

La mobilité parisienne des usagers de Twitter

LES AUTEUR.E.S

Françoise LUCCHINI, Bernard ELISSALDE, Olivier GILLET, Lény GRASSOT,
Julien BAUDRY, Armelle COUILLET

RÉSUMÉ

Cet article, fondé sur le traitement des messages Twitter géolocalisés, tente d'identifier des configurations réticulaires de l'espace urbain à partir des déplacements des émetteurs de tweets dans Paris et l'Île-de-France. Nous nous appuyons sur une méthode algorithmique de réalisation de graphes dynamiques, pour révéler des mobilités en combinant des regroupements spatiaux et des séries temporelles. Cette méthode met en évidence les variations spatiotemporelles de mobilité sous forme de polarités ou de vacuités récurrentes. Au-delà des trajectoires préférentielles empruntées par les usagers, des clusters parisiens émergent de ces graphes, révélant des hubs récurrents, des points de passage et des polarisations à l'occasion d'un événement particulier. Les graphes dynamiques permettent de traduire la « ville éphémère » et l'évolution des comportements des usagers du réseau social Twitter.

MOTS CLÉS

mobilité urbaine, Twitter, graphes dynamiques, Paris, centralité, événementiel

ABSTRACT

This paper, based on recorded geolocated Twitter messages, attempts to identify urban patterns by separating various modes of urban events from urban routine. We set out to identify relationships between types of user, types of event and the transmission of messages on the Twitter social network for Paris and the Île-de-France region. To do so, we used a graph method making it possible to combine spatial and temporal simplifications. This method highlights the spatiotemporal variations of nodes and flows over long time series of spatial situations. Beyond preferential trajectories of Twitter users, Parisian clusters appear: hotspots, crossing points, clusters during a specific event. Dynamics graph can reveal the "ephemeral city" and the behaviours development of Twitter users in Paris.

KEYWORDS

Urban mobility, Twitter, Graph dynamics, Paris, Hotspots, Events

INTRODUCTION – OBJETS CONNECTÉS ET MOBILITÉ URBAINE

La réalité des pratiques urbaines et des mobilités dans la ville est plurielle. Pour les approcher, des sources d'information, issues de différents capteurs urbains et/ou d'objets connectés, sont désormais utilisées par les géographes. Ces données numériques géolocalisées ont la capacité à révéler des formes de mobilité dans l'espace et plus généralement à identifier de nouveaux usages urbains. Les auteurs de cet article ont choisi de se pencher sur la réalité des comportements des usagers du réseau social Twitter sur Paris pendant plusieurs mois de l'année 2015. Seuls la géolocalisation des messages et leurs repères temporels d'une grande précision ont été exploités (le contenu du message n'a pas été analysé). Nous proposons dans ce travail une méthodologie fondée sur la construction de graphes capables de saisir, tout en respectant des critères d'anonymisation des données, la dynamique de mobilité des personnes connectées au réseau social Twitter sur l'espace parisien.

1. TRADUIRE LA VILLE ÉPHÉMÈRE EN GRAPHES DYNAMIQUES

Les *soft data* (Severo, 2015), sont appelées « *soft* » en opposition aux « *hard* » *data* traditionnelles produites par les institutions statistiques, constituées principalement par les informations librement disponibles sur internet issues du Web 2.0 (Facebook, Twitter, flux RSS...). Par-delà l'originalité de ces sources, se pose la question de la représentativité des données et notamment de l'équivalence entre les traces disponibles sur internet et les objets à étudier. En explorant les données du réseau social Twitter dans un périmètre francilien puis parisien, nous testons leur capacité à révéler le rapport à l'espace urbain d'une certaine catégorie de population (les *twittos*)¹. Par rapport aux travaux précédents (Lucchini *et al.*, 2014 et 2016 ; Grassot, 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 clusters (autocorrélation spatiale et méthode LISA²), la présente contribution insiste sur la mobilité et les directions des déplacements. Nous recourons ici aux techniques fondées sur l'analyse des graphes afin de détecter des flux de mobilité et de calculs de centralité des différents nœuds liés à cette mobilité selon des durées variables. Nous souhaitons ici ajouter et comparer deux méthodes d'agrégation spatiale : clustering et graphes.

