible, according to Definition 16. Since we want to evaluate the overall re-identification risk, the adversary can gather auxiliary information on the values of a set of quasi-identifying attributes for every individual who holds a record in the database. The following experiment was conducted on the Higher Education Census of 2018.

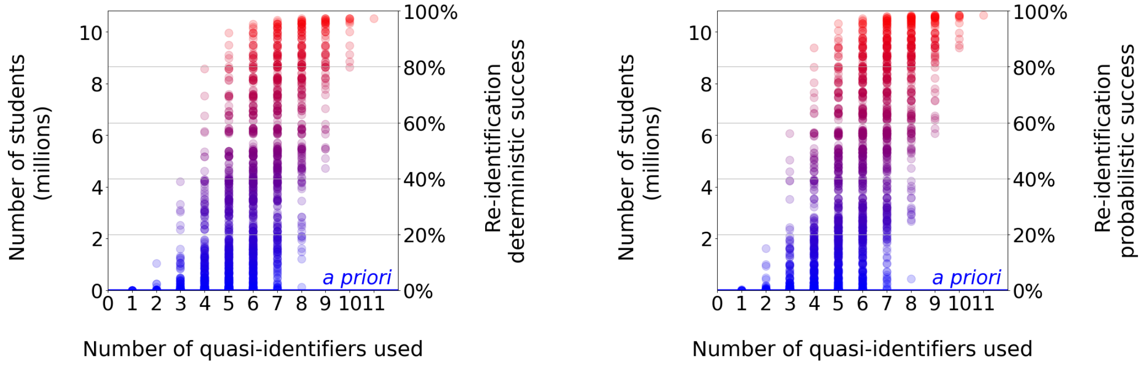
This database was released containing 12 043 994 records, which were reduced to 10 811 601 after the random selection of only one record for each data holder of multiple records. For the current experiment, we have selected as quasi-identifiers the eleven attributes listed in Table F.1. Considering an adversary could gather as auxiliary information any non-empty subset among the $2^{11} - 1 = 2\,047$ possible combinations, we have measured both the deterministic and probabilistic degradation of privacy for each combination.

The results are organized as follows.

- Table F.3 summarizes which combinations of quasi-identifiers result in the highest degradation of privacy according to the subset's size.
- Figure F.4 shows both the adversary's deterministic and probabilistic measures of success for each combination of quasi-identifying attributes.
- Figure F.5 shows the worst case histograms for the distribution of individuals according to the adversary's probabilistic measure of success for some subsets of quasi-identifying attributes with different sizes.

According to those results, we were able to conclude the following.

- **Adversary's deterministic success**. Defined in Equations 4.1 and 4.2, this metric measures the fraction of individuals in the database that can be re-identified with absolute certainty, in a scale from 0% to 100%. According to Table F.3, the adversary's *a priori* deterministic success equals 0%, i.e. no individual can be re-identified with absolute certainty without the use of auxiliary



(a) Deterministic success.

(b) Probabilistic success.

Figure F.4: Experiment 57: Adversary's success in Collective-target Re-identification Single database (CRS) attacks on the Higher Education Census of 2018. Here, the horizontal "*a priori*" line represents the adversary's *a priori* deterministic or probabilistic success, i.e. before performing the CRS attack. Each dot represents a distinct scenario defined by the selected quasi-identifying attributes among the 2 047 possibilities. The horizontal axis defines the size of the subset of quasi-identifiers, while the vertical axis defines the adversary's deterministic or probabilistic success.

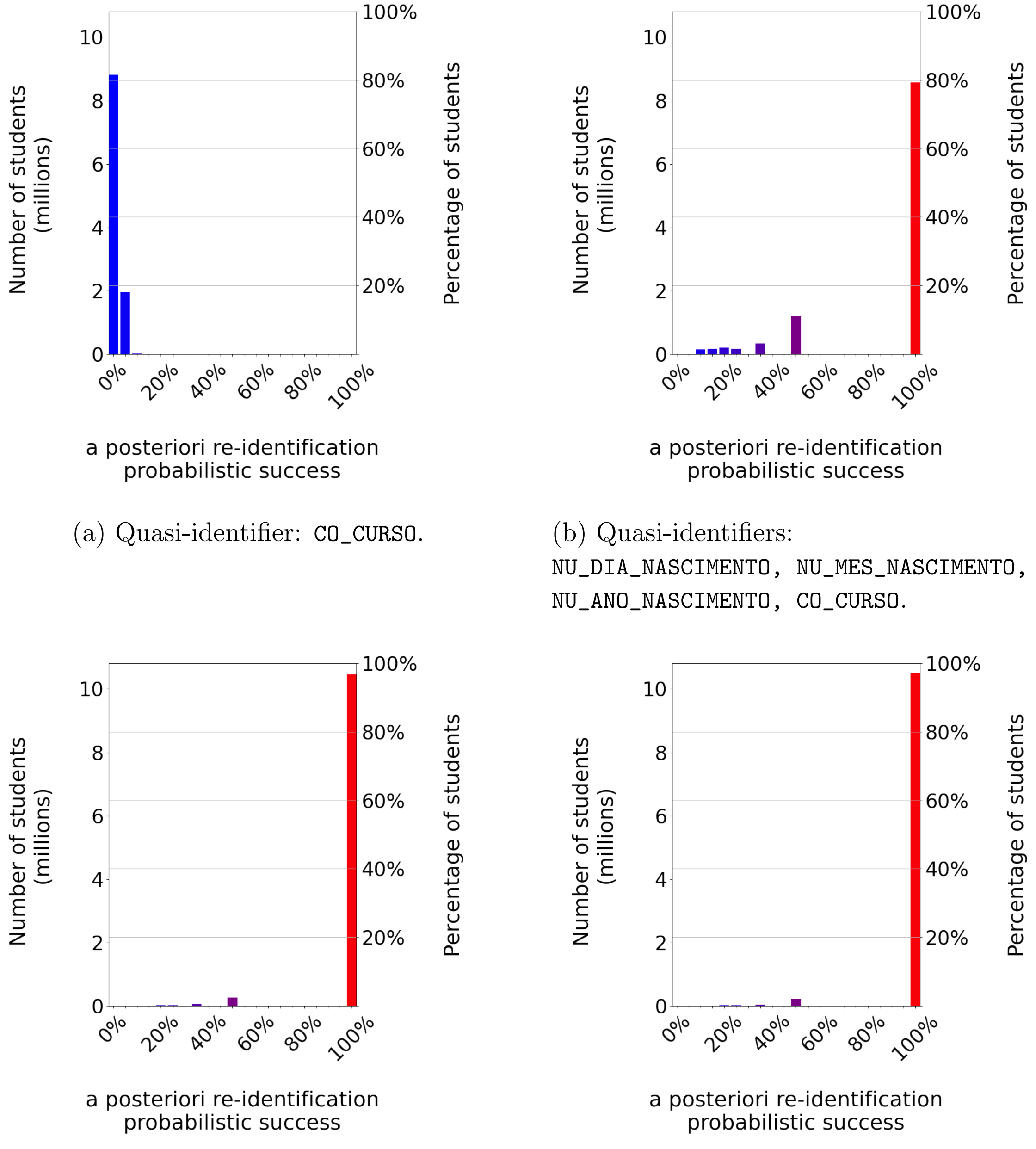
information. However, by using only three quasi-identifying attributes, the adversary can re-identify with absolute certainty up to 38.87% of the individuals in the database, while the use of four quasi-identifiers allows the adversary to re-identify up to 79.20% of the individuals. Finally, by using eight or more of the eleven quasi-identifying attributes available, the risk of re-identification increases to up to 97.22%.

- **Adversary's probabilistic success**. Defined in Equations 4.4 and 4.5, this metric measures the probability of a randomly chosen individual in the database being re-identified, in a scale from 0% to 100%. According to Table F.3, the adversary's *a priori* probabilistic success is approximately 0.000009%, i.e. the adversary's chance of re-identifying one of the 10 811 601 individuals in the database is almost zero. However, by using only three quasi-identifying attributes, the adversary increases their chance to up to 56.14%, while the use of four quasi-identifiers increases this probability to up to 86.85%. Finally, by using eight or more of the eleven quasi-identifying attributes available, the adversary's chance of re-identifying a randomly chosen individual in the database increases to up to 98.45%.

◁

		Adversary's deterministic success		Adversary's probabilistic success	
		a priori success 0.00%		*a priori* success 0.000009%	
#	**Quasi-identifiers**	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
1	CO_CURSO	0.005%	0.005%	0.35%	38 127
2	CO_MUNICIPIO_NASCIMENTO, CO_CURSO	9.71%	9.71%	14.99%	1 620 162
3	NU_DIA_NASCIMENTO, NU_ANO_NASCIMENTO, CO_CURSO	38.87%	38.87%	56.14%	6 069 562
4	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, CO_CURSO	79.20%	79.20%	86.85%	9 390 109
5	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, CO_MUNICIPIO_NASCIMENTO, CO_CURSO	92.13%	92.13%	95.42%	10 316 882
6	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO	95.32%	95.32%	97.36%	10 526 308
7	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO	96.66%	96.66%	98.15%	10 611 528
8	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO	97.22%	97.22%	98.45%	10 644 058
9	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO	97.22%	97.22%	98.45%	10 644 308
10	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO	97.22%	97.22%	98.45%	10 644 312
11	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_IES, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO	97.22%	97.22%	98.45%	10 644 312

Table F.3: Experiment 57: Quasi-identifiers with the highest degradation of privacy in Collective-target Re-identification Single database (CRS) attacks on the Higher Education Census of 2018. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* success, while the probabilistic degradation of privacy is their ratio. Table F.1 lists the English meaning of each quasi-identifying attribute.



(a) Quasi-identifier: `CO_CURSO`.

(b) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `CO_CURSO`.

(c) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `TP_SEXO`, `TP_COR_RACA`, `CO_MUNICIPIO_NASCIMENTO`, `CO_CURSO`.

(d) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `TP_SEXO`, `TP_COR_RACA`, `CO_MUNICIPIO_NASCIMENTO`, `TP_NACIONALIDADE`, `CO_PAIS_ORIGEM`, `CO_IES`, `CO_CURSO`, `TP_ESCOLA_CONCLUSAO_ENS_MEDIO`.

Figure F.5: Experiment 57: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Re-identification Single database (CRS) attacks on the Higher Education Census of 2018. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of re-identification.

We now present the CRS attacks performed on the Higher Education Census of 2019.

Experiment 58 (Collective-target Re-identification Single database (CRS) attacks on the Higher Education Census of 2019)**.** In a CRS attack, the adversary's goal is to re-identify as many individuals as possible, according to Definition 16. Since we want to evaluate the overall re-identification risk, the adversary can gather auxiliary information on the values of a set of quasi-identifying attributes for every individual who holds a record in the database. The following experiment was conducted on the Higher Education Census of 2019.

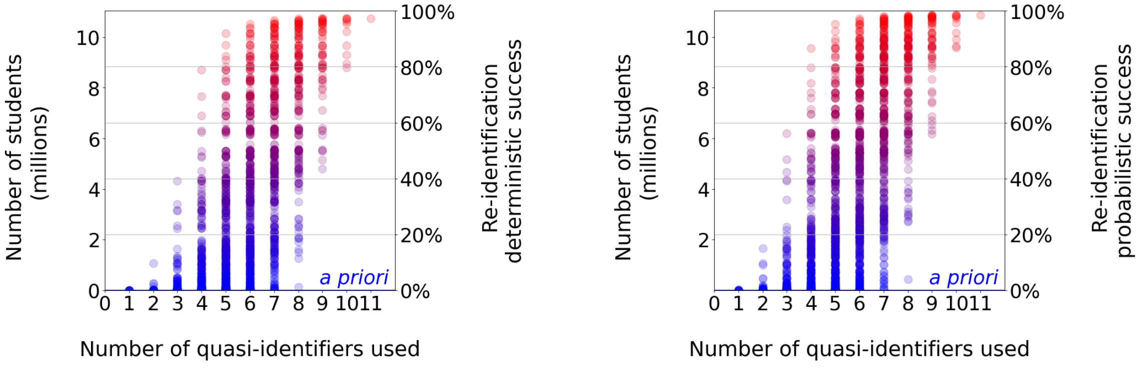
This database was released containing 12 350 832 records, which were reduced to 11 038 074 after the random selection of only one record for each data holder of multiple records. For the current experiment, we have selected as quasi-identifiers the eleven attributes listed in Table F.1. Considering an adversary could gather as auxiliary information any non-empty subset among the $2^{11} - 1 = 2\,047$ possible combinations, we have measured both the deterministic and probabilistic degradation of privacy for each combination.

The results are organized as follows.

- Table F.6 summarizes which combinations of quasi-identifiers result in the highest degradation of privacy according to the subset's size.
- Figure F.7 shows both the adversary's deterministic and probabilistic measures of success for each combination of quasi-identifying attributes.
- Figure F.8 shows the worst case histograms for the distribution of individuals according to the adversary's probabilistic measure of success for some subsets of quasi-identifying attributes with different sizes.

According to those results, we were able to conclude the following.

- **Adversary's deterministic success**. Defined in Equations 4.1 and 4.2, this metric measures the fraction of individuals in the database that can be re-identified with absolute certainty, in a scale from 0% to 100%. According to Table F.6, the adversary's *a priori* deterministic success equals 0%, i.e. no individual can be re-identified with absolute certainty without the use of auxiliary information. However, by using only three quasi-identifying attributes, the adversary can re-identified with absolute certainty up to 38.37% of the individuals in the database, while the use of four quasi-identifiers allows the adversary to



(a) Deterministic success.

(b) Probabilistic success.

Figure F.7: Experiment 58: Adversary's success in Collective-target Re-identification Single database (CRS) attacks on the Higher Education Census of 2019. Here, the horizontal "*a priori*" line represents the adversary's *a priori* deterministic or probabilistic success, i.e. before performing the CRS attack. Each dot represents a distinct scenario defined by the selected quasi-identifying attributes among the 2 047 possibilities. The horizontal axis defines the size of the subset of quasi-identifiers, while the vertical axis defines the adversary's deterministic or probabilistic success.

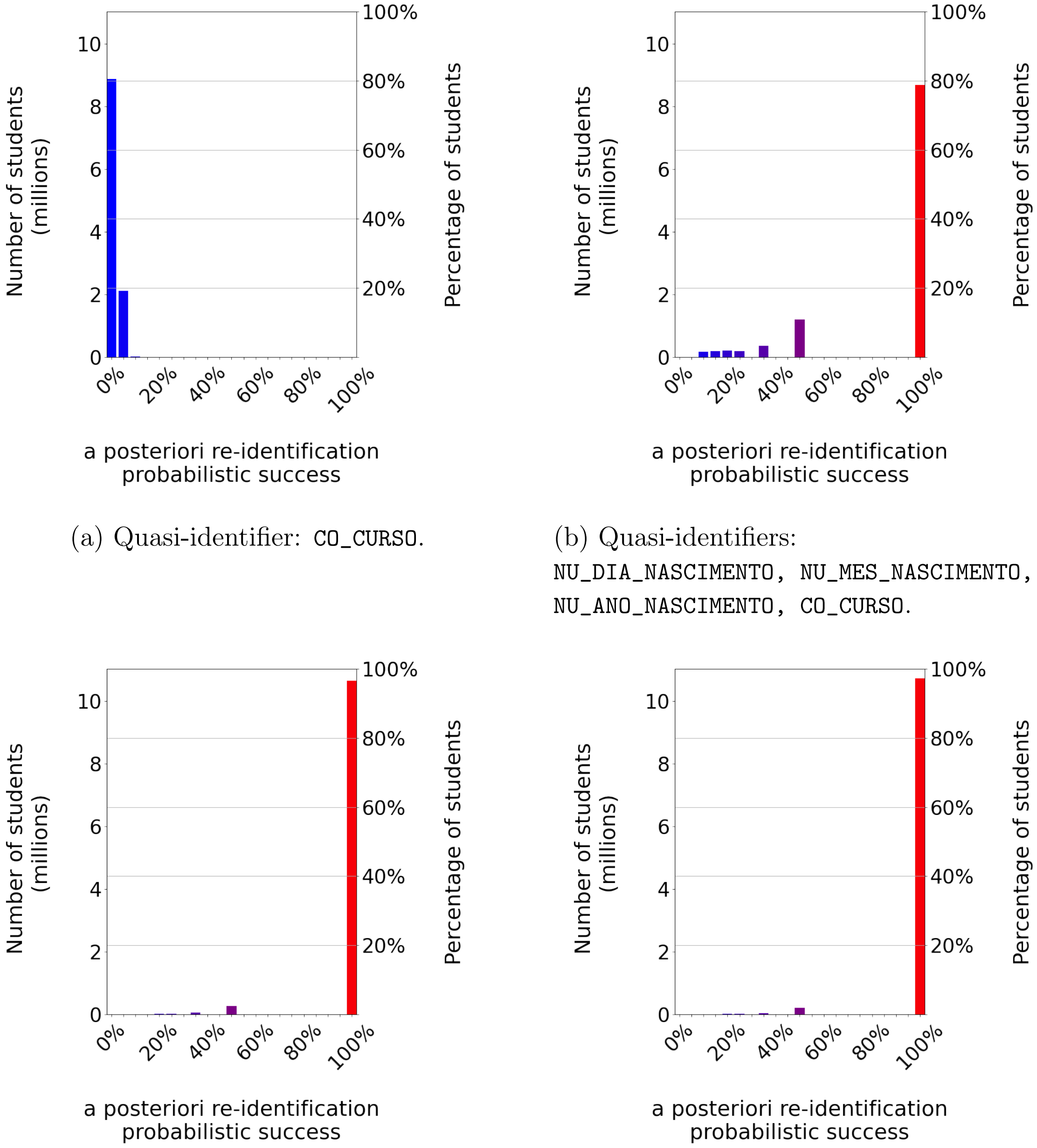
re-identify up to 79.20% of the individuals. Finally, by using eight or more of the eleven quasi-identifying attributes available, the risk of re-identification increases to up to 97.22%.

