

## À qui profite le crime ?

Bilel BENBOUZID

**Un algorithme peut-il prédire les crimes ? Les forces de police américaines s'équipent depuis plusieurs années de logiciels censés détecter les lieux des futurs crimes et délits. Leur succès tient pourtant davantage du marketing que de leur efficacité prédictive.**

Les algorithmes d'apprentissage automatique des « *big data* » trouvent des applications dans toutes les sphères de la société. Dans cette supposée « révolution des données », le secteur de la sécurité occupe une place importante. La police pourrait désormais anticiper les crimes grâce aux méthodes du *machine learning*. Parmi les nombreux acteurs qui œuvrent dans ce domaine, [Predpol](#) est l'entreprise la plus souvent mentionnée. Créée en 2012, la start-up californienne a élaboré un logiciel qui indique aux patrouilles de police la localisation des crimes à venir avec une précision époustouflante. Aux États-Unis, les forces de police sont nombreuses à s'être laissées tenter par l'installation de ce tableau de bord analytique fournissant une information prédictive quotidienne grâce à un algorithme s'inspirant des méthodes de prédiction des tremblements de terre.

À y regarder de plus près toutefois, cette solution est loin d'avoir le pouvoir prédictif que ses promoteurs lui prêtent. Nous nous sommes adressés au sismologue français David Marsan, à l'origine de l'algorithme qui a influencé le travail des développeurs de la start-up. Le professeur de l'Université de Chambéry a accepté de tester son algorithme sur les données de Chicago en accès libre. Les résultats de son travail donnent de bonnes raisons de douter de la pertinence de l'algorithme.

Predpol est avant tout le produit d'une puissante campagne de communication visant à parer une technique gestionnaire des habits neufs du *big data*. Le véritable défi que Predpol a dû affronter se situe moins sur le plan de la prédiction que celui du marketing. Autrement dit, dans le cadre du développement commercial de la start-up, la vocation est moins de se focaliser sur les savoirs et la technologie que sur ce qui fait que la police va acheter le produit. Dans ce contexte, il devient urgent d'engager une évaluation critique de cette technologie, de son efficacité comme de ses implications pour les politiques de sécurité<sup>1</sup>.

### Un plan marketing bien orchestré

La petite start-up de Santa Cruz s'est développée suivant le schéma classique de nombreuses entreprises en Californie. En 2010, deux [entrepreneurs](#), Caleb Baskin et [Ryan Coonerty](#) (par ailleurs élu démocrate du troisième arrondissement du comté de Santa Cruz),

---

<sup>1</sup> Cet article s'inscrit dans le cadre du projet de recherche INNOX (L'innovation dans l'expertise. Simulation et modélisation comme mode d'action publique) financé par le programme ANR Société Innovante (INOV2013). Sur l'enjeu de l'évaluation des algorithmes, voir D. Cardon, 2015, *À quoi rêvent les algorithmes*, Paris, Seuil (La République des idées), et ma contribution à la 17<sup>e</sup> Matinale de l'IFRIS, « [Peut-on auditer les algorithmes ?](#) ».

approchent deux chercheurs, George Mohler (maître de conférences en mathématiques appliquées) et Jeffrey Brantingham (archéologue spécialiste du paléolithique supérieur du Nord de la Chine et fils d'un célèbre couple de criminologues pionniers de la géographie du crime), en vue de convertir les fruits de leurs recherches en une entreprise rentable à fort potentiel de croissance.

Si les recherches qui ont sous-tendu Predpol ont été financées par l'acteur public, la création de la start-up a été rendue possible par les 1,3 millions de dollars investis en 2012 par une poignée de *business angels*. Suite à un processus commercial éprouvé en deux ans de manière spectaculaire, notamment grâce au travail de ses lobbyistes évoluant dans les réseaux démocrates californiens, l'entreprise se lance dans [une deuxième vague de levées](#) de fonds en capital risque (2,4 million de dollars obtenus en 2014) pour déployer davantage son activité commerciale. Ce type de développement en capitaux propres (les ressources de la société appartiennent à des actionnaires) contraignent les scientifiques de la start-up (Mohler et Brantingham, qui en sont actionnaires) à s'accorder sur les décisions stratégiques avec des investisseurs dont les objectifs sont la création maximum de valeur. L'exigence de rentabilité à court terme implique de traduire les résultats de la recherche en une campagne de communication marketing stratégique, qui soit capable de persuader des milliers d'élus locaux et de chefs de la police de l'intérêt qu'ils pourraient trouver à acheter le logiciel. Loin de résulter d'un débat politique de fond sur la réforme de l'institution, la dynamique de l'innovation policière obéit à des principes comparables à ceux qui régissent le lancement de n'importe quel produit à caractère commercial.

