Mappemonde

Revue trimestrielle sur l'image géographique et les formes du territoire

128 | 2020
Varia

Objets connectés et mobilité urbaine : visualiser les déplacements des usagers de Twitter avec des graphes dynamiques

Connected devices and urban mobility: visualizing the trajectories of Twitter users through dynamic graphs

Dispositivos electrónicos conectados y movilidad urbana: visualización de los movimientos de los usuarios de Twitter mediante gráficos de flujos dinámicos

Françoise Lucchini, Olivier Gillet, Bernard Elissalde, Lény Grassot, Julien Baudry et Armelle Couillet

Édition électronique

URL : <http://journals.openedition.org/mappemonde/4297>

DOI : 10.4000/mappemonde.4297

ISSN : 1769-7298

Éditeur

UMR ESPACE

Référence électronique

Françoise Lucchini, Olivier Gillet, Bernard Elissalde, Lény Grassot, Julien Baudry et Armelle Couillet, « Objets connectés et mobilité urbaine : visualiser les déplacements des usagers de Twitter avec des graphes dynamiques », *Mappemonde* [En ligne], 128 | 2020, mis en ligne le 01 juillet 2020, consulté le 02 septembre 2020. URL : <http://journals.openedition.org/mappemonde/4297> ; DOI : <https://doi.org/10.4000/mappemonde.4297>

Ce document a été généré automatiquement le 2 septembre 2020.

La revue *Mappemonde* est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International.

Objets connectés et mobilité urbaine : visualiser les déplacements des usagers de Twitter avec des graphes dynamiques

Connected devices and urban mobility: visualizing the trajectories of Twitter users through dynamic graphs

Dispositivos electrónicos conectados y movilidad urbana: visualización de los movimientos de los usuarios de Twitter mediante gráficos de flujos dinámicos

Françoise Lucchini, Olivier Gillet, Bernard Elissalde, Lény Grassot, Julien Baudry et Armelle Couillet

Liens pour accéder aux applications :

Lien 1 : Plateforme de visualisation des mobilités à partir des données 2015-2017 du réseau social Twitter à Paris (environnement R et hébergement sur la TGIR Huma-Num) : https://analytics.huma-num.fr/UMR_IDEES/XTERM_TCA_SOCIAL_NETWORK/

Lien 2 : Plateforme de visualisation permettant de comparer deux graphes de mobilité issus du réseau social Twitter à Paris, (environnement R et hébergement sur la TGIR Huma-Num) : https://analytics.huma-num.fr/UMR_IDEES/XTERM_TVG_SOCIAL_NETWORK/

- 1 La mobilité saisie à travers les pratiques liées aux nouvelles technologies a fait l'objet de réflexions désormais classiques de la part des sciences sociales. Pour Paul Virilio, les technologies dites « nomades », en développant la sensation d'ubiquité, permettent de garder contact avec ses proches partout où l'on se trouve (Virilio, 2009). En permettant au temps réel d'être plus rapide que l'espace réel, le citoyen emporte ses réseaux sociaux dans d'autres lieux et se connecte lui-même avec d'autres lieux en conjuguant en permanence les relations entre mobilité, coprésence et ubiquité. La mobilité

individuelle ne se résume plus à la seule capacité individuelle à se déplacer, elle relève aussi de l'interaction sociale (Massot, Orfeuil, 2005). Les données issues des objets connectés constituent ainsi une ressource nouvelle pour la compréhension des rythmes urbains et des mobilités dans la ville. La finesse spatiale et temporelle de ces données associée à l'utilisation de graphes dynamiques renouvelle les questionnements spatiotemporels sur le fonctionnement urbain. Dans cet article, nous présentons deux outils qui permettent i) d'explorer des données massives issues des réseaux sociaux ; ii) de construire des graphes de mobilité à partir de messages géolocalisés ; iii) de comparer deux configurations spatiales de graphes de mobilité. La collecte des tweets horodatés s'effectue au moyen d'une sonde qui interroge de façon régulière l'API de Twitter sur une zone de 50 miles autour du point central de l'Hôtel de Ville de Paris. Pour le travail présenté ici, notre analyse se centre sur les tweets dont les coordonnées géographiques sont renseignées sur le périmètre de Paris intra-muros en 2015 (5,6 millions de tweets géolocalisés sur un total de 49 millions). Le déplacement d'un « Twitto » correspond à l'intervalle parcouru entre deux tweets par un individu possédant le même numéro d'identifiant. En nous intéressant aux enregistrements des données Twitter sur les vingt arrondissements de Paris au cours l'année 2015, nous avons fait le choix de travailler sur le contenant et non pas sur le contenu des messages ou sur les attributs territoriaux de nature sémantique. Seule la géolocalisation des messages et leurs repères temporels d'une grande précision ont été exploités.

