



HAL
open science

La time-geography à l'heure des données issues d'internet. Applications à l'étude des mobilités quotidiennes à Bangkok

Alexandre Cebeillac

► **To cite this version:**

Alexandre Cebeillac. La time-geography à l'heure des données issues d'internet. Applications à l'étude des mobilités quotidiennes à Bangkok. CIST2020 - Population, temps, territoires, Collège international des sciences territoriales (CIST), Nov 2020, Paris-Aubervilliers, France. pp.562-566. hal-03114109

HAL Id: hal-03114109

<https://hal.science/hal-03114109>

Submitted on 18 Jan 2021

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

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

La *time-geography* à l'heure des données issues d'internet. Applications à l'étude des mobilités quotidiennes à Bangkok

AUTEUR

Alexandre CEBEILLAC

RÉSUMÉ

L'analyse des mobilités quotidiennes présente des intérêts dans bien des domaines, notamment dans la compréhension de la propagation des épidémies à l'échelle intra-urbaine. Les outils apportés par la *time-geography* depuis 50 ans sont bien adaptés à ce genre d'étude, autorisant différents niveaux de réflexion selon une approche individuelle matérialisée par des espaces d'activités, ou de manière agrégée faisant ressortir les temporalités des lieux. À l'heure où de plus en plus de données géolocalisées sont accessibles en ligne, cette communication présentera le potentiel et les différences entre des données géolocalisées provenant de Twitter et de Facebook, et la façon dont leur utilisation peut s'articuler autour d'outils fournis par la *time-geography* afin d'analyser les mobilités quotidiennes à Bangkok. Nous évaluerons aussi la possibilité de développer un modèle de mobilité individu-centré à base d'agents en tirant parti des données et suivant le concept d'espace d'activité.

MOTS CLÉS

time-geography, données en ligne, Facebook, Twitter, Bangkok

ABSTRACT

The analysis of daily mobility interests many fields, in particular in understanding the spread of epidemics on an intra-urban scale. The tools brought by Time Geography for 50 years are well suited to this kind of study, allowing different levels of reflection according to an individual approach materialised by activity spaces, or aggregated highlighting the temporalities of places. At a time when more and more geolocated data are accessible online, this communication will present the potential and the differences between geolocated data from Twitter and Facebook and how their use can be articulated around tools provided by Time Geography, in order to analyse daily mobility in Bangkok. We will also assess the possibility of developing an individual agent-based mobility model by taking advantage of the data and following the concept of activity space.

KEYWORDS

Time Geography, Online data, Facebook, Twitter, Bangkok

INTRODUCTION

L'analyse et la modélisation des mobilités intra-urbaines intéressent de nombreux secteurs, de la planification des transports publics à la compréhension de phénomènes sociaux. En contexte épidémique, le rôle des mobilités humaines en tant qu'élément primordial de diffusion des virus dans des villes de plus en plus peuplées reste notamment à éclaircir. Bangkok, capitale de la Thaïlande, est une ville de plus de 10 millions d'habitants, où des maladies vectorielles comme la dengue, bien qu'endémiques, présentent de fortes discontinuités spatio-temporelles quant aux quartiers qu'elles touchent d'une année sur l'autre. Se pose alors la question de l'étude et de l'analyse des mobilités quotidiennes dans cette mégapole au regard des différents contextes évoqués, notamment vis-à-vis des concepts et méthodes à employer ainsi que des données qui les alimentent.

L'espace d'activité: un cadre théorique toujours d'actualité

La *time-geography* introduite par Hägerstrand (1970) a apporté un arsenal d'outils et de concepts toujours utilisé aujourd'hui dans le cadre de l'étude des mobilités quotidiennes. Ainsi, l'espace d'activité d'un individu se définit comme la portion de l'espace urbain qu'il fréquente lors de ces activités quotidiennes. Hägerstrand identifie deux grands types d'activités en fonction de leur niveau de flexibilité spatio-temporel et du degré de liberté (ou contrainte) de l'individu à leur égard :

- Les activités figées s’effectuent régulièrement et/ou dans un lieu précis. ; il peut s’agir par exemple du travail, d’une école ou d’un cours de guitare le mercredi ;
- Les activités flexibles sont moins contraintes dans le temps et l’espace, moins planifiées, comme le fait « d’aller boire une bière en terrasse » parce qu’il fait beau.

