

Singularidade de dados e os limites da privacidade

Cada vez mais os dados fazem parte de nossas vidas, estamos consumindo e produzindo milhares de informações continuamente, isso é possível devido o advento de novas tecnologias, entre elas: os smartphones, computadores, *wearables gadgets*, etc. Todos esses avanços tecnológicos nos permitiram trabalhar com dados anteriormente inimagináveis, tais como informações de cartão de crédito, ou de localização. Apesar desses dados mais modernos serem extremamente valiosos em aplicação, são informação altamente ligadas às particularidades dos usuários, isso implica que cada pessoa fornece informações únicas sobre si, possibilitando que esses dados sejam facilmente identificados, o que impacta diretamente sobre os limites da privacidade dos usuários [1, 2].

A informação com alto grau de singularidade, ou seja, que possui relacionamento forte com o indivíduo que o fez, tende a ser um dado armazenado com anonimato, um tipo de anonimato simples é definido pela ausência de dados sensíveis próximo dele, como nome, CPF, *email*, etc. O problema é que por serem dados muito únicos, basta a presença de um *dataset* secundário para romper a barreira do anonimato e reidentificar os dados singulares. Por exemplo, saber o endereço do trabalho/moradia, no caso de localização, ou quanto uma pessoa paga por um produto, no caso de cartão de crédito [1, 2].

Em termos de mobilidade ou localização, pesquisas demonstram que é possível reidentificar um usuário com 95% de certeza, caso seja fornecido quatro pontos no tempo-espaço escolhidos ao acaso. A pesquisa foi realizada através de um dataset com 1.5 milhões de usuários de um país europeu com população representativa, os dados são de antenas de celulares e de serviços como foursquare, e geolocalização no twitter. O artigo demonstra que é possível descrever a singularidade de mobilidade por uma expressão matemática não-linear, dependente da resolução da informação espacial, e dos intervalos de tempo em que são coletadas, demonstrando a conexão desses fatores com a singularidade dos dados. A pesquisa revela também que os pontos escolhidos podem possuir representatividades diferentes, ou que adicionar informação de localização externa pode pouco influenciar no anonimato dos dados, mas que a adição de mais usuários, de fato, torna os pontos de localização menos singulares [2].

Em [1], vemos que dados financeiros, como o histórico de compras utilizando o cartão de crédito, mesmo tendo dados sensíveis anonimizados, podem ser utilizados para reidentificar 90% dos usuários. A pesquisa foi feita utilizando um dataset com o histórico de operações financeiras de 1.1 milhões de usuários. No texto foi proposta uma métrica para avaliar o risco de identificar um

usuário conhecendo uma quantidade n de informação externa sobre o usuário, o que chamaram de *Unicity*. Quanto maior for a quantidade externa de informação, mais identificável é um usuário. Com apenas 4 dados espaço-temporais é possível rastrear 90% dos usuários. Vemos também que as mulheres estão mais expostas a serem reidentificadas, bem como pessoas com altos ganhos financeiros.

Os avanços tecnológicos nos permitiram lidar com todo o tipo de dado, os singulares são a bola da vez. Os serviços se aproximam mais e mais dos usuários fornecendo caráter personalizado, empresas como Apple re-escrevem seus contratos para fornecer mais dados sensíveis aos usuários, mas quais os limites em que a privacidade do usuário deve ser respeitada? As informações singulares muita vezes são obtidas de forma ubíqua, e não nos damos conta de que fornecemos tais informações [1, 2]. Junto a isso, surge um novo usuário mais preocupado com suas individualidades, e junto com ele novas empresas, que cada vez mais protege seus dados, e se preocupa quando o assunto é privacidade.

[1] DE MONTJOYE, Yves-Alexandre et al. Unique in the shopping mall: On the reidentifiability of credit card metadata. **Science**, v. 347, n. 6221, p. 536-539, 2015.

[2] DE MONTJOYE, Yves-Alexandre et al. Unique in the crowd: The privacy bounds of human mobility. **Scientific reports**, v. 3, p. 1376, 2013.