Objets connectés et mobilité urbaine

- 2 Comme toute approche innovante, fondée sur des sources de données nouvelles en sciences sociales, le tournant numérique dans l'étude des interactions socio-spatiales, annoncé par R. Rogers (Rogers, 2015), implique de prendre en compte les spécificités propres à ces données. Les *soft data* (Romele et Severo, 2015) — appelées « *soft* » en opposition aux « *hard* » data traditionnelles produites par les institutions statistiques — sont constituées principalement d'informations librement disponibles sur Internet issues du Web 2.0 (Facebook, Twitter, flux RSS...), qui s'offrent aux chercheurs pour approcher les phénomènes urbains. Elles soulèvent à la fois des enjeux méthodologiques et de pertinence scientifique. À partir de l'information fournie par les données géolocalisées du réseau social Twitter dans Paris intra-muros (**figure 1**), nous souhaitons mettre en valeur leur capacité à révéler de nouveaux découpages spatiotemporels de la ville. Ces données géolocalisées ne se réduisent nullement à un simple repérage spatial des coordonnées, mais contribuent à revisiter cet objet d'étude qu'est la ville. Croisées avec des données temporelles, interprétées en termes de durée, de fréquence, de rythmes, elles permettent d'identifier des fonctionnements spatiotemporels de l'espace géographique qui ne sont pas directement détectables par des données censitaires classiques (Lucchini *et al.*, 2014, 2016).
- 3 En scandant les rythmes urbains, la finesse des découpages temporels offerte par les enregistrements légaux des données numériques librement partagées permet d'aborder la question de la variabilité du fonctionnement de la ville selon un séquençage temporel inédit. L'analyse de ces multiples temporalités permet de saisir la vie urbaine, son intensité et la dualité rythmes routiniers/situations éphémères. Ces données permettent, par exemple, d'approcher les situations urbaines temporaires, comme le déroulement de grandes manifestations populaires (l'Armada de Rouen ou la Braderie

de Lille), sportives (le tournoi de Roland-Garros) (Lucchini *et al.*, 2016 ; Grassot, 2016), mais aussi d'événements imprévus.

Figure 1. Nombre quotidien (à gauche) et hebdomadaire (à droite) d'utilisateurs et de déplacements dans Paris intra-muros

Twitter 2015

- 4 Cette alternance entre le routinier et l'éphémère est aussi vécue par les populations émettant des tweets et se traduit par des déplacements dont les trajectoires originales ou banales sont porteuses de sens. L'enjeu est donc d'interroger à la fois la manière dont on identifie ces traces dans un ensemble de données recueillies, et les seuils spatiaux ou temporels à partir desquels surgissent des mobilités originales. En particulier, le travail sur l'espace parisien tentera de mettre en relation des déplacements urbains, avec le déroulement d'événements de nature diverse, depuis des événements programmés (festivals, manifestations sportives, ludiques, etc.) jusqu'à des événements inattendus. En ce sens, cet article tentera d'identifier les pulsations de la ville (Elissalde *et al.*, 2011).
- 5 Au-delà de l'originalité de la source utilisée se pose la question de la représentativité des données, et notamment, de l'adéquation entre les traces disponibles sur Internet et les objets à étudier. Ces données, abondantes, fournies par le quadrillage systématique de la ressource par les méthodes de *datamining* conduisent, en effet, à s'interroger sur le degré d'incertitude permettant d'inférer des interprétations sur les phénomènes (sociaux, urbains, etc.) dont elles seraient le reflet (Kwan, 2016). Cette question doit être abordée de deux manières. D'abord en considérant la population prise en compte, puis du point de vue des outils mobilisés. En explorant les données du réseau social Twitter dans un périmètre parisien, il s'agit de sonder leur capacité à révéler le rapport à l'espace urbain d'une certaine catégorie de population. Le profil sociodémographique de l'utilisateur de Twitter dessine une catégorie particulière de la population. L'âge moyen est de 22 ans, et globalement, la moitié des utilisateurs ont entre 25 et 49 ans. La répartition hommes/femmes est équilibrée et 40 % des usagers du réseau social ont un diplôme universitaire. Ils résident davantage en milieu urbain, avec 67 % d'entre eux dans des villes de plus de 100 000 habitants (sources : enquête Institut IPSOS, 2013 ; Twitter France, 2015 ; Médiamétrie, mars 2017). Géographiquement, l'activité de tweeter est un phénomène principalement parisien : 33 % des Twittos captés en France émettent en Île-de-France, en y incluant les touristes nationaux ou étrangers de passage ; d'où le choix de s'intéresser à l'ensemble de la région Île-de-France et plus précisément à l'espace parisien intra-muros. Pour respecter un usage éthique de ces informations issues des réseaux sociaux, le travail sur la géolocalisation a été mené à partir d'agrégats d'individus anonymement dénombrés, et dont la proximité géographique a été calculée selon des méthodes transparentes. Par rapport aux travaux

précédents (Elissalde *et al.*, 2011 ; Lucchini *et al.*, 2015 et 2016) qui cherchaient à identifier les rythmes de la ville en observant les variations de concentrations des Twittos dans l'espace urbain et la variabilité de la taille et de la localisation des hotspots parisiens, la présente contribution recourt aux techniques de la théorie des graphes, et insiste sur la direction des déplacements et l'évaluation des centralités des différents nœuds liés à la mobilité parisienne.