La richesse du concept d’espace d’activités repose sur les deux dimensions qu’il permet de manipuler :

- Les mobilités ego-centrées, c’est-à-dire les différents types de lieu fréquenté et activité effectuée par une personne dans l’espace-temps urbain, peuvent révéler des potentiels de mobilités différents selon la localisation du domicile d’une personne et/ou son statut socio-économique, son genre et son âge. Ceci autorise à la fois des études de domaines qui relèvent de la physique et de l’analyse spatiale, notamment sur divers paramètres de dispersions, des fréquences de visites et de retours, etc., et une approche plus sociale, avec des potentiels de mobilités différents selon les individus, en étudiant le nombre de lieux fréquentés et les activités effectuées, probablement révélateurs de leur capital économique et culturel.
- Le fait d’agrèger ces trajectoires individuelles apporte une vision centrée sur les lieux, ce qui permet d’évaluer le potentiel d’attraction d’un type d’activité ou d’une zone de la ville en fonction de l’heure, faisant ressortir différents niveaux de coprésences socio-économiques et autres discontinuités spatio-temporelles dans la ville.

De plus en plus d’études sur les mobilités urbaines ont recours à la simulation et à la modélisation à base d’agent pour apprécier l’impact de divers phénomènes, comme l’exposition d’une population à la pollution (Fosset *et al.*, 2016) ou la propagation des épidémies (Huraux *et al.*, 2017). Le concept d’espace d’activité est alors utilisé pour générer des populations synthétiques mobiles pourvues d’agendas qui effectueront diverses activités dans la ville en fonction d’attributs hérités de données provenant de sources variées, en générales institutionnelles et recueillies in situ via des enquêtes de plus ou moins grande envergure (i.e. enquêtes ménages déplacement). Néanmoins ce type de donnée n’est pas disponible pour Bangkok.

Le développement des traces numériques géolocalisées

Le développement et la généralisation progressive de l’utilisation de moyens de communication en itinérance impliquent la création de traces numériques géolocalisées. Ces dernières sont largement utilisées depuis une décennie dans l’analyse des mobilités quotidiennes (Barbosa *et al.*, 2018), mais très rarement associées au concept d’espace d’activité. En effet, les données issues de la téléphonie mobile ou les messages géolocalisés provenant de la plateforme Twitter ne renseignent que les lieux visités dans le temps, sans information directe sur le type d’activité réalisée.

Cette communication examinera les potentiels et limites de l’utilisation de données géolocalisées provenant de réseaux sociaux en ligne dans le contexte d’une *time-geography* de Bangkok, en prenant l’exemple de données collectées sur Twitter et Facebook.

1. DES ESPACES D’ACTIVITÉS INDIVIDUELS À PARTIR DE TWITTER

1.1. Collecte des données sous Twitter

Twitter est une plateforme d’envoi de messages courts utilisée mensuellement par environ 300 millions de membres à travers le monde. Ces messages (ou *tweets*) sont généralement publics ; certains peuvent être géolocalisés par l’utilisateur. Ils peuvent être collectés en temps réel en passant par l’API (*application programming interface*) STREAM, une suite de protocoles permettant d’accéder à la base de données de l’entreprise via des identifiants, et ce gratuitement – tant que l’on ne dépasse pas un seuil de 1 % du volume global de la plateforme. Une collecte sur un temps suffisamment long permet de voir pour un utilisateur les différents lieux depuis lesquels il a tweeté et à quel moment, tel un échantillon de sa trajectoire dans l’espace urbain. Néanmoins, ces données sont sporadiques car non continues dans le temps et il est délicat de faire le lien entre les lieux déclarés visités en ligne de ceux réellement visités.

