



HAL
open science

Le vandalisme dans l'information géographique volontaire : apprendre pour mieux détecter ?

Quy Thy Truong, Guillaume Touya, Cyril de Runz

► To cite this version:

Quy Thy Truong, Guillaume Touya, Cyril de Runz. Le vandalisme dans l'information géographique volontaire : apprendre pour mieux détecter ?. Conférence internationale francophone Spatial Analysis and GEomatics (SAGEO), 2018, Montpellier, France. pp.61-76. hal-01924261

HAL Id: hal-01924261

<https://hal.science/hal-01924261v1>

Submitted on 15 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le vandalisme dans l'information géographique volontaire : apprendre pour mieux détecter ?

Quy Thy Truong¹, Guillaume Touya¹, Cyril de Runz²

1. Univ. Paris-Est, LASTIG COGIT, IGN, ENSG, F-94160 Saint-Mande, France

quy-thy.truong@ign.fr; guillaume.touya@ign.fr

2. Modeco, CReSTIC, University of Reims Champagne-Ardenne, CS 30012, Reims cedex 2, France

cyril.de-runz@univ-reims.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 d'explorer les différentes techniques permettant de le détecter. Tout d'abord, nous tâchons d'examiner les différentes définitions du vandalisme, mettant en avant la complexité de cette notion. Puis, nous expérimentons la détection du vandalisme des données OpenStreetMap (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. As vandalism is a serious matter for the quality of Volunteered Geographic Information, this paper aims at exploring learning techniques that enable its detection. First, a focus on the various definitions of vandalism highlights the complexity of this concept. Second, we present an experimental vandalism detection on OpenStreetMap (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

1. Introduction

Le manque de qualification de l'information géographique volontaire constitue aujourd'hui encore un frein pour les consommateurs 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 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 de 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 involontaire. 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 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 clustering est capable de retrouver des cas très simples de vandalisme artificiels, présentant des anomalies géométriques. L'interprétation de nos résultats expérimentaux ouvriront la discussion sur les pistes à envisager pour améliorer la détection.

2. Définitions théoriques et pratiques du vandalisme

2.1. Bref historique

Historiquement, le vandalisme désignait les actes de dégradation perpétrés par la tribu des Vandal, un peuple germanique réputé pour avoir pillé et sac-

cagé les oeuvres d'art et les monuments durant leur invasion de l'Occident. Pendant la période révolutionnaire, l'Abbé Grégoire reprend le terme de vandalisme pour désigner les dommages causés sur les monuments et les ouvrages d'arts 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'il 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).

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-section suivante une exploration et une analyse de différents cas de contributeurs bannis sur OSM. A la lumière de la définition

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>

précédemment dégagée, 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 (Table 1). Ici, la dégradation causée est assez évidente : l'objet géographique a complètement changé de nature puisqu'il 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, Li, 2012; Touya, 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 de doute quant à sa culpabilité. On peut donc conclure que cette contribution relève bien d'un acte de vandalisme.

TABLE 1. Historique des tags de l'objet de la Figure 1

Tag key	Version	1	2
	landuse		retail
name			:)
natural			water
water			lake

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

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

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

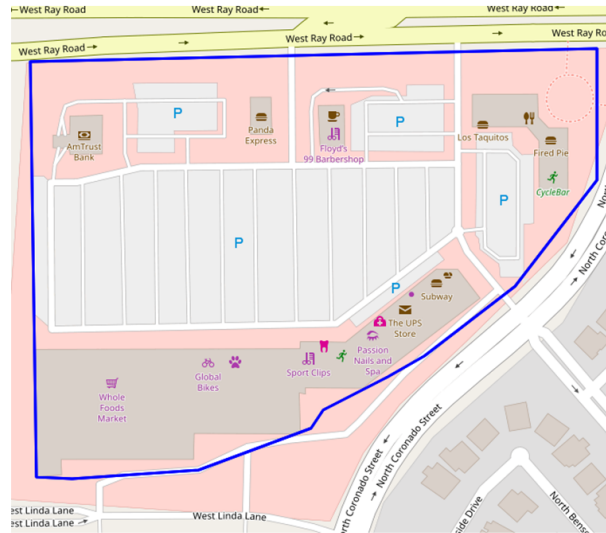


FIGURE 1. Exemple d'un cas de vandalisme : la zone commerciale est transformée en lac (cf. Table 1 pour l'historique de tags de l'objet)

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'un problème de système de coordonnées, et non d'un cas de vandalisme.

