Quoi|ProCo

Louis Béziaud[1,2], Noé Brucy[1,2], Joris Duguépéroux[1,2], et Antonin Voyez[1]

[1]Université Rennes 1
prénom.nom@etudiant.univ-rennes1.fr
[2]École normale supérieure de Rennes
prénom.nom@ens-rennes.fr

Joris Duguépéroux est porteur du projet

Pitch du projet

Dans le cadre de notre projet, nous proposons deux activités ludiques et pédagogiques et débranchées mettant en avant des aspects souvent négligés lorsque l'on aborde la problématique de la vie privée. Tout d'abord, le « Quoi À Qui ? » permet de réfléchir sur les préférences de chacun.e, par rapport à ce qu'il ou elle souhaite diffuser sur un réseau social. Par la suite, on peut se rendre compte que les alternatives proposées par les réseaux sociaux sont souvent binaires : accepter les conditions d'utilisation, ou ne rien diffuser. Ensuite, notre seconde activité, le « QuiProCo », se présente sous la forme d'un jeu qui illustre les notions de réponse randomisée et de confidentialité différentielle, qui font partie intégrante des standards actuels de la recherche en terme de protection de vie privée. Elle cherche à montrer que collecte d'information et respect de la vie privée sont réconciliables. Ces activités sont conçues pour des groupes ou classes entières de collégiens.

Constat

Lorsque l'on parle de vie privée, deux types d'argumentaires reviennent souvent pour défendre ce qui peut être perçu comme des intrusions.

Le premier, se résumant souvent par la phrase « Je n'ai rien à cacher », voudrait que l'on n'ait pas besoin de protéger sa vie privée si l'on n'est innocent.e. Le second s'appuie davantage sur le fait que les données privées fournies permettent une amélioration des services ou une diminution du coût pour l'utilisateur.trice. Ces deux assertions ne sont pas nécessairement fausses, mais elles sont simplistes et réductrices.

De même, les solutions proposées aux utilisateur.trice.s pour protéger leur vie privée se résument fréquemment à « désinscrivez-vous » ou « faites attention ».

Nous nous proposons de fournir des éléments ludiques pour une réflexion poussée issue de percées récentes de la recherche en informatique. Ceux-ci complètent et relativisent ces deux simplifications, tout en autorisant une mise en œuvre facile en salle de classe.

Concept

Quoi À Qui

La première activité consiste en une activité simple sous forme d'une unique fiche que l'élève pourra remplir par lui-même. Ces questions mettent en avant l'idée que l'on ne souhaite pas nécessairement diffuser les mêmes informations à tous les publics. Par la suite, une comparaison peut être effectuée par l'élève entre son utilisation réelle et l'utilisation

qu'il ou elle aimerait en avoir. Comme la plupart des réseaux sociaux ne proposent pas une granularité comparable à celle de notre formulaire mais un simple choix binaire (accepter les conditions ou se désinscrire), cette démarche peut également inciter à réfléchir davantage avant de poster un contenu.

QuiProCo

La seconde activité a pour but de mettre en avant des techniques alternatives pour la collecte de données, qui permettent de récupérer des statistiques fiables sur les utilisateurs sans pour autant compromettre leur vie privée.

L'idée est de séparer la classe deux : d'un côté, un.e élève qui représente l'entité souhaitant apprendre des informations sur les utilisateur.trice.s, et de l'autre, le reste de la classe, représentant lesdit.e.s utilisateur.trice.s. Les élèves formant le groupe test auront chacun.e à leur disposition une carte, symbolisant un profil d'utilisateur (ici, des chats ayant différentes particularités). L'élève à l'écart, dont le but est d'enquêter, aura deux objectifs. Le premier objectif sera générique, par exemple, connaître le nombre d'individus ayant une certaine propriété. Le second objectif sera personnel : apprendre des informations sur un individu précis, ou un profil spécifié.

