
Le vandalisme dans l'information géographique volontaire

Du concept à la détection non supervisée d'anomalie

Quy Thy Truong¹, Guillaume Touya², Cyril de Runz³

1. University Paris-Est, LASTIG MEIG, IGN, ENSG, F-94160 Saint-Mande, France
quy-thy.truong@ign.fr

2. University Paris-Est, LASTIG GEOVIS, IGN, ENSG, F-94160 Saint-Mande, France
guillaume.touya@ign.fr

3. LIFAT-Équipe BDTLN, Université de Tours, CS 32903, Blois Cedex, France
cyril.derunz@univ_tours.fr

RÉSUMÉ. Dans un contexte où le vandalisme de l'information géographique volontaire constitue une réelle menace pour la qualité des données, cet article propose une technique permettant de le détecter. Tout d'abord, nous examinons les différentes définitions du vandalisme, mettant en avant la complexité de cette notion. Une étude des cas déjà avérés de vandalisme dans OpenStreetMap (OSM) est présentée. Puis, nous expérimentons la détection du vandalisme des données OSM en utilisant une méthode de détection d'anomalie par classification non supervisée. L'analyse de nos résultats initiaux conduit alors à une discussion sur la pertinence de construire un corpus de vandalisme des données OSM utilisable dans un contexte de classification supervisée.

ABSTRACT. Since vandalism is a serious matter for the quality of Volunteered Geographic Information, this paper aims at exploring machine learning techniques that enable its detection. First, a focus on the various definitions of vandalism highlights the complexity of this concept. This focus comprises a case study on proven vandals in OpenStreetMap (OSM). Second, we present an experimental vandalism detection on OSM data using a clustering-based outlier detection. The analysis of initial results leads to a discussion about the construction of an OSM vandalism corpus that would be useful in a supervised learning context.

MOTS-CLÉS : information géographique volontaire, vandalisme, classification.

KEYWORDS: volunteered geographic information, vandalism, learning.

DOI: [10.3166/riq.2019.00073](https://doi.org/10.3166/riq.2019.00073) © 2019 Lavoisier

1. Introduction

Le manque de qualification de l'information géographique volontaire constitue aujourd'hui encore un frein pour les utilisateurs de données géographiques. En effet, dans un contexte où la science citoyenne permet à des non-spécialistes de collecter des données scientifiques, la saisie collaborative d'information géographique par des volontaires questionne actuellement les producteurs de données institutionnelles quant au mode d'acquisition de l'information spatiale et la mise à jour de leurs bases de données (Ivanovic, 2018 ; Noucher, 2014). Cependant, alors que les bases de données géographiques participatives peuvent contenir des éléments de qualité variable, le fait de permettre à n'importe quel volontaire de contribuer donne la possibilité au plus mal-intentionné de réduire la qualité des données par des actes de vandalisme cartographique (Ballatore, 2014). Des outils de détection du vandalisme de données dans Wikimédia (Potthast, 2010 ; Heindorf *et al.*, 2015) et OpenStreetMap (Neis *et al.*, 2012) ont été développés pour prévenir ce genre de dérives, or, ceux-ci n'ont pas été conçus selon la même conception du vandalisme. En effet, la frontière entre le vandalisme et la simple erreur de saisie étant très fine, certains scientifiques choisissent une définition inclusive du vandalisme, qui comprend alors les actes de dégradation involontaires. La détection du vandalisme dans l'information géographique volontaire est intéressante car elle peut constituer une étape préalable à toute utilisation de ce type de données. En effet, le filtrage de données vandalisées permet de garantir une qualité minimale à un jeu de données collaboratives. Par ailleurs, les données vandalisées obtenues présentent aussi un intérêt pour l'approfondissement des recherches sur le carto-vandalisme.

Deux questions découlent de ce constat : comment définir le vandalisme dans l'information géographique volontaire et comment le détecter pour le corriger ? Dans cet article, nous dressons un état de l'art sur le vandalisme afin de mettre en évidence les définitions formulées dans la littérature, et nous explorons les différentes approches de détection du vandalisme qui ont été expérimentées jusqu'à présent. Nous cherchons alors à détecter le vandalisme des données OpenStreetMap (OSM) en expérimentant une méthode de détection d'anomalie. Le but de notre expérience est de déterminer si un algorithme de détection d'anomalie par technique de regroupement des données (*clustering*) est capable de retrouver des cas très simples de vandalisme artificiels, présentant différents types d'anomalies. L'interprétation de nos résultats expérimentaux ouvrira la discussion sur les pistes à envisager pour améliorer la détection.

2. Définitions théoriques et pratiques du vandalisme

2.1. Contexte historique

Historiquement, le vandalisme désignait les actes de dégradation perpétrés par la tribu des Vandales, un peuple germanique réputé pour avoir pillé et saccagé les œuvres d'art et les monuments lors de l'invasion de l'Occident. À la révolution, l'abbé Grégoire reprend le terme de vandalisme pour désigner les dommages causés sur les monuments

et les ouvrages d'art du patrimoine français¹. Aujourd'hui, le terme de vandalisme renvoie plus généralement à une dégradation matérielle causée par des êtres humains. La justice française définit le vandalisme de la manière suivante² :

« L'acte de vandalisme consiste à détruire, dégrader, ou détériorer volontairement le bien d'autrui. (...) L'acte de vandalisme doit être commis sans motif légitime. Il est par exemple permis de briser une vitre pour sauver une personne en danger. »

D'après cette définition, trois conditions doivent être remplies pour pouvoir qualifier un acte de vandalisme : qu'il y ait observation d'un bien dégradé ; que la dégradation ait été voulue par son auteur ; qu'il n'y ait aucun élément de contexte justifiant la légitimité de cet acte. En effet, selon le contexte, un même acte ne sera pas forcément labélisé de la même manière : par exemple, faire dérailler un train en déposant des obstacles sur les rails sera considéré comme une farce si cela a été fait par des enfants (en supposant qu'ils n'ont pas encore conscience de l'ampleur de leurs actes), mais comme du vandalisme si cela a été commis par des adultes (Zimbardo, 1971). Dans le cas d'OSM, prenons l'exemple d'une route correctement cartographiée, qui se retrouve coupée ou supprimée par un contributeur agissant de manière totalement volontaire. Le contexte permet de déterminer si l'acte relève bien du vandalisme. S'il y a des travaux ou un sinistre sur cette route, qui justifierait qu'elle soit coupée au moment où le contributeur saisit l'information, alors cet objet n'est pas vandalisé : au contraire, il est conforme à la réalité terrain. En revanche, s'il n'y avait aucune raison de couper cette route, alors la contribution relève d'un acte de vandalisme.

Les travaux de recherche sur le vandalisme de l'information géographique volontaire ont proposé deux typologies de vandalisme cartographique. Neis *et al.* (2012) ont développé un outil de détection du vandalisme, où le vandalisme comprend tout dommage causé volontairement ou non. Cette typologie rompt avec la définition traditionnelle du vandalisme, puisqu'elle met de côté la composante intentionnalité. Ballatore (2014) propose une typologie de carto-vandalisme à partir des différentes motivations qui poussent à l'acte. Le vandalisme cartographique peut être : ludique, idéologique, fantaisiste, artistique, industriel ou spam cartographique. Bien que cette typologie soit empirique, il est difficile de l'utiliser en pratique car les différentes classes proposées ne sont pas indépendantes les unes des autres. Par exemple, le carto-vandalisme artistique peut être vu comme une sous-catégorie de carto-vandalisme fantaisiste puisque des objets cartographiés artistiquement sont également des objets fictifs.

On souligne ici que l'évaluation des trois composantes du vandalisme est ardue : elle fait appel à de véritables compétences juridiques, ce qui explique la difficulté de qualifier un acte de vandalisme. Dans le but d'appliquer cette définition du vandalisme au domaine de l'information géographique volontaire, nous présentons, dans la sous-

1. <http://www2.assemblee-nationale.fr/decouvrir-l-assemblee/histoire/grands-moments-d-eloquence/l-abbe-gregoire-31-aout-1794>

2. <https://www.service-public.fr/particuliers/vosdroits/F1514>



Figure 1. Exemple d'un cas de vandalisme : la zone commerciale est transformée en lac (cf. [tableau 1](#) pour l'historique de tags de l'objet)

section suivante, une exploration et une analyse de différents cas de contributeurs bannis sur OSM. À la lumière de la définition dégagée précédemment, nous cherchons à qualifier les cas de contributions qui relèvent du vandalisme.

