

## Les médias sociaux comme prédicteurs de la criminalité urbaine

### Social media as a predictor of urban crime

### Las redes sociales como predictoros de la criminalidad urbana

Simon Da Silva, Rémi Boivin et Francis Fortin

Volume 52, numéro 2, automne 2019

La criminologie de l'information : état des lieux et perspectives

URI : <https://id.erudit.org/iderudit/1065857ar>

DOI : <https://doi.org/10.7202/1065857ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0316-0041 (imprimé)

1492-1367 (numérique)

[Découvrir la revue](#)

Citer cet article

Da Silva, S., Boivin, R. & Fortin, F. (2019). Les médias sociaux comme prédicteurs de la criminalité urbaine. *Criminologie*, 52(2), 83–109. <https://doi.org/10.7202/1065857ar>

Résumé de l'article

La présente étude tente de déterminer l'importance d'analyser les crimes à des niveaux spatiaux et temporels de plus en plus précis. De même, une nouvelle source de données issue des médias sociaux, les messages sur Twitter, est utilisée afin de prédire la répartition des crimes à Montréal en estimant la population réelle sur le territoire, et en la caractérisant selon son humeur. Des modèles multiniveaux Poisson sont utilisés afin de prédire les crimes contre la personne et les crimes contre les biens agrégés au segment de rue selon l'heure de la journée. Les résultats montrent qu'il est primordial pour toute analyse de la criminalité à Montréal de tenir compte de la variance de la criminalité en ce qui a trait aux micro-endroits et d'y incorporer des périodes intrajournalières. La caractérisation de la population réelle de la ville a été considérée comme une avenue prometteuse pour la prédiction des crimes. Cette étude propose que l'utilisation des données de Twitter soit une avenue d'analyse concluante, mais qui reste encore à approfondir.

# Les médias sociaux comme prédicteurs de la criminalité urbaine

Simon Da Silva<sup>1</sup>

*Étudiant à la maîtrise*

*École de criminologie, Université de Montréal*

*simon\_dasilva\_@hotmail.com*

Rémi Boivin

*Professeur agrégé*

*École de criminologie, Université de Montréal*

*remi.boivin@umontreal.ca*

Francis Fortin

*Professeur agrégé*

*École de criminologie, Université de Montréal*

*francis.fortin@umontreal.ca*

**RÉSUMÉ** • La présente étude tente de déterminer l'importance d'analyser les crimes à des niveaux spatiaux et temporels de plus en plus précis. De même, une nouvelle source de données issue des médias sociaux, les messages sur Twitter, est utilisée afin de prédire la répartition des crimes à Montréal en estimant la population réelle sur le territoire, et en la caractérisant selon son humeur. Des modèles multiniveaux Poisson sont utilisés afin de prédire les crimes contre la personne et les crimes contre les biens agrégés au segment de rue selon l'heure de la journée. Les résultats montrent qu'il est primordial pour toute analyse de la criminalité à Montréal de tenir compte de la variance de la criminalité en ce qui a trait aux micro-endoirns et d'y incorporer des périodes intrajournalières. La caractérisation de la population réelle de la ville a été considérée comme une avenue prometteuse pour la prédiction des crimes. Cette étude propose que l'utilisation des données de Twitter soit une avenue d'analyse concluante, mais qui reste encore à approfondir.

**MOTS CLÉS** • Analyse de la criminalité, population flottante, analyse de l'humeur, réseaux sociaux, analyse aoristique.

---

1. École de criminologie, Université de Montréal, Pavillon Lionel-Groulx, C. P. 6128, succ. Centre-ville, Montréal (Québec), Canada, H3C 3J7.

## Introduction

La criminologie des endroits (*criminology of place*) est une branche de la criminologie qui « nous pousse à examiner et à comprendre pourquoi le crime se produit à un endroit spécifique plutôt que de concentrer nos intérêts sur les préoccupations plus traditionnelles des criminologues, soit les raisons pour lesquelles certains types spécifiques de personnes commettent des crimes » (Weisburd, 2015, p. 134, notre traduction). Weisburd (2015) indique qu'il s'agit d'une branche de la criminologie qui reçoit assez peu d'attention malgré les implications évidentes de telles études. L'auteur a analysé les études empiriques publiées dans la revue *Criminology* entre 1990 et 2014. Bien que la proportion d'articles s'intéressant aux endroits microgéographiques soit passée de 2,6 % au début des années 1990 à plus de 6 % entre 2010 et 2014, celle-ci demeure très faible. Pourtant, cette branche de la criminologie a déjà démontré l'impact significatif du *hotspot policing* (voir Eck, Chainey, Cameron, Leitner et Wilson, 2005) en matière de prévention du crime et relance la pertinence des théories qui étaient pratiquement tenues pour acquises.

Dans les prochaines sections, la loi de la concentration sera présentée ainsi que son implication pour le *hotspot policing*. Différentes stratégies visant à rendre le *hotspot policing* le plus efficace possible vont être explorées. De plus, la volonté d'étudier la répartition de la criminalité à très petite échelle de temps et d'espace va de pair avec le défi de recueillir des données permettant de le faire; cet article explorera l'utilité d'une source de telles données, le réseau social Twitter.

### *La loi de la concentration*

La criminologie des endroits est basée sur la proposition que la criminalité n'est pas répartie au hasard, mais plutôt qu'elle suit un modèle qu'il est possible d'étudier et de comprendre; le crime se concentrerait dans l'espace et, par extension, dans le temps, en fonction de facteurs identifiables. Les premières études s'intéressant au phénomène ont analysé la provenance (l'adresse) des appels d'urgence. Sherman, Gartin et Buerger (1989) ont trouvé que 3,5 % des adresses à Minneapolis produisaient 50 % des appels d'urgence pour une année. Du côté du Royaume-Uni, on a trouvé qu'aussi peu que 1 % des résidences étaient victimes de 42 % des entrées par effraction (Budd, 2001).

La loi de la concentration de la criminalité a également été testée sur le plan du segment de rue, et ce, dans plusieurs villes du monde. Par exemple, Andresen et Malleson (2011) ont observé qu'à Vancouver, environ 50% des vols de voiture ont lieu dans 5% des segments de rue de la ville. Weisburd (2015) a effectué une analyse de la concentration de différents types de crimes dans plusieurs villes aux États-Unis et en Israël, qui a permis de relever qu'entre 0,4% et 1,6% des segments de rues étaient responsables de 25% des crimes. Il a de plus analysé les changements temporels des concentrations pour des périodes de 10 ans et plus et est arrivé à la conclusion que, malgré des fluctuations significatives quant au nombre de crimes, la concentration de ceux-ci restait stable au fil des années. Curman, Andresen et Brantingham (2015) ont confirmé la stabilité temporelle (pour une période de 16 ans) de la concentration des événements criminels à Vancouver. Braga, Papachristos et Hureau (2010) ont analysé la violence avec arme à feu de 1980 à 2008 et ont observé une grande stabilité de la concentration de ces événements criminels.

Une définition de la loi de la concentration est maintenant devenue une référence pour les criminologues: «Pour une mesure définie du crime d'une unité microgéographique particulière, la concentration de la criminalité sera caractérisée par un pourcentage petit et restreint pour une proportion cumulative de crimes» (Weisburd, 2015 p. 135, notre traduction). Autrement dit, la majorité des crimes est concentrée à une minorité d'endroits.

On peut facilement penser aux implications pratiques d'une telle loi. En effet, les endroits où les crimes sont concentrés représentent des cibles parfaites pour les activités de prévention (Rosser, Davies, Bowers, Johnson et Cheng, 2017). Si le crime est concentré à des endroits spécifiques, il est naturel que les activités policières et autres activités de prévention y soient aussi concentrées (Weisburd, 2015).

### Le hotspot policing

La concentration de la criminalité dans un petit espace est souvent qualifiée de *hotspot* (point chaud) (Weisburd et Telep, 2014) et a donné naissance au *hotspot policing*, soit l'activité policière basée sur des interventions qui ciblent les endroits où les crimes se concentrent (Ratcliffe, 2004). L'efficacité du *hotspot policing* n'est plus à débattre: Weisburd et Telep (2014) affirment que de démontrer l'efficacité du *hotspot policing*

n'est plus suffisant pour se faire publier dans une revue scientifique. Il existe tout de même différentes stratégies pour s'attaquer aux *hotspots* de criminalité, et certaines sont plus efficaces que d'autres.