- **Adversary's probabilistic success**. Defined in Equations 4.4 and 4.5, this metric measures the probability of a randomly chosen individual in the database being re-identified, in a scale from 0% to 100%. According to Table F.6, the adversary's *a priori* probabilistic success is approximately 0.000009%, i.e. the adversary's chance of re-identifying one of the 11 038 074 individuals in the database is almost zero. However, by using only three quasi-identifying attributes, the adversary increases their chance to up to 56.14%, while the use of four quasi-identifiers increases this probability to up to 86.85%. Finally, by using eight or more of the eleven quasi-identifying attributes available, the adversary's chance of re-identifying a randomly chosen individual in the database increases to up to 98.45%.

◁

		Adversary's deterministic success		**Adversary's probabilistic success**	
		a priori **success** 0.00%		***a priori*** **success** 0.000009%	
#	**Quasi-identifiers**	***a posteriori*** **success**	**Privacy degradation**	***a posteriori*** **success**	**Privacy degradation**
1	`CO_CURSO`	0.005%	0.005%	0.37%	40 564
2	`CO_MUNICIPIO_NASCIMENTO, CO_CURSO`	9.70%	9.70%	14.93%	1 647 830
3	`NU_DIA_NASCIMENTO, NU_ANO_NASCIMENTO, CO_CURSO`	39.09%	39.09%	56.10%	6 192 177
4	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, CO_CURSO`	78.79%	78.79%	86.52%	9 549 875
5	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, CO_MUNICIPIO_NASCIMENTO, CO_CURSO`	91.94%	91.94%	95.29%	10 517 924
6	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO`	95.30%	95.30%	97.33%	10 743 462
7	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO`	96.61%	96.61%	98.10%	10 828 892
8	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`	97.21%	97.21%	98.43%	10 865 090
9	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`	97.21%	97.21%	98.43%	10 865 269
10	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`	97.21%	97.21%	98.43%	10 865 274
11	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_IES, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`	97.21%	97.21%	98.43%	10 865 274

Table F.6: Experiment 58: Quasi-identifiers with the highest degradation of privacy in Collective-target Re-identification Single database (CRS) attacks on the Higher Education Census of 2019. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* success, while the probabilistic degradation of privacy is their ratio. Table F.1 lists the English meaning of each quasi-identifying attribute.



(a) Quasi-identifier: `CO_CURSO`.

(b) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `CO_CURSO`.

(c) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `TP_SEXO`, `TP_COR_RACA`, `CO_MUNICIPIO_NASCIMENTO`, `CO_CURSO`.

(d) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `TP_SEXO`, `TP_COR_RACA`, `CO_MUNICIPIO_NASCIMENTO`, `TP_NACIONALIDADE`, `CO_PAIS_ORIGEM`, `CO_IES`, `CO_CURSO`, `TP_ESCOLA_CONCLUSAO_ENS_MEDIO`.

Figure F.8: Experiment 58: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Re-identification Single database (CRS) attacks on the Higher Education Census of 2019. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of re-identification.

F.1.3 Results of attribute-inference attack experiments

In this section, we present our quantitative analysis on the privacy risk of attribute-inference for individuals whose information are made public on a single database, as modeled in Section 4.1.2. We present two CAS attacks, one on the Higher Education Census of 2018, Experiment 59, and the other on the Higher Education Census of 2019, Experiment 60, both on sensitive attributes `IN_DEFICIENCIA` and `IN_FINANCIAMENTO_ESTUDANTIL`.

Experiment 59 (Collective-target Attribute-inference Single database (CAS) attacks on the Higher Education Census of 2018)**.** In a CAS attack, the adversary's goal is to infer the values of an attribute considered to be sensitive for as many individuals as possible, according to Definition 17. Since we want to evaluate the overall attribute-inference risk, the adversary can gather auxiliary information on the values of a set of quasi-identifying attributes for every individual who holds a record in the database. The following experiment was conducted on the Higher Education Census of 2018.

This database was released containing 12 043 994 records, which were reduced to 10 811 601 after the random selection of only one record for each data holder of multiple records. For the current experiment, we have selected as quasi-identifiers the eleven attributes listed in Table F.1. Considering an adversary could gather as auxiliary information any non-empty subset among the $2^{11} - 1 = 2\,047$ possible combinations, we have measured both the deterministic and probabilistic degradation of privacy for each combination.

Furthermore, we were interested in inferring the value for the attributes `IN_DEFICIENCIA` and `IN_FINANCIAMENTO_ESTUDANTIL`, both described in Table F.2 and considered by us to be sensitive.

The results are organized as follows.

- Tables F.9 and F.10 summarize which combinations of quasi-identifiers result in the highest degradation of privacy according to the subset's size for the sensitive attributes `IN_DEFICIENCIA` and `IN_FINANCIAMENTO_ESTUDANTIL`, respectively.
- Figure F.11 shows both the adversary's deterministic and probabilistic measures of success for each combination of quasi-identifying attributes, and for both sensitive attributes.
- Figures F.12 and F.13 show the worst case histograms for the distribution of individuals according to the adversary's probabilistic measure of success for some

subsets of quasi-identifying attributes with different sizes and for the sensitive attributes IN_DEFICIENCIA and IN_FINANCIAMENTO_ESTUDANTIL, respectively.

According to those results, we were able to conclude the following.

- **Adversary's deterministic success**. Defined in Equations 4.7 and 4.8, this metric measures the fraction of individuals in the database that can have their values for the sensitive attribute inferred with absolute certainty, in a scale from 0% to 100%.

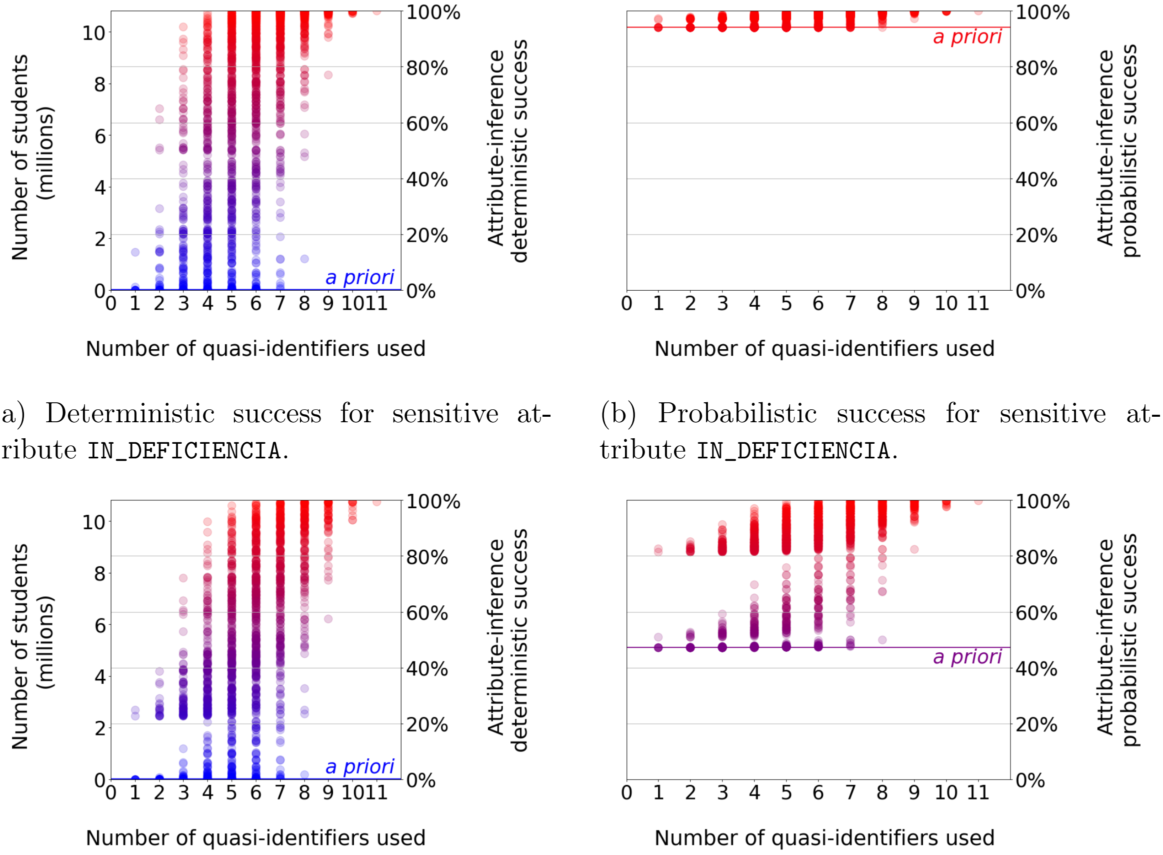
  According to Table F.9 for the sensitive attribute IN_DEFICIENCIA, the adversary's *a priori* deterministic success equals 0%, i.e. no individual can have their values for the sensitive attribute inferred with absolute certainty without the use of auxiliary information. However, by using only one quasi-identifying attribute, the adversary can infer the values with absolute certainty for up to 13.54% of the individuals in the database, while the use of two quasi-identifiers allows the adversary to infer the values for up to 65.03% of the individuals. Finally, by using four or more of the eleven quasi-identifying attributes available, the risk of re-identification increases above 99.28%.

  From Table F.10 for the sensitive attribute IN_FINANCIAMENTO_ESTUDANTIL, the adversary's *a priori* deterministic success also equals 0%. However, by using only one quasi-identifying attribute, the adversary can infer the values with absolute certainty for up to 24.93% of the individuals in the database, while the use of two quasi-identifiers allows the adversary to infer the values for up to 38.74% of the individuals. Finally, by using six or more of the eleven quasi-identifying attributes available, the risk of re-identification increases above 99.03%.

- **Adversary's probabilistic success**. Defined in Equations 4.11 and 4.12, this metric measures the probability of a randomly chosen individual in the database having their value for the sensitive attribute inferred with absolute certainty, in a scale from 0% to 100%.

  According to Table F.9 for the sensitive attribute IN_DEFICIENCIA, the adversary's *a priori* probabilistic success is already of 94.04%, i.e. the adversary's chance of inferring the value for the sensitive attribute for one of the 10 811 601 individuals in the database is already high. Furthermore, by using four or more quasi-identifying attributes, the adversary increases their chance above 99.75%.

  From Table F.10 for the sensitive attribute IN_FINANCIAMENTO_ESTUDANTIL, the adversary's *a priori* probabilistic success is of 47.28%, lower than that for the



(a) Deterministic success for sensitive attribute `IN_DEFICIENCIA`.

(b) Probabilistic success for sensitive attribute `IN_DEFICIENCIA`.

(c) Deterministic success for sensitive attribute `IN_FINANCIAMENTO_ESTUDANTIL`.

(d) Probabilistic success for sensitive attribute `IN_FINANCIAMENTO_ESTUDANTIL`.

Figure F.11: Experiment 59: Adversary's success in Collective-target Attribute-inference Single database (CAS) attacks on the Higher Education Census of 2018. Here, the horizontal "*a priori*" line represents the adversary's *a priori* deterministic or probabilistic success, i.e. before performing the CAS attack. Each dot represents a distinct scenario defined by the selected quasi-identifying attributes among the 2 047 possibilities. The horizontal axis defines the size of the subset of quasi-identifiers, while the vertical axis defines the adversary's deterministic or probabilistic success.

attribute `IN_DEFICIENCIA`. Furthermore, by using five or more quasi-identifying attributes, the adversary increases their chance above 99.20%.

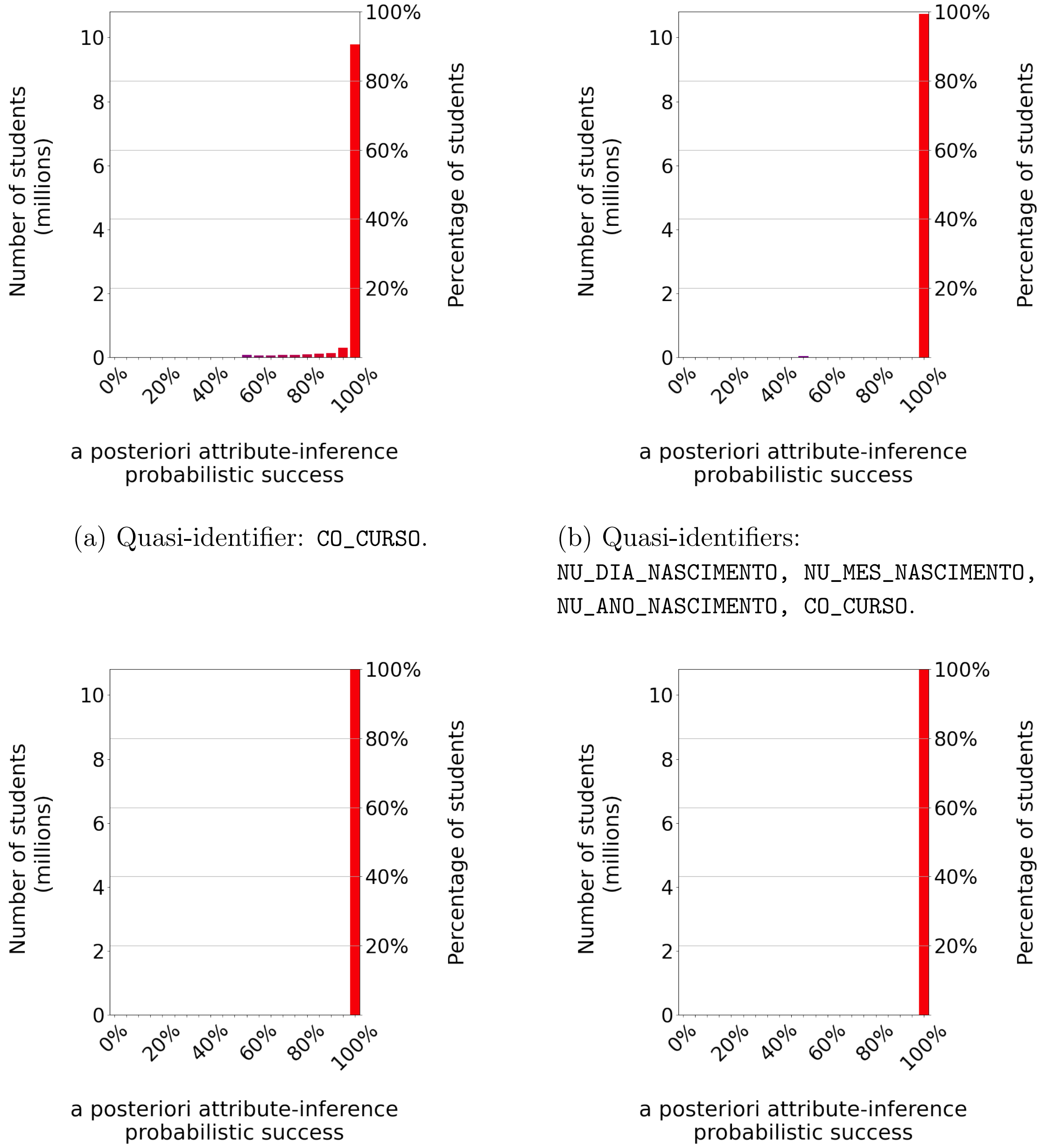
◁

		Adversary's deterministic success		**Adversary's probabilistic success**	
		a priori success 0.00%		*a priori* success 94.04%	
#	**Quasi-identifiers**	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
1	`CO_CURSO`	13.54%	13.54%	97.28%	1.0345
2	`CO_MUNICIPIO_NASCIMENTO, CO_CURSO`	50.29%	50.29%	97.80%	1.0400
2	`NU_DIA_NASCIMENTO, CO_CURSO`	65.03%	65.03%	97.45%	1.0363
3	`NU_DIA_NASCIMENTO, NU_ANO_NASCIMENTO, CO_CURSO`	94.23%	94.23%	98.90%	1.0517
4	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, CO_CURSO`	99.28%	99.28%	99.75%	1.0607
5	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, CO_MUNICIPIO_NASCIMENTO, CO_CURSO`	99.80%	99.80%	99.91%	1.0625
6	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO`	99.91%	99.91%	99.96%	1.0630
7	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO`	99.94%	99.94%	99.97%	1.0631
8	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`	99.96%	99.96%	99.98%	1.0632
9	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`	99.96%	99.96%	99.98%	1.0632
10	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`	99.96%	99.96%	99.98%	1.0632
11	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_IES, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`	99.96%	99.96%	99.98%	1.0632

Table F.9: Experiment 59: Quasi-identifiers with the highest degradation of privacy in Collective-target Attribute-inference Single database (CAS) attacks on the Higher Education Census of 2018 for the sensitive attribute `IN_DEFICIENCIA`. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* success, while the probabilistic degradation of privacy is their ratio. Table F.1 lists the English meaning of each quasi-identifying attribute.