2.2. Exploration des contributeurs bannis sur OpenStreetMap

Sur OpenStreetMap, il existe un groupe de contributeurs – le Data Working Group³ (DWG) – dont le rôle est de bannir temporairement les comptes OSM présentant des activités douteuses, notamment ceux qui vandalisent les données. Une liste des contributeurs bannis est consultable sur une page web dédiée⁴.

Le premier cas de vandalisme présenté dans la [figure 1](#) est un élément géographique correspondant initialement à une zone commerciale, qui a été transformée en lac par un second contributeur ([tableau 1](#)). Ici, la dégradation causée est assez évidente : l'objet géographique ne décrit plus du tout la même réalité, et la présence d'un lac sur une zone où se trouvent déjà un parking et des commerces paraît assez invraisemblable. La question de la vraisemblance des relations topologiques entre les objets géographiques a été soulevée pour identifier les problèmes de qualité dans OSM, qui peuvent relever de simples erreurs, d'incohérences de niveau de détail voire de vandalisme ([Goodchild et Li, 2012](#) ; [Touya et Brando, 2013](#)). De plus, le fait de nommer le lac par un *smiley* (tag "name =:)" révèle que le contributeur a certainement dégradé les données de manière intentionnelle. Enfin, en observant les autres contributions de ce compte utilisateur, ce dernier a ajouté, de la même manière, d'autres lacs en zones urbaines, ce qui laisse peu

3. https://wiki.openstreetmap.org/wiki/FR:Data_working_group

4. https://www.openstreetmap.org/user_blocks



Figure 2. Erreur de débutant : cartographie d'un hôpital dont la géométrie chevauche plusieurs îles au Nord du Canada

de doute quant à sa culpabilité. Par ailleurs, à notre connaissance, aucun élément de contexte n'a permis de justifier la saisie d'un tel élément. On peut donc conclure que cette contribution relève bien d'un acte de vandalisme.

Le cas de la [figure 2](#) correspond plutôt à une erreur de débutant qu'à un pur acte de vandalisme. En effet, bien que l'hôpital soit géométriquement mal cartographié, on note que l'objet géographique a été ajouté trois heures après la date de création du compte OSM de son auteur. De plus, le commentaire laissé par le contributeur sur cette session d'édition et ses contributions suivantes indiquent que celui-ci cherchait initialement à cartographier les éléments manquants aux alentours d'une université au Sri Lanka⁵. Ici, il s'agit probablement d'une erreur dans le système de coordonnées indiqué, et non d'un cas de vandalisme.

Tableau 1. Historique des tags de l'objet de la figure 1

Tag key	Version	
	1	2
landuse	retail	
name		☺
natural		water
water		lake

5. <https://www.openstreetmap.org/changeset/53283079>

Tableau 2. Extrait de l'historique des tags (versions n° 28-no 33) d'une ville du Kosovo dont la toponymie est sujette à controverse, l'albanais et le serbe étant les deux langues officielles. Les modifications de valeur de tag sont en gras

Tag key	Version					
	28	29	30	31	32	33
name	Malishevë-Mališevo	Malishevë	Malishevë	Malishevë/ Mališevo	Malishevë	Malishevë/ Mališevo
name:sq	Malishevë	Malishevë	Malishevë	Malishevë	Malishevë	Malishevë
name:sr-Latn		Mališevo	Mališevo	Mališevo	Mališevo	Mališevo

L'historique des modifications de tags de l'objet du [tableau 2](#) montre un désaccord entre les contributeurs sur la toponymie de la ville. Le contributeur de la version 32 de l'objet a été banni pour avoir retiré l'orthographe serbe du tag "name". Après avoir été banni, la version 32 a été annulée (*i.e.* la version 33 correspond à la version 31). Au vu du grand nombre de versions de l'objet – plus de 30 – celui-ci semble être un sujet controversé pour les contributeurs OSM. Or, précisons que les guerres d'éditions ne sont pas autorisées sur OSM⁶, celles-ci sont même considérées comme étant du vandalisme d'après le règlement du projet. Le contributeur de la version 32 a donc été banni pour non-respect des règles de bonne conduite du projet. Cependant, même si sur la version 32, l'orthographe serbe ne figure pas sur le tag "name", celui-ci reste indiqué sur le tag "name:sr-Latn". Par ailleurs, d'après l'OSCE⁷, la ville de Malishevë/Mališevo est habitée majoritairement par des albanais, ce qui constitue un élément de contexte pouvant justifier l'acte de ce contributeur. Par conséquent, on ne peut pas affirmer que sa contribution dégrade la qualité de l'objet.

Ce travail exploratoire sur la page des contributeurs bannis a été effectué manuellement, en parcourant les messages émis par les modérateurs. Ces messages peuvent contenir l'identifiant de la donnée ou du groupe de modifications (changeset) qui a mené au bannissement du contributeur, mais cette information n'est pas systématiquement renseignée. L'historique des données a été obtenu avec l'application OSM Deep History⁸. Notons que cette phase exploratoire aurait très bien pu se faire de manière automatique, en recourant à des techniques de traitement automatique de texte, pour analyser les messages entre les modérateurs et les contributeurs. Cela aurait permis de récupérer une liste plus exhaustive de cas de vandalisme. Néanmoins, notre exploration manuelle de la page a permis d'identifier rapidement quelques cas de vandalisme intéressants donnant un aperçu des raisons du bannissement de certains contributeurs. L'exploration des contributeurs bannis d'OSM montre que le bannissement ne concerne pas seulement les contributeurs ayant commis des actes de vandalisme sur les données

6. <https://wiki.openstreetmap.org/wiki/Disputes>

7. <https://www.osce.org/kosovo/13137?download=true>

8. <https://osmlab.github.io/osm-deep-history/>

OSM, mais également ceux qui ne respectent pas les règles de bonne conduite sur le projet : ne pas répondre aux commentaires adressés par d'autres contributeurs, importer massivement des données présentant une licence incompatible, ou participer à des guerres d'éditions. Finalement, le vandalisme sur OSM, selon sa définition traditionnelle, correspond à des cas très marginaux par rapport à l'ensemble des contributeurs bannis d'OSM. Les cas de vandalisme relevés à partir de la page des contributeurs bloqués correspondent aux cas les plus sérieux, néanmoins OSM encourage la communauté locale à résoudre les cas moins « sérieux » de vandalisme⁹.

3. État de l'art sur les approches de détection du vandalisme dans les bases de connaissance

La détection de vandalisme est un sujet étudié dans les projets de *crowdsourcing*, tels que Wikipédia et Wikidata (Adler *et al.*, 2011 ; Chin *et al.*, 2010 ; Heindorf *et al.*, 2015 ; Potthast, 2010 ; Tran, 2015). Les systèmes de détection cités utilisent des méthodes de classification supervisée. Or, l'apprentissage supervisé nécessite de disposer d'un corpus assez conséquent de données annotées. La construction d'un corpus de vandalisme consiste à annoter des données – par exemple, **VRAI** pour une donnée vandalisée, **FAUX** sinon – à partir desquelles un système pourra apprendre les cas de vandalisme (jeu de données d'entraînement) ou évaluer le système de détection (jeu de données test). Alors que Potthast (2010) construit un corpus de vandalisme Wikipédia par annotation humaine via le service Amazon Mechanical Turk, des corpus récents de vandalisme de données Wikidata ont été construits par annotation automatisée. Dans (Heindorf *et al.*, 2015 ; Sarabadani *et al.*, 2017), la première étape consistait à filtrer les révisions automatiques Wikidata, afin de ne conserver que les révisions manuelles. Parmi 24 millions d'éditions manuelles, Heindorf *et al.* (2015) ont considéré comme données vandalisées les contributions ayant subi une opération de révocation (*rollback*). La révocation est en réalité une opération autorisée pour un nombre restreint de contributeurs Wikidata, appelés les révocateurs¹⁰, consistant à annuler toutes les versions consécutives d'une donnée dans le but de la restaurer à un état précédant son état vandalisé. Sarabadani *et al.* (2017) adoptent une stratégie différente pour annoter automatiquement les données vandalisées. En partant du principe que les données révoquées ne sont pas les seules données vandalisées dans un jeu Wikidata, l'obtention des données vandalisées dans ce corpus résulte d'un filtrage successif où les révisions faites par des contributeurs fiables ne sont pas considérées. En parcourant uniquement les éditions faites par des contributeurs non fiables, les données issues d'une opération d'annulation – révocation ou autre opération de restauration autorisée à tous les contributeurs – sont labélisées **VRAI**.