La stratégie de lancement de Predpol repose sur trois piliers : une nouvelle forme, un slogan accrocheur et un mythe fondateur. Le premier pilier est ce qui va permettre à Predpol de s'imposer rapidement sur le marché. Les responsables du marketing de Predpol ont mis en place une stratégie assez classique dans le domaine de la mise sur le marché des technologies digitales : dans sa forme, Predpol prétend faire la différence en rendant la vie de la police plus simple qu'elle ne l'est avec les produits déjà disponibles sur le marché, tout en laissant à la police la liberté d'en définir les usages tactiques pour les patrouilles. En proposant la forme « plateforme », Predpol se distingue des logiciels existants de cartographie du crime installés sur les ordinateurs de bureau (desktop) ou sur le réseau interne de la police (l'intranet). Predpol est une offre d'analyse du crime en temps réel qui prend la forme d'un tableau de bord, téléchargeable sur une simple application. Les données sont stockées sur le *cloud*. Les forces de police peuvent ainsi externaliser le coût de l'installation et de gestion de serveurs car la location du temps de calcul est intégrée dans l'offre. Elles n'ont plus à se préoccuper des problèmes coûteux de gestion quotidienne des débogages. Elles n'ont plus à se préoccuper non plus des problèmes d'« analyse ». L'intelligence marketing de Predpol est d'avoir su cibler l'administration policière (les chefs de la police locale directement), et non pas les services spécialisés d'analyse criminelle.

Les responsables marketing de Predpol tentent de séduire les dirigeants de la police au moyen du slogan publicitaire *More Than Hotspot Tool*. Depuis le début des années 1990, le « *hotspots policing* » incarne le modèle principal d'une police proactive intervenant de manière stratégique sur les secteurs où se concentrent les crimes. Predpol dit s'inscrire dans l'ère du « *predictive policing* » car le logiciel ferait mieux que les cartes du crime classiques représentant sous la forme de *heat map* la concentration du crime sur le territoire (les fameux *hotspots*). Dans des opérations de communication de presse qui ont été très finement orchestrées, Predpol prétend se distinguer des approches traditionnelles en mentionnant les origines sismologiques de l'algorithme utilisé. En matière de crime, il en irait comme pour les

séismes : s'il est difficile de prédire avec précision l'occurrence d'un premier évènement, il est possible d'en prédire les répétitions. Predpol intégrerait ainsi dans son algorithme la dimension contagieuse de la diffusion du crime dans l'espace et le temps, d'où le slogan *More Than Hotspot*.

L'idée de la contagion du crime n'est pas nouvelle et ne vient pas de la sismologie. Elle date des années 1980 avec les premières recherches en criminologie autour de la notion de « [repeat victimization](#) ». Mais sur le plan marketing, la métaphore sismologique a un avantage sur l'explication criminologique : elle fait référence au couplage d'une science dure avec les techniques prédictives des *big data*. Ce marketing scientifique va envoyer un message simple, mais très efficace : « nous avons fait une découverte, le crime serait fongible dans les mathématiques ; nous avons enfin trouvé la solution au problème sur lequel butent l'analyse criminelle<sup>2</sup> depuis des années ». Le succès de Predpol tient en grande partie à ce mythe fondateur, coproduit par la presse et les responsables du marketing, qui présente la start-up comme la contribution de la « vraie science » à la lutte contre le crime.

### **L'émergence d'une critique sociale**

Le journaliste Darwin Bond Graham a mis en avant le problème de ce type de plan [marketing](#). Sur le site *publicintelligence.net*, il a révélé un document confidentiel qui montre les [éléments du contrat](#) qui lie Predpol à la police de Modesto. Pour pouvoir s'imposer rapidement sur le marché, Predpol propose dans sa phase de lancement sa solution à un prix soldé (un abaissement de 60%), en échange de quoi les agences de police s'engagent à communiquer de manière positive sur la marque. Les responsables du marketing de la start-up cherchent ainsi à faire exister une « communauté de marque » chargée de la promotion « *earned media* » (l'exposition publicitaire dont bénéficie gratuitement une marque sur des supports médiatiques comme les journaux où les réseaux sociaux) de Predpol. On comprend pourquoi, dans la presse, les dirigeants de la police apparaissent systématiquement comme des clients parfaitement satisfaits.