Les élèves du groupe devront répondre aux questions posées par l'élève enquêteur.trice, mais en mentant aléatoirement (voir l'annexe pour plus de détails). Ainsi, on pourra observer que l'enquêteur.trice obtiendra des statistiques proches de la réalité pour sa question générique, mais qu'il sera bien plus complexe d'inférer des informations personnelles.

Applications

Cette activité pourra être introduite par des enseignants de disciplines différentes. Par exemple, les probabilités à la base de la réponse randomisée, ou la combinatoire des cartes, pourront être développées en cours de mathématiques, alors qu'un cours d'histoire ou de sciences civiques pourra insister sur l'utilisation historique de la réponse randomisée pour l'étude des populations (sur l'avortement ou le communisme par exemple).

Valeur ajoutée

Partant du constat qu'une grande part de l'apprentissage passe par l'école, nous avons choisi pour ce projet de privilégier une approche en collaboration avec l'enseignement. Nous proposons donc des activités à effectuer en classe, s'appuyant sur un.e pédagogue, qui saura s'adapter aux besoins de ses élèves mieux qu'une activité figée.

De plus, l'utilisation de techniques pointues mais accessibles permet d'être au point avec l'état de l'art, et de ne pas répondre à la problématique de la vie privée par une défiance généralisée vis à vis de l'informatique.

Par ailleurs, ces activités permettent de combler l'absence d'animation d'informatique débranché liées à la vie privée.

Objectifs

Dans le cadre de ce projet, nous souhaitons proposer une activité qui permettent aux collégien.ne.s de réfléchir à deux principaux aspects de la vie privée. Tout d'abord, réfléchir sur leur propre perception de la vie privée, afin de savoir si leur utilisation des réseaux sociaux correspond à ce qu'ils ou elles souhaitent en faire. Ensuite, réfléchir à l'idée souvent admise selon laquelle une certaine perte de vie privée est nécessaire à la gratuité ou à l'amélioration de services : ce préjugé est en réalité souvent faux, puisque des statistiques peuvent très bien être récupérées à grande échelle sans pour autant impacter la vie privée des individus. Les élèves pourront ainsi se rendre compte qu'il existe des alternatives efficaces et respectueuses de la vie privée à la collecte massive d'informations personnelles.

Cible

Notre objectif est de cibler les collégiens dans leur ensemble. Pour atteindre un objectif aussi général, le recours à l'enseignant dans notre projet est essentiel afin de permettre une grande possibilité d'adaptation.

Le marché, la concurrence, le benchmark

À notre connaissance, il n'existe aucune activité d'informatique débranchée traitant de la protection de la vie privée. A fortiori, il n'existe pas de ressource pédagogique en lien avec les réponses randomisées ou la confidentialité différentielle. Nous nous ancrons donc dans un milieu libre de toute concurrence, mais également dépourvu de benchmark auquel se comparer.

Le produit pédagogique

Le produit pédagogique délivré, contenant la fiche à remplir pour le « Quoi À Qui » ainsi que le détail des règles et le design des cartes pour le « QuiProCo », est fourni en annexe.

Le design

Dans un premier temps, notre proposition se veut simple et rudimentaire. Bien qu'à l'avenir, nous souhaiterions proposer un design plus élaboré, ce prototype est tout à fait fonctionnel, et il serait tout à fait envisageable qu'un design alternatif soit réalisé (dans le cadre d'un cours d'art plastique avec les élèves par exemple).

La solution technique

Le prototype de test en annexe est utilisable immédiatement pour le « Quoi À Qui », et après découpage pour le « QuiProCo », à condition d'être muni d'un dé (ou d'une source quelconque d'aléas).

Des designs supplémentaires peuvent être créés en suivant le prototype fourni du « QuiProCo ».

Le plan de communication, les modalités de diffusion

Nous espérons pouvoir profiter du bouche à oreille en présentant notre projet à différents événements (Fête de la science, Forum des mathématiques, etc.) auxquels se présentent fréquemment des enseignants et autres pédagogues, et en sensibilisant ces derniers aux problématiques liées à la vie privée.

