

Etudier les traces numériques de la ville : la frontière entre en-ligne et hors-ligne

Ces dernières années, les nouvelles technologies ont changé profondément la vie de la ville. Un des changements les plus révolutionnaires concerne la diffusion des médias numériques, dont l'intérêt est que toute interaction qui les traverse laisse des traces qui peuvent être facilement enregistrées et traitées grâce aux méthodes numériques. Ces méthodes se sont montrées très adaptées aux études urbaines. Cependant, leur usage sollicite aujourd'hui plusieurs questions théoriques et méthodologiques. Cet article veut être l'occasion de réfléchir sur la valeur et l'emploi des traces numériques dans le contexte d'analyses en Science de l'Information et de la Communication (SIC), notamment concernant la ville. En particulier, nous voulons réfléchir sur comment les méthodes numériques gèrent la frontière entre en-ligne et hors-ligne.

Introduction

Ces dernières années, les nouvelles technologies ont changé profondément la vie de la ville (Boullier, 1999). Un des changements les plus révolutionnaires concerne la diffusion du numérique, et notamment du réseau Internet (Castells, 2002). Ce qui rend ce changement particulièrement intéressant est le fait qu'il affecte à la fois la ville elle-même et la façon de l'étudier et de la gérer (Benkler, 2006). En effet, l'intérêt des médias électroniques est que toute interaction qui les traverse laisse des traces qui peuvent être facilement enregistrées, massivement stockées et aisément récupérées (Rogers, 2009). Récemment, un nouveau groupe des méthodes appelées « méthodes numériques » a été développé pour traiter ce type de données. Ces méthodes se sont montrées très adaptées aux études urbaines. Cependant, leur usage aujourd'hui sollicite plusieurs questions théoriques et méthodologiques (Marres et Weltevrede, 2013).

Cet article veut être l'occasion de réfléchir sur la valeur et l'emploi des traces numériques dans le contexte d'analyses en Science de l'Information et de la Communication (SIC), notamment concernant la ville. En particulier, nous voulons réfléchir sur comment les méthodes numériques gèrent la frontière entre le numérique et l'espace physique et c'est pour cela que les études de l'espace urbain se démontrent le terrain idéal. Cet article veut, alors, se confronter sur un sujet classique de la discussion du numérique, le rapport entre hors-ligne et en-ligne. Evidemment la question du rapport entre en-ligne et hors-ligne n'est pas nouvelle et la distinction même a été à plusieurs reprises mise en discussion. Sans tomber dans les excès d'affirmer en manière absolue cette distinction ou de la rejeter à priori, nous voulons nous interroger sur le type de continuité ou discontinuité générée aujourd'hui par les traces numériques de la ville.

Le cas de la ville nous semble particulièrement intrigant pour son être essentiellement physique mais, en même temps, pour son devenir numérique. Comme le dit Helen Couclelis (2004), les villes numériques ne sont pas seulement un phénomène d'ordre technologique, mais sont une construction qui est très ancrée dans l'idéologie, la culture et les processus sociaux. Alors, quand on étudie un phénomène urbain à travers des traces que les acteurs ont laissées sur un blog ou un réseau social, sommes-nous en train d'étudier le phénomène qui a lieu dans l'espace de la ville hors-ligne ? Ou sommes-nous en train d'étudier la projection en-ligne du phénomène ? Ou cette distinction doit-elle être complètement abandonnée ?

L'article sera organisé en deux parties. Dans une première partie, nous aborderons le sujet du point de vue théorique. Après avoir présenté brièvement la question de la traçabilité de la ville, nous approfondirons les problématiques liées aux méthodes numériques. Dans une deuxième partie, nous arriverons à l'application des méthodes dans les études urbaines. Nous mettrons à l'épreuve notre pensée à travers un cas pratique. Nous avons choisi l'exemple des analyses des géolocalisations parce que, pour le fait de se trouver exactement sur la frontière entre en-ligne et hors-ligne, elles nous ont semblé le contexte le plus efficace pour observer les problématiques générées par l'application des méthodes numériques.

