

k-Anonimato

Prof. Javam Machado

LSBD/DC/UFC

Setembro/2019

Revisão

Revisão - Tipos de Atributos

- Identificadores explícitos
- Semi-identificadores
- Atributos sensíveis

Nome	Idade	Gênero	Endereço	Telefone	Conta	Saldo R\$
David C. F.	22	M	Av. L	98533 1234	2234-0	1.033,25
John T. A.	23	M	Av. K	98772 2531	7749-2	814,92
Helton B.	25	M	Av. K	98156 0092	8491-7	515,09
Maria L. Jr	32	F	Rua J	99913 9026	5723-1	2.194,79

Revisão - Tipos de Atributos

- **Identificadores explícitos**
- Semi-identificadores
- Atributos sensíveis

Nome	Idade	Gênero	Endereço	Telefone	Conta	Saldo R\$
David C. F.	22	M	Av. L	98533 1234	2234-0	1.033,25
John T. A.	23	M	Av. K	98772 2531	7749-2	814,92
Helton B.	25	M	Av. K	98156 0092	8491-7	515,09
Maria L. Jr	32	F	Rua J	99913 9026	5723-1	2.194,79

Revisão - Tipos de Atributos

- **Identificadores explícitos**
- Semi-identificadores
- Atributos sensíveis

Nome	Idade	Gênero	Endereço	Telefone	Conta	Saldo R\$
David C. F.	22	M	Av. L	98533 1234	2234-0	1.033,25
John T. A.	23	M	Av. K	98772 2531	7749-2	814,92
Helton B.	25	M	Av. K	98156 0092	8491-7	515,09
Maria L. Jr	32	F	Rua J	99913 9026	5723-1	2.194,79

Revisão - Tipos de Atributos

- Identificadores explícitos
- **Semi-identificadores**
- Atributos sensíveis

Nome	Idade	Gênero	Endereço	Telefone	Conta	Saldo R\$
David C. F.	22	M	Av. L	98533 1234	2234-0	1.033,25
John T. A.	23	M	Av. K	98772 2531	7749-2	814,92
Helton B.	25	M	Av. K	98156 0092	8491-7	515,09
Maria L. Jr	32	F	Rua J	99913 9026	5723-1	2.194,79

Revisão - Tipos de Atributos

- Identificadores explícitos
- **Semi-identificadores**
- Atributos sensíveis

Nome	Idade	Gênero	Endereço	Telefone	Conta	Saldo R\$
David C. F.	22	M	Av. L	98533 1234	2234-0	1.033,25
John T. A.	23	M	Av. K	98772 2531	7749-2	814,92
Helton B.	25	M	Av. K	98156 0092	8491-7	515,09
Maria L. Jr	32	F	Rua J	99913 9026	5723-1	2.194,79

Revisão - Tipos de Atributos

- Identificadores explícitos
- Semi-identificadores
- **Atributos sensíveis**

Nome	Idade	Gênero	Endereço	Telefone	Conta	Saldo R\$
David C. F.	22	M	Av. L	98533 1234	2234-0	1.033,25
John T. A.	23	M	Av. K	98772 2531	7749-2	814,92
Helton B.	25	M	Av. K	98156 0092	8491-7	515,09
Maria L. Jr	32	F	Rua J	99913 9026	5723-1	2.194,79

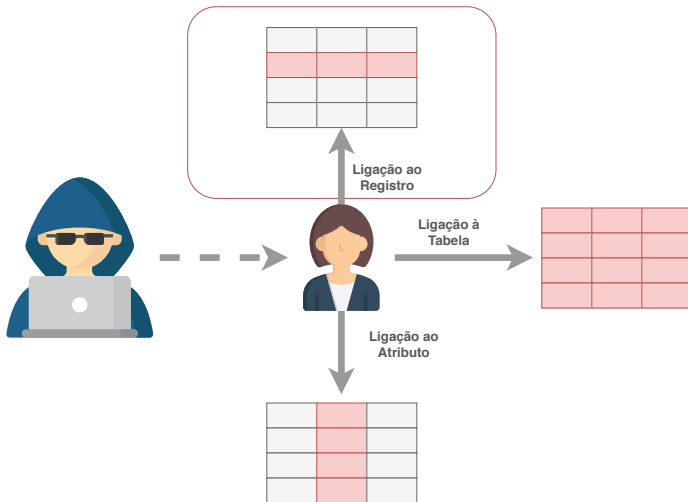
Revisão - Tipos de Atributos

- Identificadores explícitos
- Semi-identificadores
- **Atributos sensíveis**

Nome	Idade	Gênero	Endereço	Telefone	Conta	Saldo R\$
David C. F.	22	M	Av. L	98533 1234	2234-0	1.033,25
John T. A.	23	M	Av. K	98772 2531	7749-2	814,92
Helton B.	25	M	Av. K	98156 0092	8491-7	515,09
Maria L. Jr	32	F	Rua J	99913 9026	5723-1	2.194,79