		Adversary's deterministic success		**Adversary's probabilistic success**	
		***a priori* success** 0.00%		***a priori* success** 47.28%	
#	**Quasi-identifiers**	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
1	`CO_CURSO`	24.93%	24.93%	82.71%	1.7492
2	`CO_MUNICIPIO_NASCIMENTO, CO_CURSO`	38.74%	38.74%	85.26%	1.8032
3	`NU_DIA_NASCIMENTO, NU_ANO_NASCIMENTO, CO_CURSO`	72.23%	72.23%	91.28%	1.9305
4	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, CO_CURSO`	92.25%	92.25%	97.03%	2.0521
5	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, CO_MUNICIPIO_NASCIMENTO, CO_CURSO`	98.12%	98.12%	99.20%	2.0979
6	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO`	99.03%	99.03%	99.57%	2.1058
7	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO`	99.36%	99.36%	99.71%	2.1088
8	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`	99.52%	99.52%	99.78%	2.1103
9	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`	99.52%	99.52%	99.78%	2.1103
10	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`	99.52%	99.52%	99.78%	2.1103
10	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_IES, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`	99.52%	99.52%	99.78%	2.1103
11	`NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_IES, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`	99.52%	99.52%	99.78%	2.1103

Table F.10: Experiment 59: Quasi-identifiers with the highest degradation of privacy in Collective-target Attribute-inference Single database (CAS) attacks on the Higher Education Census of 2018 for the sensitive attribute `IN_FINANCIAMENTO_ESTUDANTIL`. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* success, while the probabilistic degradation of privacy is their ratio. Table F.1 lists the English meaning of each quasi-identifying attribute.



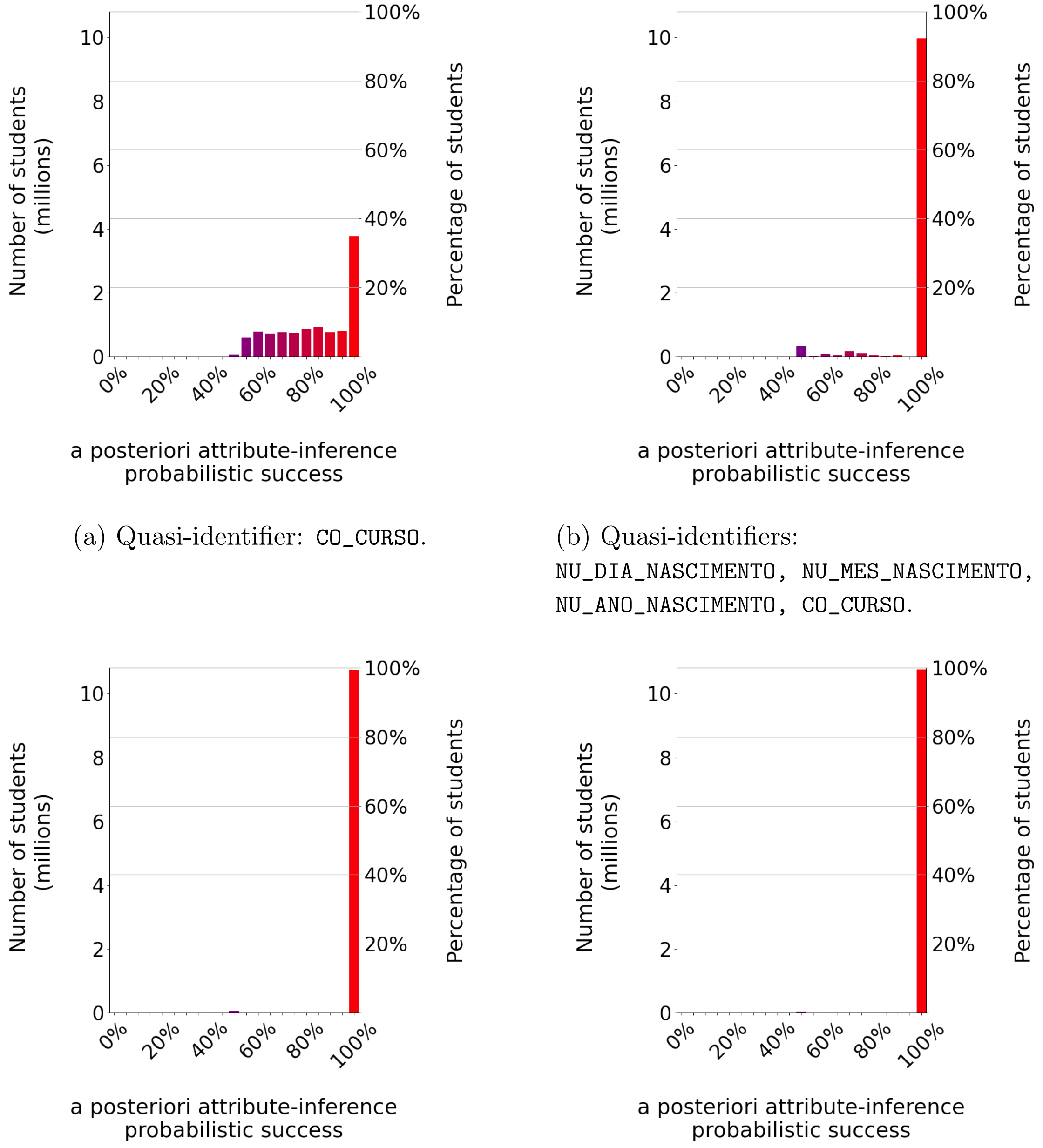
(a) Quasi-identifier: `CO_CURSO`.

(b) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `CO_CURSO`.

(c) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `TP_SEXO`, `TP_COR_RACA`, `CO_MUNICIPIO_NASCIMENTO`, `CO_CURSO`.

(d) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `TP_SEXO`, `TP_COR_RACA`, `CO_MUNICIPIO_NASCIMENTO`, `TP_NACIONALIDADE`, `CO_PAIS_ORIGEM`, `CO_IES`, `CO_CURSO`, `TP_ESCOLA_CONCLUSAO_ENS_MEDIO`.

Figure F.12: Experiment 59: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Attribute-inference Single database (CAS) attacks on the Higher Education Census of 2018 for the sensitive attribute `IN_DEFICIENCIA`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of attribute-inference.



(a) Quasi-identifier: `CO_CURSO`.

(b) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `CO_CURSO`.

(c) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `TP_SEXO`, `TP_COR_RACA`, `CO_MUNICIPIO_NASCIMENTO`, `CO_CURSO`.

(d) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `TP_SEXO`, `TP_COR_RACA`, `CO_MUNICIPIO_NASCIMENTO`, `TP_NACIONALIDADE`, `CO_PAIS_ORIGEM`, `CO_IES`, `CO_CURSO`, `TP_ESCOLA_CONCLUSAO_ENS_MEDIO`.

Figure F.13: Experiment 59: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Attribute-inference Single database (CAS) attacks on the Higher Education Census of 2018 for the sensitive attribute `IN_FINANCIAMENTO_ESTUDANTIL`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of attribute-inference.

We now present the CAS attacks performed on the Higher Education Census of 2019.

Experiment 60 (Collective-target Attribute-inference Single database (CAS) attacks on the Higher Education Census of 2019)**.** In a CAS attack, the adversary's goal is to infer the values of an attribute considered to be sensitive for as many individuals as possible, according to Definition 17. Since we want to evaluate the overall attribute-inference risk, the adversary can gather auxiliary information on the values of a set of quasi-identifying attributes for every individual who holds a record in the database. The following experiment was conducted on the Higher Education Census of 2019.

This database was released containing 12 350 832 records, which were reduced to 11 038 074 after the random selection of only one record for each data holder of multiple records. For the current experiment, we have selected as quasi-identifiers the eleven attributes listed in Table F.1. Considering an adversary could gather as auxiliary information any non-empty subset among the $2^{11} - 1 = 2\,047$ possible combinations, we have measured both the deterministic and probabilistic degradation of privacy for each combination.

Furthermore, we were interested in inferring the value for the attributes IN_DEFICIENCIA and IN_FINANCIAMENTO_ESTUDANTIL, both described in Table F.2 and considered by us to be sensitive.

The results are organized as follows.

- Tables F.14 and F.15 summarize which combinations of quasi-identifiers result in the highest degradation of privacy according to the subset's size for the sensitive attributes IN_DEFICIENCIA and IN_FINANCIAMENTO_ESTUDANTIL, respectively.
- Figure F.16 shows both the adversary's deterministic and probabilistic measures of success for each combination of quasi-identifying attributes, and for both sensitive attributes.
- Figures F.17 and F.18 show the worst case histograms for the distribution of individuals according to the adversary's probabilistic measure of success for some subsets of quasi-identifying attributes with different sizes and for the sensitive attributes IN_DEFICIENCIA and IN_FINANCIAMENTO_ESTUDANTIL, respectively.

According to those results, we were able to conclude the following.

- **Adversary's deterministic success**. Defined in Equations 4.7 and 4.8, this metric measures the fraction of individuals in the database that can have their values for the sensitive attribute inferred with absolute certainty, in a scale from 0% to 100%.

According to Table F.14 for the sensitive attribute `IN_DEFICIENCIA`, the adversary's *a priori* deterministic success equals 0%, i.e. no individual can have their values for the sensitive attribute inferred with absolute certainty without the use of auxiliary information. However, by using only one quasi-identifying attribute, the adversary can infer the values with absolute certainty for up to 15.24% of the individuals in the database, while the use of two quasi-identifiers allows the adversary to infer the values for up to 66.10% of the individuals. Finally, by using four or more of the eleven quasi-identifying attributes available, the risk of attribute-inference increases above 99.38%.

From Table F.15 for the sensitive attribute `IN_FINANCIAMENTO_ESTUDANTIL`, the adversary's *a priori* deterministic success also equals 0%. However, by using only one quasi-identifying attributes, the adversary can infer the values with absolute certainty for up to 25.06% of the individuals in the database, while the use of two quasi-identifiers allows the adversary to infer the values for up to 39.14% of the individuals. Finally, by using four or more of the eleven quasi-identifying attributes available, the risk of attribute-inference increases above 91.76%.

- **Adversary's probabilistic success**. Defined in Equations 4.11 and 4.12, this metric measures the probability of a randomly chosen individual in the database having their value for the sensitive attribute inferred with absolute certainty, in a scale from 0% to 100%.

  According to Table F.14 for the sensitive attribute `IN_DEFICIENCIA`, the adversary's *a priori* probabilistic success is already of 94.84%, i.e. the adversary's chance of inferring the value for the sensitive attribute for one of the 11 038 074 individuals in the database is already high. Furthermore, by using seven or more quasi-identifying attributes, the adversary increases their chance to up to 99.98%.

  From Table F.15 for the sensitive attribute `IN_FINANCIAMENTO_ESTUDANTIL`, the adversary's *a priori* probabilistic success is of 47.80%. Furthermore, by using only two quasi-identifying attributes, the adversary increases their chance to 86.85%, while the use of four quasi-identifiers allows the adversary to infer the values for up to 97.11% of the individuals. Finally, by using seven or more of the eleven quasi-identifying attributes available, the risk of attribute-inference increases above 99.71%.

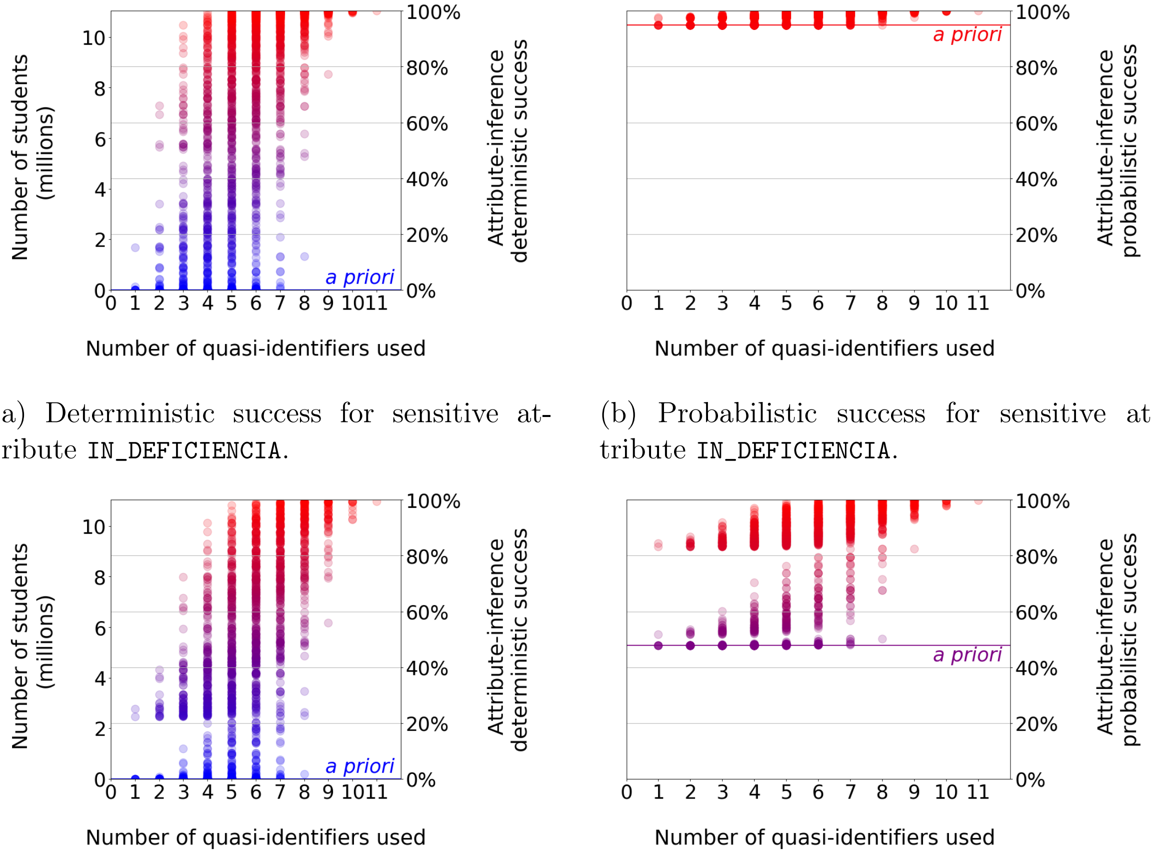
◁

		Adversary's deterministic success		**Adversary's probabilistic success**	
		a priori **success** 0.00%		***a priori*** **success** 94.84%	
#	**Quasi-identifiers**	***a posteriori*** **success**	**Privacy degradation**	***a posteriori*** **success**	**Privacy degradation**
1	CO_CURSO	15.24%	15.24%	97.58%	1.0289
2	NU_DIA_NASCIMENTO, CO_CURSO	66.10%	66.10%	97.75%	1.0307
2	CO_MUNICIPIO_NASCIMENTO, CO_CURSO	51.15%	51.15%	98.04%	1.0338
3	NU_DIA_NASCIMENTO, NU_ANO_NASCIMENTO, CO_CURSO	94.93%	94.93%	99.06%	1.0445
4	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, CO_CURSO	99.38%	99.38%	99.79%	1.0522
5	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, CO_MUNICIPIO_NASCIMENTO, CO_CURSO	99.83%	99.83%	99.93%	1.0537
6	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO	99.92%	99.92%	99.96%	1.0541
7	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO	99.95%	99.95%	99.98%	1.0542
8	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO	99.96%	99.96%	99.98%	1.0543
9	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO	99.96%	99.96%	99.98%	1.0543
10	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO	99.96%	99.96%	99.98%	1.0543
11	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_IES, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO	99.96%	99.96%	99.98%	1.0543

Table F.14: Experiment 60: Quasi-identifiers with the highest degradation of privacy in Collective-target Attribute-inference Single database (CAS) attacks on the Higher Education Census of 2019 for the sensitive attribute IN_DEFICIENCIA. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* success, while the probabilistic degradation of privacy is their ratio. Table F.1 lists the English meaning of each quasi-identifying attribute.

		Adversary's deterministic success		**Adversary's probabilistic success**	
		***a priori* success** 0.00%		***a priori* success** 47.80%	
#	**Quasi-identifiers**	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
1	CO_CURSO	25.06%	25.06%	84.43%	1.7663
2	CO_MUNICIPIO_NASCIMENTO, CO_CURSO	39.14%	39.14%	86.85%	1.8167
3	NU_DIA_NASCIMENTO, NU_ANO_NASCIMENTO, CO_CURSO	72.37%	72.37%	91.98%	1.9240
4	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, CO_CURSO	91.76%	91.76%	97.11%	2.0315
5	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, CO_MUNICIPIO_NASCIMENTO, CO_CURSO	97.98%	97.98%	99.19%	2.0749
6	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO	99.04%	99.04%	99.58%	2.0831
7	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO	99.36%	99.36%	99.71%	2.0859
8	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO	99.52%	99.52%	99.78%	2.0874
9	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO	99.52%	99.52%	99.78%	2.0874
10	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO	99.52%	99.52%	99.78%	2.0874
10	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_IES, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO	99.52%	99.52%	99.78%	2.0874
11	NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_IES, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO	99.52%	99.52%	99.78%	2.0874

Table F.15: Experiment 60: Quasi-identifiers with the highest degradation of privacy in Collective-target Attribute-inference Single database (CAS) attacks on the Higher Education Census of 2019 for the sensitive attribute IN_FINANCIAMENTO_ESTUDANTIL. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* success, while the probabilistic degradation of privacy is their ratio. Table F.1 lists the English meaning of each quasi-identifying attribute.