On note que chaque corpus présente une part de faux positifs (révisions non vandalisées annotées **VRAI**) et de faux négatifs (révisions vandalisées annotées

9. <https://wiki.openstreetmap.org/wiki/FR:Vandalisme>

10. <https://www.wikidata.org/wiki/Wikidata:Rollbackers/fr>

Tableau 3. Différentes approches de détection du vandalisme

Application	Approche	
	Non supervisée	Supervisée
Aide à la décision	Notre expérience	(Sarabadani <i>et al.</i> , 2017)
Détection automatique du vandalisme	N/A	(Potthast, 2010)

FAUX). Alors que le corpus proposé par (Heindorf *et al.*, 2015) a tendance à laisser passer de nombreux cas de vandalisme (32 % de faux négatifs), celui de (Sarabadani *et al.*, 2017) détecte un grand nombre de faux cas de vandalisme (32 % de faux positifs) en minimisant le nombre de faux négatifs (moins de 1 %). Notons ici que la détection du vandalisme dans (Sarabadani *et al.*, 2017) a pour objectif d'aider les modérateurs à corriger les données vandalisées. Le système d'apprentissage fonctionne ici comme un outil d'aide à la décision : en termes de performances, on exige en priorité un fort rappel, à la différence de la détection automatique du vandalisme, où l'on exige à la fois un fort rappel et une forte précision. Tout compte fait, la tolérance aux faux positifs dans le corpus de vandalisme dépend de l'objectif final de la détection de vandalisme.

Actuellement, il n'existe pas de corpus de vandalisme pour OSM, et les approches présentées précédemment ne sont pas toutes applicables pour les données OSM. En effet, tous les contributeurs ont les mêmes pouvoirs concernant l'édition de données, il n'existe pas de révocateurs comme dans Wikidata. Bien que le phénomène de *rollback* existe dans OSM (celui-ci est connu sous le nom de *revert*), cette opération peut être faite par n'importe qui, et pas toujours de manière justifiée. Par conséquent, il ne serait pas raisonnable de considérer toute contribution OSM ayant subi des reverts comme étant une contribution vandalisée. Comme la construction d'un corpus de vandalisme OSM est un travail qui n'a pas encore été effectué, et qui n'est pas une tâche triviale, nous cherchons à détecter le vandalisme dans OSM par une méthode non supervisée et dans le cadre d'une application d'aide à la décision (tableau 3).

Dans le cas de l'information géographique volontaire, Neis *et al.* (2012) ont proposé un système de décision multi-critère pour détecter le vandalisme. Cet outil est, pour l'instant, le plus abouti pour la détection du vandalisme dans l'information géographique volontaire. Cependant, la définition du vandalisme est différente de la définition traditionnelle, développée ici dans cet article. En effet, le vandalisme selon Neis *et al.* (2012) tient uniquement compte de la composante dégradation. Par conséquent, cette définition inclut toutes les dégradations faites à la base de données géographiques, indépendamment du fait que celles-ci aient été faites volontairement ou non. Les règles de décision de cet outil prennent en compte la réputation du contributeur et l'historique des données OSM. Par conséquent, le système détecte facilement les objets créés par de nouveaux utilisateurs OSM, sans que ceux-ci soient forcément des cas avérés de vandalisme, puisque d'une part, le système est conçu pour détecter les

erreurs volontaire et involontaire, d'autre part, les nouveaux contributeurs, à cause de leur manque d'expérience, sont plus enclins à se tromper dans leurs premières saisies. Par ailleurs, d'après leurs résultats expérimentaux, le système a tendance à retourner un grand nombre de faux positifs, dont des éditions produites par des contributeurs expérimentés. Mis à part le fait que le système marque 44 % du jeu de données test comme de possibles cas de vandalisme, il n'est pas fait mention de la quantité de faux négatifs. Cela ne remet pas en cause la méthode de décision multicritère pour détecter le vandalisme, mais il faudrait trouver un ensemble de règles permettant de réduire le nombre de faux positifs dans ce système.

À ce point, on est amené à s'interroger sur ce qui doit être considéré comme une contribution dégradée dans l'information géographique volontaire. La dégradation opérée dans le vandalisme, selon sa définition historique, s'applique à des objets, initialement en bon état, qui sont endommagés par la suite. De la même manière, la définition législative conserve l'idée de transformation d'un bien en bon état vers un état détérioré. Dans l'information géographique volontaire, si le bien dégradé portait sur la contribution, alors le vandalisme cartographique ne concernerait que les objets existants qui diminuent en qualité. Or, le carto-vandalisme fantaisiste consiste à créer des objets cartographiques totalement fictifs, c'est-à-dire que ces objets seraient dégradés dès leur création. On perd alors la notion de transformation, présente dans la définition traditionnelle du vandalisme. Nous considérons donc que le bien dégradé, dans OSM ou tout autre plateforme de saisie cartographique collaborative, correspond à l'espace cartographique. En effet, cet espace cartographique est censé être une modélisation du monde réel (avec un certain degré d'abstraction) en vue d'une ou plusieurs utilisations (affichage cartographique, analyse spatiale, etc.). Par conséquent, une contribution vandalisée dans OSM correspond à toute édition qui dégrade volontairement l'espace cartographique sans motif légitime. Cette dégradation peut se situer à deux niveaux. D'une part, celle-ci peut rendre la représentation cartographique de l'espace non fidèle au monde réel. Cela concerne, dans une certaine mesure, les contributions de mauvaise qualité. D'autre part, la dégradation peut rendre cette représentation inutilisable dans une application donnée (navigation, analyse de biodiversité, de morphologie urbaine, etc.). Cela concerne, par exemple, les données qui ne respectent pas les spécifications de saisie.

4. Le vandalisme : une anomalie à détecter

4.1. Motivations

Malgré la rareté des cas de vandalisme sur OpenStreetMap, il est en revanche techniquement permis de produire du vandalisme sur la base, puisque le mode de fonctionnement d'OSM offre la liberté au contributeur d'éditer la base de données géographiques comme bon lui semble. Toute nouvelle édition est directement chargée dans la base de données, sans aucun contrôle préalable. Par conséquent, il est important de pouvoir détecter les cas de vandalisme, au moins pour aider les contributeurs de bonne foi qui cherchent à réparer les données vandalisées.

Un acte de vandalisme peut être repéré selon différents critères. Par exemple, le cas de vandalisme présenté dans le [tableau 1](#) est repérable par une analyse des valeurs de tags, de manière à montrer que l'utilisation de smileys ou plus généralement, d'une grande quantité de caractères spéciaux dans un seul tag est inhabituelle. L'exemple de la [figure 1](#) n'est pas repérable de la même manière : là, il s'agit de voir qu'il est très inhabituel qu'un lac (ou toute zone naturelle du même type) contienne ou chevauche des bâtiments. En effet, si on mesure les intersections topologiques entre tous les lacs d'une zone et des bâtiments, la présence d'une inclusion sera anormale.

À travers ces exemples, on peut faire l'hypothèse que, en se plaçant dans le bon espace de description, les données de vandalisme sont perçues comme des anomalies. Or, en fouille de données, la détection d'anomalie consiste à identifier des motifs non conformes à un comportement attendu au sein d'un jeu de données ([Chandola et al., 2009](#)). Par conséquent, on cherche à détecter le vandalisme des données OSM au moyen de méthodes de détection d'anomalies. Néanmoins, cette méthode peut trouver ses limites pour repérer les contributions vandalisées dont il n'existe pas d'espace de description dans lequel elles seraient vues comme des anomalies. Nous reviendrons sur ce point dans notre expérimentation, présentée dans la partie suivante. La difficulté de la méthode adoptée réside donc dans l'établissement d'un espace de description adapté – lorsqu'il existe – mettant en lumière les contributions de vandalisme, et non pas des contributions qui présentent des caractéristiques particulières sans pour autant être issues d'un acte de vandalisme cartographique.

