
Cartographie interactive de matrices Origines / Destinations

Application aux systèmes de vélos en libre-service de Paris, Londres et New York

Etienne Côme, Mohamed K. El Mahrsi, Latifa Oukhellou

Université Paris-Est, IFSTTAR, COSYS-GRETTIA
F-77447 Marne-la-Vallée, France

{etienne.come,mohamed-khalil.el-mahrsi,latifa.oukhellou}@ifsttar.fr

RÉSUMÉ. Les matrices O/D (Origines / Destinations) fréquemment rencontrées dans le domaine des transports posent de nombreuses difficultés en ce qui concerne leur représentation. Ces matrices contiennent en effet une masse de données importante et sont par conséquent difficiles à cartographier. Nous proposons dans cet article une cartographie interactive permettant à l'utilisateur de changer la focale (l'origine ou la destination de référence) de manière à permettre une exploration interactive assez intuitive de ces données complexes. Cette solution s'appuie sur un lissage spatial adapté à ce type de données que nous détaillons dans cet article et permet à l'utilisateur d'explorer l'ensemble des données ou de se focaliser sur quelques points d'intérêt en particulier. La solution proposée a été mise en œuvre sur les données de flux enregistrées durant plus de six mois sur les systèmes de vélos en libre-service (VLS) de Londres, Paris et New York.

ABSTRACT. O/D (Origin / Destination) matrices are used in public transportation to model how traveler trips are distributed between stations. Due to the amount of information they contain, these matrices are often problematic to visualize and explore. We present an approach to chart such data through an interactive map that offers the possibility to intuitively switch focus between various points of interest (i.e. origin or destination reference points) and easily study how traffic flows around them. The approach is based on a spatial smoothing technique suited to the nature of the data. We also present a prototype implementation using six months' worth of data collected from bicycle sharing systems (BSS) in London, Paris, and New York.

MOTS-CLÉS : matrices O/D, transports, vélos en libre-service, visualisation, cartographie.

KEYWORDS: O/D matrices, transportation, bicycle sharing systems, visualization, cartography.

DOI:10.3166/HSP.x.1-8 © 2014 Lavoisier

1. Introduction

Les matrices O/D décrivent les flux (les migrations) de personnes ou de biens entre un ensemble fini de lieux. Elles sont fréquemment rencontrées dans le domaine des transports et sont bien connues des géographes qui s’y intéressent depuis longtemps soit pour tenter de les modéliser (cf. le modèle gravitaire (Ullman, 1980 ; Haynes, Fotheringham, 1984)), soit pour les représenter (cf. les travaux de Waldo Tobler (Tobler, 1976 ; 1987)). Celles-ci caractérisent en effet les interactions observées entre des lieux et sont donc souvent pertinentes non seulement pour concevoir, exploiter et améliorer les systèmes de transport mais également pour comprendre ces interactions.

La difficulté principale en ce qui concerne la représentation des matrices O/D provient de leur taille. En effet, celle-ci croît de manière quadratique par rapport au nombre de lieux considérés. Pour fournir un exemple d’ordre de grandeur, on peut citer le cas d’étude qui a motivé les présents travaux : la matrice O/D du système VLS de Paris (nommé Vélib’) contient plus d’un million de flux possibles (même si la plupart ne sont jamais observés) puisque le nombre de stations dans ce système est supérieur à 1000. Les flux organisés au sein de ces matrices peuvent être des flux entre zones géographiques (des communes par exemple pour les migrations domicile-travail) ou des points (par exemple de rabattement sur un système de transport tel que des stations de métros ou de VLS). Nous nous intéresserons plus particulièrement au deuxième cas de figure dans cet article. Nous supposons donc qu’en plus de la matrice O/D qui sera notée od avec od_{ij} le flux du lieu (la station) i vers le lieu (la station) j exprimé en nombre de déplacements par heure, est disponible l’ensemble des positions des stations en latitude, longitude et nous noterons celles-ci $\{s_1, \dots, s_n\}$.