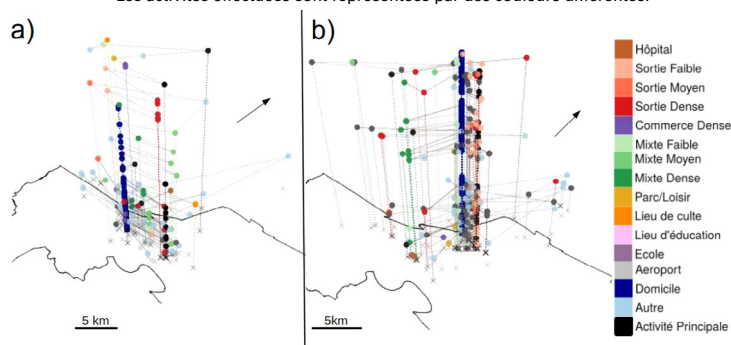
Nous avons effectué une collecte continue sur la zone de Bangkok entre juin 2014 et décembre 2015, en ne récupérant que l’identifiant, la localisation et l’heure du message. Nous avons appliqué différents filtres pour nettoyer la base de données (suppressions des bots, des touristes, etc.). À ce stade, nous n’avons qu’une succession de traces géolocalisées et datées, sans autre information sémantique.

1.2. Création d’espaces d’activités individuels

Les messages de chaque utilisateur sont alors regroupés en fonction de leur proximité pour former des lieux de l’espace d’activité, caractérisés par des fréquences et des heures de visite. Le lieu du domicile est défini par une présence régulière au même endroit entre 20h et 6h. Nous obtenons un échantillon d’environ 35000

personnes et 9 millions de traces géolocalisées. Bien que n'ayant pas d'informations socio-économiques sur ces individus, l'échantillon est représentatif spatialement au regard des données du recensement (voir Cebeillac *et al.*, 2017). De manière similaire, nous posons que l'activité principale est le lieu d'où un nombre important de messages est envoyé en journée. Finalement, afin d'associer chaque lieu fréquenté par un utilisateur à une activité, nous posons l'hypothèse qu'il existe un lien fort entre le lieu d'envoi d'un message et l'activité en cours de réalisation. Par exemple, si une personne émet un *tweet* depuis un centre commercial (*mall*), nous posons l'hypothèse qu'elle fait des courses. Les activités réalisées dans les autres lieux de l'espace d'activité d'un individu sont alors estimées en croisant une couche d'utilisation du sol réalisée à partir de données collectées sur google maps et openstreetmap (Cebeillac *et al.*, 2018). Les utilisateurs possèdent alors un ensemble de lieux associés à des activités qu'ils fréquentent plus ou moins régulièrement et à différents jours et horaires. Les outils de visualisation fournis par Hägerstrand (1970), notamment la représentation par l'aquarium spatio-temporel, permettent d'explorer ces espaces d'activité pour en observer la cohérence (fig. 1).

Figure 1. Représentation de l'espace d'activité de deux utilisateurs de Twitter à Bangkok (a et b) selon l'aquarium spatio-temporel proposé par Hägerstrand (1970). Les activités effectuées sont représentées par des couleurs différentes.



L'espace d'activité brut obtenu d'après les données laissées sur Twitter montre que l'utilisateur (a) laisse moins de messages depuis son activité principale à la fin de la période enregistrée, mais aussi qu'il fréquente de nouveaux lieux au nord-ouest de son domicile. L'utilisateur (b) a un espace d'activité plus étendu géographiquement mais laisse globalement plus de traces numériques entre son domicile et son

lieu de travail estimé. Cela permet ainsi de constater des différences entre les utilisateurs sur la fréquence d'utilisation du réseau social et sur le volume de traces numériques laissées, permettant un questionnement des données.

Une fois agrégées, ces informations font ressortir les temporalités des différents secteurs de la ville. La section suivante présentera d'autres données issues de Facebook qui serviront de point de comparaison vis-à-vis des données de Twitter.