Méthodes et traces numériques de la ville

Comme on l'a dit, un des changements les plus significatifs portés par le TIC concerne la diffusion du numérique. D'un côté, l'invention des médias électroniques a transformé radicalement les conditions d'accès à la communication de la part du citoyen (Cardon, 2010). De l'autre côté, la communication numérique a basculé les conditions mêmes de la recherche, en multipliant la disponibilité des traces des phénomènes collectifs. Ces traces sont en partie produites par le support technologique, mais surtout sont générées par de nouveaux fournisseurs de contenu tels que les membres des réseaux sociaux ou les utilisateurs de plateformes de partage du contenu. Les utilisateurs du Web 2.0 (O'Reilly, 2005) laissent des traces de leurs activités en-ligne et hors-ligne qui peuvent devenir de nouvelles sources d'information utiles pour les études concernant l'espace urbain (Douay, 2011). Dans le contexte de la ville, une des conséquences les plus intéressantes de cette double transformation est l'érosion progressive des frontières séparant ce qui est reconnu comme hors-ligne et en-ligne.

Cette traçabilité intrinsèque aux médias numériques promet, si contrôlée par une méthodologie adéquate, de fournir une source nouvelle de données pour l'étude de la vie collective (Lazer et al, 2009). Au cours de dernières années, un nouveau groupe de méthodes, appelées « méthodes numériques » (Rogers, 2013), a été développé pour traiter ce type de données. A partir de l'idée selon laquelle des phénomènes tels que la régionalisation du Web ont déterminé la « fin du virtuel » (Rogers, 2009), les méthodes numériques se présentent comme une série de techniques visant à explorer les interactions en-ligne en tant que source valide pour la compréhension de la réalité sociale. La première preuve du potentiel de ces méthodes a été fournie par la célèbre étude de Google sur les épidémies de grippe des États-Unis (Ginsberg et al, 2009). Ces méthodes sont appliquées aujourd'hui avec succès dans plusieurs terrains de recherche, tels que les études d'aménagement (Silva, 2012), de migration et des mouvements sociaux, des événements médiatiques et des controverses.

Le succès des traces numériques est notamment dû à leur pouvoir de révéler des caractéristiques des phénomènes qui ont lieu dans l'espace physique. Dans nos études précédentes (Auteur et al, 2011), on a vu notamment le cas des phénomènes d'activisme urbain sur lesquels on a rarement des données statistiques et surtout ces données sont encore plus rarement disponibles en temps réel. C'est pour cela que la présence des manifestations de protestation dans l'espace virtuel et la conséquente présence de traces numériques enthousiasment le chercheur en SHS (Venturini et Latour, 2010). En effet, à travers ces traces, le chercheur peut étudier un phénomène urbain qu'il n'aurait autrement pas pu étudier, à moins de mettre en place une démarche d'enquête de terrain beaucoup plus coûteuse en terme de ressources et de temps.

Cependant, au-delà des enthousiasmes initiaux, l'usage de ces méthodes aujourd'hui sollicite plusieurs questions théoriques et méthodologiques. Récemment, Noortje Marres et Esther Weltevrede se sont intéressées aux problèmes liés à une de ces méthodes, le *scraping*, une opération technique d'extraction d'information qui rentre dans le domaine du traitement automatisé des données. « En parlant métaphoriquement », elles disent, « on peut dire que le *scraping* structure la collection de données comme dans un "procès de distillation", ce qui implique la sélection de données formatées à partir d'un océan, relativement opaque et sous-

défini, de matériaux disponibles en-ligne » (Marres et Weltevrede, 2013, 5). Nous ne nous sommes pas intéressés ici à la manière dont le *scraping* fonctionne, mais au procès épistémologique dans lequel il s'insère. Les deux autrices affirment que le problème majeur de cette technique réside dans l'impossibilité d'établir une frontière nette entre le *medium* et son objet : « Jusqu'où devons-nous aller en prenant les outils en-ligne en compte comme éléments notables dans nos recherches ? Pouvons-nous les comprendre dans le cadre de notre "méthodologie", ou devrions-nous reconnaître qu'ils font partie de "l'objet" de notre analyse? » (Marres et Weltevrede, 2013 : 10). Nous pouvons alors dire que, en voulant regarder certains aspects de la réalité sociale, les chercheurs en SHS, à travers les méthodes numériques, ne peuvent faire autre chose qu'orienter toute leur attention sur le *medium* technique.

