

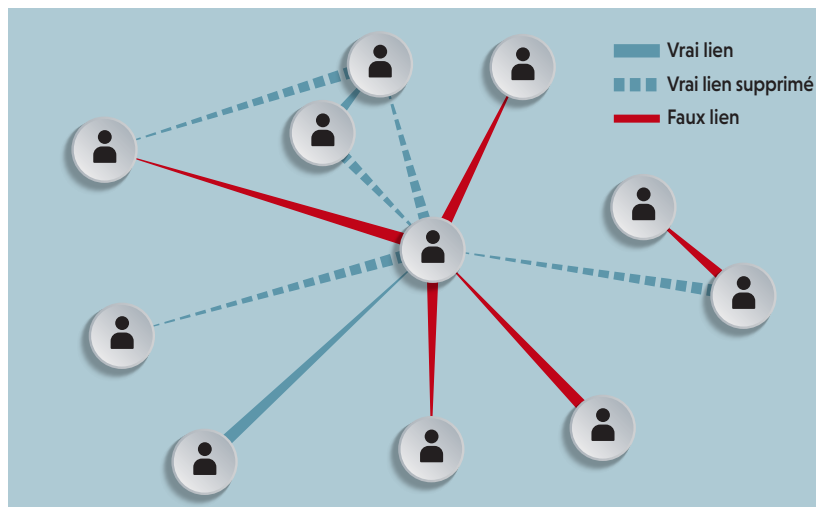
York, et ses collègues, quantifie la connaissance préalable qu'a l'attaquant de sa cible par la probabilité de trouver la valeur convoitée avant l'observation du jeu de données: si l'attaquant recherche le diagnostic médical d'une personne nommée Dupont, qu'il ne connaît de sa cible que son nom et qu'il sait que 10% des Dupont ont un cancer, il a 1 chance sur 10 de ne pas se tromper en affirmant que sa cible a un cancer. Cette connaissance est dite *a priori*. Après la découverte des données, l'attaquant a plus de chances de trouver la donnée qu'il cherche; la connaissance de sa cible qui en résulte, connaissance dite *a posteriori*, est quantifiée par la probabilité de lui associer la bonne valeur sensible.

C'est ici que la théorie bayésienne intervient, car elle permet de calculer des probabilités conditionnelles: les probabilités de déduire telle ou telle information d'un jeu de données sachant qu'on a telle ou telle connaissance préalable. La confidentialité des données publiées est caractérisée par la différence entre connaissance *a posteriori* et connaissance *a priori*, autrement dit par la connaissance supplémentaire sur la cible apportée par les données. Assurer un niveau adéquat de protection revient à limiter cette valeur. Cependant, cette technique est difficile à mettre en œuvre, car elle suppose de connaître exactement ce que l'attaquant sait de sa cible.

La *l*-diversité, méthode conçue la même année et par les mêmes chercheurs, est davantage applicable. Elle consiste à imposer la présence de *l* valeurs sensibles distinctes dans chaque groupe (voir l'encadré page 102). Un attaquant qui ne connaît de sa cible que le groupe auquel elle appartient ne peut trouver sa valeur sensible avec une probabilité supérieure à $1/l$.

De nombreuses autres techniques ont été élaborées depuis, afin de s'adapter à la diversité des données, des attaques et des attaquants. La «*t*-proximité» vise à créer des groupes au sein desquels la distribution des données est à peu près la même que dans la population globale.

Les graphes de réseau social représentent les individus par des points et leurs liens par des traits. Pour qu'ils ne dévoilent pas la vie privée des individus, on peut leur appliquer un algorithme dit de confidentialité différentielle, consistant, par exemple, à supprimer de nombreux vrais liens et à les remplacer par autant de faux liens.



La «*(c, k)*-sûreté» prend en compte la capacité de l'attaquant à formuler *k* déductions logiques du type: «Si Adrien a la grippe, alors Bettie aussi.» La «3D-confidentialité» considère que l'attaquant connaît *a priori* trois types de données: *l* valeurs sensibles que sa cible n'a pas, *k* valeurs sensibles que d'autres individus ont et *m* déductions logiques entre individus. La «*m*-invariance» protège contre les comparaisons entre les publications successives d'un même jeu de données, telles certaines informations statistiques sur un hôpital, dont les variations reflètent les arrivées, les départs et les évolutions des patients.

L'application de ces techniques commence souvent par celle de l'algorithme de Mondrian. Il sert à construire des groupes *k*-anonymes, au sein desquels on vérifie la conformité de la distribution des données sensibles vis-à-vis du modèle choisi (voir l'encadré page 102).

Volontairement ou non, tous ces modèles sont des exemples du paradigme de non-information, formulé en 1977 par le statisticien suédois Tore Dalenius: selon ce paradigme, un jeu de données est d'autant plus confidentiel qu'il renseigne peu l'attaquant sur sa cible. En 2006, Cynthia Dwork, alors chercheuse à Microsoft, a proposé une nouvelle façon de définir la confidentialité, qualifiée de confidentialité différentielle: un jeu de données serait confidentiel s'il n'est presque pas modifié par l'ajout des données d'un individu, quelles qu'elles soient.

NOYER LES DONNÉES

L'assainissement consiste alors à perturber les données de sorte que celles de chaque individu se trouvent noyées dans la perturbation (et non plus dans la foule, comme dans le *k*-anonymat). On introduit, par exemple, un grand nombre de fausses données (typiquement 100 fois plus que de vraies), afin que le jeu de données qui contient la contribution d'un individu particulier ressemble beaucoup à celui qui ne la contient pas. Certains algorithmes suppriment aussi de vraies données. La protection est assurée par l'impossibilité de distinguer les fausses données des vraies dans le jeu de données assaini, dont on ne peut même pas garantir qu'il contient la contribution d'un individu précis.

L'idée de la confidentialité différentielle est en plein essor. Elle est désormais applicable à des types variés de données, tels les graphes de réseaux sociaux. Ces graphes représentent les individus par des nœuds, connectés par des liens. La perturbation peut alors consister à supprimer de vrais liens et à en introduire de faux. Des informations tel le nombre de liens sont toujours extractibles du graphe, sans que l'on puisse déterminer les connexions précises d'un individu.

Le succès de la confidentialité différentielle s'explique en partie par sa simplicité: l'attaquant >