De plus, nous aimerions également présenter notre projet à différentes équipes et à certain.e.s chercheur.se.s s'intéressant de près à la pédagogie et à la médiation scientifique, dans le cadre de l'« informatique débranchée » par exemple, afin de leur présenter nos activités (nous pensons notamment à Marie Duflot–Kremer, du Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA), ou Martin Quinson, de l'Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA) avec qui nous sommes déjà en contact).

À terme, nous aimerions également pouvoir diffuser ces activités sur internet, sur les sites dédiés de pédagogues intéressé.e.s.

Budget

Afin de maintenir au mieux la continuité de notre projet, et de contribuer à le diffuser, plusieurs points nous semblent encore nécessaires.

— Tout d'abord, les apports de graphistes ou dessinateurs seraient pertinents, afin d'obtenir un matériel de qualité sans pour autant être confronté à des problèmes de droits. À ce titre nous visons un budget prévisionnel de 1000€.
— Ensuite, afin de promouvoir au mieux ce projet, il serait également intéressant de pouvoir organiser des rencontres d'autres personnes impliquées dans ce type de projet (par exemple Marie Duflot-Kremer du LORIA, ou Martin Quison de l'IRISA). Un financement serait donc nécessaire afin de pouvoir organiser déplacements et logements sur place. Un budget prévisionnel d'environ 550€ est envisagé pour une rencontre avec les personnes impliquées dans ces démarches à Nancy ainsi qu'à Rennes.

Un total de 1550€ serait donc nécessaire.

Planning

En terme d'organisation, nous comptons chercher des artistes afin de produire nos designs dès que nous aurons reçu des financements, et espérons terminer complètement le produit d'ici la fin de l'été. Partant de là, une visite au LORIA et à l'IRISA seraient envisageables en septembre afin de promouvoir notre projet.

L'équipe

Notre équipe est composée exclusivement d'étudiants en informatique, bien que nos profils soient variés.

— **Louis Béziaud** est étudiant en informatique, en M1 à l'École normale supérieure de Rennes et l'Université de Rennes 1
— **Noé Brucy** est étudiant en informatique, en M1 à l'École normale supérieure de Rennes et l'Université de Rennes 1
— **Joris Duguépéroux** est étudiant informatique, en M2 à Rennes, et s'intéresse de très près aux thématiques en lien avec la protection de la vie privée, dans l'optique de poursuivre un doctorat dans ce domaine.
— **Antonin Voyez** est étudiant en informatique, en L3 à l'Université de Rennes 1

Remerciements

Nous remercions Tristan Allard, maître de conférences en informatique à l'Université de Rennes 1, qui a bien voulu encadrer notre projet. Il est spécialiste des techniques d'anonymisation de données dans l'équipe Druid du laboratoire IRISA. Nous remercions par ailleurs l'équipe Druid qui a hébergé nos réunions.

Nous avons reçu le soutien de Martin Quinson, professeur à l'ENS Rennes et chercheur dans l'équipe Myriads du laboratoire IRISA, qui s'intéresse à l'enseignement de l'informatique et encadre un module de pédagogie durant lequel les élèves de l'ENS Rennes organisent des activités d'informatique débranchée dans des lycées et écoles primaires d'Ille-et-Vilaine. Ce module a été suivi par Noé Brucy, Louis Béziaud et Joris Duguépéroux.

Annexes

Le reste de ce document est composé de la fiche d'activité **Quoi À Qui ?** et du matériel nécessaire pour le **QuiProCo**.

Quoi À Qui ?

Cette fiche est totalement personnelle et ne regarde que toi, mais tu peux comparer tes résultats aux autres si tu le souhaites.
Utilise le tableau ci-contre pour t'interroger sur la façon dont tu partages les informations te concernant.