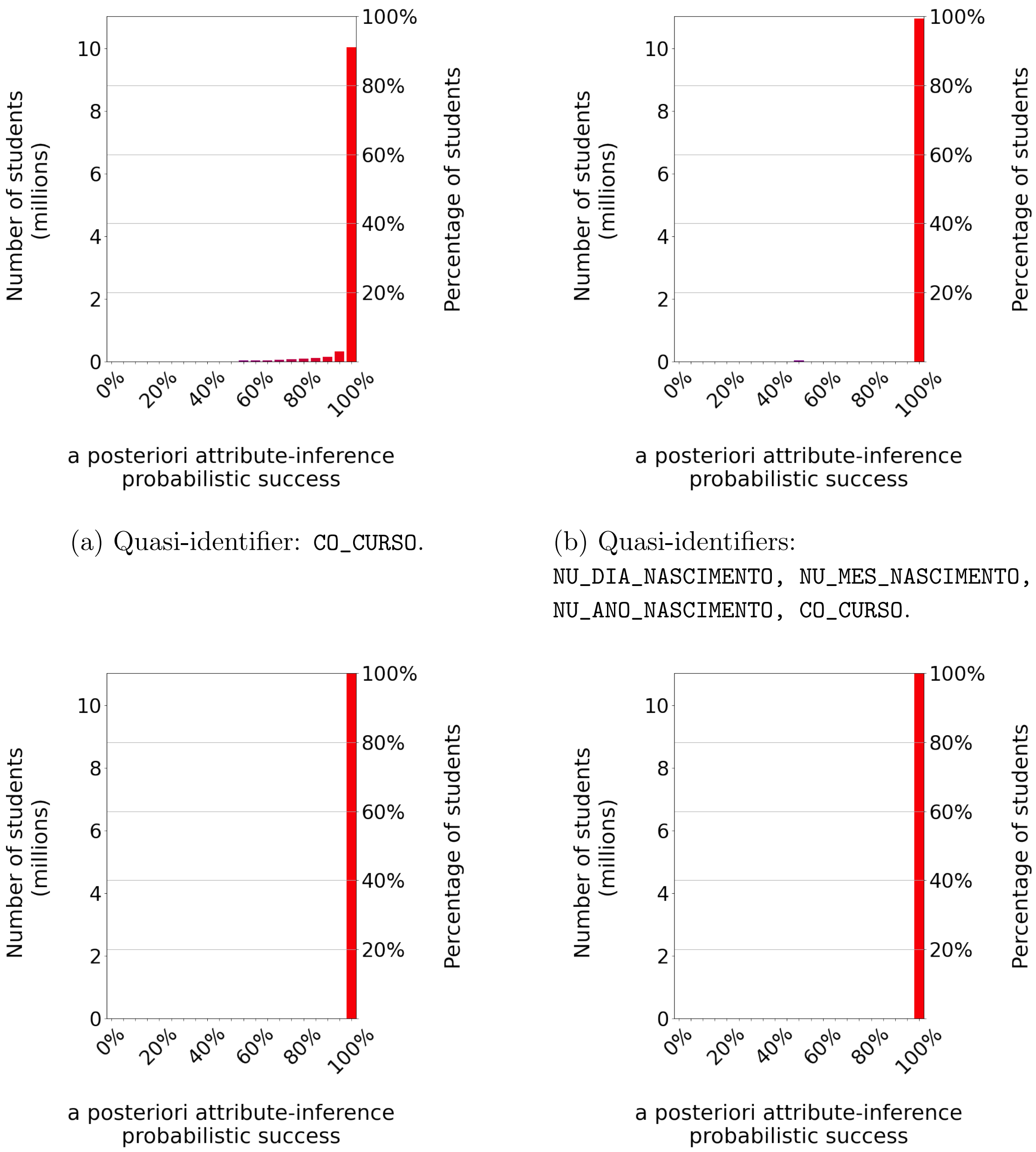
(a) Deterministic success for sensitive attribute IN_DEFICIENCIA.

(b) Probabilistic success for sensitive attribute IN_DEFICIENCIA.

(c) Deterministic success for sensitive attribute IN_FINANCIAMENTO_ESTUDANTIL.

(d) Probabilistic success for sensitive attribute IN_FINANCIAMENTO_ESTUDANTIL.

Figure F.16: Experiment 60: Adversary's success in Collective-target Attribute-inference Single database (CAS) attacks on the Higher Education Census of 2019. Here, the horizontal "*a priori*" line represents the adversary's *a priori* deterministic or probabilistic success, i.e. before performing the CAS attack. Each dot represents a distinct scenario defined by the selected quasi-identifying attributes among the 2 047 possibilities. The horizontal axis defines the size of the subset of quasi-identifiers, while the vertical axis defines the adversary's deterministic or probabilistic success.



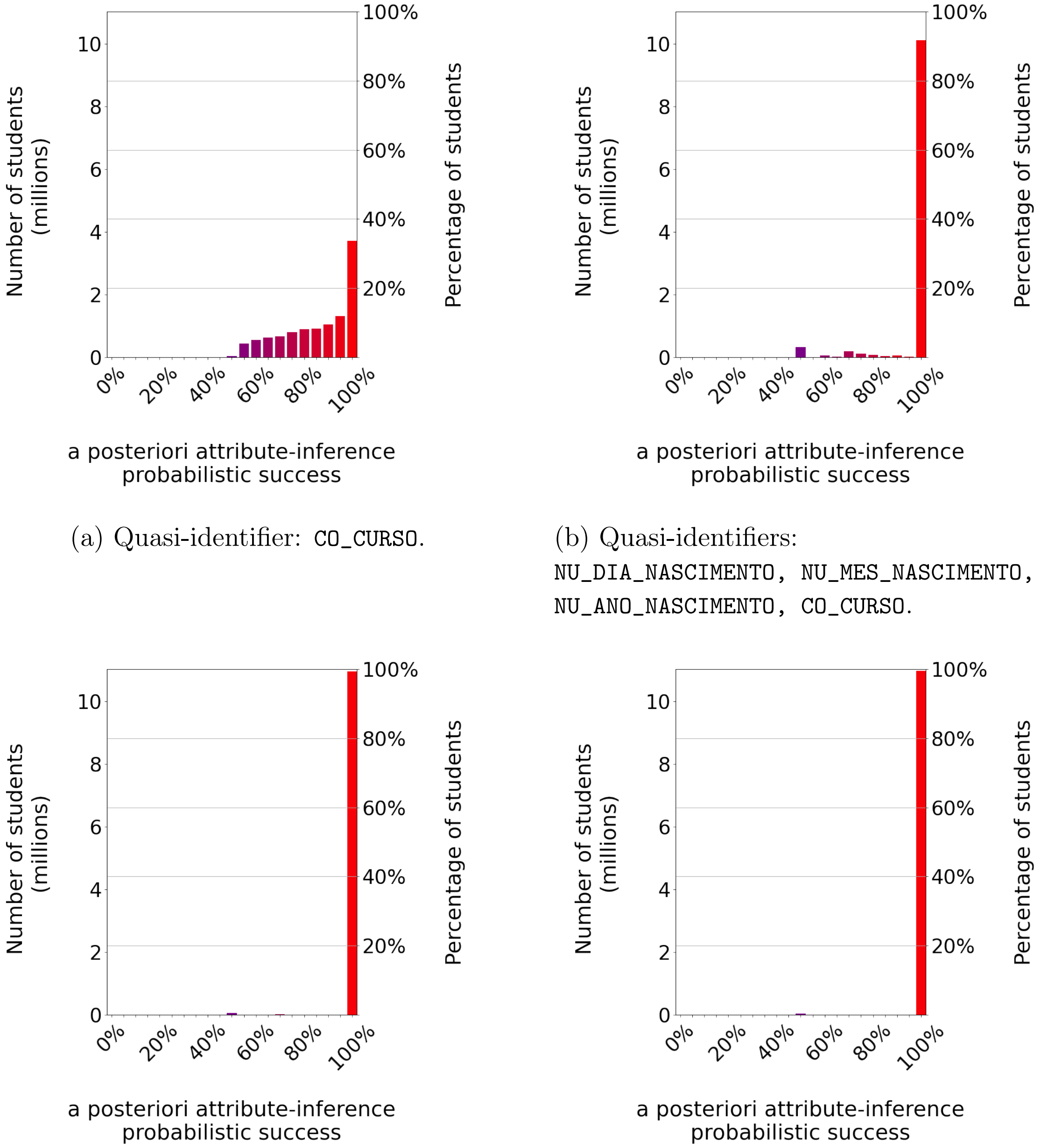
(a) Quasi-identifier: `CO_CURSO`.

(b) Quasi-identifiers: `NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, CO_CURSO`.

(c) Quasi-identifiers: `NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, CO_CURSO`.

(d) Quasi-identifiers: `NU_DIA_NASCIMENTO, NU_MES_NASCIMENTO, NU_ANO_NASCIMENTO, TP_SEXO, TP_COR_RACA, CO_MUNICIPIO_NASCIMENTO, TP_NACIONALIDADE, CO_PAIS_ORIGEM, CO_IES, CO_CURSO, TP_ESCOLA_CONCLUSAO_ENS_MEDIO`.

Figure F.17: Experiment 60: Histograms for the distribution of individuals according to the adversary's measures of success in Collective-target Attribute-inference Single database (CAS) attacks on the Higher Education Census of 2019 for the sensitive attribute `IN_DEFICIENCIA`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of attribute-inference.



(a) Quasi-identifier: `CO_CURSO`.

(b) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `CO_CURSO`.

(c) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `TP_SEXO`, `TP_COR_RACA`, `CO_MUNICIPIO_NASCIMENTO`, `CO_CURSO`.

(d) Quasi-identifiers: `NU_DIA_NASCIMENTO`, `NU_MES_NASCIMENTO`, `NU_ANO_NASCIMENTO`, `TP_SEXO`, `TP_COR_RACA`, `CO_MUNICIPIO_NASCIMENTO`, `TP_NACIONALIDADE`, `CO_PAIS_ORIGEM`, `CO_IES`, `CO_CURSO`, `TP_ESCOLA_CONCLUSAO_ENS_MEDIO`.

Figure F.18: Experiment 60: Histograms for the distribution of individuals according to the adversary's measures of success in Collective-target Attribute-inference Single database (CAS) attacks on the Higher Education Census of 2019 for the sensitive attribute `IN_FINANCIAMENTO_ESTUDANTIL`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of attribute-inference.

F.1.4 Takeaways

In this appendix, we have presented experimental results for the collective-target attacks on the Higher Education Censuses of 2018 and 2019 released by INEP, Experiments 57 and 58 for re-identification, and Experiments 59 and 60 for attribute-inference attacks, respectively.

Those results have allowed us to draw the following conclusions.

- The measures of **deterministic** and **probabilistic** successes, for both **re-identification** and **attribute-inference** attacks, have presented similar values for both the Higher Education Census of 2018 and 2019. Furthermore, the results are similar to those reported for the School Census of 2018 in Section 4.2, Experiments 18 and 19.
- The results observed for the *a priori* probabilistic attribute-inference successes for different sensitive attributes are a consequence of how much skewed the respective distributions of students are for those attributes. A more skewed distribution implies a higher *a priori* success. For instance, the attribute on whether a student uses student financing or not allows for three different values and has the less skewed distribution among all the sensitive attributes analyzed in this thesis, reporting *a priori* successes of 47.28% and 47.80% for the Higher Education Censuses of 2018 and 2019, respectively.

F.2 Attacks on longitudinal databases

In this section, we present additional quantitative analyses on the privacy risks for individuals whose information are made public on longitudinal databases. The attacks performed here were modeled according to the theory developed in Section 5.1. For illustrative individual-target experimental results, see Section C.2.2.

F.2.1 Experimental setup

As stated in Section 1.3, we have chosen to work with the databases containing information on students and released as microdata, i.e. data at the record level. In this appendix, we consider databases from the Higher Education Censuses. Also, in order to guarantee that each student holds only one record in each database, according to Assumption **AL1** from Section 5.1.1, we have randomly selected only one record for each data holder of multiple records.

Variable	Meaning
CO_IES	Higher Education Institution code.
CO_CURSO	Higher Education Institution course code to which the student is enrolled.

Table F.19: Variables from the Higher Education Census of 2014 chosen as quasi-identifiers for Collective-target Re-identification Longitudinal databases (CRL) attacks in Experiment 61 and for Collective-target Attribute-inference Longitudinal databases (CAL) attacks in Experiment 62.

Even though each database accounts for dozens of attributes, we have chosen just a few for our analysis given the computational costs, including time and memory usage. The selection criteria was as follows.

- We have chosen two quasi-identifying attributes selected according to how easily an adversary could learn them and to their variability throughout the years, i.e. whether or not they could change and hence be captured in longitudinal databases. The quasi-identifying attributes chosen for the experiments in this chapter are listed in Table F.19.
- We have also chosen two sensitive attributes selected according to the possible individual privacy breach if revealed, e.g. whether or not the individual has special needs or disabilities. The sensitive attributes chosen for the experiments in this chapter are listed in Table F.20.

Of course, the selection of those quasi-identifiers and sensitive attributes is arbitrary and can change in significance over time. Nevertheless, the results here provided illustrate possible real-life circumstances and the respective privacy risks for the data holders.

F.2.2 Results of re-identification attack experiments

In this section, we present our quantitative analyses on the privacy risk of re-identification for individuals whose information are made public on longitudinal databases, as modeled in Section 4.1.2. We present one CRL attack on the Higher Education Censuses from 2014 to 2017, Experiment 61.

Experiment 61 (Collective-target Re-identification Longitudinal databases (CRL) attack on the Higher Education Censuses)**.** In a CRL attack, the adversary's goal is to re-identify as many individuals as possible, according to Definition 21. Since we

Variable	Values	Meaning
IN_ALUNO_DEF_TGD_SUPER	0 (No) 1 (Yes) 2 (Unavailable)	Whether or not the student possesses a disability or global developmental disorder.
IN_FINANC_ESTUDANTIL	-1 (Unavailable) 0 (No) 1 (Yes)	Whether or not the student uses student financing.

Table F.20: Variables from the Higher Education Census of 2014 chosen as sensitive attributes for Collective-target Attribute-inference Longitudinal databases (CAL) attacks in Experiment 62.

want to evaluate the overall re-identification risk, the adversary can gather auxiliary information on the values of a set of quasi-identifying attributes for every individual who holds a record in the focal database. The following experiment was conducted on the Higher Education Censuses from 2014 to 2017 and had the database of 2014 as the focal one.

The focal database was released containing 10 793 936 records, which were reduced to 9 773 350 after the random selection of only one record for each data holder of multiple records. For the current experiment, we have selected as quasi-identifiers the attributes CO_IES and CO_CURSO, as specified in Table F.19. We have measured both the deterministic and probabilistic degradation of privacy for the set composed of those three attributes.

The results are organized as follows.

- Table F.21 summarizes the adversary's measures of success and privacy degradation as more databases are aggregated to the focal database.
- Figure F.22 shows the adversary's deterministic and probabilistic measures of success according to Table F.21.
- Figure F.23 shows the histograms for the distribution of individuals according to the adversary's probabilistic measure of success as more databases are aggregated to the focal database.

According to those results, we were able to conclude the following.

- **Adversary's deterministic success**. Defined in Equations 5.1 and 5.2, this metric measures the fraction of individuals in the focal database that can be re-identified with absolute certainty, in a scale from 0% to 100%. According to Table F.21, the adversary's *a priori* deterministic success equals 0%, i.e. no individual can be re-identified with absolute certainty in the focal database without the use of auxiliary information. However, by using only the attributes CO_IES and CO_CURSO as quasi-identifiers and only the database of 2014, the adversary can re-identify with absolute certainty 0.0044% of the individuals in the focal database, i.e. approximately 430 students.

As the longitudinal set of databases available to the adversary increases, allowing the use of more Higher Education Census databases as auxiliary information, so increases the fraction of individuals in the focal database that can be re-identified with absolute certainty. Particularly, the adversary's *a posteriori* deterministic success increases to 5.11% by using only the database of 2015 as auxiliary information, to 10.33% by using the databases of 2015 and 2016, and to 15.25% by using the databases from 2015 to 2017.