2. LES TEMPORALITÉS À BANGKOK

2.1. Collecte des données sur Facebook

Facebook est un réseau social utilisé quotidiennement par plus de 2 milliards de personnes. Contrairement aux pages personnelles qui sont essentiellement privées, les pages relatives à des lieux, qu'ils soient des établissements commerciaux (bar, restaurant, etc.) ou publics (école, université, administration, etc.) sont généralement visibles par tous. Une fonctionnalité de Facebook permet à chaque utilisateur de signaler dans son fil d'actualité sa présence dans une de ces lieux par un *check-in*. Cette information est visible par chacun des contacts de l'utilisateur et se manifeste par une incrémentation du nombre de personnes ayant visité le lieu sur la page de ce dernier. L'utilisation de l'API Place¹ permet de collecter ces informations sur les lieux, en plus des coordonnées géographiques et le type de lieu (école, restaurant, etc.) dans un rayon défini. En effectuant une recherche par fenêtre mobile dans une zone donnée sur une longue période, il est ainsi possible de collecter l'ensemble des lieux de la base et d'apprécier l'évolution du nombre de visites au cours du temps. Néanmoins une personne n'a pas nécessairement besoin d'être dans un lieu pour déclarer y être présent et un grand nombre d'établissements ont recours à l'achat de *check-in* pour augmenter virtuellement leur popularité. Après 108 jours d'enregistrement entre juillet et novembre 2017 et l'application de filtres pour supprimer les lieux fantaisistes et les pics de fréquentations aberrants, nous avons pu collecter un peu moins de 40 millions de ces *check-in* dans environ 128000 lieux, à un pas d'une heure.

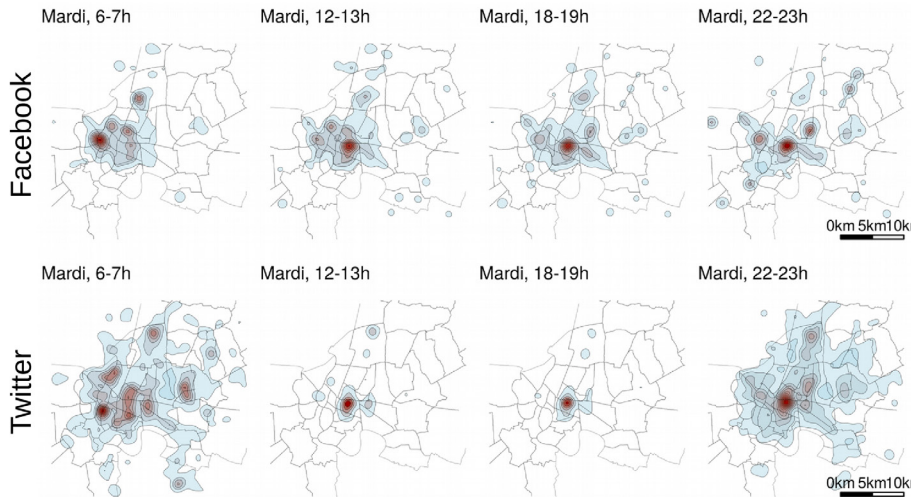
2.2. Pulsation urbaine

À partir de ces deux jeux de données, il est possible d'observer et de comparer la répartition temporelle des *check-in* et des *tweets* comme *proxy* de la population, visible sur la figure 2 qui montre ces densités à

¹ www.developers.facebook.com/docs/places/web/search

différents moments de la journée. Alors qu'une dichotomie apparaît dans la répartition des présences des utilisateurs de Twitter selon le moment de la journée, cet aspect n'est pas aussi marqué avec les données de Facebook. Concernant Twitter, nous pouvons observer clairement un effet centre périphérie, où l'activité enregistrée sur le réseau est relativement bien répartie dans l'ensemble de la ville le matin et le soir, ce qui révèle les présences des utilisateurs à leur domicile – information qui échappe aux données de Facebook qui ne renseigne pas ce type d'activité. Mais les densités observées en journée *via* Facebook semblent plus cohérentes que celles observées sur Twitter où le centre-ville apparaît surreprésenté.