4.2. Insertion de vandalisme synthétique

Notre démarche expérimentale vise à détecter automatiquement le vandalisme causé sur des bâtiments OSM dans la ville d'Aubervilliers, en banlieue parisienne. Le jeu de données ne contient *a priori* pas de contribution de vandalisée, par conséquent on ajoute manuellement des objets vandalisés avec le logiciel QGIS ([figures 3, 4 et 6](#)). Le vandalisme introduit dans la base de données peut se diviser en quatre types, selon la forme qu'il prend ([tableau 4](#)). En reprenant la typologie proposée par [Ballatore \(2014\)](#), les contributions vandalisées introduites s'apparentent à du vandalisme idéologique (type A), artistique (types B et D) et fantaisiste (type C). Bien entendu, ces contributions vandalisées comportent un certain biais car elles ne couvrent pas tous les types de vandalisme.

Le vandalisme de type A a été produit sur 42 bâtiments industriels et religieux existants auxquels nous avons supprimé le tag "name" lorsqu'il existait (c'est le cas pour les églises), ou modifié le tag en name = "...", ou "name = pfff" ([figure 3](#)). Le vandalisme de type B comprend 10 bâtiments tracés de manière artistique au milieu d'un cours d'eau. Le vandalisme de type C comprend 17 bâtiments fictifs, ajoutés dans des zones dépourvues de bâtiments (le polygone jaune de la [figure 6](#) indique que l'espace est actuellement une zone de construction). Le vandalisme de type D correspond à 3 bâtiments : deux énormes bâtiments, dont l'un se situe sur le cimetière d'Aubervilliers



Figure 3. Vandalisme synthétique de type idéologique. Une étiquette comportant la valeur du tag "name" est affichée sur les bâtiments vandalisés (en violet/gris)

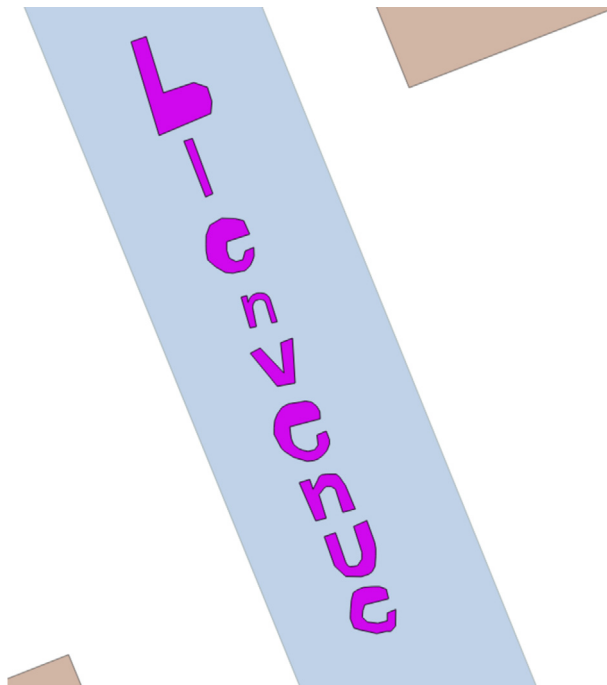


Figure 4. Vandalisme synthétique de type B



Figure 5. Vandalisme synthétique de type B (énorme bâtiment à gauche) et C

Tableau 4. Types de vandalisme introduits dans le jeu de données de bâtiments

Type	Définition
A	Bâtiments existants, souvent des magasins ou des lieux de culte, dont les tags ont été modifiés ou supprimés. La modification attributaire peut contenir des smileys
B	Bâtiments fictifs ajoutés dans des zones inhabituelles (étang, cours d'eau)
C	Bâtiments fictifs réalistes ajoutés dans des espaces vides
D	Bâtiments de forme et/ou de taille inhabituelle ajoutés dans des espaces vides

(figure 5), et un bâtiment de taille minuscule. Le jeu de données de bâtiment utilisé dans cette expérience contient, au total, 10 365 bâtiments, parmi lesquels 72 sont du vandalisme synthétique.

4.3. Expérimentations

4.3.1. Calcul des descripteurs sur les bâtiments

Comme le vandalisme synthétique introduit dans le jeu de données a été produit de différentes manières, il faut donc choisir des descripteurs sur ces données de façon à ce que les contributions vandalisées ressortent comme des anomalies selon un ou plusieurs de ces descripteurs. Par conséquent, nous avons implémenté un certain nombre de descripteurs géométriques, topologiques, sémantiques, syntaxiques, historiques, et orientés-contributeurs. Le tableau 5 donne la liste des descripteurs qui ont été calculés sur les bâtiments du jeu de données.

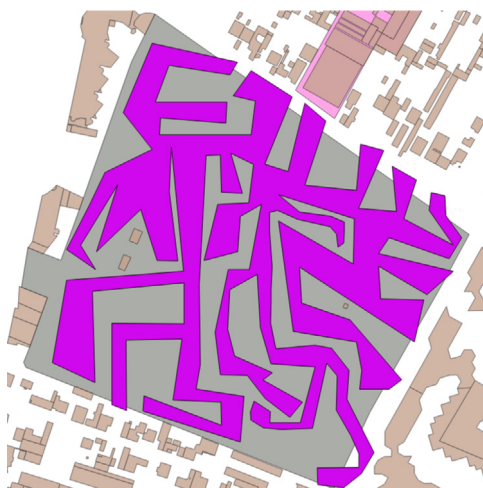


Figure 6. Vandalisme synthétique de type D

La fiabilité du contributeur OSM peut être quantifiée au moyen de différents indicateurs, comme par exemple le nombre de contributions produites dans la base (Barron *et al.*, 2014) ou le nombre d'heures passées à contribuer sur la plateforme (Geiger et Halfaker, 2013). Cependant, nous remarquons que ces indicateurs ne sont que des approximations de la fiabilité du contributeur, et peuvent ne pas la quantifier correctement. Si la fiabilité est assimilée au nombre de contributions, un contributeur produisant un faible nombre d'édits manuelles sera considéré moins fiable qu'un *bot*, c'est-à-dire un contributeur qui produit des imports massifs de données à l'aide de scripts automatiques. Or, dans le cas où le premier s'applique à faire des données manuelles de qualité et le second se trompe dans ses imports de données, l'indicateur de fiabilité ne reflètera pas fidèlement la fiabilité de l'un par rapport à l'autre. Le même raisonnement est valable pour le cas où la fiabilité est quantifiée par le nombre d'heures participées.

En conséquence, nous proposons un nouvel indicateur de fiabilité, basé sur un ensemble de métriques qui ont une influence (positive ou négative) sur la fiabilité du contributeur. Cet indicateur est calculé comme la moyenne simple de ces métriques orientées contributeur (tableau 6). Nous précisons que, pour les métriques m ayant une influence négative sur la fiabilité du contributeur, la quantité $1 - m$ a été considérée dans le calcul de la moyenne *avg*. Ainsi, plus le score *avg* est grand, plus le contributeur est censé être fiable.

Aux données de vandalisme synthétique, il a fallu y associer des profils synthétiques de contributeurs. Nous avons donc fait en sorte de créer un profil de contributeur non fiable pour les contributions vandalisées de type C. À partir de l'observation de la distribution des différentes métriques orientées contributeurs existantes, nous avons donné des valeurs de métriques suffisamment basses pour que le score *avg* soit faible, et qu'ainsi, l'auteur du vandalisme de type C ait un profil peu fiable selon ce descripteur.

Tableau 5. Liste des variables calculées sur les bâtiments OSM

Nom de la variable	Définition
<i>perimeter</i>	Périmètre du bâti
<i>shortest_length</i>	Longueur du plus petit côté
<i>median_length</i>	Valeur médiane des longueurs du polygone
<i>elongation</i>	$\frac{\text{largeur}(\text{PPRE}_{\text{bati}})}{\text{longueur}(\text{PPRE}_{\text{bati}})}$
<i>convexity</i>	$\frac{\text{aire}(\text{bati})}{\text{aire}(\text{PPRE}_{\text{bati}})}$
<i>compactity</i>	Indice de Miller : $\frac{4 * \pi * \text{aire}(\text{bati})}{\text{perimetre}(\text{bati})^2}$
<i>n_is_within_lulc</i>	Nombre de zones naturelles dans lesquelles est contenu le bâti courant
<i>n_inter_lulc</i>	Nombre de zones naturelles qui intersectent le bâti courant
<i>min_dist_surf_bati_bdtopo</i>	$\min(d_S(\text{bati}_{\text{OSM}}, \text{bati}_{\text{BDTOPO}}))$, où d_S est la distance surfacique : $d_S(A, B) = 1 - \frac{S(A \cap B)}{S(A \cup B)}$, S étant la fonction de calcul de surface
<i>max_special_char_ratio</i>	Pourcentage maximal de caractères spéciaux dans toutes les valeurs de tag
<i>n_tags</i>	Nombre de tags
<i>n_users</i>	Nombre de contributeurs uniques depuis la version 1 de l'objet
<i>timespan_to_previous</i>	Durée (en secondes) entre la version précédente de l'objet et sa version actuelle
<i>avg</i>	Score de fiabilité calculé comme moyenne simple d'indicateurs orienté-contributeurs
<i>weighted_avg</i>	Score de fiabilité calculé comme moyenne pondérée d'indicateurs orienté-contributeurs

N.B. : PPRE signifie plus petit rectangle englobant.

