Workshop on privacy for IOT

Héber HWANG ARCOLEZI
heber.hwang_arcolezi[at]univ-fcomte[point]fr

November 5, 2020

Contents

Chapter 1

Introduction to anonymization

1.1 Pseudonymization

Exercise 1.1 (Pseudonymization). *This exercise is inspired by* [1]. *The dataset of a social network given in the table 1.1 has been pseudonymized in the strongest possible way, i.e., where the name has been removed. However, additional information is available in table 1.2. In this exercise, we use this information to explore how privacy can be compromised by intelligently linking data. Suppose that all applicants are present in both databases.*

Name	Gender	Age	City of birth	Favorite TV Series	Relationship Status
*	male	19-25	Saarbrücken	Game of Thrones	single
*	female	16-18	Trier	Game of Thrones	in relationship
*	male	12-15	München	Friends!	in relationship
*	female	19-25	Berlin	Big Bang Theory	in relationship
*	female	19-25	Hamburg	Big Bang Theory	single
*	female	19-25	Saarbrücken	Game of Thrones	single
*	male	16-18	Trier	Game of Thrones	single
*	female	12-15	München	Game of Thrones	in relationship
*	male	19-25	Berlin	Big Bang Theory	single

Table 1.1: Social network dataset, sanitized by Pseudonymization

Name	Email	TV Show	Rating (1=bad, 5=great)
Alice	alice1995@email.com	Friends!	1
Bob	bobbybob@email.com	Friends!	4
Charlie	s9charchar@email.com	Friends!	2
Eve	evelyn@myhighscool.com	Friends!	1
Bob	bobbybob@email.com	Game of Thrones	1
Alice	alice1995@email.com	Game of Thrones	5
Charlie	s9charchar@email.com	Game of Thrones	5
Bob	bobbybob@email.com	Big Bang Theory	3
Charlie	s9charchar@email.com	Big Bang Theory	5
Alice	alice1995@email.com	Big Bang Theory	2
Eve	evelyn@myhighscool.com	Big Bang Theory	5

Table 1.2: Additional information

1. *Where was Alice likely born and what is her most likely marital status?*
2. *Can you also get personal information about Charlie?*
3. *Can you also learn some personal information about Bob?*

[1] http://www.infsec.cs.uni-saarland.de/teaching/16WS/Cybersecurity/exercises/exercise-11.pdf

ID	Age	Gender	Fav.Show
1	12-15	female	Friends!
2	19-25	male	Friends!
3	19-25	male	Friends!
4	12-15	female	Friends!
5	19-25	male	G.o.T.
6	19-25	male	G.o.T.
7	19-25	male	G.o.T.

ID	Age	Gender	Fav.Show
1	19-25	female	Grey's A.
2	19-25	female	Simpsons
3	19-25	female	Futurama
4	19-25	female	Friends!
5	19-25	male	G.o.T.
6	19-25	male	C.Minds
7	19-25	male	Br.Ba.

ID	Age	Gender	Fav.Show
1	19	male	Friends!
2	19	male	Friends!
3	19	male	Friends!
4	19	female	Friends!
5	20	male	G.o.T.
6	20	male	G.o.T.
7	20	male	G.o.T.

Figure 1.1: 3 small generalized datasets

1.2 k-anonymity

Exercise 1.2 (*k*-anonymity on a simple example). *This exercise is again inspired by*[1].

1. *Does the dataset 1 in Figure 1.1 satisfy the k-anonymity? If so, what is the maximum value of k?*
2. *Same question for datasets 2 and 3.*

Exercise 1.3 (Different *k*-anonymity methods on the same micro example). *We consider the dataset of table 1.3, inspired by*[2].

1. *Propose a 3-anonymous version considering the following possible generalizations:*
   - *Marital status: separated, single, widowed $\rightsquigarrow$ **
   - *Age: 20,23,24$\rightsquigarrow$[20,24], 25,28,29 $\rightsquigarrow$ [25,29], $\rightsquigarrow$ **
   - *Zip:32021,32024,32027$\rightsquigarrow$ 3202*, 32042,32045,32046$\rightsquigarrow$ 3204*, $\rightsquigarrow$ **
2. *Propose a 3-anonymous version considering Mondrian's algorithm.*

	ID	QID			Sensitive
#	Name	Marital status	Age	Zip	Crime
1	Joe	Separated	29	32042	Murder
2	Jill	Single	20	32021	Theft
3	Sue	Widowed	24	32024	Traffic
4	Abe	Separated	28	32046	Assault
5	Bob	Widowed	25	32045	Piracy
6	Amy	Single	23	32027	Indecency

Table 1.3: Crime data

[2], V., McDonagh, P., Cerqueus, T., & Murphy, L. (2014). A systematic comparison and evaluation of k-anonymization algorithms for practitioners. Transactions on data privacy, 7(3), 337-370.

3. *Calculate the values of C_{AVG} and Discernability for both methods on this simple example.*

4. *Figure 1.2 compares the usefulness of different k -anonymization methods on two different datasets. What can we conclude from these experiments?*

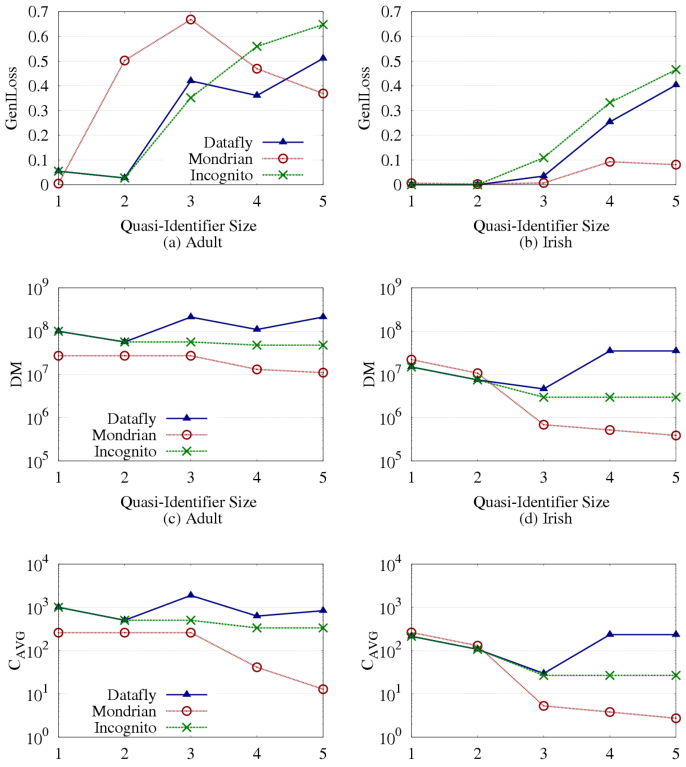

Figure 1.2: Utility measures of k-anonymity[2] algorithms