- 6 En évoquant l'idée de déplacement, nous nous référons à la définition des enquêtes CERTU, qui considèrent un déplacement comme le mouvement d'une personne d'un lieu de départ vers un lieu d'arrivée (se déplacer signifie étymologiquement « *aller de place en place* »). Le déplacement pour un usager de réseau social se repère ainsi par un tweet au lieu d'origine et un tweet à destination. Ledit déplacement peut recouvrir l'usage d'un ou plusieurs modes de transport. Certains usagers ayant une activité de tweeter stationnaire émettent plusieurs tweets au même endroit durant l'intervalle de temps sélectionné. Nous proposons ici une méthodologie qui permet de passer d'un message géolocalisé à un déplacement individuel puis à des trajectoires collectives. La **figure 2** retrace les trois étapes du processus de transformation : les tweets géolocalisés, le maillage de 600 mètres de côté pour la détection des concentrations d'utilisateurs (une maille plus large ne permet pas de voir la finesse de ces dernières), le graphe des trajectoires collectives formé par le regroupement de déplacements individuels similaires. Le maillage intermédiaire renseigne sur le niveau d'attractivité territoriale habituellement enregistré dans chacune des mailles de l'espace parisien (Maillage, intensité forte en rouge et faible en jaune, **figure 2**).

Des graphes dynamiques pour identifier les mobilités des usagers du réseau social Twitter

- 7 Le recours à des algorithmes de *clustering* permet de partitionner les données en clusters selon différents critères de proximité spatiale et temporelle, afin d'identifier des fonctionnements spatiotemporels de l'espace. Cette section présente les principaux algorithmes de clustering fondés sur la densité, comme DBSCAN, ainsi que les outils spécifiques élaborés par les auteurs de cet article pour l'exploration des données massives issues du réseau social Twitter sur l'espace parisien.

Figure 2. Des messages géolocalisés aux graphes de mobilité

L'intérêt des algorithmes de clustering DBSCAN, ST-DBSCAN et MobilityGraphs

- 8 Pour réaliser ce travail de recherche, nous nous sommes inspirés des algorithmes ST-DBSCAN (Birant et Kut, 2007) et MobilityGraphs (Landesberger *et al.*, 2016), eux-mêmes basés sur DBSCAN (Ester *et al.*, 1996) (**tableau 1**). Ce dernier algorithme partitionne les données selon une densité et requiert deux paramètres : une distance (*Eps*) et un nombre minimum de points (*MinPts*). L'idée de l'algorithme est de récupérer un nombre de points dans le voisinage d'un point donné par rapport à un seuil fixé (*MinPts*). Par la suite, l'opération est répétée pour chaque voisin du point afin de trouver l'ensemble des points du cluster. Cet algorithme simple et efficace ne requiert pas de nombre prédéfini de clusters et gère les données aberrantes. À contrario, il sollicite une bonne connaissance des données pour définir les paramètres et s'appuie exclusivement sur la proximité spatiale.
- 9 Comme mentionné précédemment, l'algorithme DBSCAN nécessite deux paramètres (*Eps* et *MinPts*). Les deux autres algorithmes, ST-DBSCAN et MobilityGraphs, se basent entièrement sur l'algorithmique DBSCAN, mais requièrent plus de paramètres. L'algorithme ST-DBSCAN introduit la notion de proximité temporelle avec un second *Eps* (« un *Eps* temporel ») ainsi qu'un paramètre non-spatial et non-temporel ($\Delta\epsilon$) pour agréger/fractionner les données en différents clusters. Ce paramètre permet de différencier des clusters adjacents qui diffèrent sur un critère d'agrégation ($\Delta\epsilon$). Les concepteurs de l'algorithme MobilityGraphs ont l'objectif de saisir les comportements de mobilité à différents pas de temps à partir des enregistrements de tweets du Grand Londres. Le volume considérable des enregistrements les a conduits à utiliser une méthode de clustering pour regrouper ces flux par grandes unités spatiales et temporelles. Leur méthode permet d'agréger des lieux et de transformer des trajectoires individuelles en flux de mobilité entre ces agrégats spatiaux. L'algorithme MobilityGraphs présente, en plus de la proximité spatiale (*Eps*), un paramètre non spatial et non temporel ainsi que plusieurs filtres pour la visualisation. Précisément, ce paramètre se base sur le nombre de liens/connexions (en valeur absolue ou relative) entre les places pour réaliser l'agrégation. L'agrégation s'opère ainsi sur des « places » proches et connectées.

Tableau 1. Quatre algorithmes de *clustering*

Algorithme	DBSCAN	MobilityGraphs	ST-DBSCAN	DigitalMob
Auteurs	Ester <i>et al.</i>	Landesberger <i>et al.</i>	Birant et Kut	Lucchini <i>et al.</i>
Année	1996	2016	2007	2017
Nombre de critères	2	2	4	4

Critères	Proximité spatiale Nombre minimum de points	Proximité spatiale Magnitude du flux	Proximité spatiale Proximité temporelle Nombre minimum de points Activité	Proximité spatiale Nombre minimum de points Activité Dynamique
----------	--	---	---	--

Création d'un algorithme spécifique aux partitionnements des données issues du réseau social Twitter