Tipos de ataques

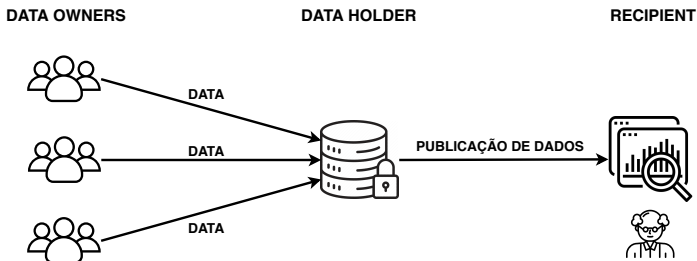
■ Ataques de ligação ao registro



Motivação

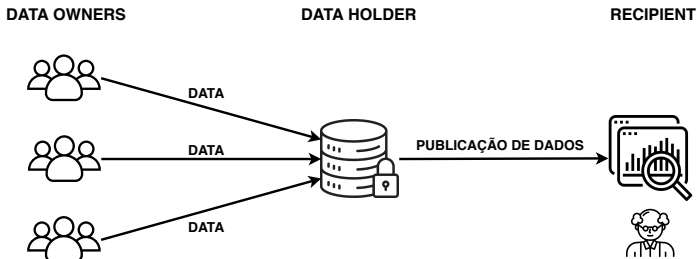
Privacidade + Publicação de dados

■ Privacy Preserving Data Publishing (PPDP)



Privacidade + Publicação de dados

■ Privacy Preserving Data Publishing (PPDP)



■ Exemplos:

- Cidadão → Censo (IBGE) → Governo ou Público em Geral
- Pacientes → Hospital → Laboratório Farmacêutico ou Pesquisador Médico

k-Anonimato

- O que é um modelo?
 - Funciona como um contrato
 - Estabelece quais garantias serão oferecidas
 - Define as regras para publicação

Modelo k-Anonimato

O registro de um indivíduo não deve ser identificado com probabilidade maior que $\frac{1}{k}$

Modelo k-Anonimato

O registro de um indivíduo não deve ser identificado com probabilidade maior que $\frac{1}{k}$



As classes de equivalência devem ter tamanho mínimo de k

Modelo k-Anonimato

O registro de um indivíduo não deve ser identificado com probabilidade maior que $\frac{1}{k}$



As classes de equivalência devem ter tamanho mínimo de k



Para cada registro, devem existir, pelo menos, k-1 registros iguais a ele

Exemplo

- Um atacante sabe acerca de Bob:
 - Participou de uma pesquisa sobre profissões em Fortaleza (Tabela E)
 - É paciente de um dado hospital que liberou dados para análises (Tabela T)
- Qual a chance do atacante reidentificar o registro de Bob?

Profissão	Idade	Cidade	Doença
Engenheiro	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	Fortaleza	AIDS
Programador	32	Quixadá	AIDS
Dentista	35	Quixadá	AIDS

Tabela: Disponibilização para realização de pesquisas (T)

Nome	Profissão	Idade	Cidade
David	Programador	29	Fortaleza
Pedro	Médico	32	Fortaleza
Bob	Programador	32	Fortaleza
Maria	Dentista	35	Fortaleza

Tabela: Dados Externos (E)

Exemplo

- Um atacante sabe acerca de Bob:
 - Participou de uma pesquisa sobre profissões em Fortaleza (Tabela E)
 - É paciente de um dado hospital que liberou dados para análises (Tabela T)
- Qual a chance do atacante reidentificar o registro de Bob?

Profissão	Idade	Cidade	Doença
Engenheiro	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	Fortaleza	AIDS
Programador	32	Quixadá	AIDS
Dentista	35	Quixadá	AIDS

Tabela: Disponibilização para realização de pesquisas (T)

Nome	Profissão	Idade	Cidade
David	Programador	29	Fortaleza
Pedro	Médico	32	Fortaleza
Bob	Programador	32	Fortaleza
Maria	Dentista	35	Fortaleza

Tabela: Dados Externos (E)

Exemplo

- Um atacante sabe acerca de Bob:
 - Participou de uma pesquisa sobre profissões em Fortaleza (Tabela E)
 - É paciente de um dado hospital que liberou dados para análises (Tabela T)
- Qual a chance do atacante reidentificar o registro de Bob?
100%!

Profissão	Idade	Cidade	Doença
Engenheiro	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Chikungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	Fortaleza	AIDS
Programador	32	Quixadá	AIDS
Dentista	35	Quixadá	AIDS

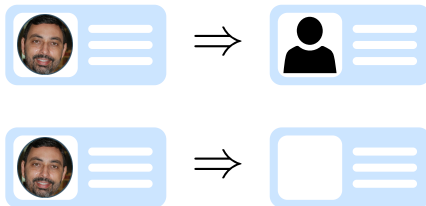
Tabela: Disponibilização para realização de pesquisas (T)

Nome	Profissão	Idade	Cidade
David	Programador	29	Fortaleza
Pedro	Médico	32	Fortaleza
Bob	Programador	32	Fortaleza
Maria	Dentista	35	Fortaleza

Tabela: Dados Externos (E)

Como anonimizar?