À la base, il y a la stratégie de la simple présence. Une patrouille peut tout simplement être présente au point chaud et le nombre de crimes et d'incidents diminuera (Sherman et Weisburd, 1995). Bien qu'efficace, cette manière de faire ne fait pas le poids contre une approche axée sur la résolution de problème, qui entraînerait des réductions de la criminalité plus de deux fois plus importantes (Braga, Papachristos et Hureau, 2012). L'approche par résolution de problème se caractérise par la détermination et l'analyse des circonstances sous-jacentes aux problèmes criminels (Eck et Spelman, 1987). Elle est considérée comme l'approche de *hotspot policing* la plus efficace pour le moment, bien que d'autres avenues prometteuses commencent à être explorées, telles que la concentration des efforts sur les délinquants connus (Ratcliffe, Groff, Haberman et Sorg, 2012).

Il est important de se rappeler que l'efficacité du *hotspot policing* a seulement été prouvée dans des grandes villes. Une des grandes limites de l'approche est que l'on connaît mal la modélisation de la loi de la concentration en milieu plus rural et, du même coup, de l'effet que pourrait avoir une activité policière plus concentrée dans ce contexte (Weisburd et Telep, 2014). Aussi, une grande limite des études s'intéressant au *hotspot policing* est qu'elles tiennent rarement compte de la temporalité « fine » des points chauds de crimes (Ratcliffe, 2004).

### *La temporalité des hotspots*

Bien que la stabilité temporelle de la concentration de la criminalité ait été prouvée sur le plan des années et sur celui des villes, les choses se compliquent lorsqu'on étudie une temporalité plus « fine ». Haberman, Sorg et Ratcliffe (2017) ont analysé la répartition des vols de rue à Philadelphie selon l'heure de la journée, la journée de la semaine et la saison. Bien que la loi de la concentration ait tenu, certains micro-endroits étaient propices aux vols en tout temps, alors que d'autres l'étaient seulement à certaines périodes. Ratcliffe (2004) incorpore la temporalité à la définition de *hotspot* et distingue trois types de *hotspots* selon leur comportement temporel au cours d'une journée :

**Points chauds diffus :** ce sont des points chauds où des événements criminels pourraient survenir à tout moment de la journée.

**Points chauds concentrés:** ce sont des points chauds où des événements criminels pourraient survenir à tout moment de la journée, mais il y a des moments où il y a plus de chances que cela arrive.

**Points chauds aigus:** ce sont des points chauds qui sont actifs à certains moments précis de la journée. Il est rare qu'un crime soit commis hors de ces moments précis.

Il est important de tenir compte de la temporalité pour les corps policiers qui utilisent le *hotspot policing*. En effet, il n'est d'aucune utilité de poster une patrouille en pleine nuit à un endroit chaud où la majorité des crimes sont commis au milieu de la journée (Ratcliffe, 2004). La manière la plus facile de représenter la temporalité à l'intérieur d'une journée, par son accessibilité dans les données policières, est d'utiliser le quart de travail des policiers, soit le jour, le soir et la nuit. Elle contient toutefois deux limites importantes. Premièrement, un crime enregistré à un certain quart n'a pas nécessairement été commis durant cette période. Par exemple, un citoyen qui entre chez lui le soir et se rend compte que quelqu'un s'y est introduit pendant son absence va appeler la police et celle-ci va enregistrer l'entrée par effraction dans le quart de soir. Ainsi, la temporalité du crime est basée sur l'enregistrement de celui-ci plutôt que sa commission, ce qui n'est pas souhaitable. Deuxièmement, les quarts de travail représentent souvent des périodes de huit heures, et il est possible de se demander si un *hotspot* identifié au quart de travail près ne pourrait pas être désagrégé à l'heure près et démontrer des concentrations différentes à l'intérieur de ces huit heures. Koper (1995) a démontré qu'une présence policière de 15 minutes à un *hotspot* de crime est le temps idéal de patrouille; on peut donc penser que d'identifier un point « critique » d'un *hotspot* précis – plus précis qu'un quart de huit heures – pourrait s'avérer intéressant pour les forces de l'ordre.

D'autres auteurs ont tenté d'utiliser des temporalités autres que le quart de travail (Haberman et Ratcliffe, 2015; Haberman *et al.*, 2017). Haberman et Ratcliffe (2015) ont analysé l'*American Time Use Survey*, qui présente les activités de l'États-Unien typique durant la journée, et ont divisé celle-ci en quatre périodes non égales qui reflètent les activités routinières de la population (6:45 à 9:59, 10:00 à 16:29, 16:30 à 21:14 et 21:15 à 6:44). L'utilisation de cette technique comporte deux limites importantes. Premièrement, elle néglige les variations entre les individus. Par exemple, certains travaillent en journée et d'autres le soir ou la nuit; il y a lieu de se demander à quel point une seule routine

représente l'ensemble des citoyens. Deuxièmement, il est difficile de savoir si ces activités sont généralisables à d'autres populations que les États-Unis. Récemment, Wheeler et Haberman (2018) ont analysé les voies de fait et les braquages à l'heure de la journée près de Seattle, et ont pu apprécier des fluctuations propres à certaines heures précises. Cela nous pousse à croire que d'analyser la criminalité à une résolution temporelle très précise permet de faire ressortir des patrons intéressants et jusqu'ici inexplorés.

En somme, l'intervention basée sur le *hotspot policing* paraît d'abord simple et efficace, mais l'intégration de la temporalité dans la compréhension complique l'analyse. De plus, si l'on veut appliquer l'approche par résolution de problème, il faut être en mesure de relever des facteurs associés à la criminalité.

#### *Expliquer les crimes sur le plan des micro-endoits et Twitter*

En ouverture de leur article, Haberman *et al.* (2017) affirment qu'il y a un manque de littérature scientifique sur le profil spatiotemporel des crimes, et plus précisément sur les variables indépendantes qui pourraient prédire les concentrations spatiales des crimes sur le plan micro et à différentes échelles de temporalité. C'est un obstacle important au développement d'interventions et de programmes sur le plan des micro-endoits; les facteurs explicatifs ont peu été étudiés.

Ce manque de variables explicatives sur le plan de l'agrégation microscopique pousse les chercheurs à en développer. Récemment, certains auteurs se sont intéressés aux données issues des médias sociaux. Plus précisément, les messages envoyés sur Twitter (les tweets) ont fait l'objet d'analyses diverses afin de dégager des informations pertinentes à la prédiction et à l'explication des crimes dans les grandes villes. Certains auteurs ont affirmé que les messages pouvaient « renforcer l'explication de l'activité criminelle dans les zones urbaines » (Bendler, Brandt, Wagner et Neumann, 2014, p. 1, notre traduction) et ont souligné leur importance dans la prédiction des voies de fait, des vols et des troubles à la paix (Bendler *et al.*, 2014) et des risques de victimisation violente (Malleon et Andresen, 2015a). Les messages sur Twitter sont disponibles en sources ouvertes, et chaque tweet, si les options de géolocalisation sont activées, est relié à des données spatiotemporelles extrêmement précises. De plus, les contenus partagés y sont différents des autres médias sociaux comme Facebook ou Snapchat

reconnus pour des messages plus personnels et privés, empêchant ainsi des analyses plus macroscopiques ou sociologiques comme sur Twitter (marketing, politique, criminalité, etc.).

Plusieurs études ont utilisé les médias sociaux dans un contexte de prédiction de la criminalité. D'abord, une première approche fut celle de relever les crimes ponctuels, mais impliquant plusieurs individus grâce aux médias sociaux. Wang, Gerber et Brown (2012) ont démontré que les délits de fuite à partir du flux Twitter pour la ville de Charlottesville pouvaient être prédits à l'intérieur d'un certain intervalle de confiance. On explore également la détection et de la description de certains phénomènes sociaux et criminels comme les tweets de cyberhaine (Awan, 2014), le racisme en ligne (Weaver, 2013) et les tensions sociales (Burnap *et al.*, 2015). Ensuite, l'approche de la prédiction d'événements a été utilisée pour saisir les variations dans les incidents de sécurité publique. Cette approche se distingue de la première puisqu'elle vise à prédire un événement à partir d'une agrégation de messages sur les médias sociaux. Par exemple, Gu, Qian et Chen (2016) ont tenté de détecter les incidents de route tels que les accidents, les blocages de circulation et les travaux de voirie dans les villes de Pittsburgh et de Philadelphie. Ils ont trouvé que l'échantillon de tweets obtenu de Twitter couvrait la plupart des incidents réels qui eurent lieu dans ces villes. Finalement, l'étude des signaux permet d'émettre l'hypothèse que «les médias sociaux fournissent des signaux socio-comportementaux pour la prédiction du crime» (Aghababaei et Makrehchi, 2016, p. 526). Ainsi, Twitter permettrait de déceler des variables prédictives pouvant influencer des changements dans les taux de criminalité. Gerber (2014) a tenté de prédire les incidents criminels à Chicago en utilisant les tweets pour les comparer aux points chauds des crimes relevés durant la même période. Pour les 25 types de crime, la modélisation a permis d'améliorer la prédiction pour 19 types de crime, les crimes liés au vol et au trafic de drogues ayant obtenu la meilleure performance. Bendler, Ratku et Neumann (2014) ont cartographié les tweets partagés à San Francisco selon les quartiers de la ville en fonction du type d'activité (zone de restaurants, centre commercial, terrains de jeu, etc.). En comparant le volume de tweets avec les crimes de chacun des quartiers, les auteurs ont déterminé quatre types de crimes pouvant être prédits à partir du flux Twitter avec un niveau de confiance acceptable: les vols par effraction, les vols de véhicules à moteur, les vols qualifiés et les vols mineurs.