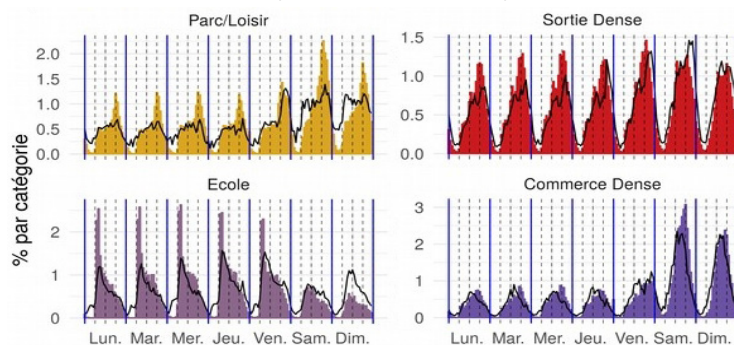
Figure 2. Pulsations urbaines à Bangkok au regard des données de Facebook et Twitter



2.3. Temporalité des activités

Il est également possible de comparer les temporalités des activités, comme présenté dans la figure 3 pour 4 types d'activités. S'il existe des différences, nous pouvons observer que les ordres de grandeurs et les principaux pics de fréquentations sont assez semblables et que les profils sont cohérents. Par exemple, les écoles sont surtout fréquentées en semaine et, bien que le pic du matin soit nettement plus important d'après Twitter que Facebook, il reste présent dans les deux jeux. Les commerces denses sont aussi bien plus fréquentés les week-ends que les jours de semaines, et les parcs plus le soir que le matin. Cela dit, il reste délicat de savoir quelle source se rapproche le plus des tendances réelles. Nous pouvons arguer qu'au regard des volumes enregistrés et du nombre d'utilisateurs Facebook serait potentiellement un meilleur indicateur pour les activités autres que l'activité principale et les présences au domicile.

Figure 3. Temporalité des activités au regard des données de Facebook (trait noir) et Twitter (histogrammes). Les traits verticaux bleus représentent minuit, et ceux en pointillé 6h, 12h et 18h.



3. PERSPECTIVE : VERS UNE MODÉLISATION DES MOBILITÉS QUOTIDIENNES

À partir de ces données et des informations qu'elles révèlent individuellement et collectivement, nous pouvons proposer des pistes pour générer des agendas synthétiques en vue d'une modélisation à base d'agent. Les espaces d'activités issus de Twitter sont discontinus dans le temps du fait de la nature

sporadique des données. L'utilisation d'hypothèses sur les niveaux de contraintes spatio-temporelles selon les types d'activités (e.g. au domicile le soir, au travail en journée) et les niveaux de régularité de fréquentation observés dans les données d'une personne (e.g. une personne se rend régulièrement dans des centres commerciaux le week-end) permettent de réajuster les données et d'obtenir des agendas continus, même s'il est délicat d'en apprécier précisément la véracité. À partir des agendas reconstruits, il est possible de générer des agendas synthétiques. D'après les données de Twitter, nous observons que les personnes résidant en périphérie ont tendance à effectuer des déplacements sur de plus longues distances et dans moins de lieux que celles résidant dans l'hyper-centre. Ainsi, selon sa localisation dans la ville, un agent se verra attribuer

un nombre de lieux et d'activités qu'il réalisera. Les séquences des activités peuvent être définies à partir de matrice de transitions entre deux activités – e.g. la probabilité d'aller dans un lieu de commerce dense plutôt que dans un parc à $t+1$ sachant que la personne est dans son activité principale à t – issues des agendas reconstruits précédemment. Il est aussi envisageable de mobiliser les temporalités des activités issues de Facebook. Par exemple, si l'agent doit choisir entre un parc et une zone commerciale, il aura plus de chance de choisir le parc si l'action se déroule en soirée, au regard des pics de fréquentation de la figure 3. Des travaux sont toujours en cours pour améliorer l'affectation d'une localisation.

4. DISCUSSION

Les concepts et représentations proposés par la *time-geography* n'ont cessé d'évoluer depuis 50 ans (Ellegård, 2019). Mais les approches initiales par l'espace d'activité et l'aquarium spatio-temporel restent des bases solides, toujours d'actualité et utilisables avec des données géolocalisées issues d'internet. Les données longitudinales provenant de Twitter associées à des couches d'utilisation du sol autorisent la création d'espaces d'activités individuels. Les données de Facebook permettent de faire ressortir les temporalités des activités dans la ville. Le niveau de représentativité des données collectées en ligne reste délicat à estimer et nous n'avons pas d'informations socio-économiques sur les utilisateurs.