- 10 Basée sur DBSCAN, notre méthode de partitionnement de données s'inspire des critères d'agrégation développés par les auteurs des algorithmes ST-DBSCAN et MobilityGraphs. Pour expliciter la spécificité de l'algorithme DigitalMob (Lucchini *et al.*, 2017) que nous avons créé, les notions de places, de noyaux et de points nécessitent d'être éclaircies. Le point est l'entité géométrique nécessaire pour faire fonctionner les algorithmes de partitionnement. Les places correspondent à des polygones, et les points correspondent aux centroïdes des entités spatiales de ce découpage territorial (quartiers, communes, mailles de matrice...). Le noyau (ou *core point* en anglais) est le point fondateur d'un cluster. Dans le cadre de notre étude, une place correspond à une cellule de 600 sur 600 mètres du maillage effectué sur l'espace parisien. Pour rappel, nous avons choisi ce maillage car il permet de voir la finesse des concentrations d'utilisateurs ; et le point est le centroïde de cette cellule. La résolution de la matrice est un compromis entre la volonté de capter le maximum de flux et la volonté d'être précis spatialement.
- 11 Notre méthode fonctionne selon quatre paramètres (**figure 3**). Les deux premiers paramètres sont l'*Eps* (proximité spatiale) et le *MinPts* (nombre minimum de points pour identifier un noyau). Le troisième paramètre est « qualitatif » et non ajustable. Il se base sur la dynamique du noyau. Ce paramètre correspond au solde de la place (dont le centroïde constitue le noyau), il peut être négatif, positif ou nul. La place attire (attraction, solde positif) ou perd (répulsion, solde négatif) des utilisateurs. À la manière de ST-DBSCAN avec $\Delta\epsilon$, le dernier paramètre est un critère non spatial et non temporel qui correspond au niveau d'activité du cluster. Il permet de gérer les clusters adjacents qui diffèrent sur ce même critère ($\Delta\epsilon$).

Figure 3. Algorithme DigitalMob d'agrégation spatiale

```

INPUTS : le jeu de données et les paramètres d'agrégation (eps, nombre minimum de points,  $\Delta\epsilon$ )

FOR place non visitée P du jeu de données
    marquer place P comme visitée
    voisins V = places dans le voisinage de P avec la fonction regionQuery(placeP, eps)
    FOR place P' dans voisins V
        supprimer les places P' de V si la dynamique est différente de P
    ENDFOR
    IF le nombre de voisins est inférieur au nombre minimum de points requis :
        marquer la place P comme bruit
    ELSE définir un nouveau cluster C
        appeler la fonction expandCluster(P, eps, nombre minimum de points, V)
    ENDFOR

expandCluster(P, eps, nombre minimum de points, V)
    ajouter place P au cluster C
    FOR place P' dans voisins
        IF place P' n'a pas été visitée
            marquer place P' comme visitée
            V' = regionQuery(P, eps)
            FOR place P' dans voisins V'
                supprimer les places P' de V' si la dynamique et l'activité ( $\Delta\epsilon$ ) sont différentes de P
            ENDFOR
            IF le nombre de V' est supérieur ou égal au nombre minimum de points requis :
                V = V U V'
            IF place P' n'appartient à aucun cluster
                ajouter P' au cluster
    ENDFOR

regionQuery(placeP, eps)
    retourner tous les points qui sont à une distance inférieure à eps de P(incluant P)

```

Mappemonde, 2020

- 12 Si la base de l'algorithme est similaire à celle de DBSCAN, nous ajoutons néanmoins des conditions sur la création des clusters. Les points susceptibles d'appartenir à un cluster sont évalués sur leur dynamique (attraction, répulsion, caractère neutre) ainsi que sur leur intensité (on peut choisir le nombre d'utilisateurs ou le nombre de tweets pour définir l'intensité du point). Par rapport à l'intensité, un point est agrégé au cluster si la différence entre la moyenne du cluster (nombre moyen d'utilisateurs ou de tweets) et la valeur du point est inférieure au seuil fixé ($\Delta\epsilon$). L'objectif est de créer spatialement des clusters de places dont l'intensité et la dynamique sont similaires. En introduisant ce critère ($\Delta\epsilon$) lors de la seconde étape, nous affirmons que le noyau (point fondateur d'un cluster) influence son environnement proche et, par conséquent, que ces places ont une activité similaire à celle du noyau.
- 13 L'ajout du paramètre $\Delta\epsilon$ est nécessaire pour éviter une trop forte simplification spatiale. L'Eps n'est pas déterminé selon un classique histogramme de distance (« knn distance histogram »), mais fixé à partir du nombre de mouvements captés. Ainsi, c'est avec un Eps à 1000 mètres que nous captions le plus de déplacements. Le dernier paramètre MinPts est fixé à 2, c'est-à-dire que deux points suffisent pour former un cluster. Nous voulons éviter la perte d'information malgré le bruit que peut générer ce seuil.
- 14 En se basant sur un critère « qualitatif », notre méthode de partitionnement de données présente l'avantage d'être fonctionnelle selon diverses temporalités (jour, semaine, mois) et échelles spatiales (région Île-de-France, Paris intra-muros). Cette méthode a tous les avantages de DBSCAN mentionnés précédemment. Si elle est très sensible à la dynamique des places, cette méthode permet en revanche de réduire la complexité visuelle de la cartographie des flux de mobilité et de prendre en compte 80 % de la totalité des déplacements.

- 15 La combinaison de ce nouveau critère « qualitatif » (inspiré du principal critère de MobilityGraphs) avec le second paramètre ($\Delta\epsilon$) issu de ST-DBSCAN, nous permet d'identifier le fonctionnement spatiotemporel de l'espace géographique. Ainsi, nous visualisons des espaces attractifs ou répulsifs pour une temporalité donnée. Cette méthode s'associe parfaitement à l'étude des temporalités urbaines. L'analyse des pulsations de la ville peut ainsi être conduite pour révéler des situations routinières de mobilité ou la survenue de mobilités inhabituelles liées à un quelconque événement et à sa diffusion dans le temps sur l'espace parisien.