Une première expérience a été menée pour détecter du vandalisme fantaisiste et artistique (Truong *et al.*, 2018b) avec DENCLUE, un algorithme de détection d'anomalie par *clustering* basé sur la densité des données (Hinneburg et Gabriel, 2007). Les expérimentations décrites ici reprennent les mêmes données de bâtiments d'Aubervilliers, en cherchant cette fois à améliorer les résultats de détection du

Tableau 6. Liste des métriques orientées contributeur, utilisées pour construire un indicateur de fiabilité du contributeur, noté *avg*

Métrique	Définition	Influence sur la fiabilité
<i>total_contributions</i>	Nombre de contributions sur la zone	+
<i>p_modif</i>	Pourcentage de contributions qui sont des éditions (<i>i.e.</i> des modifications de données existantes)	+
<i>p_delete</i>	Pourcentage de contributions qui sont des suppressions d'objets	+
<i>p_deleted</i>	Pourcentage de contributions qui sont ultérieurement supprimées	-
<i>p_modified</i>	Pourcentage de contributions qui sont ultérieurement éditées	-
<i>nbWeeks</i>	Nombre de semaines contribuées	+
<i>focalisation</i>	Focalisation des contributions sur la zone étudiée. Calcul basé sur l'étendue moyenne des changesets produits.	+

vandalisme. À la différence des expériences présentées dans (Truong *et al.*, 2018b), nous avons choisi de lancer les expériences avec DBSCAN car son implémentation sur *R* était simple d'utilisation et plus rapide à l'exécution. De plus, nous espérons améliorer nos résultats en choisissant d'autres types de descripteurs aux descripteurs purement géométriques utilisés dans l'expérience initiale avec DENCLUE. Nous explorons ainsi l'impact de l'ajout et de la suppression de certains descripteurs passés en entrée de l'algorithme de détection d'anomalie, afin de déterminer ceux qui permettent d'identifier au mieux les différents cas de vandalisme étudiés.

4.3.2. Sélection des descripteurs optimaux

Comme il a été soulevé précédemment, pour que le vandalisme soit détectable par une méthode de détection d'anomalie, il faut se placer dans un espace de description adapté. Les dimensions d'un espace de description sont définies par des variables descriptives, que l'on appellera descripteurs dans la suite de cet article. Il s'agit donc de sélectionner les descripteurs optimaux, c'est-à-dire ceux qui formeront un espace dans lequel les données vandalisées seront vues comme des anomalies. Les indicateurs de précision, de rappel et d'erreur sont utilisés pour évaluer la qualité de la détection et ainsi permettre de déterminer les descripteurs optimaux pour le problème étudié.

La précision est comprise ici comme la précision de la détection du vandalisme. Elle est définie comme la part de contributions vandalisées classées en anomalies sur l'ensemble des données classées en anomalies :

$$precision = \frac{TP}{TP + FP} \quad (1)$$

où TP correspond au nombre de vrais positifs et FP au nombre de faux positifs. Les vrais positifs sont les contributions vandalisées qui ont été classées comme anomalies par l'algorithme. Les faux positifs sont les contributions qui ne font pas partie du vandalisme synthétique mais qui ont été classées comme anomalies. Il s'agit, dans l'idéal, d'obtenir une précision maximale de 1, puisque cela signifie que $FP = 0$, c'est-à-dire qu'aucune donnée non vandalisée n'a été classée comme anomalie.

Le rappel est défini comme la portion de contributions vandalisées classées en anomalies par rapport au nombre total de contributions vandalisées :

$$rappel = \frac{TP}{TP + FN} \quad (2)$$

où FN correspond au nombre de faux négatifs, c'est-à-dire les contributions vandalisées qui n'ont pas été classées comme anomalies. Comme pour la précision, il faudrait dans l'idéal obtenir un rappel de 1, ce qui signifierait alors que $FN = 0$, c'est-à-dire que toutes les contributions du vandalisme synthétique ont été détectées comme anomalies.

L'erreur correspond au pourcentage global de données mal classées :

$$erreur = \frac{FP + FN}{TP + TN + FP + FN} \quad (3)$$

où TN correspond au nombre de vrais négatifs, c'est-à-dire les contributions non vandalisées qui n'ont pas été classées comme anomalies. Cet indicateur permet de nuancer les faibles valeurs de précision que l'on risque d'obtenir dans notre situation. En effet, comme les anomalies à détecter représentent une faible portion de l'ensemble du jeu de données, lorsque le nombre de faux positif est largement supérieur au nombre de vrais positifs, la précision calculée sera faible, tandis que ce même nombre de faux positifs peut représenter une très faible portion par rapport à l'ensemble des données non vandalisées. Contrairement à la précision et au rappel, il faut dans l'idéal minimiser l'erreur à 0 pour s'assurer qu'une classification soit globalement bonne.

La détection d'anomalies avec DBSCAN a été effectuée plusieurs fois en faisant varier la combinaison des descripteurs. Le meilleur résultat de détection du vandalisme (tableau 7) a été trouvé en introduisant les descripteurs suivants en entrée de DBSCAN :

- *min_dist_surf_bati_bdtopo*
- *perimeter*
- *max_special_char_ratio*
- *n_is_within_lulc*
- *avg*

Tableau 7. Détection du vandalisme avec DBSCAN en entrée avec l'ensemble des descripteurs suivants : {min_dist_surf_bati_bdtopo, perimeter, max_special_char_ratio, n_is_within_lulc, avg}

Précision	Rappel	Erreur	Type A	Type B	Type C	Type D
0,201	0,944	0,026	1	0,67	1	0,67

L'algorithme DBSCAN prend en entrée l'ensemble des descripteurs calculés sur les bâtiments, ainsi que deux paramètres : *eps*, la distance maximale entre deux points d'un même *cluster*, et *minPts*, le nombre minimal d'objets requis pour pouvoir former un *cluster* (Ester *et al.*, 1996). Afin de trouver les valeurs optimales de paramètres de DBSCAN – c'est-à-dire les valeurs de *eps* et *minPts* qui permettent de maximiser le rappel et la précision dans la détection d'anomalie – on lance DBSCAN (sous R, avec le *package dbscan*) en faisant varier itérativement les paramètres. Cette expérience préliminaire a permis de fixer les paramètres (*eps*, *minPts*) = (0.75,3) pour la détection des anomalies avec DBSCAN dans notre cas d'étude.

La précision de 20 % obtenue est très faible : elle s'explique par le fait que l'algorithme a classé 67 vrais positifs et 266 faux positifs. En revanche, le fort rappel traduit le fait que le système est capable de détecter en grande partie, environ 94 %, le vandalisme synthétique. L'erreur, de près de 3 %, signifie que globalement, l'algorithme se trompe très peu dans la classification du vandalisme (assimilé à l'anomalie) et du non-vandalisme. Pour confirmer l'intérêt des descripteurs utilisés dans la détection présentée dans le [tableau 7](#), on étudie par la suite l'influence des résultats avec et sans l'indicateur de fiabilité du contributeur et l'indicateur d'appariement.