Les premiers éléments analysables des tweets sont leurs coordonnées spatiales (Malleson et Andresen, 2016). En enregistrant où les tweets sont envoyés sur un territoire, il est possible de créer une distribution des tweets sur celui-ci, qui peut être utilisée comme *proxy* pour la distribution de la population; la logique étant que plus il y a de tweets envoyés à un endroit, plus on s'attend à ce qu'il y ait de personnes présentes à celui-ci (Malleson et Andresen, 2015a, 2015b).

On qualifie ces personnes présentes de population flottante (*ambient population*), soit le nombre de personnes présentes sur un territoire, sans nécessairement y résider (Boivin, 2013). Une difficulté rencontrée par les chercheurs est de générer des informations permettant de déterminer une population flottante pour un territoire. Certains auteurs, tels Felson et Boivin (2015) avec les sondages de déplacement, ont estimé la population flottante de grandes villes pour une journée typique. L'avantage des tweets est de pouvoir tenir compte des fluctuations intrajournalières de cette population, chose qui n'est pas possible avec les autres sources de données. À ce titre, ils constituent une forme de capteur de phénomènes sociaux (voir Williams, Burnap et Sloan, 2015).

Le contenu de tweets est une autre composante qui peut s'avérer intéressante. L'analyse de l'humeur issue de tweets a été tentée dans d'autres disciplines; par exemple pour prédire les résultats d'élections (Birmingham et Smeaton, 2011) ou les fluctuations des marchés boursiers (Bollen, Mao et Zeng, 2011). Malgré les différentes limites de l'analyse de sentiment tant sur le plan syntaxique (les fautes d'orthographe, l'invention de mots, l'utilisation de symboles, de structures syntaxiques inconnues, etc. – voir Eisenstein [2013]) que sémantique (sarcasme, significations spécifiques, etc.), les classificateurs ont le potentiel d'être performants pour l'analyse de contenu de tweets sur de grands ensembles (Burnap et Williams, 2015). Par exemple, en analysant des messages, Chen, Cho et Jang (2015) ont défini l'humeur selon une échelle de polarité, -1 reflétant une humeur très négative et +1 une humeur très positive.

Le but de la présente étude est d'utiliser les tweets afin de générer des facteurs explicatifs de la criminalité urbaine. Premièrement, il sera possible d'utiliser la géolocalisation de ceux-ci afin d'estimer une population flottante très précise, autant sur le plan spatial que temporel. Deuxièmement, l'analyse du contenu des tweets va permettre de faire ressortir l'humeur de ceux-ci, permettant ainsi de caractériser la population flottante.

## Méthodologie

### *Données utilisées*

Les données de crime proviennent de la base de données interne du Service de police de la Ville de Montréal (SPVM). Tous les crimes contre la personne et contre la propriété enregistrés entre 2011 et 2017 à Montréal sont utilisés pour l'étude. Au total, on dénombre 415 709 crimes, dont 99 622 contre la personne et 316 087 contre la propriété.

Les tweets renferment trois types de données essentiels aux analyses, soit des coordonnées géographiques, un temps d'envoi et des messages rattachés. Ceux-ci sont collectés à même le site de Twitter, et l'ensemble des tweets envoyés à l'échelle de la ville, de décembre 2016 à février 2017, ainsi qu'au mois d'août 2017, ont été enregistrés. Au total, on dénombre 499 510 tweets. Seuls 12,69 % ont pu être géoréférencés précisément et utilisés pour l'étude ( $n = 63\,392$ ). Cela est malgré tout supérieur aux autres études, qui présentent des proportions oscillant entre 1 et 5 % (Malleon et Andresen, 2015a). Il aurait été préférable que la période de crimes étudiée concorde avec celle des tweets. Par contre, comme il sera présenté dans cette section, l'analyse spatiotemporelle très fine de la criminalité divise les variables dépendantes en de très nombreuses unités d'analyses et un certain volume est nécessaire pour mener à bien les analyses. Le volume ne serait pas suffisant sur le plan des crimes, c'est la raison pour laquelle sept années de crimes ont été colligées. Cet anachronisme ne devrait pas poser problème, selon l'étude de Weisburd, Groff et Yang (2012), qui a montré que la configuration de crimes sur le plan des segments de rue était stable dans le temps.

Les deux premières données, soit les coordonnées spatiales et le temps d'envoi, sont des champs inclus dans les tweets collectés. Pour ce qui est des messages rattachés, on cherche à analyser l'humeur qui ressort de ceux-ci. Pour ce faire, le logiciel *R* a été utilisé, avec la librairie *sentiment*. Il est possible d'analyser chaque *tweet* et d'obtenir, à l'aide d'un algorithme, l'humeur générale de ceux-ci. En faisant suite aux études sur l'humeur, les tweets sont catégorisés selon deux polarités opposées, une positive et l'autre négative. Une démarche plus exploratoire va aussi être tentée. Ainsi, la librairie permet aussi de classifier l'humeur selon six catégories distinctes. Les *tweets* peuvent refléter de la colère, du dégoût, de la peur, de la joie, de la tristesse ou de la surprise.

La librairie *sentiment* permet les deux types d'analyse d'humeur à l'aide de lexiques contenant plusieurs milliers de mots rattachés à une polarité et à une catégorie d'humeur, mais seulement en anglais. Cela pose problème, car les tweets récoltés pour l'étude contiennent aussi des tweets en français. Pour remédier à ce problème, des lexiques anglophones conçus par le Conseil national de recherches du Canada (CNRC) (Mohammad et Turney, 2013) ont été introduits manuellement dans la librairie *R*, ainsi que leur traduction francophone (Abdaoui, Azé, Bringay et Poncet, 2016). Au final, les tweets anglophones, francophones et bilingues ont pu être analysés.

Les données issues des variables à l'étude sont agrégées sur le plan spatial et sur le plan temporel. Sur le plan spatial, les segments de rue ont été utilisés ( $n = 45\ 099$ ). Comme il a été expliqué plus tôt, les crimes sont concentrés aux micro-endroits, et donc les *hotspots* de crimes sont très locaux (Weisburd, 2015). Comme l'ont proposé Wheeler et Haberman (2018), les crimes de cette étude ont été divisés selon l'heure de la journée ( $n = 24$ ). Toutefois, contrairement à ces auteurs, qui ont étudié deux types de crimes (voies de fait et braquages) dont l'heure exacte de commission est souvent connue, cette étude s'intéresse à plusieurs types de crimes, dont certains sont parfois moins propices à avoir une heure de commission facilement identifiable. Pour remédier à ce problème, une prédiction de l'heure réelle de commission des délits à l'intérieur de l'intervalle *heure de début/heure de fin* va être calculée à l'aide de l'analyse dite *aoristique*. Elle consiste à diviser la probabilité qu'un crime soit commis par le nombre d'heures incluses dans l'intervalle début/fin, à faire la somme de ces probabilités pour tous les crimes commis et à créer une pondération temporelle pour une unité spatiale (Ratcliffe, 2000, 2002). Une courbe aoristique est donnée pour chaque catégorie de crime dans chaque unité géographique (ici le segment de rue) et permet de prédire l'heure de commission d'un délit avec l'intervalle fourni par les données policières. Ashby et Bowers (2013) ont comparé diverses méthodes de prédiction de l'heure de commission des délits et la méthode aoristique est celle qui offre la meilleure prédiction. Les crimes dont l'intervalle entre l'heure de début et l'heure de fin dépasse les 24 heures ont été retirés de la banque de données, car cela pourrait fausser la pondération donnée de l'analyse (voir Ratcliffe [2000] pour plus de détails). L'unité d'analyse est donc le segment de rue \* l'heure de la journée. Cela permet d'avoir un portrait précis de la criminalité, tant spatial que temporel.