■ Generalização e/ou Supressão



Exemplo

■ 2-anonimato com Supressão

Cidade	Idade	Renda
Fortaleza	20	1800
Sobral	25	920
Quixadá	20	1800
Sobral	22	2200

Exemplo

■ 2-anonimato com Supressão

Cidade	Idade	Renda
Fortaleza	20	1800
Sobral	25	920
Quixadá	20	1800
Sobral	22	2200



Cidade	Idade	Renda
*	20	1800
Sobral	*	*
*	20	1800
Sobral	*	*

Exemplo

■ 2-anonimato com Supressão

Cidade	Idade	Renda	\Rightarrow	Cidade	Idade	Renda
Fortaleza	20	1800		*	20	1800
Sobral	25	920		Sobral	*	*
Quixadá	20	1800		*	20	1800
Sobral	22	2200		Sobral	*	*



Esse dado parece bom para análise?

■ 2-anonimato com Generalização

Profissão	Idade	Cidade	Doença
Engenheiro	29	Fortaleza	Chinkungunya
Médico	29	Fortaleza	Chinkungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	Fortaleza	AIDS
Programador	32	Quixadá	AIDS
Dentista	35	Quixadá	AIDS

Exemplo

■ 2-anonimato com Generalização

Profissão	Idade	Cidade	Doença
Engenheiro	29	Fortaleza	Chinkungunya
Médico	29	Fortaleza	Chinkungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	Fortaleza	AIDS
Programador	32	Quixadá	AIDS
Dentista	35	Quixadá	AIDS

Exemplo

■ 2-anonimato com Generalização

Profissão	Idade	Cidade	Doença
Engenheiro	29	Fortaleza	Chinkungunya
Médico	29	Fortaleza	Chinkungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	Fortaleza	AIDS
Programador	32	Quixadá	AIDS
Dentista	35	Quixadá	AIDS



Profissão	Idade	Cidade	Doença
Profissional	[25-35]	CE	Chinkungunya
Médico	29	Fortaleza	Chinkungunya
Médico	29	Fortaleza	Hepatite C
Programador	32	CE	AIDS
Programador	32	CE	AIDS
Profissional	[25-35]	CE	AIDS

Atividade A

- Criar generalizações para os seguintes dados de forma a garantir o modelo 2-anonimato

45 35 22 24 28 27 31 18 65 42

Atividade B

- Criar classes de equivalência capaz de garantir o modelo 2-anonimato
 - semi-identificadores: idade, renda
 - atributo sensível: doença

Idade	Renda	Doença
20	920	Gripe
25	1500	Dengue
30	1800	Gastrite
35	2200	Pneumonia

Atividade C

- Utilizando Supressão e Generalização aplique o modelo 2-anonimato
 - Utilize supressão por cada nível de data

Placa	Motorista	CPF	Data de Nascimento	Data da Infração	Tipo de Multa	Valor da Multa
HXR-1542	José Pereira	258.568.856	14/03/1977	03/01/2013	1	170
HTS-5864	Jorge Cury	566.548.584	04/03/1977	03/01/2013	2	250
HUI-5846	Paula Maria	384.987.687	24/05/1977	03/01/2013	1	170
HTR-5874	Joatan Lima	054.864.576	20/04/1978	04/01/2013	1	170
HOI-6845	José Sá	244.684.876	22/05/1978	04/01/2013	2	250
HQO-5846	Kilvia Mota	276.684.159	13/05/1978	05/01/2013	2	250
HUY-8545	José Pereira	538.687.045	15/05/1978	05/01/2013	1	170

Para pensar...

- Dado um conjunto de dados
 - Registro possui mais de um atributo sensível
- Dado o valor de k
- Dadas as hierarquias de generalização
- Dada uma métrica de utilidade

Como gerar as classes de equivalência de forma a otimizar a utilidade?

Atividade C

Atividade C - Passo Intermediário

Placa	Motorista	CPF	Data de Nascimento	Data da Infração	Tipo de Multa	Valor da Multa
*	*	*	03/1977	01/2013	1	170
*	*	*	03/1977	01/2013	2	250
*	*	*	05/1977	01/2013	1	170
*	*	*	04/1978	01/2013	1	170
*	*	*	05/1978	01/2013	2	250
*	*	*	05/1978	01/2013	2	250
*	*	*	05/1978	01/2013	1	170

Atividade C - Possível Solução

Placa	Motorista	CPF	Data de Nascimento	Data da Infração	Tipo de Multa	Valor da Multa
*	*	*	03/1977	01/2013	1	170
*	*	*	03/1977	01/2013	2	250
*	*	*	*	*	*	*
*	*	*	*	*	*	*
*	*	*	05/1978	01/2013	2	250
*	*	*	05/1978	01/2013	2	250
*	*	*	05/1978	01/2013	1	170