4.3.3. Influence d'un indicateur de fiabilité du contributeur

On lance DBSCAN sans l'indicateur de fiabilité du contributeur, puis successivement avec d'autres indicateurs de fiabilité du contributeur. En effet, comme présenté dans la partie 4.3.1, la fiabilité du contributeur étant un descripteur qualitatif, il peut exister différents descripteurs pour quantifier cette variable. Il s'agit, d'une part, de vérifier que la prise en compte d'un indicateur de fiabilité permet d'améliorer la détection du vandalisme avec DBSCAN. D'autre part, il s'agit de déterminer la meilleure manière de modéliser la fiabilité du contributeur dans le cadre du problème posé. On s'intéresse en particulier à la détection du vandalisme de type C, car celui-ci n'est pas censé être vu comme une anomalie par les autres variables descriptives. On cherche donc à observer l'influence d'un indicateur de fiabilité sur la détection du vandalisme en considérant trois indicateurs de fiabilité différents ([tableau 8](#)), que l'on

Tableau 8. Différents indicateurs de fiabilité du contributeur

Indicateur de fiabilité	Définition
<i>total_contributions</i>	Nombre total d'éditions d'un contributeur sur la zone d'étude
<i>avg</i>	Moyenne des indicateurs sur le contributeur
<i>n_semesters_auberv</i>	Nombre de semestres contribués sur la zone d'étude

ajoute à l'ensemble initial de descripteurs (tableau 5). Pour chaque indicateur, on relance DBSCAN en calculant des indicateurs de performance, ainsi que la proportion de vandalisme de type C détecté comme anomalie (tableau 9).

De manière générale, le nombre de faux positifs augmente avec la prise en compte d'un indicateur de fiabilité, quel qu'il soit. Par conséquent, la précision diminue et l'erreur augmente. En revanche, on observe une augmentation du rappel, ce qui signifie que le système parvient à classer plus de contributions vandalisées correctement (*i.e.* comme anomalies). Quel que soit le descripteur de fiabilité sélectionné, on observe que le vandalisme de type C est mieux détecté qu'en l'absence totale d'indicateur de fiabilité, puisque sans cet indicateur, aucun vandalisme de type C n'est détecté. Avec chaque indicateur de fiabilité, les indicateurs de précision et d'erreur sont moins bons que sans cet indicateur, tandis que le rappel est meilleur : cela traduit une augmentation du nombre de faux positifs et une diminution du nombre de faux négatifs. Or, notre objectif est de fournir un système d'alerte pour de l'aide à la décision, donc une légère hausse du nombre de faux positifs reste tolérable tant que le nombre de faux négatifs se trouve, quant à lui, réduit. Par conséquent, ces résultats montrent l'intérêt de prendre en compte la fiabilité du contributeur dans la détection de vandalisme, en particulier pour détecter les cas les moins évidents.

Parmi les trois indicateurs de fiabilité introduits, on constate qu'avec le descripteur *avg*, le système détecte le vandalisme de type C en intégralité. En comparaison avec

Tableau 9. Résultats de détection d'anomalies avec différents indicateurs de fiabilité du contributeur en entrée de DBSCAN

Indicateur de fiabilité	Précision	Rappel	Erreur	Vandalisme C détecté
∅	0,231	0,563	0,015	0
<i>total_contributions</i>	0,146	0,592	0,026	0,067
<i>n_semesters_auberv</i>	0,086	0,958	0,070	1
<i>avg</i>	0,201	0,944	0,026	1

Tableau 10. Résultats de la détection d'anomalies avec et sans indicateur d'appariement en entrée de DBSCAN

	Sans appariement	Avec appariement
Précision	0,234	0,201
Rappel	0,916	0,944
Erreur	0,021	0,026
Type A détecté	1	1
Type B détecté	0,67	0,67
Type C détecté	0,87	1
Type D détecté	0,67	0,67

total_contributions et *n_semesters_auberv*, la prise en compte de *avg* donne une précision maximale et une erreur minimale. Quant au rappel de 94 %, celui-ci est moins élevé que pour *n_semesters_auberv* (de 96 %) mais il est bien plus satisfaisant que celui de 56 %, obtenu sans indicateur de fiabilité. Par conséquent, l'indicateur *avg* nous semble le plus adapté pour modéliser la fiabilité du contributeur dans le cadre de la détection du vandalisme.

4.3.4. Validité d'un indicateur d'appariement

Il s'agit à présent de vérifier l'intérêt de l'indicateur d'appariement pour la détection du vandalisme. L'appariement suppose des données de référence, qui ne sont pas toujours accessibles. Dans notre expérience, on utilise comme données de référence la couche de bâti de la BD TOPO de l'IGN. Puis, on relance la détection d'anomalies avec et sans cet indicateur (tableau 10). Avec l'indicateur d'appariement, la précision diminue de 3 %, mais le rappel augmente de 3 %. En d'autres termes, le système parvient à mieux détecter le vandalisme mais devient plus tolérant face aux faux positifs, d'où une plus grande erreur avec l'appariement. En observant les différents types de vandalisme détectés, on constate que l'indicateur d'appariement n'a pas d'effet sur la détection du vandalisme de type A, B et D. En revanche, il permet de détecter intégralement le vandalisme de type C, alors que celui-ci n'était détecté qu'à 87 % sans appariement. Ces résultats permettent donc de valider la pertinence de ce descripteur pour détecter le vandalisme.

5. Discussion

Dans le cas d'un outil d'aide à la décision, le système de détection du vandalisme doit, en priorité, minimiser le nombre de faux négatifs *i.e.* éviter de laisser des cas de

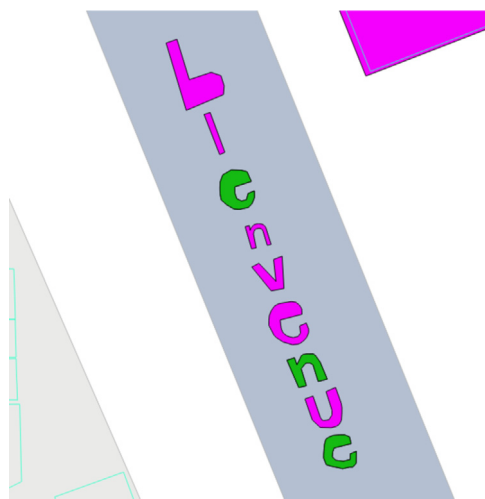


Figure 7. Faux négatifs du vandalisme de type B (bâtiments en vert)

vandalisme non détectés. En cela, nos premiers résultats sont plutôt satisfaisants, puisque notre expérience détecte le vandalisme avec un rappel de 94 %, et près de 5 % de faux négatif (quatre de nos bâtiments vandalisés n'ont pas été détectés). Parmi les faux négatifs, un minuscule bâtiment fictif (qui appartient au vandalisme de type D) n'a pas été reconnu comme une anomalie car le jeu de données contenait initialement un grand nombre de petits bâtiments. En effet, les données OSM françaises sont en général issues de l'import massif de données du cadastre. Par conséquent, de nombreux éléments OSM cartographiés comme bâtiments sont en réalité des petits morceaux de bâtiments de formes singulières. Les autres faux négatifs sont trois cas de vandalisme de type B (figure 7). Ici, on peut se demander pourquoi ces trois bâtiments n'ont pas été détectés alors qu'ils sont similaires (vu la valeur de leurs descripteurs) aux autres bâtis tracés dans le cours d'eau, qui eux, ont bien été marqués comme anomalies par DBSCAN. La détection partielle de ce vandalisme permet, malgré tout, de localiser une zone sur laquelle se trouveront potentiellement d'autres cas de vandalisme non détectés par le système.

En prenant en compte les descripteurs d'appariement et de fiabilité du contributeur, le système détecte correctement le vandalisme avec une précision de 20 % : on compte 261 faux positifs, ce qui représente 2,5 % du jeu de données étudié. En observant les faux positifs (figure 8), on remarque que ce sont des bâtiments relativement grands par rapport à la distribution globale du périmètre de l'ensemble des bâtis non vandalisés (figure 9). Pour que le système soit véritablement une aide à la prise de décision, une réduction du nombre de faux positifs s'impose. Dans notre situation, un faux positif correspond à un bruit que l'on cherche à éliminer. Pour cela, on pourrait filtrer les bâtiments normaux qui risquent d'être détectés par le système, en pré-qualifiant les objets présentant des spécificités géométriques, comme par exemple le centre commercial d'Aubervilliers, remarquable par sa grande taille ; ou des immeubles



Figure 8. Bâtiments détectés comme faux positifs par DBSCAN (en jaune, gris)