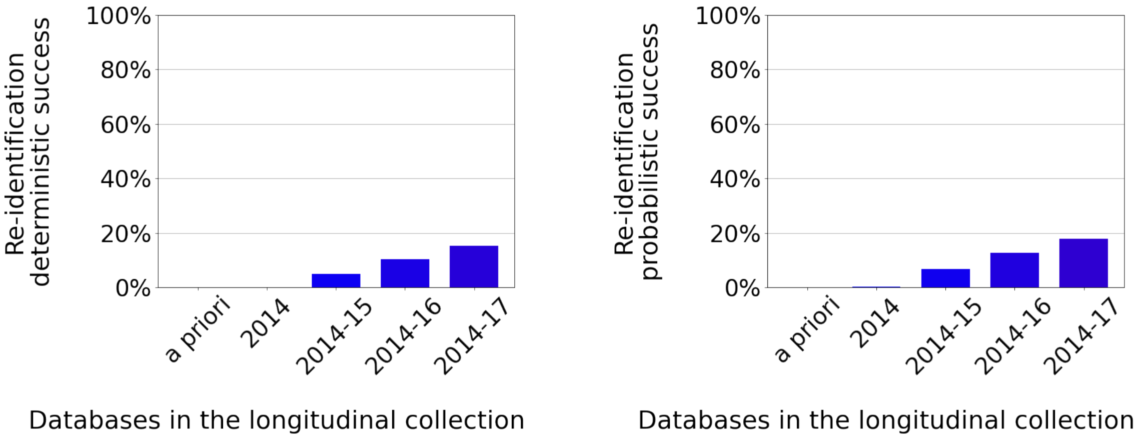
- **Adversary's probabilistic success**. Defined in Equations 5.4 and 5.5, this metric measures the probability of a randomly chosen individual in the focal database being re-identified, in a scale from 0% to 100%. According to Table F.21, the adversary's *a priori* probabilistic success is approximately 0.00001%, i.e. the adversary's chance of re-identifying one of the 9 773 350 individuals in the focal database is almost zero. However, by using only the attributes CO_IES and CO_CURSO as quasi-identifiers and only the database of 2014, the adversary increases their chance to 0.34%.

  As the longitudinal set of databases available to the adversary increases, so increases the adversary's chance of re-identifying a randomly chosen individual in the focal database. Particularly, the adversary's *a posteriori* probabilistic success increases to 6.72% by using only the database of 2015 as auxiliary information, to 12.64% by using the databases of 2015 and 2016, and to 17.95% by using the databases from 2015 to 2017.

◁

	Adversary's deterministic success		**Adversary's probabilistic success**	
	***a priori* success** 0.00%		***a priori* success** 0.00001%	
Databases in the longitudinal aggregation	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
`2014`	0.0044%	0.0044%	0.34%	33 197
`2014 to 2015`	5.11%	5.11%	6.72%	656 790
`2014 to 2016`	10.33%	10.33%	12.64%	1 235 822
`2014 to 2017`	15.25%	15.25%	17.95%	1 754 487

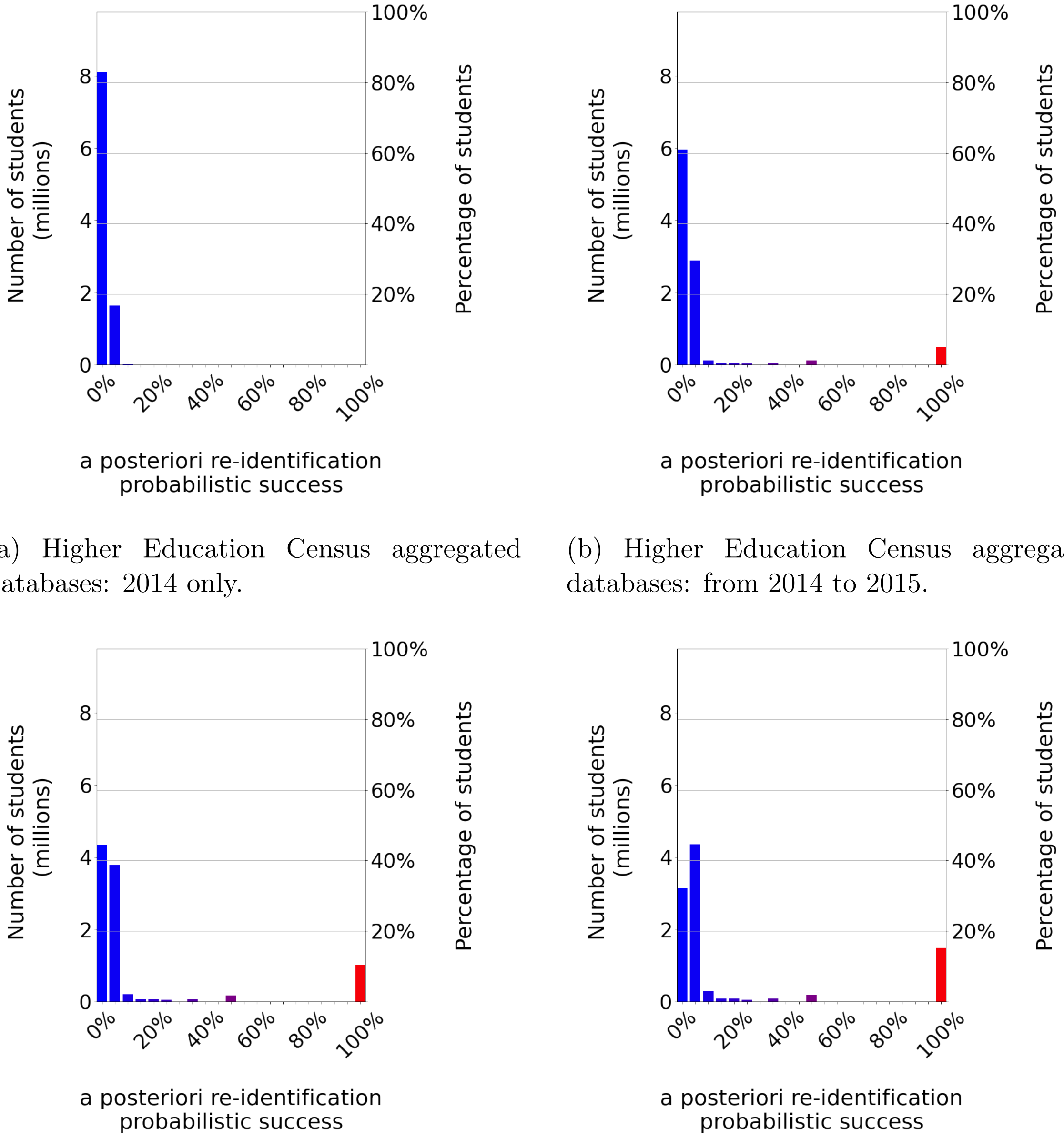
Table F.21: Experiment 61: Privacy degradation in Collective-target Re-identification Longitudinal databases (CRL) attacks on the Higher Education Censuses from 2014 to 2017 using the quasi-identifiers `CO_IES` and `CO_CURSO`. Here the focal database is that for the Higher Education Census of 2014. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* success, while the probabilistic degradation of privacy is their ratio.



(a) Deterministic success.

(b) Probabilistic success.

Figure F.22: Experiment 61: Adversary's success in Collective-target Re-identification Longitudinal databases (CRL) attacks on the Higher Education Censuses from 2014 to 2017 using the quasi-identifiers `CO_IES` and `CO_CURSO`. Here, each bar represents a different longitudinal aggregation of databases, always having the Higher Education Census of 2014 as the focal one, except for the bar with label *a priori*, which indicates the adversary's *a priori* success relying only on the focal database. The height of each bar represents the adversary's deterministic or probabilistic success.



(a) Higher Education Census aggregated databases: 2014 only.

(b) Higher Education Census aggregated databases: from 2014 to 2015.

(c) Higher Education Census aggregated databases: from 2014 to 2016.

(d) Higher Education Census aggregated databases: from 2014 to 2017.

Figure F.23: Experiment 61: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Re-identification Longitudinal databases (CRL) attacks on the Higher Education Censuses from 2014 to 2017 using the quasi-identifiers `CO_IES` and `CO_CURSO`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of re-identification.

F.2.3 Results of attribute-inference attack experiments

In this section, we present our quantitative analysis on the privacy risk of attribute-inference for individuals whose information are made public on longitudinal databases, as modeled in Section 4.1.2. We present one CAL attack on the Higher Education Censuses from 2014 to 2017, on sensitive attributes `IN_ALUNO_DEF_TGD_SUPER` and `IN_FINANC_ESTUDANTIL`.

Experiment 62 (Collective-target Attribute-inference Longitudinal databases (CAL) attack on the Higher Education Censuses)**.** In a CAL attack, the adversary's goal is to infer the values of an attribute considered to be sensitive for as many individuals as possible, according to Definition 22. Since we want to evaluate the overall attribute-inference risk, the adversary can gather auxiliary information on the values of a set of quasi-identifying attributes for every individual who holds a record in the focal database. The following experiment was conducted on the Higher Education Censuses from 2014 to 2017 and had the database for year 2014 as the focal one.

The focal database was released containing 10 793 936 records, which were reduced to 9 773 350 after the random selection of only one record for each data holder of multiple records. For the current experiment, we have selected as quasi-identifiers the attributes `CO_IES` and `CO_CURSO`, as specified in Table F.19. We have measured both the deterministic and probabilistic degradation of privacy for the set composed of those three attributes.

Furthermore, we were interested in inferring the value for the attributes `IN_ALUNO_DEF_TGD_SUPER` and `IN_FINANC_ESTUDANTIL`, both described in Table F.20 and considered by us to be sensitive.

The results are organized as follows.

- Tables F.24a and F.24b summarize the adversary's measures of success and privacy degradation as more databases are aggregated to the focal database for the sensitive attributes `IN_ALUNO_DEF_TGD_SUPER` and `IN_FINANC_ESTUDANTIL`, respectively.
- Figure F.25 shows both the adversary's deterministic and probabilistic measures of success according to Table F.24, for both sensitive attributes.
- Figures F.26 and F.27 show the histograms for the distribution of individuals according to the adversary's probabilistic measure of success as more databases are aggregated to the focal database for the sensitive attributes `IN_ALUNO_DEF_TGD_SUPER` and `IN_FINANC_ESTUDANTIL`, respectively.

According to those results, we were able to conclude the following.

- **Adversary's deterministic success**. Defined in Equations 5.7 and 5.8, this metric measures the fraction of individuals in the focal database that can have their values for the sensitive attribute inferred with absolute certainty, in a scale from 0% to 100%.

  According to Table F.24a for the sensitive attribute `IN_ALUNO_DEF_TGD_SUPER`, the adversary's *a priori* deterministic success equals 0%, i.e. no individual can have their values for the sensitive attribute inferred with absolute certainty in the focal database without the use of auxiliary information. However, by using only the attributes `CO_IES` and `CO_CURSO` as quasi-identifiers and only the database of 2014, the adversary can infer the values with absolute certainty for up to 19.89% of the individuals, i.e. approximately 1 943 919 students.

  As the longitudinal collection of databases available to the adversary increases, allowing the use of more Higher Education Census databases as auxiliary information, so increases the fraction of individuals in the focal database that can have their values for the sensitive attribute `IN_ALUNO_DEF_TGD_SUPER` inferred with absolute certainty. Particularly, the adversary's *a posteriori* deterministic success increases to 37.90% by using only the database of 2015 as auxiliary information, to 50.29% by using the databases of 2015 and 2016, and to 58.82% by using the databases from 2015 to 2017.

  From Table F.24b for the sensitive attribute `IN_FINANC_ESTUDANTIL`, the adversary's *a priori* deterministic success also equals 0%. However, by using only the attributes `CO_IES` and `CO_CURSO` as quasi-identifiers and only the database of 2014, the adversary can infer the values with absolute certainty for up to 24.35% of the individuals in the focal database, i.e. approximately 2 379 811 students.

  As the longitudinal collection of databases available to the adversary increases, so increases the fraction of individuals in the focal database that can have their values for the sensitive attribute `IN_FINANC_ESTUDANTIL` inferred with absolute certainty. Particularly, the adversary's *a posteriori* deterministic success increases to 31.90% by using only the database of 2015 as auxiliary information, to 37.09% by using the databases of 2015 and 2016, and to 41.63% by using the databases from 2015 to 2017.

- **Adversary's probabilistic success**. Defined in Equations 5.11 and 5.12, this metric measures the probability of a randomly chosen individual in the focal database having their value for the sensitive attribute inferred with absolute certainty, in a scale from 0% to 100%.

According to Table F.24a for the sensitive attribute IN_ALUNO_DEF_TGD_SUPER, the adversary's *a priori* probabilistic success is already of 92.09%, i.e. the adversary's chance of inferring the value for the sensitive attribute for one of the 9 773 350 individuals in the focal database without the use of auxiliary information is already high. Furthermore, by using only the attributes CO_IES and CO_CURSO as quasi-identifiers and only the database of 2014, the adversary increases their chance to up to 97.61%.

As the longitudinal collection of databases available to the adversary increases, allowing the use of more Higher Education Census databases as auxiliary information, so increases the adversary's chance of inferring the value for the sensitive attribute IN_ALUNO_DEF_TGD_SUPER. Particularly, the adversary's *a posteriori* probabilistic success increases to 97.98% by using only the database of 2015 as auxiliary information, to 98.21% by using the databases of 2015 and 2016, and to 98.39% by using the databases from 2015 to 2017.

From Table F.24b for the sensitive attribute IN_FINANC_ESTUDANTIL, the adversary's *a priori* probabilistic success is of 52.10%. Furthermore, by using only the attributes CO_IES and CO_CURSO as quasi-identifiers and only the database of 2014, the adversary increases their chance to up to 83.47%.

As the longitudinal collection of databases available to the adversary increases, so increases the adversary's chance of inferring the value for the sensitive attribute IN_FINANC_ESTUDANTIL. Particularly, the adversary's *a posteriori* deterministic success increases to 85.23% by using only the database of 2015 as auxiliary information, to 86.40% by using the databases of 2015 and 2016, and to 87.50% by using the databases from 2015 to 2017.

◁

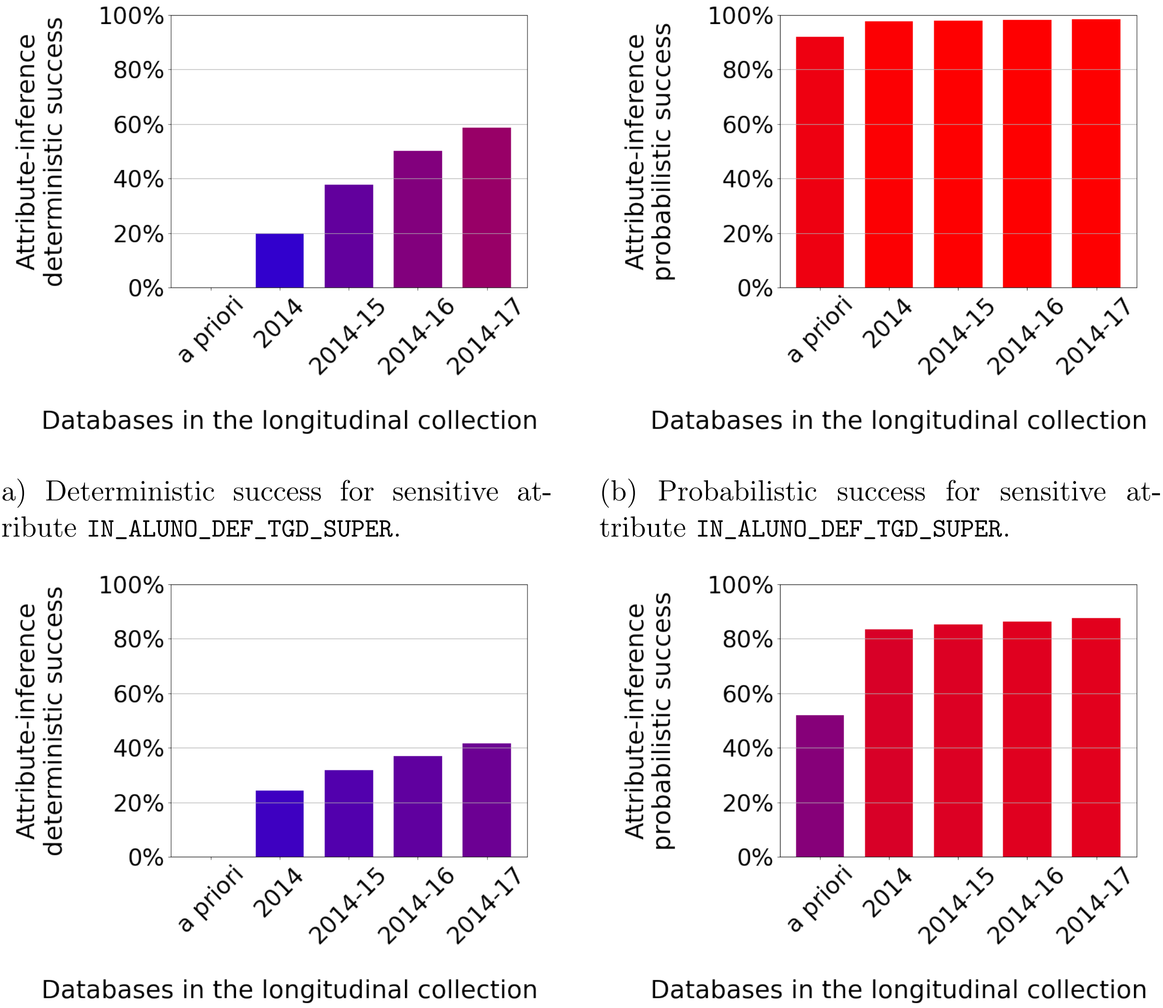
	Adversary's deterministic success		**Adversary's probabilistic success**	
	***a priori* success** 0.00%		***a priori* success** 92.09%	
Databases in the longitudinal aggregation	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
`2014`	19.89%	19.89%	97.61%	1.0599
`2014 to 2015`	37.90%	37.90%	97.98%	1.0640
`2014 to 2016`	50.29%	50.29%	98.21%	1.0664
`2014 to 2017`	58.82%	58.82%	98.39%	1.0684

(a) Sensitive attribute: `IN_ALUNO_DEF_TGD_SUPER`.