Fais une croix dans chaque case si tu accepterais de partager l'information en ligne à la personne en colonne.

	Réseau social	Parents	Enseignants	Amis	Famille éloignée	Amis de tes amis	Inconnus rencontrés sur internet	Inconnus rencontrés dans la rue
Nom								
Prénom								
Pseudo								
Âge								
Adresse postale								
Adresse e-mail								
Goûts musicaux								
Résultats scolaires								
Numéro de téléphone								
Photos sans mon visage								
Photos avec mon visage								
Relations amoureuses								

Une fois le tableau rempli, demande-toi si ton utilisation actuelle d'Internet respecte ce que tu as indiqué.

Si tu constates beaucoup de différences, pense à ce tableau la prochaine fois que tu partages une information sur Internet.

QuiProCo

QUI est-ce, ré**P**onse **R**and**O**misée, **CO**nfidentialité différentielle

Suite au développement d'internet et des outils de stockage d'information, il est aujourd'hui possible d'accumuler de gigantesques masses de données personnelles. Cette accumulation de données est à la fois une chance et un danger. D'une part, la collecte de données permet d'acquérir des informations sur la population et de mener des enquêtes. On peut ainsi traquer l'évolution d'une maladie, mesurer le niveau de pauvreté, optimiser les moyens de transports, etc. D'un autre coté, la collecte et le stockage de ces informations attentent à la vie privée des personnes concernées qui doivent révéler des informations personnelles et parfois sensibles, comme par exemple leur maladie ou leur niveau de richesse.

La confidentialité différentielle permet de rendre possible des enquêtes globales, tout en limitant les informations que l'on peut obtenir d'une personne particulière.

Le but de cette activité, basée sur le jeu *Qui est-ce ?*, est d'illustrer une technique de confidentialité différentielle : la réponse randomisée. Le principe est simple : lorsqu'une question est posée à une personne, cette dernière choisit aléatoirement de mentir ou de dire la vérité. Selon la probabilité de mentir, les résultats globaux de l'enquête seront plus ou moins perturbés mais il sera difficile de connaître les réponses d'un individu particulier.

Matériel

- 24 cartes personnages, avec 3 caractéristiques variables
- 1 dé, ou autre source d'aléas
- 4 cartes objectifs globaux
- 4 cartes objectifs personnels

Déroulement de l'activité

Tout d'abord, une personne est désignée dans le groupe et sera « l'enquêteur ». C'est elle qui posera les questions aux autres joueurs, qui seront les « répondants ».

L'enquêteur tire une carte « objectif global » et une carte « objectif personnel ». L'objectif global est correspond au résultat d'une enquête sur l'ensemble des répondants, l'objectif personnel est une information privée que l'enquêteur doit trouver sur un répondant en particulier.

Chaque répondant va ensuite choisir secrètement un nombre dans sa tête et le marquer sur un bout de papier. Il ne dit ni ne montre ce nombre à personne.

L'activité se déroule en deux phases durant lesquelles l'enquêteur a droit à un nombre limité de questions (en fonction du temps disponible) pour réaliser ses deux objectifs.

Lors de la première phase, un dé est lancé à chaque question. Si le dé correspond au nombre qu'il a choisi, le répondant ment. Sinon il dit la vérité.

Lors de la deuxième phase, tout le monde dit la vérité.

Objectif pédagogique et complément scientifique

Le but de cette activité est d'illustrer une technique de protection de la vie privée. Normalement l'objectif global devrait être rempli tandis que l'objectif personnel devrait échouer. Il s'agit de sensibiliser au fait qu'il est possible de collecter des données utiles tout en limitant l'impact sur la vie privée.

Réponse randomisée

Le mécanisme de réponse randomisée a été introduit par le sociologue Stanley Warner en 1965 [5] afin de mener des enquêtes sur des sujets sensibles, par exemple estimer la proportion de tricheurs parmi les étudiants [4], ou de recours à l'avortement [1].

