CRM 2022 WORKSHOP
BNC Challenge

Défis

- Mesurer la performance, la sécurité et la confidentialité d'une approche anonymisée
- Accélérer l'accès aux informations personnelles en utilisant des données anonymisées ou synthétisées avec la confiance d'avoir le contrôle sur le risque
- Trouver le meilleur équilibre possible entre la vie privée des utilisateurs et l'utilité des données

Facteurs à considérer pour l'évaluation des critères

- La valeur de l'information : est-ce que l'information est confidentielle ou prévisible? Si l'information est confidentielle, elle sera considérée comme étant plus à risque.

- Les probabilités d'accès : si une information est facile à obtenir, alors elle ne doit pas permettre d'identifier un individu.

Critères pour évaluer la qualité de l'anonymisation (Risque VS Utilité)

RISQUE :

- La distance entre observations : chaque observation doit avoir plusieurs autres observations qui lui ressemblent dans la zone de confusion (logique du k-anonymat).

- L'inférence : il ne doit pas être possible de déduire, de façon quasi certaine, de nouvelles informations sur un individu.

- *Differential privacy* : la suppression d'une observation dans une base de données ne devrait pas influencer significativement les résultats des analyses.

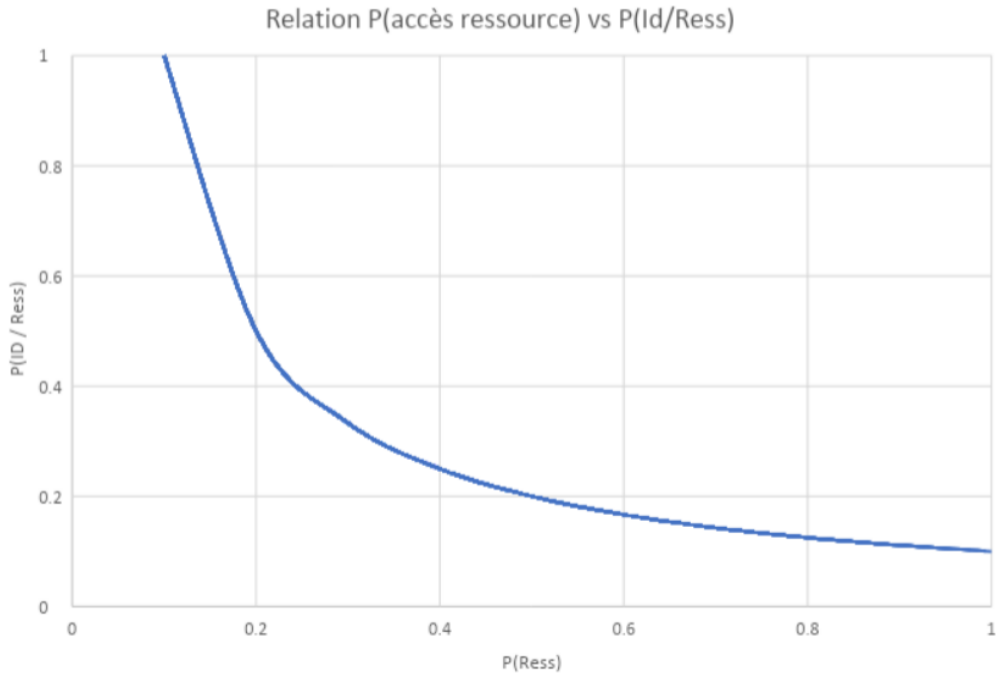
UTILITÉ :

- La corrélation : une bonne anonymisation doit préserver globalement les corrélations entre les variables.

Relation

Probabilité d'identifier un individu =

Probabilité d'obtenir des attributs * Probabilité d'identifier un individu connaissant ces attributs (calcul de la valeur)