Pour ce faire, nous nous sommes inspirés du programme Mobility Graphs (Landesberger *et al.*, 2015) appliqué aux enregistrements de tweets dans le Grand Londres, afin de saisir les comportements de mobilité à travers les volumes de mouvement (*flow magnitude*) et les différents pas de temps. Le volume considérable des enregistrements de déplacements a conduit ces auteurs à utiliser des méthodes de *clustering* pour regrouper ces flux par grandes unités spatiales et temporelles. Leur méthode permet d'agréger des lieux et de transformer des flux entre les lieux individualisés en flux entre des agrégats spatiaux (c'est-à-dire des quartiers ou des lieux précis). Un point important de méthode porte sur la procédure de *clustering* mobilisée (Kisilevich *et al.*, 2009). Dans Mobility Graphs, la méthode d'agrégation spatiale s'appuie sur l'algorithme DBSCAN en y ajoutant les caractéristiques de proximité spatiale et la force des flux. Cet algorithme capture et met en évidence des mouvements à l'échelle de l'ensemble de l'agglomération londonienne. Nous avons également suivi Derya et Alp (2007) dans l'extension qu'ils proposent (ST-DBSCAN). Contrairement aux algorithmes existants de regroupement basé sur la densité, ST-DBSCAN peut regrouper les données spatio-temporelles en fonction aussi bien de leurs attributs non spatiaux, que spatiaux et temporels. De plus cet algorithme résout le problème posé par les contrastes de valeurs des objets situés en frontière de deux clusters voisins. Par ailleurs ce travail a tenu compte des interrogations qui se sont faites jour sur la valeur heuristique des algorithmes utilisés dans le traitement des *big data* (Kwan, 2016).

2. UNE ACTIVITÉ FRANCILIENNE MODESTE CONTRE UNE INTENSE ACTIVITÉ PARIS INTRA-MUROS

Le traitement du corpus de 1 300 000 messages géolocalisés à l'échelle de la région Île-de-France – englobant plus de 1 500 communes autour de Paris – révèle un fort contraste entre Paris intra-muros et le reste de l'espace francilien, en lien avec les choix de localisation et de

1 Géographiquement, l'activité de twittage 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. Le profil sociodémographique de l'utilisateur de Twitter dessine, lui aussi, une catégorie particulière de la population. L'âge moyen est de 22 ans et, globalement, 61 % des utilisateurs ont moins de 35 ans (source : enquête institut IPSOS 2013).

2 *Local Indicators of Spatial Association* (Anselin 1995).

déplacements préférentiels des twittos. Si l'activité journalière du réseau social Twitter couvre plus de 70 % de Paris intra-muros, en revanche elle n'anime qu'un quart du territoire francilien. Les niveaux d'activité mensuels et hebdomadaires confirment cette tendance totalisant entre les deux cinquièmes et la moitié du territoire francilien concerné par le réseau social Twitter, contre une couverture de plus de 80 % pour Paris intra-muros. En volume, les déplacements des usagers du réseau social Twitter s'élèvent à plus de 5 000 mouvements par semaine pour une moyenne de 10 000 utilisateurs géolocalisés. Au fil des semaines la structure du graphe des déplacements reste relativement stable (environ 70 nœuds chaque semaine), excepté pour la semaine du 13 novembre 2015³. La multiplicité des centres attracteurs lui confère une forte connectivité, confirmée par l'indice bêta de niveau élevé (supérieur à 8) avec des *maxima* lors des semaines 25 (du 21 au 26 juin) et 37 (mi-septembre).