### *Stratégie d'analyse*

L'analyse multivariée des crimes sera effectuée à l'aide de modèles Poisson, plutôt qu'avec les modèles linéaires couramment utilisés. Avec près de 1,1 million d'unités d'analyses (45 099 segments de rue \* 24 heures de la journée), plusieurs de celles-ci ne contiennent pas de crime et, en général, leur valeur est très petite. Ainsi, la distribution des deux types de crimes est caractérisée par une asymétrie positive, un mode à zéro et de petites valeurs, et est donc mieux prédite par un modèle de type Poisson (Osgood, 2000).

Une approche multiniveau, soit tenant compte d'un deuxième niveau – les segments de rue – regroupant les unités d'analyses, a été préférée pour les analyses, et ce, pour deux raisons principales. Premièrement, l'approche multiniveau permet de respecter le postulat d'indépendance des unités d'analyses. Ainsi, les segments de rue selon l'heure de la journée ne sont pas complètement indépendants entre eux. Certaines unités d'analyses sont liées au même segment de rue, et ce lien peut expliquer le nombre de crimes commis. En introduisant le segment de rue comme deuxième niveau (soit qui regroupe des unités du premier niveau), on tient compte de la non-indépendance des unités du premier niveau qui appartiennent au même segment.

Deuxièmement, l'utilisation d'un modèle multiniveau permet d'avoir un portrait plus complet de la criminalité. À l'aide du modèle des composantes de la variance, on note que pour les crimes contre la propriété, la majorité de la variance (60,95 %) totale se trouve au niveau des segments de rue, mais qu'une part non négligeable est attribuée aux segments de rue selon l'heure de la journée (39,05 %). Le constat est semblable pour les crimes contre la personne, avec des pourcentages de 59,16 % et de 40,84 %, respectivement. Il est ainsi possible de conclure que pour avoir un portrait complet du phénomène criminel au niveau des segments de rue à Montréal, il est primordial de tenir compte du temps de la journée. Aussi, les parts de variances sont départagées entre les deux niveaux hiérarchiques, ce qui confirme l'intérêt d'utiliser un modèle multiniveau. L'analyse de l'effet des tweets est produite au premier niveau. D'autres prédicteurs sont introduits au deuxième niveau afin d'avoir un portrait plus complet de la criminalité de la ville.

*Variables indépendantes au deuxième niveau*

## La désorganisation sociale

Des prédicteurs issus de la théorie de la désorganisation sociale sont utilisés. Cette théorie a d'abord été introduite par Shaw et McKay (1942), et a comme prémisse que les facteurs sociaux d'une communauté urbaine peuvent perturber la structure et l'organisation de celle-ci et du même coup influencer la criminalité. Trois facteurs ont été mis de l'avant : l'hétérogénéité ethnique, le revenu et la mobilité résidentielle. Dans cette étude, les familles monoparentales sont ajoutées, et sont liées à un faible degré de supervision parentale et un plus haut niveau de désavantage social (Boivin et Ouellet, 2011).

Les données sociodémographiques liées à la théorie de la désorganisation sociale ont été colligées à l'aide des données issues du recensement de 2016 de Statistique Canada. Pour chaque secteur de recensement ( $n = 530$ ) de Montréal, soit des petites zones géographiques qui comptent une population entre 2500 et 8000 habitants (Statistique Canada, 2018), les quatre facteurs ont été modélisés. On inclut donc le pourcentage de familles monoparentales, le pourcentage de la population sous le seuil de la pauvreté, le pourcentage de minorités visibles et le pourcentage de personnes ayant déménagé au cours des cinq années précédentes (mobilité résidentielle). La population résidente est aussi introduite comme une variable indépendante et est elle aussi issue du recensement de 2016 de Statistique Canada.

Il est à noter que les variables du recensement utilisées ne sont disponibles qu'au secteur de recensement près. Afin de les introduire dans le deuxième niveau de notre modèle, soit le segment de rue, les données sont désagrégées. Ainsi, tous les segments de rue appartenant au même secteur de recensement ont les mêmes valeurs en ce qui a trait aux variables de désorganisation sociale.

## Le lag spatial

Une dernière variable est ajoutée au deuxième niveau, soit le lag spatial. Introduit par Anselin (1988), le lag spatial représente la moyenne des valeurs avoisinantes pour une valeur spatiale donnée. Ainsi, il est souvent constaté que des variables définies spatialement sont corrélées entre elles, et donc qu'elles ne respectent pas le postulat d'indépendance. Ici, on se demande si les crimes étudiés sont corrélés en ce qui a trait au segment de rue. Pour répondre à cette question, le I de Moran est

utilisé, soit une mesure d'autocorrélation qui va permettre de déterminer si le nombre de crimes d'un segment est corrélé au nombre de crimes des segments avoisinants.

Pour déterminer si un segment est avoisinant, des matrices de distance sont utilisées. Ainsi, tous les segments situés dans un rayon de  $x$  mètres ou moins sont considérés avoisinants. Le  $I$  de Moran a été calculé pour chaque incrément de 50 mètres jusqu'à une distance maximale de 500 mètres et bien qu'une corrélation significative ait été trouvée pour chaque, la plus importante était à 100 mètres, et ce, autant pour les crimes contre la personne que les crimes contre la propriété. La corrélation était de 0,169 pour les crimes contre la personne et de 0,229 pour les crimes contre la propriété ( $p < 0,001$ ). Les corrélations positives et modérées indiquent que les segments de rue sont entourés d'autres segments avec des niveaux semblables de crimes. Cela soutient l'idée que les crimes sont concentrés au niveau de segments de rue et confirme qu'il y a un problème de dépendance spatiale au deuxième niveau d'agrégation. Pour résoudre ce problème, on introduit un lag spatial comme variable indépendante au deuxième niveau. Un lag spatial est généré pour chaque catégorie de crime, et la limite de 100 mètres, là où l'autocorrélation est la plus élevée, est déterminée. La dépendance spatiale est donc contrôlée et le postulat d'indépendance est respecté.

Pour l'ensemble des modèles et pour les deux grandes catégories de crimes, le lag spatial était significatif. Ainsi, cette influence des segments avoisinants sur le segment étudié démontre la dépendance entre les unités spatiales avoisinantes et le besoin de tenir compte du lag spatial pour toute analyse contenant des données spatiales.

### *Les modèles*

Au final, des modèles Poisson multiniveaux multivariés sont mis de l'avant. Au deuxième niveau, les variables liées à la désorganisation sociale et le lag spatial sont introduits. Au premier niveau, les variables liées aux tweets, soit les polarités, les six catégories d'humeurs et le nombre de tweets sont introduits, ainsi que les heures de la journée.

Trois modèles sont présentés: 1) Le premier modèle contient les variables de deuxième niveau, le nombre de tweets, les variables de polarité et les heures de la journée; 2) Le deuxième modèle est identique au premier, mis à part que les variables de polarité sont remplacées par

les variables liées aux six catégories d'humeurs; 3) Le troisième modèle n'inclut aucune variable de la polarité ou de catégorie d'humeurs, et introduit l'interaction entre le nombre de tweets et les heures de la journée.

On remarque que les variables du deuxième niveau ainsi que les heures de la journée sont présentes dans tous les modèles et permettront de tirer des conclusions sur la désorganisation sociale, sur l'utilité du lag spatial et sur l'influence de l'heure de la journée sur le nombre de crimes. Dans les premier et deuxième modèles, différentes variables d'humeur sont présentes et permettront de mieux comprendre la relation entre l'humeur et la criminalité. Le nombre de tweets est présent dans les trois modèles. Dans le troisième modèle, il est possible de considérer son effet par lui-même, et dans les deux premiers modèles, il est possible de considérer son effet lorsque l'humeur est prise en compte. Finalement, l'effet des tweets selon l'heure de la journée est exploré à l'aide du troisième modèle.

## Résultats

Une première analyse porte sur les crimes contre la personne. Tous les prédicteurs du deuxième niveau sont statistiquement significatifs. La population résidente et le pourcentage de minorités visibles ont une relation négative avec les crimes contre la personne, alors que les autres prédicteurs sociodémographiques ont une relation positive. Pour ce qui est des trois prédicteurs liés aux tweets, seule la proportion de tweets positifs est significative, et elle a une relation négative avec les crimes contre la personne. Finalement, les catégories de la variable heure sont incluses dans le modèle, mais pour montrer leur effet de façon plus claire, elles sont présentées dans la section suivante sous forme de graphique, plutôt que dans le tableau ci-dessous.