Des outils dédiés à l'exploration des données

- 16 Deux applications web ont été développées en R via le package « shiny » (Chang *et al.*, 2017) et sont hébergées par la TGIR Humanités numériques.
- 17 La première application génère les graphes de déplacements suivant l'algorithme DigitalMob (Lucchini *et al.*, 2017) que nous avons spécifiquement développé (voir § précédent) pour révéler les flux préférentiels de mobilité et les clusters d'utilisateurs à partir des données du réseau social Twitter. La seconde application permet de comparer, d'un pas de temps à l'autre, deux graphes de déplacement, et d'en évaluer le niveau de ressemblance / dissemblance.
- 18 Avec la première application (**figure 4**), en partant des données massives de Twitter enregistrées à Paris, on peut générer des graphes de déplacements préférentiels quotidiens, hebdomadaires ou mensuels. Cette application offre la possibilité d'explorer de multiples combinaisons de paramètres d'agrégation, et permet, comme dans l'exemple parisien développé dans cet article, de déterminer la meilleure combinaison de paramètres en fonction du pas de temps envisagé.

Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://journals.openedition.org/mappemonde/4297>

- 20 La seconde application (**figure 5**) mesure le degré de similarité entre deux graphes de mobilité d'un pas de temps à l'autre. Deux indicateurs locaux sont pour cela mobilisés : le premier indicateur porte sur la similarité de la distribution spatiale des nœuds et des liens du graphe ; et le second indicateur observe la similarité dans la hiérarchie des positions occupées par les clusters de mobilité (voir ci-dessous « Comparer des graphes de mobilité : mesurer la variabilité spatiotemporelle des mobilités »).

Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://journals.openedition.org/mappemonde/4297>

- 22 La section suivante présente en détail comment les tweets parisiens en 2015 peuvent se traduire sur un plan spatiotemporel en graphes de mobilité, et comment on peut observer la similarité/dissimilarité et interpréter les situations de mobilités observées.

Traduction spatiotemporelle des données Twitter

- 23 Les regroupements spatiaux et temporels apportés par notre méthode de clustering et de visualisation interactive des données Twitter en 2015 permettent une lecture rapide

d'une grande masse d'informations géolocalisées. Pour conduire l'observation des mobilités et les interpréter, nous avons choisi d'avoir la possibilité d'interroger plusieurs échelles temporelles. Selon le pas de temps choisi — de la journée au mois —, la taille, l'organisation et la configuration spatiale du graphe varie et se complexifie (**figure 6**). On observe ici que les émissions de messages des utilisateurs de Twitter en 2015 regroupées sous forme de graphes dessinent des mobilités dont les configurations sont porteuses de sens.

Figure 6. Graphe quotidien, hebdomadaire et mensuel

Visualiser des mobilités

- 24 Les croisements spatiaux et temporels proposés par les graphes de mobilité nous éclairent sur la grande variabilité et sur la complexité des mobilités des usagers de Twitter à Paris. Ils révèlent en outre des lieux incontournables pour ces usagers parisiens et des déplacements préférentiels.
- 25 Contrairement à des déplacements qui s'effectueraient sur un réseau unique, fixe et constant (le réseau ferré par exemple), le graphe de mobilité parisienne formé par les déplacements des usagers de ce réseau social est évolutif dans le temps et dans l'espace. Cette variabilité du graphe de mobilité reflète l'étendue de la complexité des déplacements des usagers, et le nombre de leurs interactions avec le territoire parisien. Parvenir à visualiser cette variabilité sur le nombre de nœuds et sur le nombre de liens du graphe est aussi un gage de la bonne réactivité de notre méthode de *clustering* pour rendre compte de la complexité des déplacements.
- 26 Pour autant, s'il existe une variabilité et une complexité inhérentes aux mobilités des usagers de Twitter, les graphes générés rendent visibles des polarisations-clés et des flux préférentiels sur l'espace parisien, à la journée, à la semaine ou au mois. Des lieux incontournables sont détectés, tandis que d'autres lieux conservent une influence plus modeste et périphérique. Il arrive également que des clusters momentanés surgissent à la faveur d'événements sociaux, culturels ou sportifs, et, que ces derniers, en se déployant sur l'espace urbain, introduisent des modifications du graphe routinier.