3. LE TRAITEMENT DES TWEETS SERT DE RÉVÉLATEURS AUX PHÉNOMÈNES DE CENTRALITÉ

Les usagers de ce réseau social n'investissent pas de manière ubiquiste l'espace parisien, et associent des formes de *hotspots* et des lieux à fréquentation sporadique, des axes préférentiels et d'autres temporaires. Derrière l'apparent mouvement brownien des itinéraires suivis par les usagers de Twitter, les lieux d'émission se concentrent sur un petit nombre de sommets récurrents. Il arrive cependant que le dispositif habituel du graphe soit bousculé par des noyaux moins fréquents, qui s'intercalent en révélant une activité émergente et momentanée. Les classements les plus fréquents sur Paris intra-muros proposent quatre types de polarisations majeures :

- Deux quartiers concentrent 90 % de tous les déplacements : quartier Châtelet à l'Hôtel de ville de Paris, abords de Notre-Dame de Paris et du Marais, et quartier de la Tour Eiffel et du Champ de Mars.
- Des polarisations dont la fréquence correspond à plus de 30 % : on y trouve l'Arc de Triomphe, les grands hôtels parisiens autour des Champs-Élysées et le faubourg St-Honoré, ainsi que le quartier de la Chaussée d'Antin, proche des grands magasins Galerie Lafayette et Printemps.
- Le troisième type de polarisations forme un ensemble d'une dizaine de sommets de moindre fréquence d'apparition (12 à 25 % de la période observée), émergeant à la faveur de concentrations plus éphémères. On y trouve des zones touristiques importantes (Sacré Cœur et Montmartre, Louvre et Tuileries), des manifestations culturelles ou sportives, et d'autres événements plus tragiques comme les rassemblements aux lendemains des attentats de novembre 2015 (place de la République, et le Bataclan lui-même).
- Un dernier type de polarisation apparaît dans les graphes uniquement de manière occasionnelle : St-Germain-des-Prés, Hippodrome de Longchamp, Père Lachaise présentent des indices de centralité faibles, et n'émergent qu'à l'occasion d'un événement particulier.

4. DES CONFIGURATIONS DE GRAPHE VARIANT SELON LE TYPE D'ÉVÈNEMENT

Les lieux successifs d'émission de messages des utilisateurs de Twitter regroupés sous forme de graphes dessinent des mobilités à mettre en relation avec d'éventuels événements urbains (Lucchini *et al.*, 2016) ou d'un jour particulier (jours fériés et lendemain des attentats). Les indicateurs servant aux diagnostics reprennent les indices définis par K. Kansky (1963), repris par le GDR Réseaux⁴.

3 Un pic de communication inhabituel, correspondant aux attentats du Bataclan, est identifiable durant la semaine 46 (9 au 15 novembre) avec 40 000 tweets échangés sur l'ensemble de l'Île-de-France.

4 groupefmr.hypotheses.org

En règle générale, les jours fériés produisent une contraction des graphes de déplacements observables par une diminution globale du nombre de nœuds et d'arêtes de l'ordre de de 20 à 50 %, accompagnée d'une forte réduction de l'intensité des flux (indice *thêta*). Par contre, les particularités propres à chaque jour férié ne transforment pas les graphes de mobilité de la même façon. D'un indicateur à l'autre, la réactivité des graphes en matière d'intensité (indice *thêta*) de connexité (indice *bêta*) et de connectivité (indices *alpha* et *gamma*) varie fortement selon les journées étudiées. Selon le type d'événement (commémorations, manifestations), selon la durée des congés, et selon la manière dont cet événement investit l'espace public (rassemblements, défilés) et les formes de localisation qu'il prend (ubiquiste, mono-site ou pluri-sites), le graphe résultant variera sensiblement.