Dans le modèle des crimes contre la propriété, tous les prédicteurs du deuxième niveau sont significatifs à  $p < 0,001$ . Comme pour les crimes contre la personne, les coefficients sont tous positifs mis à part pour la population résidentielle et le pourcentage de minorités visibles. Pour ce qui est des variables reliées aux tweets, seul le nombre de tweets est non significatif; les deux polarités sont significatives à  $p < 0,001$  et ont une relation positive avec le nombre de crimes.

TABLEAU 1

**Modèle multiniveau Poisson 1 prédisant les crimes au segment de rue  
(coefficients bêta non standardisés et [intervalles de confiance])**

Variable	Crimes contre la personne	Crimes contre la propriété
Population résidente	-0,000** (1,000; 1,000)	-0,000*** (1,000; 1,000)
Familles monoparentales (%)	0,033*** (1,029; 1,037)	0,016*** (1,012; 1,021)
Pop sous le seuil pauvreté (%)	0,019*** (1,015; 1,024)	0,026*** (1,021; 1,032)
Minorités visibles (%)	-0,002* (0,997; 1,000)	-0,008*** (0,991; 0,994)
Mobilité résidentielle (%)	0,020*** (1,018; 1,023)	0,025*** (1,023; 1,028)
Lag spatial	0,088*** (1,082; 1,102)	0,022*** (1,015; 1,029)
Nombre de tweets	-0,003 (0,992; 1,002)	0,009 (0,999; 1,020)
P. de tweets négatifs	-0,000 (0,999; 1,001)	0,002*** (1,001; 1,003)
P. de tweets positifs	-0,001* (0,999; 1,000)	0,001*** (1,001; 1,002)

\*\*\* p < 0,001, \*\* p < 0,01, \* p < 0,05.

À noter que la variable Heure, contenant 24 catégories et dont l'heure de référence = 6, est incluse dans le modèle 1. Elle est plutôt présentée à la section suivante sous forme de figure afin de faciliter la compréhension du lecteur.

TABLEAU 2

**Modèle multiniveau Poisson 2 prédisant les crimes au segment de rue  
(coefficients bêta non standardisés et [intervalles de confiance])**

Variable	Crimes contre la personne	Crimes contre la propriété
Population résidente	-0,000** (1,000; 1,000)	-0,000*** (1,000; 1,000)
Familles monoparentales (%)	0,033*** (1,029; 1,037)	0,016*** (1,012; 1,021)
Pop sous le seuil pauvreté (%)	0,019*** (1,015; 1,024)	0,026*** (1,021; 1,032)
Minorités visibles (%)	-0,002* (0,997; 1,000)	-0,008*** (0,991; 0,994)
Mobilité résidentielle (%)	0,020*** (1,018; 1,023)	0,026*** (1,023; 1,028)
Lag spatial	0,088*** (1,082; 1,102)	0,022*** (1,015; 1,029)
Nombre de tweets	-0,003 (0,992; 1,002)	0,010 (0,999; 1,022)
P. de tweets de colère	0,000 (0,998; 1,002)	0,003** (1,001; 1,004)
P. de tweets de dégoût	-0,000 (0,997; 1,002)	0,001 (1,000; 1,002)
P. de tweets de peur	-0,000 (0,998; 1,001)	0,002** (1,001; 1,004)
P. de tweets de joie	-0,001 (0,998; 1,000)	0,001** (1,000; 1,002)
P. de tweets de tristesse	-0,001 (0,998; 1,001)	0,001** (1,001; 1,002)
P. de tweets de surprise	-0,001 (0,998; 1,001)	0,001*** (1,000; 1,002)

\*\*\* p < 0,001, \*\* p < 0,01, \* p < 0,05.

À noter que la variable Heure, contenant 24 catégories et dont l'heure de référence = 6, est incluse dans le modèle 2. Elle est plutôt présentée à la section suivante sous forme de figure afin de faciliter la compréhension du lecteur.



Le modèle 2 incorpore les variables sociodémographiques et le lag spatial au deuxième niveau, ainsi que le nombre de tweets, les six catégories d'humeurs de tweets et les heures de la journée au premier niveau. Tout comme le modèle 1, les catégories d'heures sont présentées à la section suivante.

Pour le modèle prédisant les crimes contre la personne, aucune des variables liées aux tweets n'est significative. Ce résultat diffère quelque peu de la logique du premier modèle où la proportion de tweets positifs est significative. Les variables du deuxième niveau et la variable heure sont identiques au premier modèle.

Dans le modèle des crimes contre la propriété, la proportion de tweets de surprise est significative à  $p < 0,001$ , les proportions de tweets de tristesse, joie, peur et colère sont significatives à  $p < 0,01$ , et la proportion de tweets de dégoût ainsi que le nombre de tweets sont non significatifs. On note, comme pour le premier modèle, qu'un plus grand spectre d'humeurs semble lié aux crimes contre la propriété, comparativement aux crimes contre la personne. Les variables du deuxième niveau sont identiques au premier modèle.

Un troisième modèle est conçu, afin d'analyser plus en détail la relation entre le nombre de tweets et les crimes. Selon les deux modèles présentés ci-haut, la prédiction du nombre de crimes par le nombre de tweets est non significative, autant pour les crimes contre la personne que les crimes contre la propriété. Le troisième modèle inclut l'interaction Nombre de tweets \* Heure, afin d'analyser la relation entre le nombre de tweets et le nombre de crimes selon l'heure de la journée.

Dans le troisième modèle, seules les catégories d'heures significatives ont été incluses dans le tableau afin d'en faciliter la lecture. Bien que les autres catégories ne soient pas présentées, elles font partie du modèle. Pour le modèle des crimes contre la personne, on remarque premièrement que le nombre de tweets n'est pas significatif. Par contre, la variable d'interaction nous indique que le nombre de tweets est significatif, mais seulement à l'heure 3, 4 et 5.

Dans le modèle prédisant les crimes contre la propriété, on remarque que le coefficient de la variable Nombre de tweets est positif et significatif à  $p < 0,001$ . Ainsi, le nombre de tweets semble avoir une valeur ajoutée, mais les variables de tweets liées à l'humeur (dans les modèles 1 et 2) la rendent non significative. Finalement, on voit que l'interaction est seulement significative à certaines heures de la journée (9/23).

TABLEAU 3

Modèle multiniveau Poisson 3 prédisant les crimes au segment de rue  
(coefficients bêta non standardisés et [intervalles de confiance])

Variable	Crimes contre la personne	Crimes contre la propriété
Population résidente	-0,000** (1,000; 1,000)	-0,000*** (1,000; 1,000)
Familles monoparentales (%)	0,033*** (1,030; 1,037)	0,016*** (1,012; 1,021)
Pop sous le seuil pauvreté (%)	0,019*** (1,015; 1,024)	0,026*** (1,021; 1,032)
Minorités visibles (%)	-0,002* (0,997; 1,000)	-0,008*** (0,991; 0,994)
Mobilité résidentielle (%)	0,020*** (1,018; 1,023)	0,026*** (1,023; 1,028)
Lag spatial	0,088*** (1,082; 1,102)	0,022*** (1,015; 1,030)
Nombre de tweets	-0,019 (0,923; 1,043)	0,021*** (1,010; 1,033)
Interaction Heure 0 * Nombre de tweets	0,038 (0,976; 1,105)	0,013** (1,004; 1,023)
Interaction Heure 1 * Nombre de tweets	0,060 (0,990; 1,137)	0,057** (1,026; 1,093)
Interaction Heure 2 * Nombre de tweets	0,086* (1,006; 1,181)	0,099** (1,044; 1,167)
Interaction Heure 3 * Nombre de tweets	0,119* (1,016; 1,250)	0,103* (1,001; 1,227)
Interaction Heure 4 * Nombre de tweets	0,094* (1,021; 1,182)	0,025 (0,963; 1,092)
Interaction Heure 5 * Nombre de tweets	0,034 (0,973; 1,100)	0,006* (1,001; 1,012)
Interaction Heure 7 * Nombre de tweets	0,029 (0,969; 1,094)	-0,021* (0,961; 0,997)
Interaction Heure 21 * Nombre de tweets	0,033 (0,972; 1,099)	0,005* (1,000; 1,009)
Interaction Heure 22 * Nombre de tweets	0,031 (0,971; 1,097)	0,004* (1,001; 1,008)
Interaction Heure 23 * Nombre de tweets	0,034 (0,972; 1,101)	0,014** (1,006; 1,023)

\*\*\* p < 0,001, \*\* p < 0,01, \* p < 0,05.

À noter que la variable Heure, contenant 24 catégories et dont l'heure de référence = 6, est incluse dans le modèle 3. Elle est plutôt présentée à la section suivante sous forme de figure afin de faciliter la compréhension du lecteur.