Des activistes se sont mobilisés pour dénoncer la start-up et ses méthodes communication très agressives. Sur le site californien *Indibay.org*, un groupe pour la protection des droits civiques de la ville d'Oakland s'adresse au maire de la ville (alors sur le point de souscrire à l'offre de Predpol) dans une [lettre ouverte](#) réclamant un véritable débat public et une recherche indépendante sur l'efficacité de cette technologie. Le groupe rappelle les nombreux doutes sur l'efficacité de Predpol déjà émis dans la [presse](#) par le journaliste Darwin Bond Graham. Il soupçonne notamment les évaluations positives de Predpol de manquer d'objectivité, celles-ci se fondant uniquement sur des mesures d'efficacité réalisées par des chercheurs actionnaires de Predpol. Du reste, selon les activistes, pour d'autres produits similaires, lorsque des essais contrôlés ont été réalisés par des organismes externes, les résultats sont systématiquement moins spectaculaires.

En France, le Collectif de Recherche Transdisciplinaire Esprit Critique et Science (CORTECS) a publié une [note critique](#) qui compare les performances de l'algorithme de Predpol avec celles d'autres algorithmes, à partir des données ouvertes de la ville de Chicago. La note critique indique que des algorithmes standard ont réalisé des scores de prédiction assez proches de ceux de Predpol, ce qui remet largement en question le caractère innovant de

---

<sup>2</sup> L'analyse criminelle aux États-Unis est une pratique professionnelle exercée dans les services de police locaux. Elle repose sur l'usage de l'analyse de données et la mobilisation de concepts criminologiques pour aider les responsables de la police à mettre en place des stratégies. Il ne faut pas confondre cette activité avec celle du profilage criminel.

l'offre de la start-up. Predpol ne souhaitant pas rendre public son algorithme, CORTECS n'a pas pu le tester directement. La critique tient à la seule comparaison des « scores » d'efficacité (seuls les scores entre différents algorithmes sont comparés), ce qui limite la portée de la critique du collectif.

Predpol n'est pas la seule entreprise à éviter la discussion publique sur la qualité et les impacts sociaux des technologies qu'elles diffusent auprès de la police. L'association de tutelle *Leadership Conference on Civil Right* a ainsi publié le 31 août 2016 [une pétition](#), signée par 17 organisations non gouvernementales, dénonçant « les défauts systémiques, la partialité inhérente, et le manque de transparence endémique de la police prédictive et des fournisseurs de produits qui lui sont associés ».

### Évaluer l'algorithme de Predpol

Comment dès lors évaluer le logiciel, puisque Predpol empêche l'accès au code source ? Si les chercheurs de la start-up revendiquent de s'inspirer d'un algorithme utilisé en sismologie, il peut être intéressant de consulter directement les travaux des sismologues qui ont développé et utilisent cet algorithme. Les travaux qui ont influencé Predpol sont en particulier ceux de [David Marsan](#), professeur au laboratoire de science de la terre de l'Université de Savoie, à Chambéry, spécialisé dans les études sur les répliques des tremblements de terre. Il a publié en 2008 dans [la revue Science](#) un article dans lequel il traite du fait que les tremblements de terre, quelle que soit leur taille, peuvent en déclencher d'autres. Les secousses principales d'un séisme déclenchent des répliques qui en retour déclenchent leur propre séquence de tremblements de terre. Ceci produit une cascade de déclenchements qui étend la portée du choc initial. C'est un problème ancien en sismologie que de déterminer si les séismes sont liés les uns aux autres, soit directement, soit indirectement. Marsan et son collaborateur cherchent à montrer dans leur papier que cette structure en cascade peut-être modélisée sous forme de probabilités sans faire d'hypothèse sur les mécanismes (sans intégrer les contraintes physiques propres à la réalisation d'un séisme), et sans qu'il soit besoin de tester au préalable les paramètres propres aux modèles (on parle d'une statistique sans paramètre *a priori*).