- 27 Parfaitement visibles et identifiables, les polarisations majeures du graphe de mobilité font ressortir dans l'espace parisien un triangle de flux préférentiels et un ensemble de *hotspots* avec par exemple des lieux remarquables comme l'Hôtel de Ville, Notre-Dame de Paris, la tour Eiffel, l'Étoile et l'Arc de Triomphe, le Louvre, mais aussi les Champs-Élysées, les grands hôtels et les grands magasins, le Sacré Cœur, les portes de Clignancourt et de Saint-Denis, la place de la République, les gares... Le niveau de concentration des connexités vers les nœuds occupant une position hiérarchique majeure (Top 10) est tel que 10 % des sommets les plus importants du graphe concentrent 20 à 40 % de tous les liens du graphe. En général, une fois cette polarisation majeure et récurrente détectée sur le graphe de mobilité, on constate une extrême variabilité pour le reste des niveaux inférieurs. On observe ainsi des graphes de mobilité avec une double facette : une très forte polarisation au sommet et une variabilité importante pour les niveaux inférieurs.
- 28 À l'échelle de la journée, cette nouvelle manière de représenter les mobilités dans Paris permet de visualiser la variabilité introduite par des journées particulières (une victoire sportive du PSG en ligue 1 au Parc des Princes, les attentats de novembre 2015, **figure 7**) ou de suivre les perturbations introduites par un événement dont les effets se diffusent plusieurs jours durant sur le graphe de mobilité (exemple de plusieurs journées du tournoi international de tennis à Roland-Garros, **figure 8**). Les attentats du 13 novembre 2015 ont eu une répercussion sensible sur la mobilité des usagers de Twitter, avec un double effet visible dès la journée suivante : d'un côté, une chute du volume de déplacements, et de l'autre côté, une convergence de ces derniers vers l'Hôtel de Ville de Paris (**figure 7**), alors que le volume habituel d'usagers se multipliait exceptionnellement par cinq (**figure 1**).

Figure 7. Lendemain des attentats du 13 novembre 2015

- 29 Dans le cas de l'Open de tennis français de Roland-Garros, qui prend place la dernière semaine de mai et la première semaine de juin chaque année à la porte d'Auteuil, il est possible d'observer la création d'un cluster attractif à la porte d'Auteuil pendant le tournoi (exemples du vendredi, deux jours avant la finale du tournoi, et du dimanche, jour de la finale, **figure 8**), alors que ce cluster attractif disparaît une fois le tournoi révolu (deux jours après la finale, **figure 8**). Cette perturbation du graphe de mobilité parisienne des usagers de Twitter est également perceptible à l'échelle de la semaine (**figure 9**).
- 30 Les exemples développés ci-dessus illustrent le fait que des événements programmés sur un calendrier festif, social, sportif ou culturel, ou bien d'autres événements inattendus peuvent modifier en partie le graphe de mobilité parisienne des usagers du réseau social Twitter sur une temporalité donnée. Ceci mobilise la question de la

détection d'un événement à partir de la modification observée sur un graphe de mobilité. Pour aller plus loin sur ce point, nous avons développé une mesure des niveaux de ressemblance et de dissemblance de deux graphes de déplacements préférentiels d'un pas de temps à l'autre.

Figure 8. Impact de l'évènement Roland-Garros sur le graphe de mobilité quotidienne

Comparer des graphes de mobilité : mesurer la variabilité spatiotemporelle des mobilités

- 31 Nous mobilisons ici une mesure du niveau de ressemblance/dissemblance entre deux graphes de mobilité, pour approcher plus efficacement la variabilité spatiotemporelle des mobilités. Ce travail participe aussi d'une approche qui vise à lire de manière plus aboutie la permanence et le changement dans les mobilités observées. Plus globalement, cette approche rend possible l'analyse des pulsations de la ville, ou de ce que l'on nomme « la ville éphémère », qui correspond à la variation des interactions entre des lieux et des individus, fruit de l'ensemble de nos mobilités et de nos multiples activités sur l'espace urbain.
- 32 La mesure de la similarité porte à la fois sur les nœuds et les liens du graphe de mobilité d'un pas de temps à l'autre. Nous nous appuyons ici sur un premier indicateur de similarité, qui nous informe sur la ressemblance/dissemblance spatiale de la distribution des nœuds et des liens du graphe sur l'espace parisien. Cet indicateur permet d'identifier une récurrence spatiale, c'est-à-dire la continuité dans le temps de la polarisation spatiale intra-urbaine des zones d'intérêt (présence des mêmes clusters aux mêmes endroits d'une séquence temporelle à une autre, information exprimée en pourcentage ; 100 % de similarité spatiale correspond à deux semaines spatialement identiques). À partir de ces nœuds communs, la permanence des liens est aussi analysée en observant leur croissance/décroissance et leur intensité. Un second indicateur de similarité nous permet ensuite d'observer la ressemblance/dissemblance dans la hiérarchie des positions occupées par les clusters de mobilité (présence des mêmes Top 10 aux mêmes endroits d'une séquence temporelle à une autre, information exprimée en pourcentage ; 100 % de similarité hiérarchique correspond à deux

semaines hiérarchiquement identiques). Cette récurrence hiérarchique, ou continuité dans la hiérarchie des polarisations intra-urbaines, s'envisage à partir de la présence du même positionnement hiérarchique de clusters et *hotspots* dans les quartiers parisiens d'une séquence temporelle à une autre.

Figure 9. Impact de l'évènement Roland-Garros sur le graphe de mobilité hebdomadaire

- 33 En reprenant l'exemple du tournoi international de Roland-Garros, la **figure 10** compare deux semaines consécutives (**figure 9**), l'une pendant le tournoi, l'autre après le tournoi. Un changement dans les mobilités observées d'une semaine sur l'autre est en effet particulièrement visible : le niveau de similarité spatiale n'est que de 40 % pour les nœuds du graphe de mobilité, et les liens sont dissemblables entre les deux semaines à plus de 80 %. Les cinq premières polarisations qui ressortent la première semaine sont, dans l'ordre, les clusters d'intensité (au choix, nombre d'utilisateurs ou nombre de tweets) formés à l'Hôtel de Ville, à la tour Eiffel, au Louvre, à Roland-Garros et à Saint-Michel ; tandis que la seconde semaine, les cinq premiers clusters d'intensité sont formés autour de la tour Eiffel, de l'Hôtel de Ville, du Louvre, de l'Étoile et de Beaubourg. La position hiérarchique des Top 5 se modifie entre les deux semaines, en révélant l'impact sensible de l'évènement Roland-Garros.