L’une des solutions naturelles pour représenter de telles données s’appuie sur une représentation de type graphe, des arcs orientés étant dessinés pour chacun des flux, le volume du flux pouvant être encodé à l’aide d’un attribut visuel tel que la largeur de la flèche (ou sa transparence). Ces cartographies permettent de visualiser l’ensemble des données (tous les flux sont représentés) dans leur contexte géographique puisque les positions des lieux étudiés sont utilisées. Cependant, lorsque le nombre de lieux augmente ces approches mènent rapidement à des représentations très difficilement exploitables, comme le montre la Figure 1. Celle-ci présente le résultat obtenu avec une approche de ce type pour une matrice O/D estimée sur les données Vélib’. L’effet de superposition et donc d’occlusion lié au nombre très important de flèches rend la lecture de la carte extrêmement difficile.

Pour résoudre ce problème, différentes solutions ont été proposées. Il est par exemple possible de filtrer certains flux : en supprimant les flux faibles, ou en ne conservant que les flux originaires d’un sous-ensemble restreint d’origines (Phan *et al.*, 2005). Il est également possible de regrouper les flux proches géographiquement de manière à diminuer les superpositions, cette technique déjà utilisée par Charles-Joseph Minard en 1864 pour représenter les exportations de vin Français (voir Figure 2, a) a depuis été automatisée grâce aux techniques de regroupement d’arêtes telles que celles proposées par (Cui *et al.*, 2008) et (Holten, Van Wijk, 2009) (voir Figure 2, c). D’autres

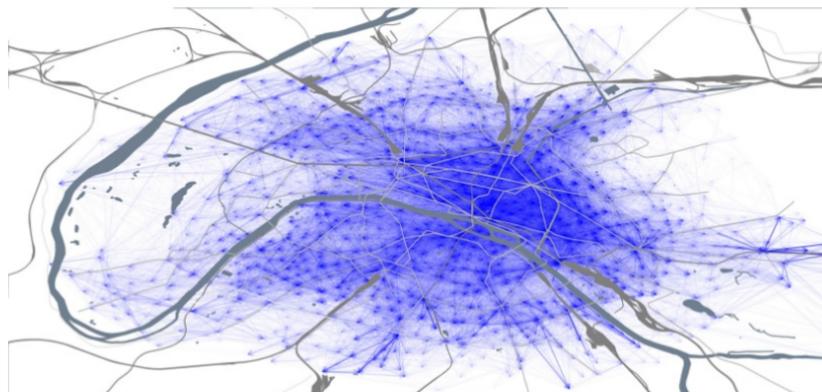


Figure 1. Représentation de la matrice O/D des flux de Vélib' estimée sur la tranche horaire 16h-19h des jours de semaines sous forme de graphe orienté, le volume de chaque flux est encodé grâce à la largeur et à la transparence des flèches.

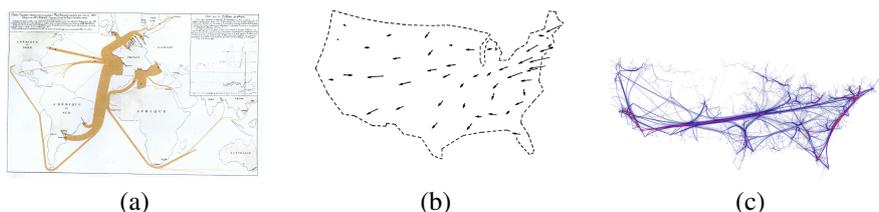


Figure 2. (a) Représentation des exportations de Vin Français (Charles-Joseph Minard, 1864), (b) Champ vectoriel associés aux flux d'étudiant Américain (Waldo Tobler, 1976), (c) Représentation des flux migratoires Américains (Danny Holten et Jark Van Wijk, 2009).

solutions essayent d'attaquer ce problème en tentant d'extraire les caractéristiques saillantes de la matrice O/D avant de les représenter.