Dans la tentative de donner une réponse à cette limite des méthodes numériques, Marres et Weltevrede affirment qu'il faudrait se référer à l'idée, ressortant des études en STS (sciences, technologies et société), selon laquelle, si nous abordons la recherche du point de vue de son équipement, certaines distinctions habituelles deviennent vraiment difficiles à soutenir, comme celles entre les techniques, les méthodes et les objets de recherche.

Le rapport en-ligne et hors-ligne à l'épreuve de la géolocalisation

Face à la difficulté de déterminer les frontières entre techniques, méthodes et objets dans ce domaine, deux voies sont possibles. La première consiste tout simplement à nier la possibilité de cette distinction, soit selon une approche optimiste, soit selon une approche pessimiste par rapport au potentiel des méthodes numériques. Dans la deuxième, par contre, il s'agit d'« expliquer plus, pour comprendre mieux », et c'est précisément cette « voie longue » que nous nous apprêtons à parcourir. En particulier, dans cette deuxième partie nous nous intéresserons au cas, de plus en plus commun, des traces numériques géoréférencées. C'est surtout dans la ville que nous assistons à la géoréférenciation de certaines informations en-ligne (Hecht et Stephen, 2014). Ce qui rend ce type de traces intéressant pour notre recherche est le fait qu'elles se trouvent sur la frontière entre en-ligne et hors-ligne, dans le sens que ces traces sont spatialement (et temporellement) situées.

Quand on laisse un message sur Twitter avec notre portable, (si le capteur GPS est activé) le message sera associé à notre position géographique, c'est-à-dire aux coordonnées géographiques du lieu d'où on a envoyé le message. C'est seulement un exemple de comment nos activités sur Internet sont de plus en plus géoréférencées. Ce phénomène est dû surtout à une évolution technologique récente liée au Web 2.0 (Tulloch, 2007) : nombreuses plateformes de réseautage social et de partage de contenu offrent la possibilité de faire une déclaration de position géographique, c'est-à-dire de lier notre action dans l'espace numérique à un point dans l'espace physique à travers l'attribution volontaire ou involontaire (générée en manière automatique par la plateforme qu'on utilise) de coordonnées géographiques.

La déclaration de position peut être faite dans deux contextes d'action différents. Le cas le plus commun est la déclaration ou partage de notre géolocalisation, c'est-à-dire notre position dans l'espace. L'exemple typique est le *check-in* sur Facebook ou sur d'autres réseaux sociaux. Qu'on arrive dans un restaurant, pour bon nombre de personnes le premier réflexe est de partager la position géographique avec leurs contacts pour leur faire savoir qu'ils sont en train de profiter d'un bon dîner.

Autrement, on peut lier un contenu (un texte, une photo, une vidéo, etc.) qu'on est en train de publier sur la Toile à une position géographique. On peut conseiller un hôtel sur une plateforme de conseil touristique, on peut participer à des initiatives de « net activisme » en signalant des abus, on peut signaler un problème urbain en contribuant à une plateforme de cartographie participative. Dans les deux cas, à travers ces actions, on produit des traces précieuses pour le chercheur ou pour le décideur public qui pourra trouver des informations ponctuelles sur le territoire qu'il étudie et/ou gère.