C'est une approche qui correspond en apparence à la «*data science*» dont se revendique Predpol car la manière de modéliser proposée par Marsan n'implique ni modèle *a priori*, ni paramètres. En revanche la méthode utilisée, soit le processus ponctuel auto-excitatif, qui correspond à l'hypothèse selon laquelle le crime déclencherait le crime<sup>3</sup>, impose une contrainte importante pour l'étude du crime. Elle implique de considérer ce phénomène social comme n'importe quel phénomène physique, dont la modélisation dépend de la structure spatiale inhérente aux entités qui constituent le phénomène étudié. Par exemple, une file d'attente est modélisable avec ce type de méthode, car son processus d'évolution auto-excitatif dépend de la structure spatiale spécifique de la «*forme file*». Que la propagation des séismes dépende de leur structure spatiale propre est facilement envisageable. En revanche, il est difficile d'imaginer une telle approche structurelle de l'espace pour un phénomène aussi contingent que le crime. Le déterminisme de l'algorithme de Predpol est la négation (assumée) de tout ce qui ne peut être décrit physiquement.

---

<sup>3</sup> Les statistiques de processus ponctuels sont des méthodes classiques, très souvent utilisées en statistique spatiale pour modéliser la distribution d'un ensemble de points sur une surface finie, autrement dit pour décrire la répartition de points sur un plan. Dans notre cas, chaque point correspond à un crime et le processus ponctuel concerne les instants et les lieux où se réalisent les crimes.

Pour nous donner à comprendre ce tour de force intellectuel réalisé par Predpol, nous avons contacté directement David Marsan qui a accepté de tester son algorithme à partir des données de la ville de Chicago, en particulier pour le cas des cambriolages. Que pense le sismologue du comportement de son algorithme sur les données du crime ? David Marsan nous a communiqué le carnet de notes qu'il a tenu pendant la semaine où il a testé son propre algorithme. La note mérite d'être rendue publique. Dans la mesure où le développement de Predpol dépend d'un plan marketing dans la presse grand public, et non pas dans l'espace scientifique, nous avons choisi de communiquer [ces notes](#) à un journal de grande audience comme *La Vie des Idées* en espérant susciter une réponse de la part des chercheurs actionnaires de la start-up. Nous avons aussi choisi de rendre publique cette note en raison de la réticence de l'entreprise Predpol à mettre en débat l'efficacité et les impacts sociaux du logiciel.

Le formalisme mathématique de ces notes en limite la compréhension totale aux seuls experts, cependant la conclusion de David Marsan est accessible à tous. Il est intéressant de constater que le sismologue émet de nombreux doutes sur la capacité de son algorithme à faire mieux que les cartes classiques de *hotspots* :

Ces résultats donnent de bonnes raisons de douter de la capacité des modèles proposés à faire mieux que les simples cartes de *hotspots*. La contribution de la contagion (*the triggering contribution*) pour expliquer la survenance d'événements futurs est faible (elle ne représente que 1,7% pour le meilleur modèle). La part de la « mémoire » dans le processus ne peut donc fournir qu'une très modeste contribution à l'efficacité du système de prédiction. Plus important encore, il est supposé que la dynamique du processus reste le même au fil du temps. La possible non-stationnarité du processus est clairement un problème car elle limite l'utilisation des informations passées pour prédire l'avenir. En 2015, les cambriolages ne se distribuent pas (dans le temps et dans l'espace) comme en 2014. Cette non-stationnarité est probablement due à des évolutions incontrôlées de la façon dont les actes criminels sont accomplis. Elle peut être aussi due à la mise en place de nouveaux algorithmes de prédiction qui, une fois exploités par les patrouilles de police, provoqueraient des réactions chez les cambrioleurs. Contrairement à des processus naturels tels que les tremblements de terre, des analyses comme celles présentées ici pourraient donc modifier le processus observé, ce qui rend plus difficile de prévoir correctement les événements futurs.<sup>4</sup>

David Marsan montre tout d'abord que l'algorithme ne fait guère plus que du *hotspot mapping*. Pour comprendre cette remarque, il faut expliquer que l'algorithme de Predpol calcule l'intensité du risque en fonction de l'espace et du temps en additionnant deux éléments : la part de la concentration et celle de la contagion. La note de David Marsan