Le concept d'espace d'activité est suffisamment flexible pour être compatible avec des approches d'autres disciplines (*ibid.*), tant en épidémiologie et géographie de la santé (Perchoux *et al.*, 2013), qu'en informatique et modélisation à base d'agents (Banos *et al.*, 2005). La souplesse de la méthode permet également de s'adapter à différents types de données, autorisant un croisement des sources. Des enquêtes de terrains supplémentaires semblent cependant nécessaires pour valider et/ou enrichir le modèle de déplacement. Aussi associer un modèle de mobilité à des modèles épidémiologiques et environnementaux permettrait-il de mieux comprendre la diffusion des épidémies dans l'espace urbain (Daudé *et al.*, 2015).

RÉFÉRENCES

- Banos A., Chardonnel S., Lang C., Marilleau N., Thévenin T., 2005, « Une approche multi-agents de la ville en mouvement », *Modèles et systèmes multi-agents pour la gestion de l'environnement et des territoires (SMAGET)*, Les Arcs, p. 1-17.
- Barbosa H., Barthelemy M., Ghoshal G., James C.R., Lenormand M., Louail T., Menezes R., Ramasco J.J., Simini F., Tomasini M., 2018, "Human Mobility: Models and Applications"; *Phys. Rep.*, n° 734, p. 1-74 [en ligne: doi.org/10.1016/j.physrep.2018.01.001].
- Cebeillac A., Daudé É., Vaguet A., 2018, "Spatial Discontinuities, Health and Mobility"; *Rev. Int. Géomat.*, n° 28, p. 389-407 [en ligne: doi.org/10.3166/riq.2018.00062].
- Cebeillac A., Hureau T., Daudé É., 2017, "Where? When? and How Often? What Can We Learn about Daily Urban Mobilities from Twitter Data and Google Map in Bangkok (Thailand) and What Are the Perspectives for Dengues Studies?"; *Netcom*, 31(3/4), p. 283-308 [en ligne: doi.org/10.4000/netcom.2725].
- Daudé É., Vaguet A., Paul R., 2015, « La dengue, maladie complexe », *Nat. Sci. Sociétés*, n° 23, p. 331-342 [en ligne: doi.org/10.1051/nss/2015058].
- Ellegård K., 2019, *Time Geography: Concepts, Methods and Applications*, Abingdon, Routledge, coll. "Routledge Studies in Human Geography".
- Fosset P., Banos A., Beck E., Chardonnel S., Lang C., Marilleau N., Piombini A., Leysens T., Conesa A., André-Poyaud I., Thévenin T., 2016, "Exploring Intra-Urban Accessibility and Impacts of Pollution Policies with an Agent-Based Simulation Platform: GaMiroD"; *Systems* 4, n° 5 [en ligne: doi.org/10.3390/systems4010005].
- Hägerstrand T., 1970, "What About People in Regional Science?"; *Pap. Reg. Sci. Assoc.*, n° 24, p. 7-21.
- Hureau T., Cebeillac A., Misslin R., Daudé É., 2017, « MOMOS: modélisation à base d'agents des mobilités quotidiennes en milieu urbain pour la simulation spatiale », communication au 13^{es} rencontres de Théo Quant, Besançon [en ligne: www.thema.univ-fcomte.fr/theoq/pdf/2017/communications/2017_Hureau_MOMOS_modelisation_base_agents_mobilites_quotidiennes_milieu_urbain_simulation_spatiale.pdf].
- Perchoux C., Chaix B., Cummins S., Kestens Y., 2013, "Conceptualization and Measurement of Environmental Exposure in Epidemiology: Accounting for Activity Space Related to Daily Mobility"; *Health Place*, n° 21, p. 86-93 [en ligne: doi.org/10.1016/j.healthplace.2013.01.005].

L'AUTEUR

Alexandre Cebeillac

Université de Rouen Normandie — IDEES

a.cebeillac@hotmail.fr