	Adversary's deterministic success		**Adversary's probabilistic success**	
	***a priori* success** 0.00%		***a priori* success** 52.10%	
Databases in the longitudinal aggregation	***a posteriori* success**	**Privacy degradation**	***a posteriori* success**	**Privacy degradation**
`2014`	24.35%	24.35%	83.47%	1.6022
`2014 to 2015`	31.90%	31.90%	85.23%	1.6358
`2014 to 2016`	37.09%	37.09%	86.40%	1.6582
`2014 to 2017`	41.63%	41.63%	87.50%	1.6794

(b) Sensitive attribute: `IN_FINANC_ESTUDANTIL`.

Table F.24: Experiment 62: Privacy degradation in Collective-target Attribute-inference Longitudinal databases (CAL) attacks on the Higher Education Censuses from 2014 to 2017 using the quasi-identifiers `CO_IES` and `CO_CURSO`. Here the focal database is that for the Higher Education Census of 2014. The deterministic degradation of privacy is the difference between the *a posteriori* and *a priori* success, while the probabilistic degradation of privacy is their ratio.



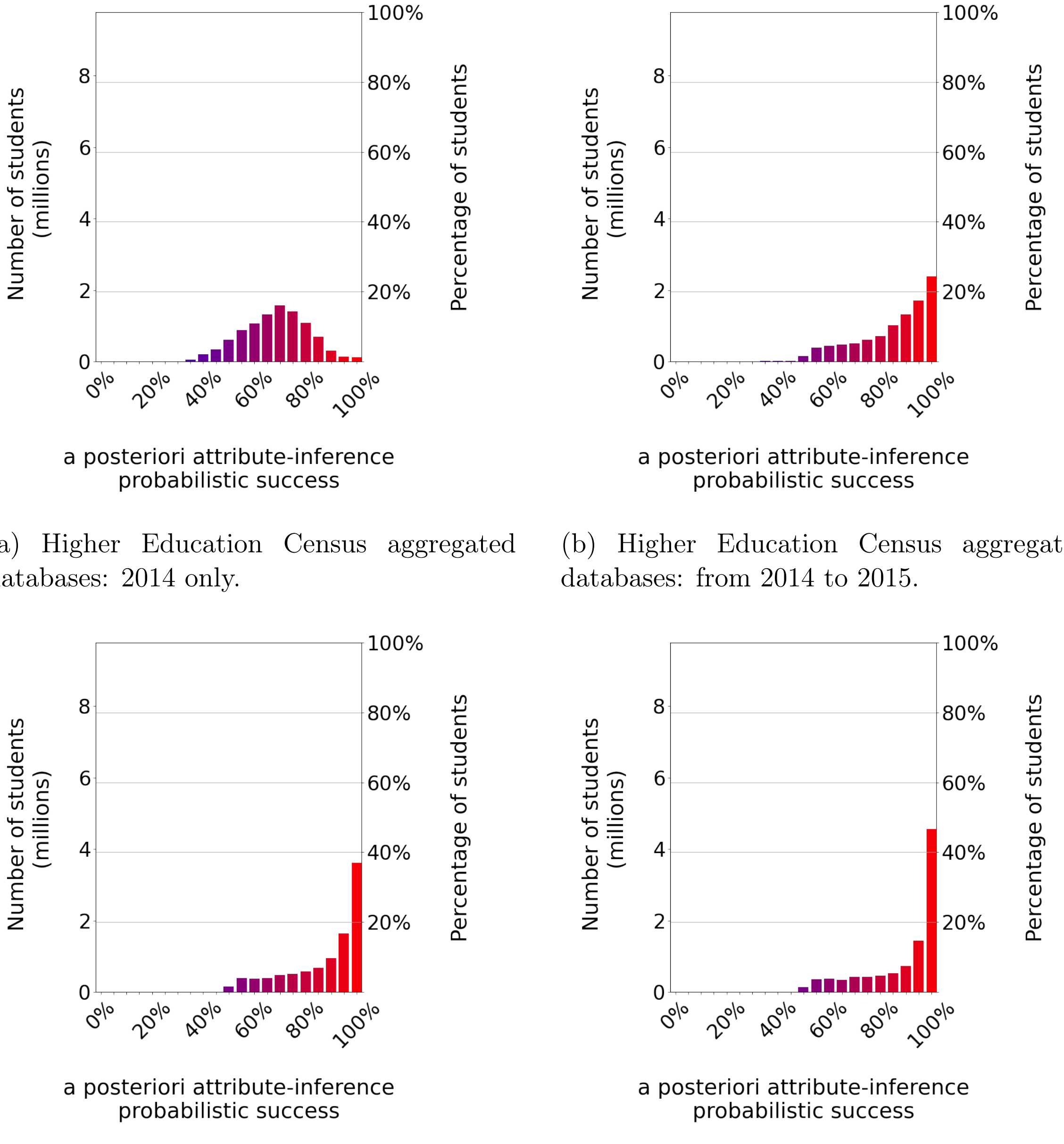
(a) Deterministic success for sensitive attribute `IN_ALUNO_DEF_TGD_SUPER`.

(b) Probabilistic success for sensitive attribute `IN_ALUNO_DEF_TGD_SUPER`.

(c) Deterministic success for sensitive attribute `IN_FINANC_ESTUDANTIL`.

(d) Probabilistic success for sensitive attribute `IN_FINANC_ESTUDANTIL`.

Figure F.25: Experiment 62: Adversary's success in Collective-target Attribute-inference Longitudinal databases (CAL) attacks on the Higher Education Censuses from 2014 to 2017 using the quasi-identifiers `CO_IES` and `CO_CURSO`. Here, each bar represents a different longitudinal aggregation of databases, always having the Higher Education Census of 2014 as the focal one, except for the bar with label *a priori*, which indicates the adversary's *a priori* success relying only on the focal database. The height of each bar represents the adversary's deterministic or probabilistic success.



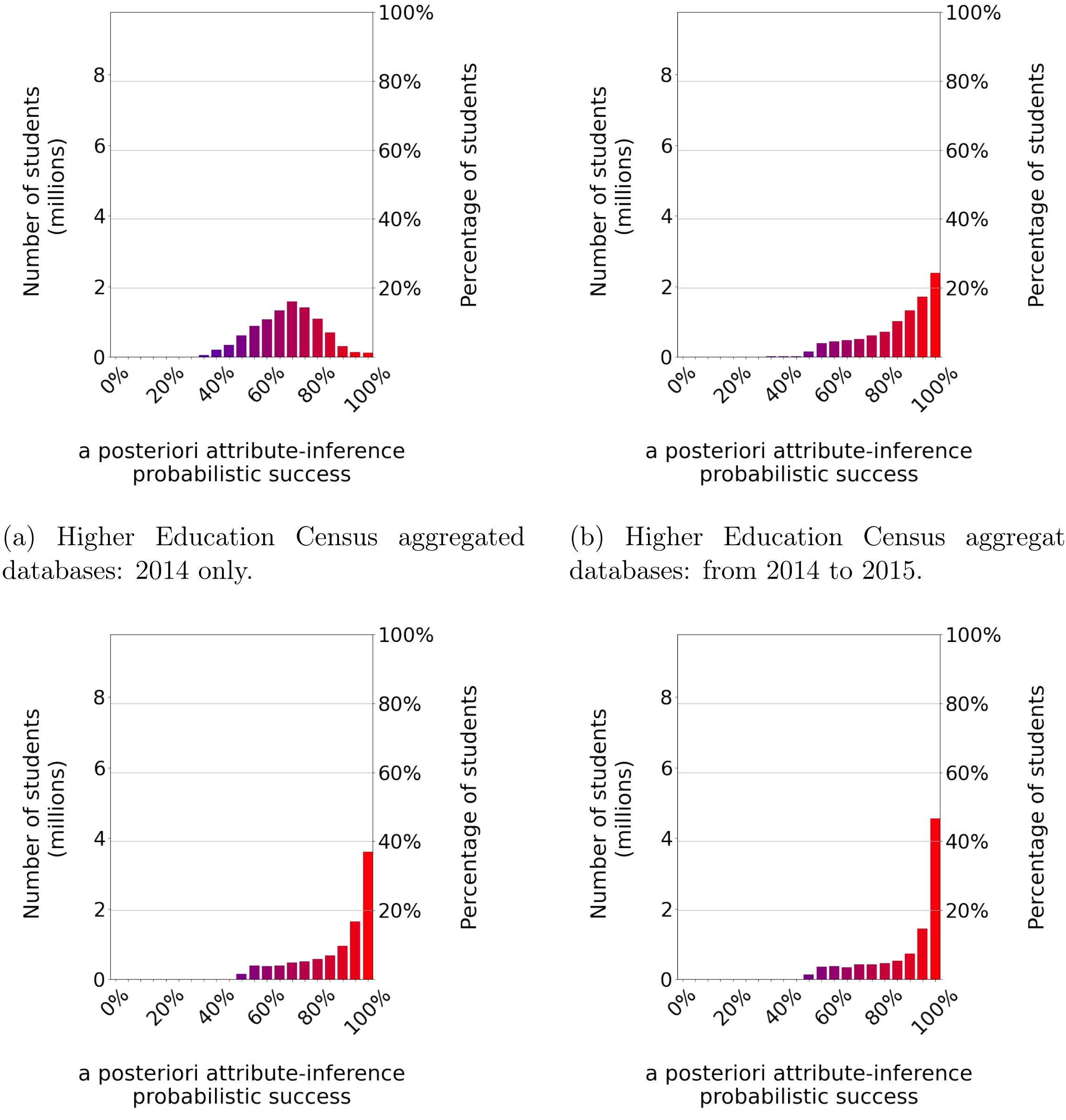
(a) Higher Education Census aggregated databases: 2014 only.

(b) Higher Education Census aggregated databases: from 2014 to 2015.

(c) Higher Education Census aggregated databases: from 2014 to 2016.

(d) Higher Education Census aggregated databases: from 2014 to 2017.

Figure F.26: Experiment 62: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Attribute-inference Longitudinal databases (CAL) attacks on the Higher Education Censuses from 2014 to 2017 using the quasi-identifiers `CO_IES` and `CO_CURSO` for the sensitive attribute `IN_ALUNO_DEF_TGD_SUPER`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of attribute-inference.



(a) Higher Education Census aggregated databases: 2014 only.

(b) Higher Education Census aggregated databases: from 2014 to 2015.

(c) Higher Education Census aggregated databases: from 2014 to 2016.

(d) Higher Education Census aggregated databases: from 2014 to 2017.

Figure F.27: Experiment 62: Histograms for the distribution of individuals according to the adversary's probabilistic measure of success in Collective-target Attribute-inference Longitudinal databases (CAL) attacks on the Higher Education Censuses from 2014 to 2017 using the quasi-identifiers `CO_IES` and `CO_CURSO` for the sensitive attribute `IN_FINANC_ESTUDANTIL`. The horizontal axis defines the possible values for the adversary's *a posteriori* probabilistic success while the vertical axis defines the number of individuals in the database subject to that risk of attribute-inference.

F.2.4 Takeaways

In this appendix, we have presented experimental results for the collective-target attacks on the Higher Education Censuses of 2018 and 2019 released by INEP, Experiments 61 and 62 for re-identification and for attribute-inference attacks, respectively.

Those results have allowed us to draw the following conclusions.

- The measures of **deterministic** and **probabilistic** successes, for both **re-identification** and **attribute-inference** attacks, have presented similar values to those reported for the School Census in Section 5.2, Experiments 23 and 24. The lower values for the Higher Education Census results are due to the use of one less quasi-identifying attribute, hence expected.

- The results observed for the *a priori* probabilistic attribute-inference successes for different sensitive attributes are a consequence of how much skewed the respective distributions of students are for those attributes. A more skewed distribution implies a higher *a priori* success.

Appendix G

Proofs of results in Chapter 6

In this appendix, we present some remarks, propositions, and proofs on the theoretical model developed in Chapter 6, for the oblivious mechanism in Section G.1, and for the local mechanism in Section G.2.

G.1 Oblivious mechanism

In this section, we provide the detailed derivations of the formulas from Definition 30 for an oblivious mechanism.

Remark 63 (Distribution induced by an oblivious mechanism)**.** Here we explain why the joint probability distribution $p^{\Gamma^{obv}} : \mathbb{D}(\mathit{records}(\mathcal{A}) \times \mathit{dom}(a_s) \times \mathbb{N} \times \mathbb{N})$ of an oblivious mechanism is given by Equation (6.7) in Definition 30.

First notice that, by the chain rule, for every record $x \in \mathit{records}(\mathcal{A})$, sensitive value $s \in \mathit{dom}(a_s)$, real count $u \in \mathbb{N}$, and reported count $u' \in \mathbb{N}$, we have that $p^{\Gamma^{obv}}(x^\star{=}x, x^\star[a_s]{=}s, \texttt{count}_q(D{\cup}x^\star){=}u, M^{obv}_{\texttt{count}_q}(D{\cup}x^\star){=}u')$ can be broken into the product:

$$
\begin{aligned}
&p^{\Gamma^{obv}}(x^\star{=}x) \cdot p^{\Gamma^{obv}}(x^\star[a_s]{=}s \mid x^\star{=}x) \cdot p^{\Gamma^{obv}}(\texttt{count}_q(D{\cup}x^\star){=}u \mid x^\star{=}x, x^\star[a_s]{=}s) \cdot \\
&p^{\Gamma^{obv}}(M^{obv}_{\texttt{count}_q}(D{\cup}x^\star){=}u' \mid x^\star{=}x, x^\star[a_s]{=}s, \texttt{count}_q(D{\cup}x^\star){=}u) \ ,
\end{aligned}
$$

where

- $p^{\Gamma^{obv}}(x^\star{=}x)$ is given by $\pi^\star_x$, since the probability of any record x being added to D is given by the prior distribution $\pi^\star$ on records;
- $p^{\Gamma^{obv}}(x^\star[a_s]{=}s \mid x^\star{=}x)$ is given by $\delta_{x[a_s]}(s)$;

- $p^{\Gamma^{obv}}(\mathtt{count}_q(D \cup x^\star){=}u \mid x^\star{=}x, x^\star[a_s]{=}s)$ is equal to $p^{\Gamma^{obv}}(\mathtt{count}_q(D \cup x^\star){=}u \mid x^\star{=}x)$, since the real count is independent of any secret value given the database, and is given by $\delta_{\mathtt{count}_q(D \cup x)}(u)$;
- $p^{\Gamma^{obv}}(M^{obv}_{\mathtt{count}_q}(D \cup x^\star){=}u' \mid x^\star{=}x, x^\star[a_s]{=}s, \mathtt{count}_q(D \cup x^\star){=}u)$ is equal to $p^{\Gamma^{obv}}(M^{obv}_{\mathtt{count}_q}(D \cup x^\star){=}u' \mid \mathtt{count}_q(D \cup x^\star){=}u)$, since the mechanism is oblivious, and is given by $R^{obv}(u' \mid u)$.