---

<sup>4</sup> David Marsan, Notes de travail, avril 2015 : "These results cast strong doubts on the capacity of the models proposed here to outperform simple hotspot maps obtained by smoothing, for the dataset analyzed. The triggering contribution [la contagion] to the occurrence of future events is small (it accounts only for 1.7 % for the best model). Accounting for memory in the system therefore can only provide a very modest contribution to the effectiveness of the prediction scheme. More importantly, it is assumed that the dynamics of the process stays the same over time. Possible non-stationarity of the process is thus clearly an issue, as it will prevent the use of past information to predict the future. This is for example experienced in this analysis, as 2015 burglary events are clearly not distributed (in time and in space) as they were in 2014. This non stationarity is likely due to uncontrolled evolutions in the way these acts are performed, but, in situations where new prediction algorithms are set up and exploited by police patrols, could also be a response by burglars to such a change. Unlike natural processes like earthquakes, analyses like the one presented here could therefore have the ability to modify the observed process, making it more difficult to, correctly predict future events."

indique que la contribution de la contagion dans la réalisation du processus existe, mais qu'elle est extrêmement faible. Elle est même négligeable. Ensuite, David Marsan pose le problème de l'absence de stationnarité qui signifie que la structure du processus sous-jacent évolue avec le temps (le crime peut évoluer suivant un processus auto-excitatif en 2013 et d'une autre manière en 2014). Pour le dire autrement, le crime n'a pas la même structure sous-jacente d'une année sur l'autre. Cette absence de stationnarité tient à l'interaction complexe entre le phénomène lui-même (le cambriolage) et des forces externes (le travail de la police notamment). Ceci fait une grande différence avec l'activité sismologique, dont la structure sous-jacente au XXe siècle est équivalente à celle du XXIe siècle. En toute rigueur, cette absence de stationnarité empêche d'appréhender le phénomène de manière standard, à partir d'un « processus ponctuel auto-excitatif », commercialisable à l'infini. Les phénomènes non stationnaires impliquent de mobiliser d'autres méthodes et d'intégrer des variables externes dans le processus d'apprentissage statistique. Les scientifiques qui ont œuvré au développement de Predpol connaissent évidemment ces limites, qu'ils considèrent comme des questions encore ouvertes (cf. [l'article](#) que les chercheurs de la start-up ont spécialement consacré à ce problème, sans qu'ils aient pour autant pu le résoudre entièrement). Les contraintes marketing dans lesquels les chercheurs actionnaires sont pris les empêchent de mettre en avant ces questions, qui sont pourtant cruciales pour que puisse se tenir un débat public sur la commercialisation d'un produit destiné à être utilisé par un service public. Si l'intérêt de Predpol devait être discuté seulement à l'aune son algorithme, la start-up n'aurait pas lieu d'être.

Quand bien même les forces de police comprendraient les limites de l'algorithme, elles ne renonceraient pas nécessairement à l'offre de Predpol, car l'efficacité prédictive de l'algorithme n'est pas la question cruciale aux yeux de la police. Pour celle-ci, l'enjeu principal est moins la prédiction des crimes qu'un management simplifié de l'action policière. Les études de Predpol ont montré qu'en passant seulement 5% de leur temps disponible dans les zones identifiées par Predpol, les patrouilles de police sont deux fois plus efficaces que lorsqu'elles patrouillent sur les points chauds classiquement identifiés par les analystes. La réalité statistique de ce que prétend Predpol n'a pas beaucoup d'importance – c'est pourquoi, soit dit en passant, l'efficacité de l'algorithme de Predpol n'a pas été contrôlée par des organisations indépendantes. Ce qui compte, c'est de pouvoir optimiser et surtout contrôler ce temps de présence tactique dans l'espace à risques. Pour ce faire, Predpol intègre les données des systèmes de suivi GPS placés dans les voitures de police, ce qui permet ainsi d'optimiser le dosage de la présence des patrouilles de police selon les secteurs de la ville : le carré prédictif reste rouge sur la carte tant que la police n'y a pas patrouillé, il tourne ensuite au bleu lors des premiers passages, puis il apparaît en vert lorsque le policier a passé le temps suffisant et optimal calculé selon les ressources disponibles (par exemple 5% du temps de la journée de travail d'un agent de police). Pour un responsable de secteur, Predpol apparaît comme un bon outil pour s'assurer que les policiers font bien leur travail préventif, souvent par la simple présence dissuasive, de manière aléatoire, mais sur une durée optimisée, dans les zones où le risque est estimé le plus haut. L'enjeu du *predictive policing* est de gérer, selon des critères gestionnaires, l'offre publique de vigilance quotidienne.