Les méthodes numériques ont retrouvé dans ces données un terrain idéal. Non seulement l'information géolocalisée fournit une référence apparemment indiscutable entre le en-ligne et le hors-ligne, mais également elle rend la donnée sociale beaucoup plus riche et précise en permettant d'observer le développement de phénomènes collectifs dans l'espace. Pour ce lien serré entre espace numérique et lieu physique, l'analyse des données géolocalisées du Web 2.0 à travers les méthodes numériques est devenue de plus en plus populaire non seulement dans le contexte académique, mais aussi dans les usages d'autres acteurs publics et privés. La plupart de travaux se sont concentrés sur l'analyse du partage de localisation sur les réseaux sociaux en-ligne (Nan et Gauling, 2010) et encore plus sur des plateformes de réseautage social comme Foursquare et Gowalla (fermé en 2012) dont le but principal est le partage de localisations favorites (Berjani et Strufe, 2011). Plusieurs se sont intéressés sur les espaces urbains et notamment sur les mécanismes de recommandation et d'identification de points d'intérêt (Ying et al, 2012).

Ce type d'études a mené à la diffusion du concept de « urban computing » proposé par Yuang et al. (2011), selon lequel chaque appareil, personne, véhicule, bâtiment de la rue dans la zone urbaine peut être un capteur pour comprendre la dynamique de la ville. Cependant, bien que ces méthodes ont été beaucoup utilisées dans les études urbaines, rarement l'enquêteur a soulevé le problème du sens des résultats qu'il obtient en analysant ces traces et notamment la question du rapport entre ces résultats et la ville. Il faut réfléchir sur ce que le chercheur en SHS peut vraiment tirer de ces traces et quelle validité peut-on leur reconnaître.

a. La ville numériquement comprise

A bien y regarder, on peut identifier deux couches de complexité dans nos réflexions sur la géolocalisation. Avant d'arriver aux problèmes générés par l'application des méthodes numériques, il est important en effet de réfléchir sur les limites éventuellement liées à l'opération de géolocalisation en elle-même.

Il faut d'abord s'interroger alors sur les problèmes liés aux intentions de l'utilisateur. Même s'il faut apprécier la remontée de l'amateur (Flichy, 2010) à travers le numérique et par conséquent l'abondance des traces qu'il peut produire, on doit reconnaître en effet que la trace en elle-même nous donne rarement des informations sur le point de vue de l'utilisateur qui laisse la trace, ni sur la signification qu'il est en train de donner à son action. Est-il en train d'entreprendre une action dans l'espace physique ou numérique ?

De plus, la géolocalisation sur une plateforme du Web 2.0 peut être une action volontaire mais souvent elle est le résultat d'un artefact technologique : la plateforme récupère en automatique la position GPS de la personne qui envoie un message sans qu'elle s'en rende compte, ou, action encore plus « violente », la plateforme peut convertir le nom d'une localité (par exemple « Paris ») dans une longitude et latitude précises (les coordonnées de Notre-Dame). Cela signifie que la référence à l'espace de la ville générée par l'action de géolocalisation d'une personne ne peut plus être considérée comme une information sûre et précise sur l'action de cette personne dans l'espace de la ville comme cela pouvait le sembler à première vue.

Après avoir présenté la complexité qui est cachée dans l'opération technique de la géolocalisation, on peut maintenant ajouter le deuxième niveau lié à l'action des méthodes numériques. A titre d'exemple, on peut prendre une des méthodes les plus répandues, la cartographie des localisations collectées sur Twitter. Twitter est en général un des terrains où les géolocalisations sont les plus exploitées. Cela est dû principalement à la facilité d'extraction et de traitement de ces données. En outre, c'est un contenu simple où le sujet est clairement identifiable grâce aux *hashtag*. Cette méthode a trouvé un succès particulier dans l'étude des désastres (Murthy, 2013). En effet, le succès de Twitter pendant un tsunami ou un tremblement de terre joue aujourd'hui un rôle crucial pour la couverture médiatique de ces événements et pour l'organisation des secours. Ces tweets ne portent pas beaucoup plus d'information que

d'autres sources d'informations plus classiques (comme par exemple les images satellitaires) sur l'évolution du désastre, mais ce qui est intéressant est que ces données garantissent une couverture totale d'événement en temps réel (Murthy, 2013).