Enfin, l'historique des modifications de tags de l'objet de la Table 2 montre un désaccord entre les contributeurs sur la toponymie de la ville. Le contributeur de la version n°32 de l'objet a été banni pour avoir retiré l'orthographe serbe du tag `name`. Après avoir été banni, la version n°32 a été annulée (i.e. la version n°33 correspond à la version n°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 n°32 a donc été banni pour non respect des règles de bonne conduite du projet. Cependant, même si sur la version n°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

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

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

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

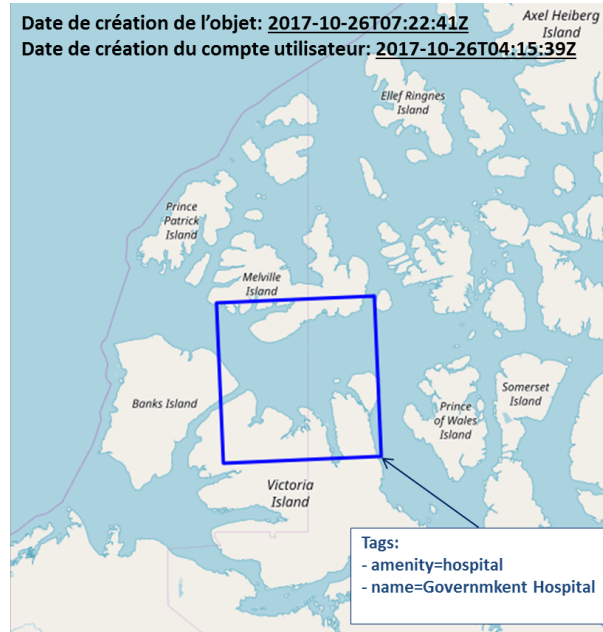


FIGURE 2. Erreur de débutant : cartographie d'un hôpital dont la géométrie chevauche plusieurs îles au Nord du Canada.

qui pourrait justifier l'acte de ce contributeur. Par conséquent, on ne peut pas affirmer que sa contribution dégrade la qualité de l'objet.

TABLE 2. Extrait de l'historique des tags (versions n°28-n°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 bleu.

Version Tag key	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

Ce travail exploratoire a été effectué manuellement, en parcourant sur la page des contributeurs bannis les messages émis par les modérateurs dans lesquels on retrouve parfois 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 obtenue avec l'application OSM Deep History⁸. Bien qu'il soit envisageable de fouiller systématiquement ces données, cela demande un travail conséquent d'analyse automatique de texte qui, selon nous, n'était pas primordial pour étudier les cas de vandalisme OSM. 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 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. Les approches de détection du vandalisme dans les bases de connaissance

Actuellement, l'outil de détection du vandalisme cartographique le plus abouti sur OSM est un système de décision multicritère (Neis *et al.*, 2012). 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 qu'ils soient forcément des cas avérés de vandalisme. 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. 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.

La détection de vandalisme dans les projets de crowdsourcing – tels que Wikipédia et Wikidata – est un sujet largement étudié dans la littérature (Adler *et al.*, 2011 ; Heindorf *et al.*, 2015). Les systèmes de détection proposés utilisent tous des méthodes de classification supervisée. Or, il n'existe pas de corpus de vandalisme pour les données spatiales, c'est pourquoi jusqu'à présent, l'approche supervisée n'a pas encore été envisagée pour détecter le vandalisme dans l'information géographique volontaire. Cette section explore les différentes approches possibles pour détecter le vandalisme dans les bases de connaissance. Tout d'abord, dans la section 3.1 nous analysons les travaux menés sur la détection du vandalisme dans les bases de connaissances (type Wikidata) par

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

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

classification supervisée. Puis nous décrivons dans la section 3.2 notre démarche expérimentale pour détecter le vandalisme par une approche non-supervisée de détection d'anomalie.

3.1. Analyse des corpus de vandalisme

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) suit une stratégie différente pour annoter automatiquement les données labélisées. 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 **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, c'est-à-dire qu'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 forte précision. Tout compte fait, la tolérance aux faux

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

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 les données OSM, par conséquent, nous expérimentons la détection du vandalisme OSM par une méthode non-supervisée et dans le cadre d'une application d'aide à la décision (Table 3).

TABLE 3. *Différentes approches de détection du vandalisme*

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

3.2. *Expérimentation préliminaire*

Malgré la rareté des cas de vandalisme sur OpenStreetMap, le risque de provoquer le vandalisme est, en revanche, loin d'être nul 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.