Objectif : Trouver un équilibre entre le risque et l'utilité

Recommandations pour évaluer les critères

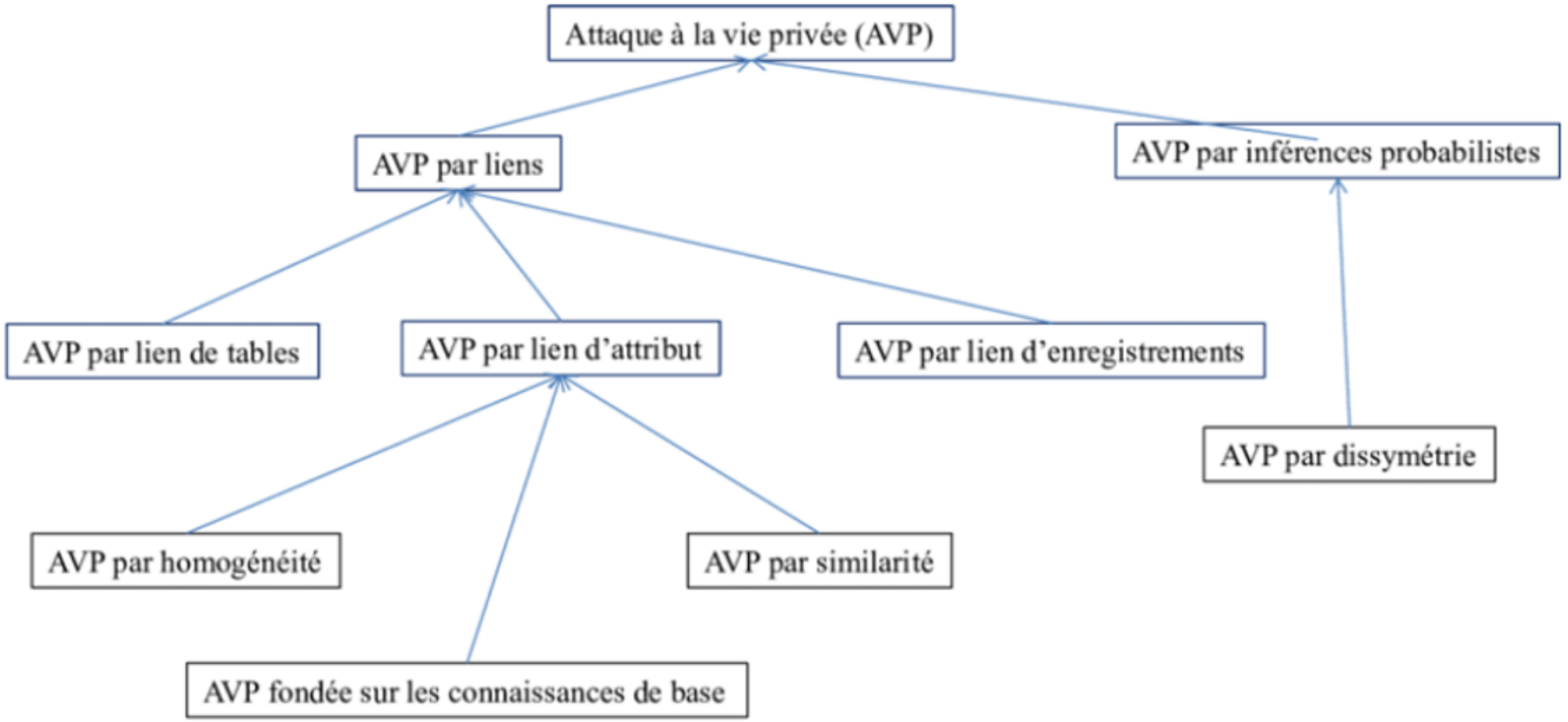
- Technique des composantes principales
- Évaluation : trouver un couplage parfait (notion de théorie des graphes) entre les données originelles et les données anonymisées
- Distance entre les observations : technique du plus proche voisin
- Tenir compte de la gouvernance des données

TabFormer : Exemples de *patterns*

- *Zip* + *Merchant name* = Localisation de l'entreprise
- *Merchant state* + *Merchant city* + *Merchant name* = Localisation de l'entreprise
- *Zip* + *Merchant name* + *UseChip* = Ville du client
- *Merchant state* + *Merchant city* + *Merchant name* + *Date* = Possibilité d'identifier le client et d'obtenir l'*account number*
- *Merchant state* + *Merchant city* + *Merchant name* + *Date* + Habitudes d'un consommateur (réseaux sociaux) = Possibilité de trouver l'*account number* d'un client et ses transactions (vol d'identité)

Questions ?

Processus d'une attaque de la vie privée



https://www.researchgate.net/figure/Taxonomie-des-modeles-dattaque-de-la-vie-privee-Le-modele-dattaque-par-liens-sapplique_fig2_324938181