Plusieurs chercheurs ont essayé de transposer cette méthode à l'analyse de contextes ordinaires et, entre autres, à l'étude de la ville. Un exemple intéressant est l'étude de Wakamiya et al. (2011) qui a identifié les caractéristiques des villes sur la base de l'activité Twitter géolocalisée dans ces villes. En mesurant la quantité des tweets envoyés dans les différents moments de la journée, ils arrivent à construire une typologie des zones urbaines (ville dortoir, ville bureau, ville nocturne, ville multifonctionnelle). Ce qu'il est pour nous vraiment intéressant de remarquer est que cette technique définit l'image de la ville simplement sur la base des activités de foule qui sûrement ne peuvent pas être considérées comme les seules activités qui ont lieu dans une ville. En effet, tant qu'on considère des contextes exceptionnels comme les désastres ou d'autres contextes ordinaires comme la vie quotidienne de la ville, Twitter permet de relever et d'interpréter seulement les pics d'activités et non les vides qui sont également une partie importante d'un phénomène. L'étude de Wakamiya et al. (2011) est alors particulièrement utile pour notre argumentation, parce que si dans les intentions de ses auteurs est mise en évidence toute la puissance des méthodes numériques pour étudier la ville hors-ligne, les représentations urbaines obtenues sont efficaces mais tout à fait partielles.

Conclusion

Dans le paragraphe précédent, nous avons fait part de nos doutes face à la possibilité de connaître de manière approfondie la ville hors-ligne à travers les méthodes numériques appliquées aux données en-ligne géoréférencées. Notre intention n'est pas tout simplement déconstructive ou, encore pire, destructive. Plutôt, nous avons l'intention de transformer cette limite en opportunité pour penser plus et voir autrement le potentiel de la compréhension numérique de la ville.

En effet, nous ne sommes pas les seuls à avoir pris une position partiellement critique face au potentiel du numérique et, plus précisément, du numérique géoréférencé, pour la connaissance – et la construction – des lieux urbains. Selon Serge Wachter, la ville, les lieux et les morphologies qui la composent n'enregistrent que des changements mineurs face au numérique et ils ne sont pas vraiment touchés par les poussées de la société de l'information (Wachter, 2010, 17). Pour cette raison, Nicolas Nova écrit que l'espace numérique offre seulement une représentation partielle du territoire (Nova, 2009, 149). Le même auteur se demande alors : « Comment représenter ce qui n'est pas enregistrable ? Comment, également, cartographier l'imprévu ? Comment prendre en compte les imperfections qui feront disparaître de la carte les capteurs abimés, en panne ou volontairement détruits ? » (Nova, 2009, 150). Similairement, Eric Gordon et Adriana de Souza e Silva affirment que, en croyant de contrôler le monde à travers les cartes et les supports de géolocalisation, nous finissons par le perdre : « le monde c'est trop avec nous » (Gordon et de Souza e Silva, 2011, 37).

Du même, nous ne sommes pas les premiers à souligner les limites des analyses des activités Twitter à travers les méthodes numériques. Déjà Richard Rogers (2013b), un des chercheurs qui a le plus contribué au développement et à la promotion des méthodes numériques a mis en avant le besoin de débanaliser le recours à Twitter. Néanmoins, notre intention était d'aller plus loin au-delà de ces considérations. D'un côté, nous avons insisté, plus que ne l'ont fait souvent les chercheurs en SHS intéressés aux méthodes numériques, sur les limites du numérique pour la compréhension du territoire urbain. De l'autre côté, nous n'avons pas cédé au pessimisme de ceux qui souhaitent une renonciation immédiate au « mythe de la trace » numérique (Casati, 2013). Nous espérons plutôt avoir fourni du matériel qui donne à penser. En tout cas, ceci sera le point de départ de nos recherches futures.