$\triangleleft$

Proposition 64 (Privacy loss in the context of an attack on an oblivious mechanism)**.** *Consider $\Gamma^{obv} = \langle D, \mathcal{A}, a_s, a_u, \mathtt{count}_q, \pi^\star, M^{obv}_{\mathtt{count}_q} \rangle$, in which $M^{obv}_{\mathtt{count}_q}$ is an oblivious mechanism operating on the extended database $D \cup x^\star$ and using an oblivious randomization function $R^{obv} : \mathbb{N} \to \mathbb{D}(\mathbb{N})$, as in Definition 30. Then, the corresponding privacy loss is given by:*

$$\textit{privacy-loss}(\Gamma^{obv}) = \frac{\textit{post-vul}(\Gamma^{obv})}{\textit{prior-vul}(\Gamma^{obv})} \ ,$$

where

$$\textit{post-vul}(\Gamma^{obv}) = \sum_{u' \in \mathbb{N}} \max_{s \in dom(a_s)} \sum_{\substack{x \in records(\mathcal{A}), \\ x[a_s]=s}} \pi^\star_x \cdot R^{obv}(u' \mid \mathtt{count}_q(D \cup x)) \ ,$$

and

$$\textit{prior-vul}(\Gamma^{obv}) = \max_{s \in dom(a_s)} \sum_{\substack{x \in records(\mathcal{A}) \\ x[a_s]=s}} \pi^\star_x \ .$$

Proof. Note that *privacy-loss*(Γ^{obv}) is given by the ratio from Equation (6.3). In the case of an oblivious mechanism, the posterior Bayes vulnerability of Equation (6.4) specializes to:

$$\begin{aligned}
&\textit{post-vul}(\Gamma^{obv}) = \\
&= \sum_{u' \in \mathbb{N}} \max_{s \in dom(a_s)} p^{\Gamma^{obv}}(x^\star[a_s]{=}s, M^{obv}_{\mathtt{count}_q}(D \cup x^\star){=}u') && \text{(Eq. (6.4))} \\
&= \sum_{u' \in \mathbb{N}} \max_{s \in dom(a_s)} \sum_{\substack{x \in records(\mathcal{A}), \\ u \in \mathbb{N}}} p^{\Gamma^{obv}} \begin{pmatrix} x^\star{=}x, x^\star[a_s]{=}s, \\ \mathtt{count}_q(D \cup x){=}u, \\ M^{obv}_{\mathtt{count}_q}(D \cup x){=}u' \end{pmatrix} && \text{(marginal)}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{u' \in \mathbb{N}} \max_{s \in dom(a_s)} \sum_{\substack{x \in records(\mathcal{A}),\\ u \in \mathbb{N}}} \begin{matrix}\pi^\star_x \cdot \delta_{x[a_s]}(s) \cdot \delta_{\mathtt{count}_q(D \cup x)}(u) \cdot \\ R^{obv}(u' \mid u)\end{matrix} && \text{(Eq. (6.7))} \\
&= \sum_{u' \in \mathbb{N}} \max_{s \in dom(a_s)} \sum_{\substack{x \in records(\mathcal{A}),\\ x[a_s]=s}} \pi^\star_x \cdot R^{obv}(u' \mid \mathtt{count}_q(D \cup x)) && \text{(simplifying)}
\end{aligned}
$$

Furthermore, the prior Bayes vulnerability of Equation (6.5) specializes to:

$$
\begin{aligned}
&\mathit{prior\text{-}vul}(\Gamma^{obv}) = \\
&= \max_{s \in dom(a_s)} p^{\Gamma^{obv}}(x^\star[a_s]{=}s) && \text{(Eq. (6.5))} \\
&= \max_{s \in dom(a_s)} \sum_{x \in records(\mathcal{A})} p^{\Gamma^{obv}}(x^\star{=}x, x^\star[a_s]{=}s) && \text{(marginal)} \\
&= \max_{s \in dom(a_s)} \sum_{x \in records(\mathcal{A})} p^{\Gamma^{obv}}(x^\star{=}x) \cdot p^{\Gamma^{obv}}(x^\star[a_s]{=}s \mid x^\star{=}x) && \text{(chain rule)} \\
&= \max_{s \in dom(a_s)} \sum_{x \in records(\mathcal{A})} \pi^\star_x \cdot \delta_{x^\star[a_s]}(s) && \text{(Eq. (6.7))} \\
&= \max_{s \in dom(a_s)} \sum_{\substack{x \in records(\mathcal{A})\\ x[a_s]=s}} \pi^\star_x && \text{(simplifying)}
\end{aligned}
$$

We obtain our final result by taking the ratio of these values. □

Proposition 65 (Utility in the context of an attack on an oblivious mechanism)**.** *Consider $\Gamma^{obv} = \langle D, \mathcal{A}, a_s, a_u, \mathtt{count}_q, \pi^\star, M^{obv}_{\mathtt{count}_q} \rangle$, in which $M^{obv}_{\mathtt{count}_q}$ is an oblivious mechanism operating on the extended database $D \cup x^\star$ and using an oblivious randomization function $R^{obv} : \mathbb{N} \to \mathbb{D}(\mathbb{N})$, as in Definition 30. Then, the corresponding utility is given by:*

$$
\mathit{utility}(\Gamma^{obv}) = \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} \sum_{\substack{x \in records(\mathcal{A}),\\ \mathtt{count}_q(D \cup x)=u}} \pi^\star_x \cdot R^{obv}(u' \mid u) \ .
$$

Proof. Note that in the case of an oblivious mechanism, the expected utility of Equation (6.6) specializes to:

$$
\begin{aligned}
&\mathit{utility}(\Gamma^{obv}) = \\
&= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} p^{\Gamma^{obv}}(\mathtt{count}_q(D \cup x^\star){=}u, M^{obv}_{\mathtt{count}_q}(D \cup x^\star){=}u') && \text{(Eq. (6.6))}
\end{aligned}
$$

$$= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} \sum_{\substack{x \in \mathit{records}(\mathcal{A}), \\ s \in \mathit{dom}(a_s)}} p^{\Gamma^{obv}} \begin{pmatrix} x^\star{=}x, x^\star[a_s]{=}s, \\ \mathtt{count}_q(D \cup x^\star){=}u, \\ M^{obv}_{\mathtt{count}_q}(D \cup x^\star){=}u' \end{pmatrix} \qquad \text{(marginal)}$$

$$= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} \sum_{\substack{x \in \mathit{records}(\mathcal{A}), \\ s \in \mathit{dom}(a_s)}} \begin{matrix} \pi^\star_x \cdot \delta_{x[a_s]}(s) \cdot \delta_{\mathtt{count}_q(D \cup x)}(u) \cdot \\ R^{obv}(u' \mid u) \end{matrix} \qquad \text{(Eq. (6.7))}$$

$$= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} \sum_{x \in \mathit{records}(\mathcal{A})} \begin{matrix} \pi^\star_x \cdot \delta_{\mathtt{count}_q(D \cup x)}(u) \cdot \\ R^{obv}(u' \mid u) \end{matrix} \cdot \sum_{s \in \mathit{dom}(a_s)} \delta_{x[a_s]}(s) \qquad \text{(reorganizing)}$$

$$= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} \sum_{\substack{x \in \mathit{records}(\mathcal{A}), \\ \mathtt{count}_q(D \cup x) = u}} \pi^\star_x \cdot R^{obv}(u' \mid u) \qquad \text{(simplifying)}$$

□

G.2 Local Mechanism

In this section, we provide the detailed derivations of the formulas from Definition 31 for a local mechanism.

Remark 66 (Distribution induced by a local mechanism)**.** Here we explain why the joint probability distribution $p^{\Gamma^{loc}} : \mathbb{D}(\mathit{records}(\mathcal{A}) \times \mathit{dom}(a_s) \times \mathbb{N} \times \mathbb{N})$ of a local mechanism is given by Equation (6.10) in Definition 31.

First notice that, by the chain rule, for every record $x \in \mathit{records}(\mathcal{A})$, sensitive value $s \in \mathit{dom}(a_s)$, real count $u \in \mathbb{N}$, and reported count $u' \in \mathbb{N}$, we have that $p^{\Gamma^{loc}}(x^\star{=}x, x^\star[a_s]{=}s, \mathtt{count}_q(D \cup x^\star){=}u, M^{loc}_{\mathtt{count}_q}(D \cup x^\star){=}u')$ can be broken into the product:

$$p^{\Gamma^{loc}}(x^\star{=}x) \cdot p^{\Gamma^{loc}}(x^\star[a_s]{=}s \mid x^\star{=}x) \cdot p^{\Gamma^{loc}}(\mathtt{count}_q(D \cup x^\star){=}u \mid x^\star{=}x, x^\star[a_s]{=}s) \cdot$$
$$p^{\Gamma^{loc}}(M^{loc}_{\mathtt{count}_q}(D \cup x^\star){=}u' \mid x^\star{=}x, x^\star[a_s]{=}s, \mathtt{count}_q(D \cup x^\star){=}u) \ ,$$

where

- $p^{\Gamma^{loc}}(x^\star{=}x)$ is given by $\pi^\star_x$, since the probability of any record x being added to D is given by the prior distribution $\pi^\star$ on records;
- $p^{\Gamma^{loc}}(x^\star[a_s]{=}s \mid x^\star{=}x)$ is given by $\delta_{x[a_s]}(s)$;

- $p^{\Gamma^{loc}}(\texttt{count}_q(D{\cup}x^\star){=}u \mid x^\star{=}x, x^\star[a_s]{=}s)$ is equal to $p^{\Gamma^{loc}}(\texttt{count}_q(D{\cup}x^\star){=}u \mid x^\star{=}x)$, since the real count is independent of any secret value given the database, and is given by $\delta_{\texttt{count}_q(D\cup x)}(u)$;
- $p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(D{\cup}x^\star){=}u' \mid x^\star{=}x, x^\star[a_s]{=}s, \texttt{count}_q(D{\cup}x^\star){=}u)$ is equal to $p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(D{\cup}x){=}u')$, since the reported answer is independent of both the real answer and the secret value given the database.

However, we need to specify $p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(d{\cup}y){=}c)$ for every database $d \in \mathcal{D}(\mathcal{A})$, record $y \in \mathit{records}(\mathcal{A})$, and reported answer $c \in \mathbb{N}$. In order to do so, recall that $M^{loc}_{\texttt{count}_q}$ works as follows: first the local randomization function R^{loc} is applied to the useful value of every record in $d{\cup}y$ to produce a randomized database, then the counting query $\texttt{count}_q$ is applied to the randomized database to produce the reported answer u'. Hence, we can provide the following recursive definition of $p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(d{\cup}y){=}c)$:

Base case 1: when the reported count is a value $c < 0$ or $c > |d{\cup}y|$. It is impossible that a counting query operating on a database with $|d{\cup}y| > 0$ records returns such values, hence:

$$p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(d{\cup}y){=}c) = 0, \qquad\qquad \text{if } c < 0 \text{ or } c > |d{\cup}y|;$$

Base case 2: when $0 \leq c \leq |d{\cup}y|$ and $d{\cup}y = \emptyset$. In this case the input database has no records and, hence, no record satisfies the property q used in the query $\texttt{count}_q$. Consequently, the only possible reported answer in the valid range is $c = 0$:

$$p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(d{\cup}y){=}c) = 1, \qquad\qquad \text{if } d{\cup}y = \emptyset \text{ and } c = 0;$$

Base case 3: when $0 \leq c \leq |d{\cup}y|$ and $d{\cup}y = \{\!\{y\}\!\}$, i.e. when the only record in $d{\cup}y$ is y itself because $d = \emptyset$, and the reported count can only be $c = 0$ or $c = 1$. In this case, the mechanism $M^{loc}_{\texttt{count}_q}$ will first apply the local randomization function R^{loc} to the useful value $y[a_u]$ of y itself, resulting in a distribution $R^{loc}(w \mid y[a_u])$ for every possible value $w \in \mathit{range}(a_u)$ for the useful attribute. Each w yields a randomized database $\{\!\{w\}\!\}$ with only one record, and the probability that $M^{loc}_{\texttt{count}_q}(d{\cup}y)$ returns the reported count $c = 1$ is just the expectation that the randomized database $\{\!\{w\}\!\}$ returns a count of 1.

But notice that $\texttt{count}_q(\{\!\{w\}\!\})$ returns a count of 1 iff property *prop* is satisfied by w. Similarly, the probability that $M^{loc}_{\texttt{count}_q}(d{\cup}y)$ returns reported count $c = 0$ is just the expectation that the randomized database $\{\!\{w\}\!\}$ returns a count of 0, which happens iff property *prop* is not satisfied by w. Hence, we get:

$$p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(d{\cup}y){=}c) =$$
$$\begin{cases} \displaystyle\sum_{\substack{w\in dom(a_u),\\ q(w)=\texttt{false}}} R^{loc}(w \mid y[a_u]), & \text{if } d{\cup}y = \{\!\{y\}\!\} \text{ and } c = 0;\\ \displaystyle\sum_{\substack{w\in dom(a_u),\\ q(w)=\texttt{true}}} R^{loc}(w \mid y[a_u]), & \text{if } d{\cup}y = \{\!\{y\}\!\} \text{ and } c = 1;\end{cases}$$

Recursive case: when $0 \leq c \leq |d{\cup}y|$ and $\emptyset \neq d{\cup}y \neq \{\!\{y\}\!\}$, i.e. when neither d or y are empty, hence $|d{\cup}y| \geq 2$ and the reported count c is in the valid range, i.e. a non-negative integer smaller than or equal to the total number of records in $d{\cup}y$. In this case, the key insight is that the local randomization function R^{loc} is applied to record y in an independent manner from the application of R^{loc} to every record in d.

Therefore, there are two ways in which $M^{loc}_{\texttt{count}_q}(d{\cup}y)$ returns exactly the reported count c: either it is the case that $M^{loc}_{\texttt{count}_q}(\{\!\{y\}\!\}){=}0$ and $M^{loc}_{\texttt{count}_q}(d){=}c$, or it is the case that $M^{loc}_{\texttt{count}_q}(\{\!\{y\}\!\}){=}1$ and $M^{loc}_{\texttt{count}_q}(d){=}c-1$. Hence, we get:

$$\begin{aligned} &p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(d{\cup}y){=}c) =\\ &= p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(\{\!\{y\}\!\}){=}1)\cdot p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(d){=}c{-}1)+\\ &\quad p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(\{\!\{y\}\!\}){=}0)\cdot p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(d){=}c),\\ &\qquad\qquad\qquad\qquad\qquad \text{if } 0 \leq c \leq |d{\cup}y|\\ &\qquad\qquad\qquad\qquad\qquad \text{and } \emptyset \neq d{\cup}y \neq \{\!\{y\}\!\}\ . \end{aligned}$$

The recursive definition of $p^{\Gamma^{loc}}(M^{loc}_{\texttt{count}_q}(d{\cup}y){=}c)$ provided in Equation (6.11) is finally obtained by considering all the cases described above.

◁

Proposition 67 (Privacy loss in the context of an attack on a local mechanism)**.** *Consider $\Gamma^{loc} = \langle D, \mathcal{A}, a_s, a_u, \texttt{count}_q, \pi^\star, M^{loc}_{\texttt{count}_q}\rangle$, in which $M^{loc}_{\texttt{count}_q}$ is a local mechanism operating on the extended database $D{\cup}x^\star$ and using a local randomization function*

$R^{loc} : dom(a_u) \to \mathbb{D}(a_u)$, as in Definition 31. Then, the corresponding privacy loss is given by:

$$\mathit{privacy\text{-}loss}(\Gamma^{loc}) = \frac{\mathit{post\text{-}vul}(\Gamma^{loc})}{\mathit{prior\text{-}vul}(\Gamma^{loc})} \ ,$$

where

$$\mathit{post\text{-}vul}(\Gamma^{loc}) = \sum_{u' \in \mathbb{N}} \max_{s \in dom(a_s)} p^{\Gamma^{loc}}(x^\star[a_s]{=}s, M^{loc}_{\mathtt{count}_q}(D \cup x^\star){=}u') \ .$$

and

$$\mathit{prior\text{-}vul}(\Gamma^{loc}) = \max_{s \in dom(a_s)} \sum_{\substack{x \in records(\mathcal{A}) \\ x[a_s]=s}} \pi^\star_x \ .$$

Proof. Note that $\mathit{privacy\text{-}loss}(\Gamma^{loc})$ is given by the ratio from Equation (6.3). In the case of a local mechanism, the posterior Bayes vulnerability of Equation (6.4) specializes to:

$$\begin{aligned}
&\mathit{post\text{-}vul}(\Gamma^{loc}) = \\
&= \sum_{u' \in \mathbb{N}} \max_{s \in dom(a_s)} p^{\Gamma^{loc}}(x^\star[a_s]{=}s, M^{loc}_{\mathtt{count}_q}(D \cup x^\star){=}u') && \text{(Eq. (6.4))} \\
&= \sum_{u' \in \mathbb{N}} \max_{s \in dom(a_s)} \sum_{\substack{x \in records(\mathcal{A}), \\ u \in \mathbb{N}}} p^{\Gamma^{loc}} \begin{pmatrix} x^\star{=}x, x^\star[a_s]{=}s, \\ \mathtt{count}_q(D \cup x){=}u, \\ M^{loc}_{\mathtt{count}_q}(D \cup x){=}u' \end{pmatrix} && \text{(marginal)} \\
&= \sum_{u' \in \mathbb{N}} \max_{s \in dom(a_s)} \sum_{\substack{x \in records(\mathcal{A}), \\ u \in \mathbb{N}}} \begin{matrix} \pi^\star_x \cdot \delta_{x[a_s]}(s) \cdot \delta_{\mathtt{count}_q(D \cup x)}(u) \cdot \\ p^{\Gamma^{loc}}(M^{loc}_{\mathtt{count}_q}(D \cup x){=}u') \end{matrix} && \text{(Eq. (6.10))} \\
&= \sum_{u' \in \mathbb{N}} \max_{s \in dom(a_s)} \sum_{\substack{x \in records(\mathcal{A}), \\ x[a_s]=s, \\ \mathtt{count}_q(D \cup x)=u}} \pi^\star_x \cdot p^{\Gamma^{loc}}(M^{loc}_{\mathtt{count}_q}(D \cup x){=}u') && \text{(simplifying)}
\end{aligned}$$

Furthermore, the prior Bayes vulnerability of Equation (6.5) specializes to:

$$\begin{aligned}
&\mathit{prior\text{-}vul}(\Gamma^{loc}) = \\
&= \max_{s \in dom(a_s)} p^{\Gamma^{loc}}(x^\star[a_s]{=}s) && \text{(Eq. (6.5))} \\
&= \max_{s \in dom(a_s)} \sum_{x \in records(\mathcal{A})} p^{\Gamma^{loc}}(x^\star{=}x, x^\star[a_s]{=}s) && \text{(marginal)}
\end{aligned}$$

$$
\begin{aligned}
&= \max_{s \in dom(a_s)} \sum_{x \in records(\mathcal{A})} p^{\Gamma^{loc}}(x^\star{=}x) \cdot p^{\Gamma^{loc}}(x^\star[a_s]{=}s \mid x^\star{=}x) && \text{(chain rule)}\\
&= \max_{s \in dom(a_s)} \sum_{x \in records(\mathcal{A})} \pi^\star_x \cdot \delta_{x^\star[a_s]}(s) && \text{(Eq. (6.10))}\\
&= \max_{s \in dom(a_s)} \sum_{\substack{x \in records(\mathcal{A})\\ x[a_s]=s}} \pi^\star_x && \text{(simplifying)}
\end{aligned}
$$