À noter que les 24 catégories de la variable d'interaction, dont la catégorie de référence = Heure 6 \* Nombre de tweets, sont incluses dans le modèle. Dans un souci d'en faciliter la lecture, seules les catégories significatives ont été présentées dans le tableau.

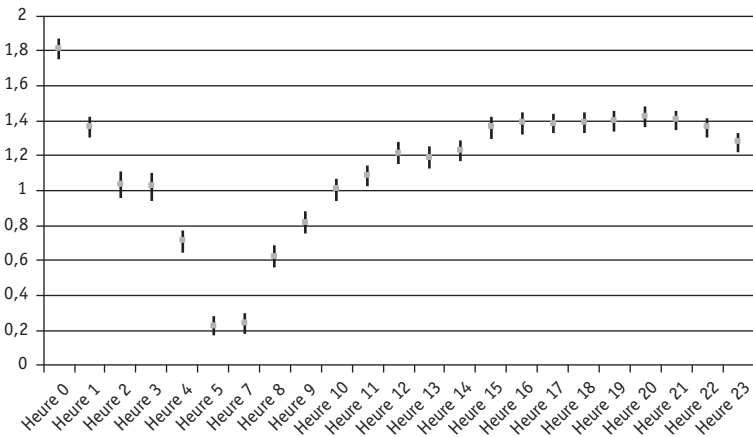
Cela renforce l'idée que d'analyser le crime selon des périodes temporelles précises est important. Il serait intéressant de comparer les différentes heures de la journée entre elles, plutôt que seulement par rapport à l'heure de référence (Heure = 6).

### *La prédiction de la criminalité selon l'heure de la journée*

Comme mentionné précédemment, la variable Heure, divisée en 24 catégories avec l'heure 6 comme variable de référence, était incluse dans les trois modèles présentés. La variable est significative pour l'ensemble des catégories des modèles multivariés. Dans cette section, les différences entre les crimes contre la propriété et la personne sont analysées à l'aide de figures. Aussi, une comparaison des coefficients des heures est faite afin de pouvoir apprécier les différences entre ceux-ci plutôt que seulement par rapport à l'heure de référence.

**FIGURE 1**

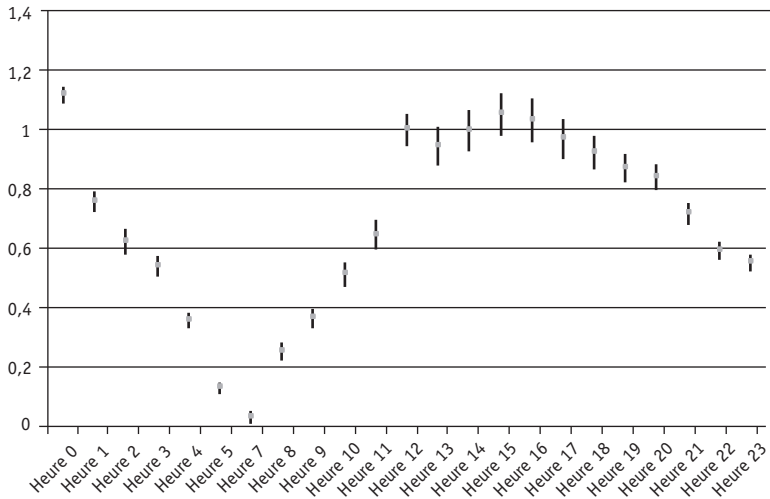
**Les coefficients bêta non standardisés ( $\pm 2$  erreurs standards)  
des catégories de la variable Heure des modèles prédisant les crimes  
contre la personne (Heure de référence = 6)**



Le graphique ci-dessus présente les différents coefficients des catégories de la variable Heure pour les modèles prédisant les crimes contre la personne. L'utilisation de minimums et de maximums correspondant à  $\pm 2$  erreurs standards permet d'apprécier des différences significatives entre les coefficients. On note un coefficient très élevé à l'heure 0. Des conclusions claires ne peuvent être prises pour les coefficients reliés à Heure = 0, et ce, pour l'ensemble des modèles présentés. Ainsi, il a été démontré dans d'autres études montréalaises que certains policiers attribuaient « minuit » à certains crimes dont ils n'étaient pas capables de déterminer l'heure. L'heure 0 est donc biaisée par cette pratique et

difficilement interprétable. Si on examine la distribution des coefficients des autres heures, on note une diminution jusqu'à l'heure de référence (6), suivie d'une augmentation de forme exponentielle jusqu'à 15 h, en finissant par une stabilité jusqu'à 23 h.

**FIGURE 2**  
**Les coefficients bêta non standardisés ( $\pm 2$  erreurs standards)**  
**des catégories de la variable Heure des modèles prédisant les crimes**  
**contre la propriété (Heure de référence = 6)**



L'évolution dans le temps des coefficients pour les modèles des crimes contre la propriété est semblable à celle des crimes contre la personne, mis à part pour la partie de 12 h à 23 h. Le maximum est atteint plus rapidement pour les crimes contre la propriété (12 h plutôt que 15 h), et les coefficients sont moins stables. Ainsi, on note une diminution graduelle de 15 h à 23 h.

En analysant les deux graphiques, on observe qu'en plus d'être différents de l'heure de référence Heure = 6, les coefficients des heures de la journée sont différents entre eux. Aussi, quelques différences sont observables entre les patrons des modèles des crimes contre la personne et ceux des crimes contre la propriété. On en conclut qu'il est important de s'attarder à l'évolution des crimes selon le temps de la journée, et que cette évolution semble quelque peu différente entre les crimes contre la personne et ceux contre la propriété.

## Discussion et conclusion

Cette étude s'inscrit dans une nouvelle lignée de la criminologie, la criminologie des endroits, où l'on tente de comprendre les phénomènes criminels à des niveaux spatiotemporels de plus en plus précis et où l'on analyse l'effet d'endroits plutôt que de personnes spécifiques. On démontre la grande concentration des crimes dans les grandes villes, où une minorité d'endroits sont la scène d'une majorité des crimes. Les forces de l'ordre et autres acteurs de sécurité publique ont donc intérêt à concentrer leurs efforts sur des *hotspots* très précis de crimes pour avoir un retour sur investissement maximal. Deux limites sont apparentes dans les études des dernières années. Premièrement, alors que l'on tente d'être toujours de plus en plus précis sur le plan spatial, on ne tient pas compte des fluctuations temporelles précises des crimes urbains. Deuxièmement, il est difficile de trouver des prédicteurs de la criminalité à un niveau très micro, et cela limite notre capacité à expliquer la variance entre les endroits. Cette étude a donc évalué l'importance des fluctuations intrajournalières des crimes à Montréal, et a utilisé les tweets envoyés par la population pour générer deux facteurs explicatifs de la criminalité à des niveaux très précis, soit la population flottante et une caractéristique de celle-ci, l'humeur.

### *Les tweets et la population flottante*

Cette étude s'est penchée sur les deux grandes catégories de crimes, soit les crimes contre la personne et les crimes contre la propriété. L'influence des tweets sur ces catégories était assez différente; le nombre de *tweets* et l'humeur prédisaient significativement et positivement le nombre de crimes contre la propriété, tandis que c'était un peu moins clair pour les crimes contre la personne.

Ces derniers ne pouvaient être prédits ni par le nombre de tweets ni par les catégories d'humeurs. Du côté de la polarité, la proportion de tweets positifs influence négativement le nombre de crimes. On peut penser que plus les gens ont du plaisir et partagent des expériences positives, moins ils sont portés à faire preuve de violence et donc à commettre des crimes contre la personne. Il est difficile de confirmer cette hypothèse, car les autres études qui ont lié les crimes et l'analyse de sentiment ne se sont pas spécifiquement penchées sur les crimes contre la personne. Chen *et al.* (2015) ont analysé les vols simples et sont venus à la conclusion que la polarité, qu'elle soit positive ou négative,

tive, était liée à une augmentation de crimes. Kim, Cha et Sandholm (2014) ont analysé l'ensemble des crimes et seuls les tweets négatifs y étaient liés. Les différences de résultats entre les deux grandes catégories de crimes de la présente étude, ainsi que les différences avec les autres études indiquent ultimement que les facteurs explicatifs entre différents types ou catégories de crimes peuvent être différents (Haberman, 2017). La généralisation et la comparaison des résultats doivent donc se faire avec prudence.