Puisque les cas de vandalisme sur OSM sont très rares, qu'ils ne sont pas tous détectés et ne sont pas localisables, on tente de lancer une détection d'anomalie par une méthode de clustering sur les données OSM, où les anomalies doivent être des cas de vandalisme OSM à détecter. En effet, comme expliqué dans la partie 2.2., si les contributions vandalisées sont très marginales par rapport à l'ensemble des contributions des utilisateurs bannis, elles le sont encore plus par rapport à l'intégralité de la base cartographique. En outre, elles se distinguent des autres contributions car leur mauvaise qualité les rendent nuisibles à l'utilisation (navigation, analyse de biodiversité, analyse de morphologie urbaine, etc.). Par conséquent, une contribution vandalisée peut être considérée comme une anomalie à détecter dans un jeu de données OSM. Notre expérience vise à détecter automatiquement le vandalisme causé aux bâtiments OSM dans la ville d'Aubervilliers, en banlieue parisienne. Le jeu de données ne contient a priori pas de cas de vandalisme, par conséquent on ajoute manuellement des objets vandalisés (Figures 3 et 4) : 10 bâtiments sont tracés de manière artistique au milieu d'un cours d'eau et du cimetière d'Aubervilliers et 17 bâtiments fictifs de différentes tailles sont ajoutés dans des zones dépourvues de bâtiments (le polygone jaune de la Figure 4 indique que l'espace est actuellement une zone de construction).

Avant de lancer un algorithme de clustering sur le jeu de données ainsi vandalisé, on calcule un certain nombre de descripteurs sur les bâtiments OSM.

TABLE 4. Liste des variables géométriques calculées sur les bâtiments OSM
 N.B. : PPRE signifie Plus Petit Rectangle Englobant.

Nom de la variable	Formule
<i>perimeter_out_of_max</i>	$\frac{\text{perimetre}(\text{bati})}{\max_{\text{dataset}}(\text{perimetre})}$
<i>area_out_of_max</i>	$\frac{\text{aire}(\text{bati})}{\max_{\text{dataset}}(\text{aire})}$
<i>shortest_length_out_of_perimeter</i>	$\frac{\text{length}(\text{shortest_edge}_{\text{bati}})}{\text{perimetre}(\text{bati})}$
<i>median_length_out_of_perimeter</i>	$\frac{\text{length}(\text{median_edge}_{\text{bati}})}{\text{perimetre}(\text{bati})}$
<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>	Miller's index: $\frac{4*\pi*\text{aire}(\text{bati})}{\text{perimetre}(\text{bati})^2}$

paramètre m correspondant au niveau de bruit, l'algorithme commence par construire un modèle de clustering à partir des variables d'entrée, puis prédit pour chaque bâtiment une classe selon le modèle construit. Un bâtiment particulier peut être classé dans un cluster à un élément ou directement comme une anomalie (*outlier*) si ses descripteurs sont complètement incohérents avec le modèle de clustering.

 TABLE 5. Résultats expérimentaux ($\sigma = 0.005, m = 10250$)

	Anomalies	Outlier	Partition à un élément
Vandalisme			
Artistique (total : 10)		1	9
Fictif (total : 17)		0	15
Non-vandalisme (total : 10315)		115	6080

On observe que 25 cas de vandalisme synthétique – artistique ou simplement fictif – sur les 27 ont été détectés : cela représente 92% des cas de vandalisme synthétique, ce qui constitue un résultat plutôt remarquable. En revanche, 60% des bâtiments a priori normaux ont été détectés comme des anomalies (*outliers* ou partitionnés dans un cluster de taille 1).

4. Discussion

Dans le cas d'un outil d'aide à la décision, le système de détection du vandalisme doit minimiser en priorité le nombre de faux négatifs i.e. les données vandalisées qui ne sont pas labélisées comme telles. En cela, nos premiers résultats sont plutôt satisfaisants puisque notre expérience détecte moins de 8% de faux négatifs. Parmi ces faux négatifs, un minuscule bâtiment fictif 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. Afin de distinguer les petits morceaux du cadastre de cas réels de vandalisme, il faudrait, dans l'idéal, faire en sorte d'utiliser des attributs géométriques qui ne fassent pas ressortir les spécificités géométriques des objets cartographiés. Ici, l'attribut *area_out_of_max* n'était peut-être pas pertinent pour détecter les anomalies. De la même manière, il aurait peut-être fallu considérer une variable de densité de polygone comptant le nombre de sommets dans chaque bâtiment pour détecter tous les vandalismes artistiques.

Les variables d'entrée du clustering ne contenaient aucun attribut topologique qui aurait pu indiquer la présence de bâtiments vandalisés tracés dans un cours d'eau, dans une zone de construction et sur un cimetière. L'ajout de variables topologiques permettrait de détecter les bâtiments situés à des endroits inhabituels. D'autres descripteurs de type historiques, sémantiques et contributeurs auraient pu être ajoutés en entrée de l'algorithme, mais ceux-ci n'ont pas été implémentés. Dans un premier temps, nous cherchions à voir si les erreurs géométriques, dont nos actes de vandalisme, pouvaient être détectées comme des anomalies à partir de descripteurs purement géométriques.