We obtain our final result by taking the ratio of these values. □

Proposition 68 (Utility in the context of an attack on a local mechanism)**.** *Consider $\Gamma^{loc} = \langle D, \mathcal{A}, a_s, a_u, \mathtt{count}_q, \pi^\star, M^{loc}_{\mathtt{count}_q} \rangle$, in which $M^{loc}_{\mathtt{count}_q}$ is a local mechanism operating on the extended database $D \cup x^\star$ and using a local randomization function $R^{loc} : dom(a_u) \to \mathbb{D}(a_u)$, as in Definition 31. Then, the corresponding utility loss is given by:*

$$
utility(\Gamma^{loc}) = \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} \sum_{\substack{x \in records(\mathcal{A}),\\ \mathtt{count}_q(D \cup x)=u}} \pi^\star_x \cdot p^{\Gamma^{loc}}(M^{loc}_{\mathtt{count}_q}(D \cup x){=}u') \ .
$$

Proof. Note that in the case of a local mechanism, the expected utility of Equation (6.6) specializes to:

$$
\begin{aligned}
&utility(\Gamma^{loc}) =\\
&= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} p^{\Gamma^{loc}}(\mathtt{count}_q(D \cup x^\star){=}u, M^{loc}_{\mathtt{count}_q}(D \cup x^\star){=}u') && \text{(Eq. (6.6))}\\
&= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} \sum_{\substack{x \in records(\mathcal{A}),\\ s \in dom(a_s)}} p^{\Gamma^{loc}} \begin{pmatrix} x^\star{=}x, x^\star[a_s]{=}s,\\ \mathtt{count}_q(D \cup x^\star){=}u,\\ M^{loc}_{\mathtt{count}_q}(D \cup x^\star){=}u' \end{pmatrix} && \text{(marginal)}\\
&= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} \sum_{\substack{x \in records(\mathcal{A}),\\ s \in dom(a_s)}} \begin{matrix} \pi^\star_x \cdot \delta_{x[a_s]}(s) \cdot \delta_{\mathtt{count}_q(D \cup x)}(u) \cdot \\ p^{\Gamma^{loc}}(M^{loc}_{\mathtt{count}_q}(D \cup x){=}u') \end{matrix} && \text{(Eq. (6.10))}\\
&= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} \sum_{x \in records(\mathcal{A})} \begin{matrix} \pi^\star_x \cdot \delta_{\mathtt{count}_q(D \cup x)}(u) \cdot \\ p^{\Gamma^{loc}}(M^{loc}_{\mathtt{count}_q}(D \cup x){=}u') \end{matrix} \cdot \sum_{s \in dom(a_s)} \delta_{x[a_s]}(s) && \text{(reorganizing)}\\
&= \sum_{u' \in \mathbb{N}} \max_{u \in \mathbb{N}} \sum_{\substack{x \in records(\mathcal{A}),\\ \mathtt{count}_q(D \cup x)=u}} \pi^\star_x \cdot p^{\Gamma^{loc}}(M^{loc}_{\mathtt{count}_q}(D \cup x){=}u') && \text{(simplifying)}
\end{aligned}
$$

□

Bibliography

[1] Aggarwal, C. C. (2005). On k-anonymity and the curse of dimensionality. In *VLDB*, volume 5, pages 901--909.

[2] Alvim, M. S., Chatzikokolakis, K., McIver, A., Morgan, C., Palamidessi, C., and Smith, G. (2014). Additive and Multiplicative Notions of Leakage, and Their Capacities. In *Proc. of CSF*, pages 308--322. IEEE.

[3] Alvim, M. S., Chatzikokolakis, K., McIver, A., Morgan, C., Palamidessi, C., and Smith, G. (2016). Axioms for Information Leakage. In *IEEE 29th Computer Security Foundations Symposium, CSF 2016, Lisbon, Portugal, June 27 - July 1, 2016*, pages 77--92.

[4] Alvim, M. S., Chatzikokolakis, K., McIver, A., Morgan, C., Palamidessi, C., and Smith, G. (2019). An axiomatization of information flow measures. *Theoretical Computer Science*, 777:32--54.

[5] Alvim, M. S., Chatzikokolakis, K., McIver, A., Morgan, C., Palamidessi, C., and Smith, G. (2020a). *The Science of Quantitative Information Flow*. Springer.

[6] Alvim, M. S., Fernandes, N., McIver, A., and Nunes, G. H. (2020b). On Privacy and Accuracy in Data Releases (Invited Paper). In *31st International Conference on Concurrency Theory, CONCUR 2020, September 1-4, 2020, Vienna, Austria (Virtual Conference)*, volume 171 of *LIPIcs*, pages 1:1--1:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

[7] Brenner, H. and Nissim, K. (2010). Impossibility of Differentially Private Universally Optimal Mechanisms. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 71–80.

[8] Bu, Y., Fu, A., Wong, R. C. W., Chen, L., and Li, J. (2008). Privacy preserving serial data publishing by role composition. *Proceedings of the VLDB Endowment*, 1(1):845.

[9] Byun, J.-W., Sohn, Y., Bertino, E., and Li, N. (2006). Secure anonymization for incremental datasets. In *Workshop on secure data management*, pages 48--63. Springer.

[10] Dalenius, T. (1977). Towards a methodology for statistical disclosure control. *statistik Tidskrift*, 15(429-444):2--1.

[11] Dalenius, T. (1986). Finding a needle in a haystack or identifying anonymous census records. *Journal of official statistics*, 2(3):329.

[12] Ding, B., Kulkarni, J., and Yekhanin, S. (2017). Collecting telemetry data privately. *arXiv preprint arXiv:1712.01524*.

[13] Dinur, I. and Nissim, K. (2003). Revealing information while preserving privacy. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 202--210.

[14] Dwork, C. (2011). A firm foundation for private data analysis. *Communications of the ACM*, 54(1):86--95.

[15] Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265--284. Springer.

[16] Dwork, C., Roth, A., et al. (2014). The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211--407.

[17] ECLAC (2011). Code of good practice in statistics for Latin America and the Caribbean. https://www.cepal.org/en/publications/16423-code-good-practice-statistics-latin-america-and-caribbean-november-2011.

[18] El Emam, K. (2013). *Guide to the de-identification of personal health information*. CRC Press.

[19] El Emam, K. and Arbuckle, L. (2013). *Anonymizing health data: case studies and methods to get you started*. O'Reilly Media.

[20] Erlingsson, Ú., Pihur, V., and Korolova, A. (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054--1067.

[21] EUROSTAT (2017). European Statistics Code of Practice. https://ec.europa.eu/eurostat/en/web/products-catalogues/-/ks-02-18-142.

[22] EUROSTAT (2019). Quality Assurance Framework of the European Statistical System. https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646.

[23] Fung, B. C. M., Wang, K., Fu, A. W.-C., and Yu, P. S. (2010). *Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques*. Chapman & Hall/CRC, 1st edition. ISBN 1420091484.

[24] Ganta, S. R., Kasiviswanathan, S. P., and Smith, A. (2008). Composition attacks and auxiliary information in data privacy. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 265--273.

[25] Garfinkel, S. (2000). *Database Nation: The Death of Privacy in the 21st Century*. O'Reilly & Associates, Inc., USA. ISBN 1565926536.

[26] Garfinkel, S., Abowd, J. M., and Martindale, C. (2018a). Understanding database reconstruction attacks on public data. *Queue*, 16(5):28--53.

[27] Garfinkel, S., Abowd, J. M., and Powazek, S. (2018b). Issues Encountered Deploying Differential Privacy. In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, WPES'18, pages 133--137. Association for Computing Machinery.

[28] Ghosh, A., Roughgarden, T., and Sundararajan, M. (2009). Universally utility-maximizing privacy mechanisms. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 351--360.

[29] Gkoulalas-Divanis, A. and Loukides, G. (2015). *Medical Data Privacy Handbook*. Springer, 1st edition.

[30] Government of Australia (1905). Census and Statistics Act 1905. https://www.legislation.gov.au/Details/C2016C01005.

[31] Government of Australia (1975). Australian Bureau of Statistics Act 1975. https://www.legislation.gov.au/Details/C2017C00096.

[32] Government of Australia (1988). Privacy Act 1988. https://www.legislation.gov.au/Details/C2015C00598.

[33] Government of Brazil (1968). Lei 5.534, de 14 de novembro de 1968. http://www.planalto.gov.br/ccivil_03/leis/L5534.htm.

[34] Government of Brazil (1973). Decreto 73.177, de 20 de novembro de 1973. http://www.planalto.gov.br/ccivil_03/decreto/Antigos/D73177.htm.

[35] Government of Brazil (1988). Constitution of the Federative Republic of Brazil. https://www2.senado.leg.br/bdsf/handle/id/243334.

[36] Government of Brazil (1997). Lei 9.448, de 14 de março de 1997. http://www.planalto.gov.br/ccivil_03/LEIS/L9448.htm.

[37] Government of Brazil (2007a). Decreto 6.317, de 20 de dezembro de 2007. http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2007/Decreto/D6317.htm.

[38] Government of Brazil (2007b). Lei Complementar 131, de 27 de maio de 2009, Lei Capiberibe. http://www.planalto.gov.br/ccivil_03/LEIS/LCP/Lcp131.htm.

[39] Government of Brazil (2011). Lei 12.527, de 18 de novembro de 2011, Lei de Acesso à Informação (LAI). http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm.

[40] Government of Brazil (2018). Lei 13.709, de 14 de agosto de 2018, Lei Geral de Proteção de Dados Pessoais (LGPD). http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/Lei/L13709.htm.

[41] Government of the United States of America (1954). U.S. Code, Title 13 - CENSUS. https://www.law.cornell.edu/uscode/text/13.

[42] Government of the United States of America (1974). U.S. Code, Title 5, Section 552a - Records maintained on individuals. https://www.law.cornell.edu/uscode/text/5/552a.

[43] Government of the United States of America (2002). Confidential information protection and statistical efficiency act (cipsea). https://www.eia.gov/cipsea/cipsea.pdf.

[44] Government of the United States of America (2005). Statistical Policy Working Paper 22 - Report on Statistical Disclosure Limitation Methodology. https://nces.ed.gov/FCSM/pdf/spwp22.pdf.

[45] Government of the United States of America (2015). Overview of the Privacy Act of 1974. https://www.justice.gov/opcl/overview-privacy-act-1974-2015-edition.

[46] Instituto Brasileiro de Geografia e Estatística (2013). Código de Boas Práticas das Estatísticas do IBGE. https://ftp.ibge.gov.br/Informacoes_Gerais_e_Referencia/Codigo_de_Boas_Praticas_das_Estatisticas_do_IBGE.pdf.

[47] Instituto Brasileiro de Geografia e Estatística (2018). Confidencialidade no IBGE: procedimentos adotados na preservação do sigilo das informações individuais nas divulgações de resultados das operações estatísticas. https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101636.

[48] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2017). Portaria 91, de 2 de fevereiro de 2017. https://download.inep.gov.br/educacao_basica/censo_escolar/legislacao/2017/portaria_inep_91_02022017_principios_fundamentais_estatisticas_eduacionais.pdf.

[49] Jaynes, E. T. (2003). *Probability theory: The logic of science.* Cambridge University Press, Cambridge.

[50] Kifer, D. and Machanavajjhala, A. (2011). No Free Lunch in Data Privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data*, SIGMOD '11, pages 193–204. Association for Computing Machinery.

[51] Kohlmayer, F., Prasser, F., Eckert, C., Kemper, A., and Kuhn, K. A. (2012). Flash: efficient, stable and optimal k-anonymity. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, pages 708--717. IEEE.

[52] Li, N., Li, T., and Venkatasubramanian, S. (2007). t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. volume 2, pages 106 – 115.

[53] Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007). L-Diversity: Privacy beyond k-Anonymity. *ACM Transactions on Knowledge Discovery from Data*, 1(1):3–es. ISSN 1556-4681.

[54] McIver, A., Morgan, C., Smith, G., Espinoza, B., and Meinicke, L. (2014). Abstract channels and their robust information-leakage ordering. In *International Conference on Principles of Security and Trust*, pages 83--102. Springer.

[55] Meindl, B., Kowarik, A., and Templ, M. (2021). sdcMicro - Statistical Disclosure Control Methods for Anonymization of Microdata and Risk Estimation. https://sdctools.github.io/sdcMicro/index.html.

[56] Narayanan, A. and Shmatikov, V. (2008). Robust De-anonymization of Large Sparse Datasets. In *Proc. of S&P*, pages 111–125. ISSN 1081-6011.

[57] Nugent, C. (2020). Brazil's Government Accused of a 'Statistical Coup' After it Limited Publishing of COVID-19 Data. https://time.com/5849959/brazil-covid-data/.

[58] Phillips, D. (2020). Brazil stops releasing Covid-19 death toll and wipes data from official site. https://www.theguardian.com/world/2020/jun/07/brazil-stops-releasing-covid-19-death-toll-and-wipes-data-from-official-site.

[59] Prasser, F., Bild, R., Eicher, J., Spengler, H., Kohlmayer, F., and Kuhn, K. A. (2016). Lightning: Utility-driven anonymization of high-dimensional data. *Trans. Data Priv.*, 9(2):161--185.

[60] Prasser, F., Eicher, J., Spengler, H., Bild, R., and Kuhn, K. A. (2020). Flexible data anonymization using ARX - Current status and challenges ahead. *Software: Practice and Experience*, 50(7):1277–1304.

[61] Prasser, F. and Kohlmayer, F. (2021). ARX - Data Anonymization Tool. https://arx.deidentifier.org/.

[62] Prasser, F., Kohlmayer, F., Lautenschlaeger, R., and Kuhn, K. A. (2014). ARX - a comprehensive tool for anonymizing biomedical data. In *AMIA Annual Symposium Proceedings*, volume 2014, page 984. American Medical Informatics Association.

[63] Queiroz, M. and Motta, G. (2015). Privacidade e Transparência no Setor público: Um Estudo de Caso da Publicação de Microdados do INEP. In *XV Simposio Brasileiro em Seguranca da Informacao e de Sistemas Computacionais-SBSeg*.

[64] Raskhodnikova, S., Smith, A., Lee, H. K., Nissim, K., and Kasiviswanathan, S. P. (2008). What can we learn privately. In *Proceedings of the 54th Annual Symposium on Foundations of Computer Science*, pages 531--540.

[65] Samarati, P. and Sweeney, L. (1998). Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression.

[66] Smith, G. (2009). On the Foundations of Quantitative Information Flow. In *Proc. of FOSSACS*, volume 5504 of *LNCS*, pages 288--302. Springer.

[67] Sweeney, L. (2000). Simple Demographics Often Identify People Uniquely. https://kilthub.cmu.edu/articles/Simple_Demographics_Often_Identify_People_Uniquely/6625769/1.

[68] Templ, M., Kowarik, A., and Meindl, B. (2015). Statistical disclosure control for micro-data using the R package sdcMicro. *Journal of Statistical Software*, 67(4):1--36.

[69] Thakurta, A. G., Vyrros, A. H., Vaishampayan, U. S., Kapoor, G., Freudiger, J., Sridhar, V. R., and Davidson, D. (2017a). Learning new words. US Patent 9,594,741.

[70] Thakurta, A. G., Vyrros, A. H., Vaishampayan, U. S., Kapoor, G., Freudinger, J., Prakash, V. V., Legendre, A., and Duplinsky, S. (2017b). Emoji frequency detection and deep link frequency. US Patent 9,705,908.

[71] The European Parliament and the Council of the European Union (1995). Directive 95/46/EC. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046.

[72] The European Parliament and the Council of the European Union (2016). Regulation (EU) 2016/679. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679.

[73] United Nations (1994). Official Statistics: Principles and Practices, Organization and Management. https://unstats.un.org/unsd/methods/statorg/default.htm.

[74] United Nations (2014a). Fundamental Principles of Official Statistics (A/RES/68/261 from 29 January 2014). https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx.

[75] United Nations (2014b). Principles Governing International Statistical Activities. https://unstats.un.org/unsd/methods/statorg/Principles_stat_activities/principles_stat_activities.htm.

[76] United States Census Bureau (2019a). Disclosure Avoidance Techniques Used for the 1960 Through 2010 Decennial Censuses of Population and Housing Public Use Microdata Samples. https://www.census.gov/library/working-papers/2019/adrm/six-decennial-censuses-da.html.

[77] United States Census Bureau (2019b). Legacy Techniques and Current Research in Disclosure Avoidance at the U.S. Census Bureau. https://www.census.gov/library/working-papers/2019/adrm/legacy-da-techniques.html.

[78] Warner, S. L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 60(309):63--69. PMID: 12261830.

[79] World Health Organization. Coronavirus disease 2019 (COVID-19) - Situation Report - 51. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf.

[80] World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

[81] Xiao, X. and Tao, Y. (2007). M-invariance: towards privacy preserving re-publication of dynamic datasets. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 689--700.