Bibliographie

- Beaude B. (2012), *Internet, changer l'espace, changer la société : les logiques contemporaines de synchronisation*, FYP éditions.
- Benkler Y. (2006), *The wealth of networks : how social production transforms markets and freedom* (1st ed.), New Haven, Conn: Yale University Press, p. 528.
- Berjani B., Strufe T. (2011), « A recommendation system for spots in location-based online social networks », *Proceedings of the 4th Workshop on Social Network Systems*.
- Boullier D. (1999), *L'urbanité numérique. Essai sur la troisième ville en 2100*, Paris, L'Harmattan, 1999, p. 184.
- Cardon D. (2010), *La démocratie Internet. Promesses et limites*, Seuil, coll. « La république des idées », p. 102.
- Casati, R. (2012), *Contre le colonialisme numérique*, Paris, Albin Michel.
- Castells M. (2002), *The Internet Galaxy: Reflections on the Internet, Business, and Society*, Oxford University Press.
- Couclelis H (2004), « The construction of the digital city », *Environment and Planning B: Planning and Design*, n° 31, p.5 -19.
- Douay N. (2011), « Urbanisme 2.0, de la contestation à la participation », *TechniCités*, n°214 (supplément spécial pour la 32e rencontre nationale des agences d'urbanisme "Territoires et Projets Les outils de la gouvernance"), p. 46-49.
- Flichy, P. (2010), *Le sacre de l'amateur. Sociologie des passions ordinaires à l'ère numérique*, Paris, Seuil.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., Brilliant, L. (2009) « Detecting influenza epidemics using search engine query data », *Nature*, n°. 457, p. 1012-4.
- Gordon, E., De Souza e Silva, A. (2011), *Net Locality: Why Location Matters in a Networked World*, Wiley-Blackwell.
- Hecht B., Stephens M. (2014), « A Tale of Cities: Urban Biases in Volunteered Geographic Information », *ICWSM*.
- Lazer D, Pentland A, Adamic L, Aral S, Barabasi A-L, Brewer D, Christakis N et al (2009), « Computational social science », *Science*, n°. 323, p. 721-3.
- Marres N., Weltevrede E. (2013), « Scraping the social ? », *Journal of Cultural Economy*, n.°3.
- Marres, N. (2012), « The redistribution of methods: on intervention in digital social research, broadly conceived », *The sociological review*, n°.60, p. 139-165.
- Murthy, D. (2013), *Twitter: Social Communication in the Twitter Age*, Cambridge : Polity Press.
- Nan L., Guanling C. (2010), « Sharing location in online social networks », *Network, IEEE* , n°.5, p. 20-25.
- Nova N. (2009), *Les médias géolocalisés : comprendre les nouveaux espaces numériques*, Pearson.
- O'Reilly T. (2005), « What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software », *Web 2.0 Conference 2005*.
- Rogers R. (2013), *Digital methods*, MIT Press.
- Rogers, R. (2013b), « Debanalizing Twitter: The Transformation of an Object of Study », *Proceedings of ACM Web Science 2013*, Paris, Mai.
- Rogers R. (2009), *The End of the Virtual: Digital Methods*, Amsterdam University Press.

- Silva, C.N. (2012), *Online Research Methods in Urban and Planning Studies: Design and Outcomes*. IGI Global.
- Tulloch, D. L. (2007), « Many, many maps: Empowerment and online participatory mapping », *First Monday*, n°. 2, février.
- Venturini T., Latour B. (2010), « The Social Fabric : Digital Traces and Quali-quantitative Methods », *Proceedings of Future En Seine 2009*, Cap Digital, Paris.
- Wakamiya S., Ryong L, and Kazutoshi S. (2011), « Urban Area Characterization Based on Semantics of Crowd Activities in Twitter », *Proc. of Fourth International Conference on Geospatial Semantics (GeoS 2011)*, Brest, France, p. 108-123.
- Wachter, S. (2010), *La ville interactive : L'architecture et l'urbanisme au risque du numérique et de l'écologie*, Paris : L'Harmattan.
- Ying J.J., Hsueh-Chan Lu E., Kuo W.N. and Tseng V.S. (2012), « Urban Point-of-Interest Recommendation by Mining User », *Check-in Behaviors UrbComp'12*, Beijing, Chine.
- J. Yuan, Y. Zheng, and X. Xie (2011), « Urban Computing with Taxicabs », *Proceedings of 13th ACM International Conference on Ubiquitous Computing (UbiComp 2011)*, Beijing, Chine.
- Auteur et al (2011), « Citizen protest in the online networks: the case of the China's bloody map », *7th UK Social Networks Conference*, Julliet, Londres.