Figure 10. Changement de mobilité pendant et après le tournoi de Roland-Garros

- 34 Précisément pour l'évènement Roland-Garros, une fois le tournoi révolu, on observe une perte de puissance de l'attraction des usagers autour du cluster localisé à la porte d'Auteuil (**figure 10**), et une bascule de l'intérêt vers d'autres points attracteurs du territoire parisien, aux alentours de Radio France, du Trocadéro, du Pont de l'Alma et de la tour Eiffel, de la porte-Maillot, de Clichy, de Montparnasse, d'Austerlitz, et de Montmartre. Certains déplacements préférentiels nord-sud sur la rive droite de la Seine perdent en puissance (tracés bleus, **figure 10**), y compris sur l'axe principal de l'Étoile

vers le Louvre et l'Hôtel de Ville, en passant par les Champs Élysées. D'autres déplacements préférentiels se renforcent au contraire (tracés rouges, **figure 10**), entre le Trocadéro, la tour Eiffel et le Louvre, mais aussi sur une direction Sud-Ouest/Nord de Paris, du Trocadéro à Montmartre, ainsi qu'en direction de la rive gauche depuis le centre parisien vers Montparnasse et Austerlitz.

- 35 Ces éléments issus des mesures de similarité spatiale et hiérarchique concourent à illustrer l'ampleur du changement des mobilités observées d'une semaine sur l'autre et aussi à rendre compte de formes de récurrences.

Conclusion et perspectives

- 36 Cette recherche conduite sur les messages géolocalisés des usagers du réseau social Twitter, enregistrés durant l'année 2015 à Paris, participe d'une approche scientifique qui interroge la capacité des *soft data* et des données issues des objets connectés à constituer une ressource nouvelle pour la compréhension des rythmes urbains et des mobilités dans la ville.
- 37 Nous avons montré qu'en mobilisant un algorithme de *clustering* dédié, il est possible de gérer des données massives et de générer des graphes de mobilité des usagers d'un réseau social. Pour visualiser ces mobilités, notre méthodologie permet de passer d'un message géolocalisé à un déplacement individuel puis à des trajectoires collectives. Dans cet article, nous avons présenté un outil de visualisation des mobilités parisiennes réalisé en nous appuyant sur des graphes dynamiques, et également un outil qui permet de comparer deux graphes de mobilité d'une séquence temporelle à une autre — deux applications interactives sous R -. Ces mobilités parisiennes rendues visibles au travers des graphes dynamiques illustrent les déplacements préférentiels des usagers du réseau social dans Paris, ainsi que des zones de concentration et pôles d'intérêt.
- 38 Avoir la possibilité de visualiser et de comparer la ressemblance /dissemblance des flux de déplacements préférentiels et des pôles d'intérêt invite à approfondir le champ de l'analyse des rythmes urbains, et l'étude de la variabilité des mobilités. Car les formes de permanence et de changement de mobilités observées, leur variabilité inhérente et l'accentuation de l'incertitude au déplacement, ont permis d'approcher la complexité de la vie urbaine. Ici commence le questionnement de ce qu'on nomme « la ville éphémère », c'est-à-dire la variation des interactions entre des individus et des lieux sur des pas de temps qui rythment la vie urbaine, et qui révèle, au-delà de l'apparent mouvement brownien des déplacements, à la fois de l'attendu et de l'incertain.

BIBLIOGRAPHIE

BIRANT D., KUT A. (2007). "ST-DBSCAN: An algorithm for clustering spatial-temporal data". *Data & Knowledge Engineering*, "Intelligent Data Mining", vol. 60, n° 1, p. 208-221. <https://doi.org/10.1016/j.datak.2006.01.013>