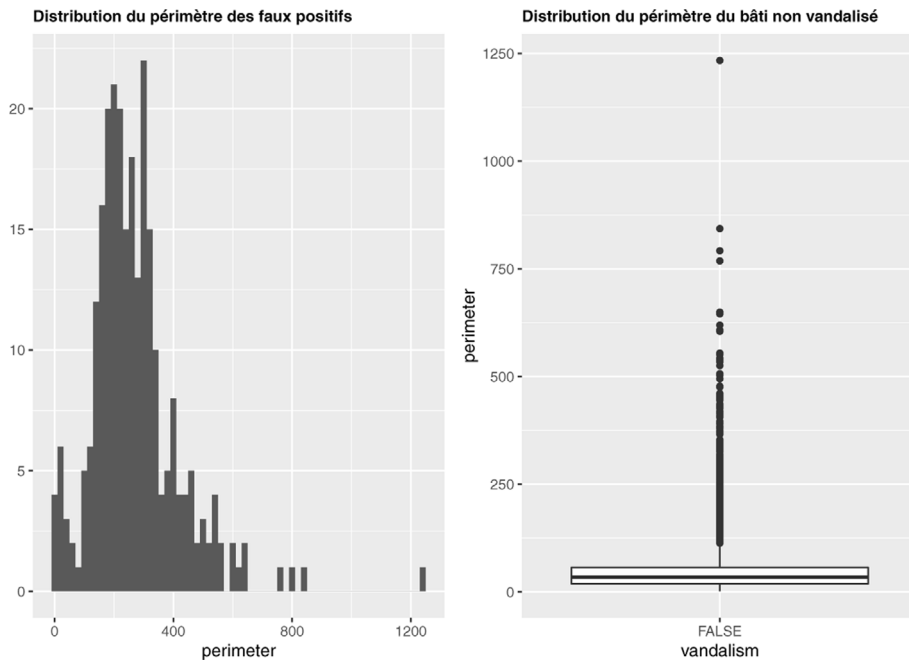


Figure 9. Distribution du périmètre des faux positifs (à gauche) et du bâti non vandalisé (à droite)

géométriquement remarquables par leur architecture originale. Cela nécessite une connaissance *a priori* du jeu de données initial. Cette étape de filtrage des faux positifs peut également être faite en sortie de l'algorithme de clustering : en prenant l'ensemble des données détectées comme anomalies, on pourrait filtrer les faux positifs par une seconde détection d'anomalie en prenant d'autres descripteurs en entrée. Dès lors que le nombre de faux positifs aura diminué, notre système de détection de vandalisme par classification non supervisée sera plus précis, et il pourra éventuellement s'étendre à une application de détection entièrement automatisée, mais cela ne pourra être envisagé que si l'on parvient à réduire drastiquement le nombre de faux positifs.

La détection du vandalisme avec une précision de 20 % et un rappel de 94 % a permis de détecter entièrement le vandalisme de type A. Pour cela, nous avons mis en place le descripteur syntaxique *max_special_char_ratio* pour récupérer les anomalies de ce type de vandalisme. Or, le vandalisme synthétique de type A correspond à des bâtiments sur lesquels le tag name a pu être supprimé ou modifié par la valeur "name = pfff" (voir partie 4.2). Le descripteur *max_special_char_ratio* calculé pour ces bâtiments-là ne mettra en avant aucune anomalie puisque le tag ne contient aucun caractère spécial. On est face ici à l'existence de « faux » vrais positifs, c'est-à-dire des éléments marqués comme anomalies pas le système, mais qui, au vu des descripteurs choisis, n'auraient pas dû être détectés comme tels : finalement, ces bâtiments peuvent être considérés comme des faux positifs. Pour traiter ce problème, il est essentiel de toujours s'interroger sur la pertinence des descripteurs choisis en entrée de l'algorithme de détection d'anomalie.

Dans notre expérience, nous avons choisi d'utiliser un descripteur topologique, *n_is_within_lulc*, qui mettait en valeur l'inclusion du bâti dans des espaces naturels. Or, on pourrait également considérer un indicateur qui calculerait le nombre d'intersections du bâti avec ces espaces naturels, voire d'autres éléments géographiques, tels que les routes. L'ajout de ces variables topologiques permettrait de détecter d'autres bâtiments situés à des endroits inhabituels. L'ajout d'autres descripteurs de type « analyse historique » ou sémantiques pourraient peut-être améliorer encore les résultats, mais ceux-ci n'ont pas été implémentés.

Par ailleurs, il est important de discuter de la manière dont ont été construits certains descripteurs. Dans le cas du descripteur d'appariement, nous avons pris le parti d'utiliser le minimum de la distance surfacique, pour des raisons de simplicité. Or, il existe d'autres manières de concevoir un indicateur d'appariement (Fan *et al.*, 2014 ; Xu *et al.*, 2017), qui mériteraient d'être considérées pour déterminer la méthode d'appariement la plus adaptée pour détecter les anomalies dues au vandalisme. En adoptant une telle démarche pour le descripteur de fiabilité de l'utilisateur, nous avons obtenu des résultats très intéressants. D'une part, nous avons montré que le fait d'approcher la fiabilité du contributeur à sa quantité de contributions (en nombre d'éditions ou nombre d'heures) n'est pas satisfaisante dans notre situation. Le calcul d'un indicateur de fiabilité du contributeur par la moyenne d'indicateurs intermédiaires a permis d'améliorer la détection de cas de vandalisme plus difficiles à détecter, tels que le vandalisme fantaisiste, qui est imperceptible selon les autres descripteurs. D'autre part, grâce à ce

nouveau descripteur, nous avons montré l'intérêt de prendre en compte la variable contributeur pour qualifier les contributions OSM.

La détection du vandalisme des données OSM par une classification supervisée nécessite de construire un corpus de contributions annotées qui, pour l'instant, n'existe pas dans le cas des données OSM. Dans le cadre d'une aide à la décision, on peut reprendre la piste proposée par (Sarabadani *et al.*, 2017) qui consiste à filtrer les contributions Wikidata faites par des contributeurs de confiance. Ainsi, les contributions douteuses de notre jeu de données seront récupérées pour être labélisées comme étant du vandalisme. Or, contrairement, à Wikidata, il n'y a pas de privilège accordée aux contributeurs OSM au niveau des opérations permises pour l'édition de données. Par conséquent, il sera nécessaire de mettre au point une méthode de qualification des contributeurs pour permettre d'évaluer leur niveau de fiabilité (Keßler et Groot, 2013 ; Truong *et al.*, 2018a).

Ce corpus de données annotées pourra contenir des cas de vandalisme synthétique similaires à ceux que nous avons produits manuellement (figures 4 et 6), mais également d'autres types de vandalisme synthétique que nous n'avons pas considérés dans notre expérience. Par exemple, on pourrait cartographier des bâtiments qui – bien que fictifs – respecteraient les lois géographiques, c'est-à-dire qu'ils seraient géométriquement cohérents entre eux et avec les autres objets de la carte (Goodchild et Li, 2012). Ce genre de vandalisme fictif mais réaliste pourrait être détecté en changeant l'échelle d'étude, et ainsi, au lieu d'analyser les objets géographiques individuellement, il s'agirait d'analyser la cohérence entre des groupes d'objets pour identifier des cartographies imaginaires de villes ou de quartiers (Landrieu et Simonovsky, 2017).

Par ailleurs, dans les expériences présentées, le vandalisme synthétique a été créé manuellement par un opérateur humain qui, à lui seul, ne peut pas rendre compte de tous les cas de vandalisme possibles et imaginables, mais introduit un biais dans les données créées. En effet, le vandalisme créé repose sur sa conception personnelle du carto-vandalisme. Bien qu'il ne soit pas réaliste de couvrir tous les cas de vandalisme réels par des contributions synthétiques, ce biais peut être atténué en faisant appel à plusieurs opérateurs « vandales » différents, humains ou non, pour dégrader la base. En effet, le vandalisme de type spam cartographique décrit par Ballatore (2014) se fait au moyen d'édits automatiques (*bots*¹¹). Aussi, une possibilité serait de produire du carto-vandalisme aléatoire à partir de scripts automatiques.

Enfin, la détection totalement automatique du vandalisme OSM par classification – qu'elle soit supervisée ou non – n'est pas encore envisageable. En effet, celle-ci suppose de construire un corpus solide de vandalisme OSM, qui ne laisse pas de place à l'approximation : toutes les contributions OSM du corpus devront être annotées de manière à ce que le nombre de faux positifs et le nombre de faux négatifs soient nuls à l'issue de la détection. Dans ce cadre-là, l'annotation automatique proposée par Sarabadani *et al.* (2017) n'est pas assez satisfaisante pour construire un tel corpus. Il

11. <https://wiki.openstreetmap.org/wiki/Bot>

faudra alors chercher un moyen de faire mieux que les méthodes proposées jusqu'à présent dans la littérature.