Le cas d'étude présenté peut sembler simpliste puisque les cas de vandalisme synthétiques à détecter sont uniquement de type artistique et fantaisiste. Nous n'avons pas inséré dans le jeu de données OSM tous les types de vandalisme mentionnés dans la section 2.1 car l'objectif principal n'était pas de parvenir immédiatement à détecter tous les cas de vandalisme existants, mais plutôt d'évaluer la capacité d'un algorithme de détection d'anomalie par clustering à retrouver les cas de vandalisme artificiels, aussi simples soient-ils. Enfin, on ne s'attendait pas à détecter absolument tous les cas de vandalisme uniquement grâce à un partitionnement sur des variables géométriques. Ces premiers résultats sont donc plutôt encourageants pour la suite de nos recherches.

Par ailleurs, 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 peut 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 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 clas-

sification non-supervisée pourra éventuellement s'étendre à une application de détection entièrement automatisée, mais cela ne pourra être envisagé que si l'on arrive à réduire drastiquement le nombre de faux positifs.

L'utilisation d'un algorithme d'apprentissage non-supervisé pour détecter avec précision le vandalisme semble être *a priori* une piste bien optimiste, surtout avec une définition où l'évaluation de l'intentionnalité, du dommage causé et du contexte, permet de le distinguer de l'erreur involontaire. Notre approche cherche à valider ou non l'hypothèse selon laquelle les contributions vandalisées présentent des caractéristiques aberrantes dans l'espace des données. En d'autres termes, on émet l'hypothèse que les cas de vandalisme sont des anomalies à détecter. La vérification de cette hypothèse s'effectue donc en choisissant les descripteurs les plus représentatifs de l'ensemble du jeu de données, de manière à ce que les contributions anormales ressortent à travers les valeurs de ces descripteurs. L'aspect intentionnel d'une erreur de saisie pourrait se modéliser par des descripteurs portant sur les contributeurs, mais ce modèle doit être vérifié expérimentalement.

La détection du vandalisme des données OSM par une classification supervisée nécessite de construire un corpus de contributions annotées. 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, Groot, 2013; Truong *et al.*, 2018). Ce corpus de données annotées pourra contenir des cas de vandalisme synthétique similaires à ce que nous avons produits manuellement (Figures 3 et 4) mais également envisager de provoquer d'autres types de vandalisme synthétique. 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, Li, 2012). Ce genre de vandalisme fictif mais réaliste pourrait être détecté en changeant d'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, Simonovsky, 2017).

Enfin, l'éventualité de passer à une détection totalement automatique du vandalisme OSM par classification supervisée est encore loin. 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 de faux négatifs soit nul. Dans ce cadre-là, l'annotation automatique proposée par Sarabadani

et al. n'est pas assez satisfaisante pour construire un tel corpus. Il faudra alors chercher un moyen de faire mieux que les méthodes proposées jusqu'à présent dans la littérature.

5. 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 automatique. L'état de l'art, plutôt riche quant à la détection de vandalisme des 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, il s'agira d'améliorer les résultats de notre expérience initiale, en faisant varier dans un premier temps l'influence des paramètres σ et m de l'algorithme DENCLUE. Dans un second temps, on pourra tester d'autres algorithmes de clustering, tels que DBScan ou BIRCH, pour vérifier si ceux-ci sont plus efficaces pour détecter le vandalisme. De plus, pour sortir du contexte français des données OSM où les bâtiments sont cartographiés par morceaux, on prévoit 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. 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 sera aussi étudiée, telle que le vandalisme par édition de tag ou par suppression d'objet. Dans ce cas, il sera nécessaire d'enrichir les descripteurs de données OSM par d'autres types d'attributs (historique, topologique, contributeur, sémantique) et lorsque ce sera possible, des indicateurs d'appariement avec des données de référence. A 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, à partir du moment où les résultats expérimentaux présenteront beaucoup moins de faux positifs, la construction d'un corpus de données de vandalisme des données OSM pourra être envisagée, en poursuivant la réflexion sur les différents points abordés dans la discussion de cet article. Ce

corpus pourra alors être utilisé pour détecter le vandalisme des données OSM par une approche supervisée.

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>
- Goodchild M. F., Li L. (2012, 28 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. Cromptvoets (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 - www '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.

Truong Q.-T., Touya G., Runz C. de. (2018). Building social networks in volunteered geographic information communities: What contributor behaviours reveal about crowdsourced data quality. In P. Fogliaroni, A. Ballatore, E. Clementini (Eds.), *Proceedings of workshops and posters at the 13th international conference on spatial information theory (cosit 2017)*, p. 125–131. Springer International Publishing.
Consulté sur http://dx.doi.org/10.1007/978-3-319-63946-8_25

Zimbardo P. G. (1971). *A Social-Psychological analysis of vandalism: Making sense of senseless violence*. Rapport technique. Stanford University, Department of Psychology.