Pour ce qui est du nombre de tweets, il n'était pas significatif de façon globale, ce qui corrobore les résultats d'autres études montréalaises (Boivin et D'Elia, 2017), mais pas d'autres études canadiennes (Andresen, 2011). Deux hypothèses peuvent être mises de l'avant. Premièrement, le rôle des gardiens, issus de la théorie des activités routinières, pourrait expliquer pourquoi les crimes contre la personne ne semblent pas être influencés par le nombre de personnes présentes. Alors que le nombre de personnes à un endroit augmente, le nombre de victimes et d'auteurs de crimes potentiels augmente et il ne faut pas oublier que le nombre de gardiens, formels ou informels, augmente également. Il est possible qu'à Montréal les personnes présentes aient un plus grand effet dissuasif sur le crime, qui supplanterait l'influence de l'augmentation des auteurs et victimes potentiels. Et cet effet des gardiens ne serait pas aussi notable pour les crimes contre la propriété, car ceux-ci n'attirent pas autant l'attention des personnes autour que les crimes contre la personne. Ce serait une hypothèse à confirmer dans de prochaines études. Deuxièmement, les grandes catégories de crimes renferment plusieurs types de crimes qui peuvent tenir de comportements assez différents. Par exemple, Boivin (2013) a trouvé que la population flottante était associée plus fortement au nombre de voies de fait entre étrangers qu'aux violences conjugales, un type de crime typiquement plus « local ». On peut penser que d'analyser des types de crimes plutôt que des grandes catégories pourrait être plus intéressant et plus précis, surtout d'un point de vue pratique.

Les résultats liés aux tweets étaient plus probants pour les crimes contre la propriété. Ainsi, il a été constaté qu'une augmentation du nombre de tweets – ou de la population flottante – prédisait une augmentation du nombre de crimes. Par contre, lorsque l'on tenait compte des humeurs, l'effet du nombre de personnes perdait sa significativité. On peut conclure, en se basant sur ces résultats, que le nombre de personnes présentes sur un segment de rue perd de son importance

lorsqu'on tient compte de ses caractéristiques, ici l'humeur. Des études ont tenté de caractériser les visiteurs en fonction de l'intention de leur visite (Boivin et D'Elia, 2017 ; Felson et Boivin, 2015), soit pour le travail, pour magasiner, pour un loisir ou pour l'école, et ont obtenu des résultats intéressants. Selon les résultats présentés dans ce rapport, il y aurait intérêt à concentrer les efforts des prochaines études à caractériser la population flottante plutôt qu'à simplement la compter. Cela ouvre la possibilité à d'autres analyses qualifiant les tweets et les usagers de la plateforme qui permettraient de caractériser une population. Par exemple, Sloan, Morgan, Burnap et Williams (2015) présentent deux outils qui estiment les données démographiques (âge, occupation, etc.) à partir des descriptions de profils des usagers de Twitter.

### *L'importance de la temporalité intrajournalière*

Plusieurs démarches et résultats de l'étude ont permis d'évaluer l'importance des fluctuations intrajournalières des crimes et force est de constater qu'elles sont indispensables à la compréhension du phénomène criminel urbain.

Le modèle des composantes de la variance a mis en compétition la variance explicable sur le plan spatial des segments de rue à celle sur le plan temporel des heures de la journée. Bien que la variance soit légèrement plus élevée sur le plan des segments, elle est également élevée sur le plan des heures de la journée. Les variations temporelles sont donc presque aussi importantes que les variations géographiques.

L'introduction d'une variable d'interaction dans le troisième modèle multivarié présenté, soit entre le nombre de tweets et les heures de la journée, laisse supposer que la force des prédicteurs qui ont une dimension temporelle peut changer dans le temps, et donc que les chercheurs qui s'intéressent à ceux-ci dans le cadre de leur étude devraient tester leur variabilité temporelle et tenter de comprendre les facteurs expliquant celle-ci. Ce résultat alimente aussi la littérature sur la population flottante. Ainsi, la variabilité temporelle du nombre de tweets laisse entendre que la composition, et donc les caractéristiques de la population flottante, change au cours de la journée. Les prochaines études devraient tenter de déterminer les facteurs explicatifs de ces fluctuations, et surtout utiliser une donnée de population flottante très précise.

### *L'utilisation des tweets pour la prédiction de la criminalité*

L'utilité des tweets pour définir une population flottante précise autant sur le plan spatial que temporel a été démontrée grâce à l'étude. Aussi, l'analyse de l'humeur a ouvert la porte à l'analyse du contenu des tweets pour caractériser la population flottante. L'analyse de tweets semble une avenue prometteuse pour la prédiction de la criminalité urbaine.

On peut se demander si les utilisateurs de Twitter à Montréal sont représentatifs de la population réelle de la ville. Plusieurs études se sont penchées sur la question (Mellon et Prosser, 2017 ; Steiger *et al.*, 2015 ; Sui et Goodchild, 2011) et s'accordent pour affirmer que les utilisateurs de la plateforme ne représentent pas parfaitement la population générale. Une étude britannique (Mellon et Prosser, 2017) a démontré que les utilisateurs de Twitter sont notamment plus jeunes, plus éduqués et les hommes y sont surreprésentés. De plus, il existe des différences importantes entre les utilisateurs de différents pays (Poblete *et al.*, 2011). Des études de représentativité devraient être réalisées au Canada afin de mieux baliser et définir la population utilisant Twitter.

Dans les prochaines études, il y aurait lieu de se pencher sur d'autres méthodes d'analyse du contenu des tweets qui permettraient de caractériser la population flottante. Outre les messages eux-mêmes, une avenue intéressante serait d'extraire les données démographiques des usagers à partir des descriptions de profil de ceux-ci (Sloan *et al.*, 2015).

## Références

- Abdaoui, A., Azé, J., Bringay, S. et Poncelet, P. (2015). Collaborative content-based method for estimating user reputation in online forums. *International Conference on Web Information Systems Engineering*, 292-299.
- Aghababaei, S. et Makrehchi, M. (2016). Mining social media content for crime prediction. *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 526-531. <https://doi.org/10.1109/WI.2016.0089>
- Andresen, M. A. (2011). The ambient population and crime analysis. *The Professional Geographer*, 63(2), 193-212.
- Andresen, M. A. et Malleon, N. (2011). Testing the stability of crime patterns: implications for theory and policy. *Journal of Research in Crime and Delinquency*, 48(1), 58-82.
- Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht, Pays-Bas: Kluwer Academic.
- Ashby, M. P. et Bowers, K. J. (2013). A comparison of methods for temporal analysis of aoristic crime. *Crime Science*, 2(1), 1-16.



- Awan, I. (2014). Islamophobia and Twitter: A typology of online hate against Muslims on social media. *Policy & Internet*, 6(2), 133-150.
- Bendler, J., Brandt, T., Wagner, S. et Neumann, D. (2014). Investigating crime-to-Twitter relationships in urban environments-facilitating a virtual neighborhood watch. *Proceedings of the European Conference on Information Systems (ECIS)*.
- Bendler, J., Ratku, A. et Neumann, D. (2014). Crime mapping through geo-spatial social media activity. *Proceedings of 14th International Conference on Information Systems (ICIS14)*.
- Bermingham, A. et Smeaton, A. (2011). On using Twitter to monitor political sentiment and predict election results. *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*, 2-10.
- Boivin, R. (2013). On the use of crime rates. *Canadian Journal of Criminology and Criminal Justice*, 55(2), 263-277.
- Boivin, R. et D'Elia, M. (2017). A network of neighborhoods: Predicting crime trips in a large Canadian city. *Journal of research in crime and delinquency*, 54(6), 824-846.
- Boivin, R. et Ouellet, F. (2011). La dynamique de la criminalité à Montréal: l'écologie criminelle revisitée. *Canadian Journal of Criminology and Criminal Justice*, 53(2), 189-216.
- Bollen, J., Mao, H. et Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1), 1-8.
- Braga, A., Papachristos, A. et Hureau, D. (2012). Hot spots policing effects on crime. *Campbell Systematic Reviews*, 8(8), 1-96.
- Braga, A. A., Papachristos, A. V. et Hureau, D. M. (2010). The concentration and stability of gun violence at micro places in Boston, 1980-2008. *Journal of Quantitative Criminology*, 26(1), 33-53.
- Budd, T. (2001). *Burglary: Practice messages from the British crime survey*. Great Britain Home Office, Policing and Reducing Crime Unit.
- Burnap, P., Rana, O. F., Avis, N., Williams, M., Housley, W., Edwards, A.,... Sloan, L. (2015). Detecting tension in online communities with computational Twitter analysis. *Technological Forecasting and Social Change*, 95, 96-108.
- Burnap, P. et Williams, M. L. (2015). Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy & Internet*, 7(2), 223-242.
- Chen, X., Cho, Y. et Jang, S. Y. (2015). Crime prediction using Twitter sentiment and weather. *Systems and Information Engineering Design Symposium (SIEDS)*, 63-68.
- Curman, A. S., Andresen, M. A. et Brantingham, P. J. (2015). Crime and place: A longitudinal examination of street segment patterns in Vancouver, BC. *Journal of Quantitative Criminology*, 31(1), 127-147.
- Eck, J. E. et Spelman, W. (1987). *Problem-solving: Problem-oriented policing in Newport News*. Washington, DC: National Institute of Justice.
- Eck, J. E., Chainey, S., Cameron, J. G., Leitner, M. et Wilson, R. E. (2005). *Mapping crime: Understanding hot spots*. Washington, DC: National Institute of Justice.