Cependant, quelles que soient les différentes mesures de la centralité (Shimbel, *betweenness centrality*) utilisées, rares sont les événements organisés susceptibles de déformer cette structure pérenne. La situation du samedi 14 novembre 2015, lendemain des attentats du Bataclan, est à cet égard tout à fait atypique. Confrontés au choc des attentats, les usagers de Twitter, comme le reste de la population, ont réduit leurs déplacements (cf. indices), mais pas leurs émissions de tweets. Le besoin d'échanger et le sursaut de la société face à ce drame se sont traduits par une croissance exponentielle (quatre fois plus ce jour-là) du nombre de tweets envoyés. Le cas du tournoi de tennis de Roland Garros est également original. La capacité du quartier de la porte d'Auteuil à « déformer » le graphe d'ensemble est atypique au point de totaliser certains jours 28 % des tweets géolocalisés enregistrés dans Paris. Pour autant, le volume des spectateurs pour un événement n'est pas un gage d'omothétie en matière de twittage. Ainsi le festival de musique rock des journées Solidays sur l'hippodrome de Longchamp regroupe moins de twittos émetteurs que le quartier du Marais ou le quartier de Charonne dans l'Est parisien sans événement particulier.

CONCLUSION

La traduction spatiale et temporelle des informations géolocalisées du réseau social Twitter révèle de fortes disparités dans l'animation et la fréquentation des quartiers franciliens et parisiens. La structure des graphes des déplacements des usagers de Twitter présente des variabilités spatiales et temporelles considérables dans les itinéraires suivis. Cette variabilité semble inhérente au degré d'incertitude contenu dans l'acte de twittage au niveau de l'individu. Pour autant, les agrégats spatiaux et temporels produits par l'algorithme ont permis d'identifier la récurrence des polarisations majeures et des trajectoires préférentielles dans l'espace parisien. Seuls quelques rares événements sportifs ou dramatiques, et quelques phénomènes éphémères dans des quartiers, semblent susceptibles de corriger de leur empreinte cette régularité. Étudier les unes et les autres permet d'en saisir le développement et la réactivité des comportements.

RÉFÉRENCES

- Birant D., Kut A., 2007, « ST-DBSCAN: An algorithm for clustering spatial-temporal data », *Data & Knowledge Engineering, Intelligent Data Mining*, n° 60, p. 208-221.
- Grassot L., 2016, *Mobilités événementielles et espace urbain. Exploitation des données de téléphonie mobile pour la modélisation des grands événements urbains*, thèse de géographie, université de Rouen.
- Kansky K.J., 1963, *Structure of Transportation Networks: Relationships Between Network Geometry and Regional Characteristics*, Chicago, University of Chicago.
- Kisilevich S., Mansmann F., Nanni M., Rinzivillo S., 2010, « Spatio-temporal clustering data », in O. Maimon et L. Rokach (dir.), *Data Mining and Knowledge Discovery Handbook*, Boston, MA, Springer, p. 855-874.

- Kwan M.P., 2016, « Algorithmic geographies: Big data, algorithmic uncertainty, and the production of geographic knowledge », *Annals of the American Association of Geographers*, n° 106, p. 274-282.
- Landesberger T., Von Brodtkorb 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*, n° 22, p. 11-20.
- Lucchini F., Elissalde B., Grassot L., 2014, « Urban events and emerging phenomena », communication à *JCCSA 2014*, Le Havre.
- Lucchini F., Elissalde B., Grassot L., 2015, « Mobile phone use, events and emergent phenomena », communication à *ECTQG 2015*, Bari, Italie.
- Severo M., Romele A., 2017, *Traces numériques et territoires*, Paris, Presses des Mines [en ligne : www.gis-cist.fr/portfolio/traces-numeriques-et-territoires].
- Weiler A., Grossniklaus M., Scholl M.H., 2017, « Survey and Experimental Analysis of Event Detection Techniques for Twitter », *The Computer Journal*, n° 60, p. 329-346.

LES AUTEUR.E.S

Françoise Lucchini

Université de Rouen – IDEES
francoise.lucchini@univ-rouen.fr

Bernard Elissalde

Université de Rouen – IDEES
bernard.elissalde@univ-rouen.fr

Olivier Gillet

Université de Rouen – IDEES
olivier.gillet@univ-rouen.fr

Lény Grassot

Université de Lyon – LAET
l.grassot@gmail.com

Julien Baudry

Université du Havre – Litis
julien.baudry@univ-lehavre.fr

Armelle Couillet

Université de Rouen – IDEES
armelle.couillet@cnrs.fr