Reste que les critères de distribution de cette offre de sécurité dans la population ne sont jamais évoqués dans le plan marketing de Predpol. La question est pourtant cruciale : à quelle politique de justice sociale d'accès à la sécurité publique l'algorithme de Predpol soumet-il la population concernée ? L'algorithme étant façonné uniquement par les données issues des plaintes des victimes (et non pas par celles des arrestations, pour éviter que l'algorithme ne reflète l'activité discriminatoire des policiers), il oriente l'offre de sécurité

uniquement vers les publics qui portent plainte auprès des autorités (hormis pour les homicides). Or, les enquêtes de victimation (des collectes de données auprès de la population générale, qui interrogent les personnes sur les crimes qu'elles auraient subis) montrent que la distribution des plaintes n'est pas homogène dans la population car certaines victimes pensent que la police ne peut rien faire pour régler leurs problèmes ou qu'il ne vaut pas la peine de déposer plainte. Le fait que les victimes ne recourent pas à la police s'explique par leur position sociale, leurs expériences passées avec la police, leur lieu de résidence et leur propension à agir dans l'intérêt de la vie du quartier. Les non-signalements sont des phénomènes sociaux en tant que tels, qui échappent complètement à l'apprentissage statistiques par les données enregistrées par la police. En n'ajustant pas le calcul de l'intensité du risque en fonction de ce taux de non-signalements<sup>5</sup>, l'algorithme de Predpol produit un biais qui peut avoir des impacts sociaux graves : il peut recommander de concentrer l'offre de sécurité sur une partie de la population, au détriment des personnes dont la participation active à la préservation de la qualité de vie de leur quartier est la plus faible. Le problème de l'algorithme de Predpol n'est pas la discrimination policière, comme beaucoup ont pu le craindre, mais celui de l'exclusion d'une partie de la population à l'offre de sécurité publique. Autrement dit, sur le temps long, le strict suivi des recommandations de l'algorithme de Predpol pourrait creuser les inégalités d'accès à la sécurité. C'est à ce prisme qu'il faudrait évaluer la fiabilité des modèles de Predpol.

Predpol est un cas parmi d'autres d'un mouvement général de commercialisation de savoirs privatisés à destination de l'action publique. Le problème n'est pas tant que des entreprises privées à but lucratif diffusent dans le secteur des technologies analytiques. Que des start-up dynamiques s'appuient sur des savoirs en sciences sociales et se développent sur un plan commercial est plutôt salubre. Le danger se trouve plutôt dans l'absence totale de contrôle de la mise sur le marché de ce type de machines prédictives. Les techniques de marketing utilisées réduisent les discussions à de la persuasion. Pour faire de l'acteur public un consommateur éclairé, il faut imaginer des méthodes de rétroingénierie sociale, comme celles que propose la Fondation Internet Nouvelle Génération (FING) avec le projet [Nos Systèmes](#). Après la gouvernance de la statistique publique, puis celle des données publiques, il faut désormais imaginer les acteurs institutionnels qui pourront composer une véritable gouvernance de l'algorithmique des données publiques.

---

<sup>5</sup> Pour une critique d'un usage aveugle du *machine learning* dans l'étude du crime, cf. l'article de Robert Sampson et de ses collaborateurs : O'Brien, Daniel, Robert J. Sampson, and Christopher Winship. 2015. "Ecometrics in the Age of Big Data: Measuring and Assessing 'Broken Windows' Using Large-scale Administrative Records." *Sociological Methodology* 45: 101-147. En mobilisant les traces laissées en ligne par les habitants de la ville de Boston, les chercheurs poursuivent le travail de modélisation de « l'efficacité collective » – une théorie classique en sociologie urbaine selon laquelle le manque de cohésion entre voisins et de participation active à la préservation de la qualité de vie (le manque d'efficacité collective) serait corrélé à l'insécurité dans certains secteurs. De nombreuses villes américaines permettent aux citoyens de signaler, à partir d'applications sur leurs terminaux mobiles, une grande variété de dysfonctionnements urbains : graffitis, objets encombrants abandonnés, éclairage public défectueux, carrefours dangereux, nids-de-poule. Mais tous les habitants n'utilisent pas ces services de manière égale. L'équipe de R. Sampson montre comment tirer profit de ces données massives pour l'étude des désordres urbains, tout en ajustant son modèle par rapport au taux de réponses civiques – une maîtrise des biais dans les données qui s'avère essentielle lorsqu'on s'intéresse aux inégalités dans la lutte contre le crime.

Publié dans [laviedesidees.fr](http://laviedesidees.fr), le 13 septembre 2016

© **laviedesidees.fr**