6. Conclusion et perspectives

Dans cet article, notre travail s'est attaché à la définition même de ce qu'est le vandalisme et en quoi l'évaluation de ses trois composantes – dommage causé, intentionnalité et contexte – devient un réel défi pour développer sa détection de manière totalement automatique. L'état de l'art, plutôt riche quant à la détection de vandalisme dans les bases de connaissances participatives par approche supervisée, a permis de révéler un manque de recherche sur la détection par classification non supervisée. Les premières expériences de détection par classification non supervisée ont permis de rendre compte de l'importance du choix des descripteurs en entrée de l'algorithme de *clustering*. En effet, ces attributs ne peuvent pas être de simples indicateurs de qualité mais doivent être spécialement conçus pour révéler la particularité des cas de vandalisme. Nos résultats expérimentaux nous ont encouragés à réfléchir à des moyens de réduire le nombre de faux positifs et de faux négatifs. En améliorant nos résultats expérimentaux, nous pourrions alors envisager d'étendre la détection de vandalisme pour répondre à des objectifs plus ambitieux tels que la construction d'un corpus de vandalisme, voire une détection « tout automatique » du vandalisme.

Dans nos travaux futurs, pour sortir du contexte français des données OSM où les bâtiments sont cartographiés par morceaux, nous prévoyons de lancer la même expérience sur un jeu de données provenant d'un autre pays tel que l'Allemagne où les bâtiments sont généralement cartographiés manuellement et en entier. Dans ce nouveau jeu de données, les cas de vandalisme qui n'étaient pas détectés parmi les bâtiments d'Aubervilliers pourraient l'être parmi ceux d'une ville allemande. Comme il a été mentionné précédemment, la détection d'autres types de vandalisme, tels que le vandalisme par édition de tag ou par suppression d'objet, sera aussi étudiée. Dans ce cas, il sera nécessaire d'enrichir les descripteurs de données OSM par d'autres d'attributs. À nouveau, il faudra choisir judicieusement ces descripteurs, en prenant soin d'en limiter le nombre, afin de ne pas se trouver piégé par la malédiction de la grande dimension. Enfin, étant donné le nombre non négligeable de faux positifs détectés avec notre approche, il serait intéressant de se tourner vers une approche supervisée dans le but de réduire le nombre de faux positifs.

Bibliographie

- Adler Alfaro L., Mola-Velasco S.M., Rosso P., West A.G. (2011). Wikipedia vandalism detection: Combining natural language, metadata, and reputation features. In A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing*, vol. 6609, p. 277-288. Berlin, Heidelberg, Springer Berlin Heidelberg. Consulté sur http://dx.doi.org/10.1007/978-3-642-19437-5_23

- Ballatore A. (2014, 1^{er} août). Defacing the map: Cartographic vandalism in the digital commons. *The Cartographic Journal*, vol. 51, n° 3, p. 214-224. Consulté sur <http://dx.doi.org/10.1179/1743277414y.0000000085>
- Barron C., Neis P., Zipf A. (2014, 1^{er} décembre). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, vol. 18, n° 6, p. 877-895. Consulté sur <http://dx.doi.org/10.1111/tgis.12073>
- Chandola V., Banerjee A., Kumar V. (2009, juillet). Anomaly detection: A survey. *ACM Comput. Surv.*, vol. 41, n° 3, p. 15:1-15:58. Consulté sur <http://doi.acm.org/10.1145/1541880.1541882>
- Chin S.C., Street W.N., Srinivasan P., Eichmann D. (2010). Detecting wikipedia vandalism with active learning and statistical language models. In *Proceedings of the 4th workshop on information credibility*, p. 3-10. New York, NY, USA, ACM. Consulté sur <http://dx.doi.org/10.1145/1772938.1772942>
- Ester M., Kriegel H.-P., Jorg S., Xu X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of 2nd international conference on kdd*, p. 226-231. Consulté sur <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.71.1980>
- Fan H., Zipf A., Fu Q., Neis P. (2014, 3 avril). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, vol. 28, n° 4, p. 700-719. Consulté sur <http://dx.doi.org/10.1080/13658816.2013.867495>
- Geiger R.S., Halfaker A. (2013). Using edit sessions to measure participation in wikipedia. In *Proceedings of the 2013 conference on computer supported cooperative work - cscw '13*, p. 861-870. New York, NY, USA, ACM Press. Consulté sur <http://dx.doi.org/10.1145/2441776.2441873>
- Goodchild M.F., Li L. (2012, mars). Assuring the quality of volunteered geographic information. *Spatial Statistics*.
- Heindorf S., Potthast M., Stein B., Engels G. (2015). Towards vandalism detection in knowledge bases. In *Proceedings of the 38th international acm sigir conference on research and development in information retrieval - sigir '15*, p. 831-834. ACM Press. Consulté sur <http://dx.doi.org/10.1145/2766462.2767804>
- Hinneburg A., Gabriel H.H. (2007). DENCLUE 2.0: Fast clustering based on kernel density estimation. In M.R. Berthold, J.S. Taylor, N. Lavrac, M.R. Berthold, J.S. Taylor, N. Lavrac (Eds.), *Ida*, vol. 4723, p. 70-80. Springer. Consulté sur http://dx.doi.org/10.1007/978-3-540-74825-0_7
- Ivanovic S. (2018). *Une approche basée sur la qualité pour mettre à jour les bases de données géographiques de référence à partir de traces GPS issues de la foule*. Thèse de doctorat non publiée, Université Paris Est.
- Keßler C., Groot R., de. (2013). Trust as a proxy measure for the quality of volunteered geographic information in the case of OpenStreetMap. In D. Vandenbroucke, B. Bucher, J. Crompvoets (Eds.), *Geographic information science at the heart of europe*, p. 21-37. Springer International Publishing. Consulté sur http://dx.doi.org/10.1007/978-3-319-00615-4_2
- Landrieu L., Simonovsky M. (2017). *Large-scale point cloud semantic segmentation with superpoint graphs* (vol. abs/1711.09869). Published 2017 in ArXiv.

- Neis P., Goetz M., Zipf A. (2012, 22 novembre). Towards automatic vandalism detection in OpenStreetMap. *ISPRS International Journal of Geo-Information*, vol. 1, n° 3, p. 315-332. Consulté sur <http://dx.doi.org/10.3390/ijgi1030315>
- Noucher M. (2014, septembre). *L'information géographique volontaire : vers un renouvellement de la fabrique cartographique institutionnelle ?*
- Potthast M. (2010). Crowdsourcing a wikipedia vandalism corpus. In *Proceedings of the 33rd international acm sigir conference on research and development in information retrieval*, p. 789-790. New York, NY, USA, ACM. Consulté sur <http://dx.doi.org/10.1145/1835449.1835617>
- Sarabadani A., Halfaker A., Taraborelli D. (2017). Building automated vandalism detection tools for wikidata. In *Proceedings of the 26th international conference on world wide web companion - ww w '17 companion*, p. 1647-1654. ACM Press. Consulté sur <http://dx.doi.org/10.1145/3041021.3053366>
- Touya G., Brando C. (2013). *Detecting Level-of-Detail inconsistencies in volunteered geographic information data sets*, vol. 48, n° 2.
- Tran K.-N.D. (2015). *Detecting vandalism on wikipedia across multiple languages*. Thèse de doctorat non publiée, Australian National University.
- Truong Q.T., De Runz C., Touya G. (2018a). Analysis of collaboration networks in openstreetmap through weighted social multigraph mining. *International Journal of Geographical Information Science*, p. 1-32. Consulté sur <https://doi.org/10.1080/13658816.2018.1556395>
- Truong Q.T., Touya G., De Runz C. (2018b). Towards Vandalism Detection in OpenStreet-Map Through a Data Driven Approach (Short Paper). In S. Winter, A. Griffin, M. Sester (Eds.), *10th international conference on geographic information science (gis science 2018)*, vol. 114. Dagstuhl, Germany, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. Consulté sur <http://drops.dagstuhl.de/opus/volltexte/2018/9389>
- Xu Y., Chen Z., Xie Z., Wu L. (2017, 19 juin). Quality assessment of building footprint data using a deep autoencoder network. *International Journal of Geographical Information Science*, vol. 31, n° 10, p. 1929-1951. Consulté sur <http://dx.doi.org/10.1080/13658816.2017.1341632>
- Zimbardo P.G. (1971). *A Social-Psychological analysis of vandalism: Making sense of senseless violence*. Rapport technique. Stanford University, Department of Psychology.