- Eisenstein, J. (2013). What to do about bad language on the internet. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics*, 359-369.
- Felson, M. et Boivin, R. (2015). Daily crime flows within a city. *Crime Science*, 4(1).
- Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. *Decision Support Systems*, 61, 115-125. <https://doi.org/10.1016/j.dss.2014.02.003>
- Gu, Y., Qian, Z. S. et Chen, F. (2016). From Twitter to detector: Real-time traffic incident detection using social media data. *Transportation Research Part C: Emerging Technologies*, 67, 321-342.
- Haberman, C. P. et Ratcliffe, J. H. (2015). Testing for temporally differentiated relationships among potentially criminogenic places and census block street robbery counts. *Criminology*, 53(3), 457-483.
- Haberman, C. P. (2017). Overlapping Hot Spots? *Criminology & Public Policy*, 16(2), 633-660.
- Haberman, C. P., Sorg, E. T. et Ratcliffe, J. H. (2017). Assessing the validity of the law of crime concentration across different temporal scales. *Journal of Quantitative Criminology*, 33(3), 547-567.
- Kim, J., Cha, M. et Sandholm, T. (2014). SocRoutes: Safe routes based on tweet sentiments. *Proceedings of the 23rd International Conference on World Wide Web*, 179-182.
- Koper, C. S. (1995). Just enough police presence: Reducing crime and disorderly behavior by optimizing patrol time in crime hot spots. *Justice Quarterly*, 12(4), 649-672.
- Malleson, N. et Andresen, M. A. (2015a). Spatio-temporal crime hotspots and the ambient population. *Crime science*, 4(1), 10-17.
- Malleson, N. et Andresen, M. A. (2015b). The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science*, 42(2), 112-121.
- Malleson, N. et Andresen, M. A. (2016). Exploring the impact of ambient population measures on London crime hotspots. *Journal of Criminal Justice*, 46, 52-63.
- Mellon, J. et Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, 4(3), 2053168017720008.
- Mohammad, S. M. et Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29, 436-465.
- Osgood, D. W. (2000). Poisson-based regression analysis of aggregate crime rates. *Journal of Quantitative Criminology*, 16(1), 21-43.
- Poblete, B., Garcia, R., Mendoza, M. et Jaimes, A. (2011). Do all birds tweet the same?: characterizing twitter around the world. *Proceedings of the 20th ACM international conference on Information and knowledge management*, 1025-1030. ACM.

- Prathap, B. R. et Ramesha, K. (2018). Twitter sentiment for analyzing different types of crimes. *2018 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, 483-488.
- Ratcliffe, J. H. (2000). Aoristic analysis: The spatial interpretation of unspecified temporal events. *International Journal of Geographical Information Science*, 14(7), 669-679.
- Ratcliffe, J. H. (2002). Aoristic signatures and the spatio-temporal analysis of high volume crime patterns. *Journal of Quantitative Criminology*, 18(1), 23-43.
- Ratcliffe, J. H. (2004). The *hotspot* matrix: A framework for the spatio-temporal targeting of crime reduction. *Police practice and research*, 5(1), 5-23.
- Ratcliffe, J., Groff, E., Haberman, C. et Sorg, E. (2012). *Smart policing initiative final report*. Philadelphie, PA: Temple University Center for Security and Crime Science.
- Rosser, G., Davies, T., Bowers, K. J., Johnson, S. D. et Cheng, T. (2017). Predictive crime mapping: Arbitrary grids or street networks? *Journal of Quantitative Criminology*, 33(3), 569-594.
- Shaw, C. R. et McKay H. D. (1942). *Juvenile delinquency and urban areas: A study of rates of delinquents in relation to differential characteristics of local communities in American cities*. Chicago, IL: University of Chicago Press.
- Sherman, L. W., Gartin, P. R. et Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, 27(1), 27-56.
- Sherman, L. W. et Weisburd, D. (1995). General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial. *Justice quarterly*, 12(4), 625-648.
- Sloan, L., Morgan, J., Burnap, P. et Williams, M. (2015). Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PLoS one*, 10(3), e0115545.
- Statistique Canada. (2018). Secteur de recensement: définition en langage simple. Repéré à <https://www150.statcan.gc.ca/n1/pub/92-195-x/2011001/geo/ct-sr/ct-sr-fra.htm>
- Steiger, E., Westerholt, R., Resch, B. et Zipf, A. (2015). Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems*, 54, 255-265.
- Sui, D. et M. Goodchild. (2011). The convergence of GIS and social media: Challenges for GIScience. *International Journal of Geographical Information Science* 25(11), 1737-1748.
- Wang, X., Gerber, M. S. et Brown, D. E. (2012). Automatic crime prediction using events extracted from Twitter posts. Dans S. J. Yang, A. M. Greenberg et M. Endsley (dir.), *Social Computing, Behavioral-Cultural Modeling and Prediction* (vol. 7227, p. 231-238). [https://doi.org/10.1007/978-3-642-29047-3\\_28](https://doi.org/10.1007/978-3-642-29047-3_28)
- Weaver, S. (2013). A rhetorical discourse analysis of online anti-muslim and anti-semitic jokes. *Ethnic and Racial Studies*, 36(3), 483-499.
- Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology*, 53(2), 133-157.

- Weisburd, D., Groff, E. R. et Yang, S. M. (2012). *The criminology of place: Street segments and our understanding of the crime problem*. Oxford University Press.
- Weisburd, D. et Telep, C. W. (2014). Hot spots policing: What we know and what we need to know. *Journal of Contemporary Criminal Justice*, 30(2), 200-220.
- Wheeler, A. et Haberman, C. (2018). Modeling the spatial patterns of intra-day crime trends. <https://dx.doi.org/10.2139/ssrn.3136030>
- Williams, M. L., Burnap, P. et Sloan, L. (2016). Crime sensing with big data: The affordances and limitations of using open source communications to estimate crime patterns. *British Journal of Criminology*, 57(2), 320-340. <http://doi.org/10.1093/bjc/azw031>

## Social media as a predictor of urban crime

**ABSTRACT** • *This study attempts to determine the importance of analyzing crimes at increasingly precise spatial and temporal levels. Messages from the social media source Twitter are used to predict the distribution of crimes in Montreal by estimating the actual population in the area and characterizing it by mood. Multi-level Poisson models are used to predict violent and property crimes for a particular street segment according to the time of day. The results show that any crime analysis in Montreal must take into account variance in crime at the microsite level and incorporate intraday periods. Also, the characterization of a city's actual population has been identified as a promising avenue for crime prediction. This study suggests that the analysis of Twitter data makes it possible to draw some conclusions, but it still needs to be further developed.*

**KEYWORDS** • *Crime analysis, ambient population, sentiment analysis, social media, aoristic analysis.*

## Las redes sociales como predictores de la criminalidad urbana

**RESUMEN** • *Este estudio intenta determinar la importancia de analizar los crímenes en niveles espaciales y temporales cada vez más precisos. De este modo, una nueva fuente de datos proveniente de las redes sociales, los mensajes de Twitter, es utilizada con el fin de predecir la repartición de los crímenes en Montreal, estimando la población real en el territorio y caracterizándola según su humor. Unos modelos de multiniveles Poisson son utilizados con el fin de predecir los crímenes contra la persona y los crímenes contra los bienes y agregados al segmento de calle según la hora del día. Los resultados demuestran que para todo análisis sobre la criminalidad en Montreal, es primordial tener en cuenta la varianza de la criminalidad al nivel de los micro-sitios, e incorporar períodos intra-diarios. Igualmente, la caracterización de la población real de la ciudad fue identificada como una avenida prometedoras para la predicción de los crímenes. Este estudio sugiere que la utilización de los datos de Twitter es una avenida de análisis concluyente, pero que toca todavía profundizar.*

**PALABRAS CLAVE** • *Análisis de la criminalidad, población flotante, análisis del humor, redes sociales, análisis aoristic.*