Une description de ce mécanisme se résume par la suite d'étapes ci-dessous :

- L'enquêteur pose une question fermée (dont la réponse est oui ou non) au participant
- Le ou la participant.e répond la vérité avec une certaine probabilité p connue de l'enquêteur, et ment sinon (par exemple, il peut jeter un dé, et mentir si le résultat est 1 et dire la vérité sinon ; on aura alors $p = \frac{5}{6}$)
- L'enquêteur récupère la réponse sans savoir si elle est vraie (le ou la participant.e peut toujours nier sa réponse du fait du protocole)
- Avec un nombre suffisant de réponses, des statistiques fiables peuvent quand même être établies

Confidentialité différentielle

La confidentialité différentielle est un formalisme récent de la notion de vie privée, proposé puis complété par Cynthia Dwork depuis une dizaine d'années [3, 2]. Cette notion constitue la référence actuel en matière de confidentialité et de protection de la vie privée.

L'idée générale est que la réponse à une question sur un ensemble d'individus n'est que très faiblement liée à la présence ou à l'absence d'un individu spécifique. En particulier, la réponse randomisée répond à cette définition.

Description du prototype fourni

Cartes personnages

Un jeu de 24 cartes avec 3 caractéristiques

- La couleur du chat (marron, roux, blanche)
- La forme des yeux (petit, allongé, grand)
- Un accessoire (le chapeau, walkman, bandana, noeud papillon)

Objectifs globaux

Quelques objectifs statistiques pouvant être réalisés avec ces personnages

- Quel est le nombre de chats roux écoutant de la musique ?
- Est-ce que les chats blancs ont plus de chance d'avoir des petits yeux que les autres ?
- Y a-t-il plus de chats marrons portant un nœud papillon ou de chats roux avec les yeux allongés ?
- Quelle est la couleur la plus représentée parmi les chats qui portent un bandana ?

Objectifs personnels

Quelques objectifs personnels pouvant être réalisés avec ces personnages

- Quel joueur a le chat roux aux grand yeux et écoutant de la musique ?
- Quel joueur a le chat blanc aux petits yeux et portant un bandana ?
- Choisissez un joueur. Quelles sont les caractéristiques du chat de ce joueur ?
- Identifiez les chats marrons à gros yeux et n'ayant pas de chapeau.

Références

[1] James R Abernathy, Bernard G Greenberg, and Daniel G Horvitz. Estimates of induced abortion in urban north carolina. *Demography*, 7(1) :19–29, 1970.

[2] Cynthia Dwork. Differential privacy. In *33rd International Colloquium on Automata, Languages and Programming, part II (ICALP 2006)*, volume 4052, pages 1–12, Venice, Italy, July 2006. Springer Verlag.

[3] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9 :211–407, 2014.

[4] NJ Scheers and C Mitchell Dayton. Improved estimation of academic cheating behavior using the randomized response technique. *Research in Higher Education*, 26(1) :61–69, 1987.

[5] Stanley L. Warner. Randomized response : A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309) :63–69, 1965.

Cartes personnages à découper

Images : `https://github.com/NoahDragon/cat-generator-avatars`

Cartes objectifs globaux à découper

Quel est le nombre de chats roux écoutant de la musique ?	Est-ce que les chats blancs ont plus de chance d'avoir des petits yeux que les autres ?
Y a-t-il plus de chats marrons portant un nœud papillon ou de chats roux avec les yeux allongés ?	Quelle est la couleur la plus présente parmi les chats qui portent un bandana ?

Cartes objectifs personnels à découper

Quel joueur a le chat roux aux grand yeux et écoutant de la musique ?	Quel joueur a le chat blanc aux petits yeux et portant un bandana ?
Choisissez un joueur. Quelles sont les caractéristiques du chat de ce joueur ?	Identifiez les chats marrons à gros yeux et n'ayant pas de chapeau.