- CERTU. (2008). *Résultats des enquêtes nationales Transport. Glossaire*. Ministère du Développement durable, 228 p.
- CHANG W., CHENG J., ALLAIRE J. J., XIE Y., MCPHERSON J. (2017). shiny : Web Application Framework for R. En ligne : <https://CRAN.R-project.org/package=shiny>
- ELISSALDE B., LUCCHINI F., FREIRE-DIAZ S. (2011). « Mesurer la ville éphémère ». In PUMAIN D., MATTEI M-F. (éd.), *Données Urbaines*, n° 6, Paris : Anthropos, p. 105-122.
- ESTER M., KRIEGLER H.-P., SANDER J., XU X. (1996). A Density-based Algorithm for Discovering Clusters. A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Presented at the Second International Conference on Knowledge Discovery and Data Mining, KDD'96, Portland, Oregon, USA, 1996.
- GRASSOT L., (2016). *Mobilités événementielles et espace urbain : Exploitation des données de téléphonie mobile pour la modélisation des grands événements urbains*. Rouen : Université Rouen Normandie, thèse de doctorat de géographie, 556 p.
- KWAN M.P. (2016). "Algorithmic geographies: Big data, algorithmic uncertainty, and the production of geographic knowledge". *Annals of the American Association of Geographers*, vol. 106, n° 2, p. 274-282.
- LANDESBERGER VON T., BRODKORB F., ROSKOSCH P., ANDRIENKO N., ANDRIENKO G., KERREN A. (2016). "MobilityGraphs: Visual Analysis of Mass Mobility Dynamics via Spatio-Temporal Graphs and Clustering". *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, n° 1, p. 11-20. <https://doi.org/10.1109/TVCG.2015.2468111>
- LUCCHINI F., ELISSALDE B., GRASSOT L., SALAMAND Y. (2014). "Urban events and emerging phenomena". Proceedings of ICCSA 2014, The 4th International Conference on Complex Systems and Applications, Le Havre, France, 2014, <http://iccsa2014.univ-lehavre.fr/iccsa2014proceedings.pdf>.
- LUCCHINI F., ELISSALDE B., GRASSOT L., BAUDRY J. (2015). *Mobile phone use, events and emergent phenomena*. Proceedings of the European Colloquium on Theoretical and Quantitative Geography, Bari, Italie, 2015.
- LUCCHINI F., ELISSALDE B., GRASSOT L., BAUDRY J., (2016), « Paris Tweets. Données numériques géolocalisées et événements urbains ». *Netcom*, vol. 30, n° 3-4. En ligne : <https://netcom.revues.org/2579>.
- LUCCHINI F., ELISSALDE B., GILLET O., GRASSOT L., BAUDRY J., (2017), *Social networks, events and cities*. Proceedings of the European Colloquium on Theoretical and Quantitative Geography, York 2017.
- MASSOT M.-H., ORFEUIL J.-P. (2005). « La mobilité au quotidien, entre choix individuel et production sociale ». *Cahiers internationaux de Sociologie*, Presses universitaires de France, CXVIII, p. 81-100.
- SEVERO M., ROMELE A. dir., (2015). *Traces numériques et territoires*. Paris : Presses des Mines, coll. « Territoires numériques », 270 p.
- ROGERS R. (2015). « Au-delà de la critique big data. La recherche sociale et politique à l'ère numérique ». In SEVERO M., ROMELE A. (dir.), *Traces numériques et Territoires*, coll. « Territoires numériques ». Paris : Presses des Mines, p. 13-32.
- VIRILIO P. (2009). *Le Futurisme de l'instant : Stop-Eject*. Paris : Éditions Galilée, 104 p.

RÉSUMÉS

Avec le tournant numérique, les données issues des objets connectés constituent une ressource nouvelle pour la compréhension des rythmes urbains et des mobilités dans la ville. À l'aide des messages géolocalisés (1 540 000 tweets) des usagers du réseau social Twitter enregistrés durant l'année 2015, cet article présente un outil de visualisation des mobilités parisiennes en s'appuyant sur des graphes dynamiques. Nous montrons comment un algorithme de *clustering* nous permet de passer d'un message géolocalisé à un déplacement individuel puis à des trajectoires collectives. Ces dernières illustrent les mobilités préférentielles des usagers du réseau social dans Paris, ainsi que des zones de concentration et pôles d'intérêt.

With the digital turn, data from connected devices are a new resource for understanding urban rhythms and mobility. Using the geolocated messages (1, 540, 000 tweets) of Twitter social network users recorded in 2015, this article uses dynamic graphs to visualize Parisian mobilities. Using a clustering algorithm, we show how to move from a geolocated message to an individual displacement and on to collective trajectories. These trajectories illustrate the preferential mobility of users of the social network in Paris, as well as the areas of concentration and poles of interest.

La transformación o giro digital facilita el acceso a la información de los dispositivos electrónicos conectados. Es un nuevo recurso para comprender los ritmos urbanos y la movilidad en la ciudad. Este artículo presenta una herramienta con gráficos de flujos dinámicos para visualizar la movilidad en París. Para ello utiliza los mensajes de los usuarios de la red social de Twitter registrados y geolocalizados durante el año 2015 (1,540,000 tweets), y mediante un algoritmo de agrupamiento convierte en flujos dinámicos colectivos un mensaje geolocalizado con una trayectoria individual. De este modo se puede visualizar movilidades preferidas, áreas de interés y de concentración de los usuarios de redes sociales en París.

INDEX

Thèmes : Enjeux de l'information territoriale locale ouverte et participative

Palabras claves : movilidad urbana, Twitter, gráficos de flujos dinámicos, París, acontecimientos

Keywords : urban mobility, Twitter, dynamic graphs, Paris, events

Mots-clés : mobilité urbaine, Twitter, graphes dynamiques, Paris, évènements

AUTEURS

FRANÇOISE LUCCHINI

UMR CNRS IDEES, Université Rouen Normandie

OLIVIER GILLET

UMR CNRS IDEES, Université Rouen Normandie

BERNARD ELISSALDE

UMR CNRS IDEES, Université Rouen Normandie

LÉNY GRASSOT

INSERM, Centre Léon Bérard

JULIEN BAUDRY

EA LITIS, université du Havre

ARMELLE COUILLET

UMR CNRS IDEES, Université Rouen Normandie