La solution classique de W. Tobler (Tobler, 1976) (voir Figure 2, b) est de ce type puisqu'elle repose sur le calcul d'un champ vectoriel résumant l'ensemble des interactions. D'autres travaux plus récents tels que ceux de (Rae, 2009) peuvent également être mis dans cette catégorie. Pour finir ce survol non exhaustif des solutions proposées nous citerons les travaux de L. Wilkinson (Wilkinson, 1979) et ceux de J. Bertin (Bertin, 1967) qui évitent l'écueil précédent en ne travaillant que sur la matrice O/D et en abstrayant celle-ci de son contexte géographique. La représentation est alors matricielle (Wilkinson, Friendly, 2009). Dans ce type d'approche l'objectif principal consiste à rechercher une permutation pertinente des lignes et des colonnes de la matrice permettant de mettre en évidence des groupes de lieux échangeant un volume important.

Après cette introduction rapide aux problématiques liées à la représentations des matrices O/D, nous détaillons dans la section suivante la solution de cartographie interactive proposée. Nous évoquerons ensuite comment de telles matrices peuvent être estimées dans le contextes particulier des systèmes de VLS. Ces systèmes sont en effets non stationnaires et il est donc intéressant de s'intéresser à différentes matrices couvrant différentes tranches horaires. Finalement, le prototype développé pour mettre en oeuvre cette idée sur les données de VLS de Londres, Paris et New York sera présenté ainsi que quelques observations effectuées sur les différentes matrices O/D estimées.

2. Cartographie interactive : focus et lissage

L'un des angles d'attaque face aux données de type O/D non exploré dans les propositions précédentes est l'interactivité. Avec l'interactivité, il n'est plus nécessaire de devoir représenter tout à la fois sur une seule carte. En donnant à l'utilisateur la possibilité d'interagir avec la carte, celui-ci peut cibler des points d'intérêt en particulier et seules les informations pertinentes par rapport à ces derniers sont visualisées. Les outils d'exploration de graphes, par exemple, mettent fréquemment en oeuvre cette stratégie en déclenchant la mise en évidence des échanges directs d'un noeud lors du simple survol de celui-ci par la souris.

Nous proposons dans le contexte des matrices O/D de permettre à l'utilisateur de déplacer une focale (une destination ou une origine de référence) par rapport à laquelle les flux (entrants ou sortants) seront calculés et représentés. En effet, lorsqu'une origine x ou une destination y de référence est fixée, le flux f peut être représentée à l'aide d'une carte de chaleur superposée au plan de la ville, ce qui permet d'analyser les flux dans leurs contexte géographique et de fournir à l'utilisateur une représentation intuitive sous forme de carte de chaleur. L'idée est alors d'offrir une vision continue du phénomène sous-jacent aux données observées (la matrice O/D). Ceci permet en effet d'assurer une continuité dans les transitions lors du changement de « focus », améliorant ainsi le confort d'exploration.

De plus, même si le nombre de lieux (stations) sur lesquelles des données sont observées est fini, les déplacements sous-jacents peuvent sans difficulté être pensés et représentés comme un phénomène continu. En effet, même si un déplacement est observé entre la station i et la station j , le point d'origine du déplacement de l'utilisateur dans le cas des systèmes de transports est différent de la position de la station. L'utilisateur a dû marcher quelques mètres pour trouver une station. De même, la destination réelle d'un déplacement est elle aussi différente de la station de destination. Nous noterons x et y ces points de départ et d'arrivée. Pour essayer d'estimer le flux entre $(x, y) \in R^2 \times R^2$ à partir de la matrice od et des positions des stations, on peut faire l'hypothèse que x est proche de i et y de j , plus formellement on peut utiliser un noyau (Gaussien ou triangulaire par exemple) pour estimer la probabilité qu'un utilisateur partant de x utilise la station i , dans ce cas on a :

$$\pi_i(x) = \frac{K_\beta(x, s_i)}{\sum_k K_\beta(x, s_k)},$$

avec K_β un noyau de paramètre β normalisé ($\int_x K_\beta(x, y)dx = 1, \forall y \in R^2$). Le paramètre β permet de régler la distance que les utilisateurs sont prêts à parcourir pour rejoindre une station. De même, la probabilité d'aller réellement en y sachant que l'on arrive à la station j peut être simplement modélisé par un noyau centré en s_j que nous pouvons noter $K_\sigma(y, s_j)$. En utilisant ces approximations nous proposons d'estimer le flux de x à y par :

$$f(y|x) = \sum_i \sum_j \pi_i(x) od_{ij} K_\sigma(y, s_j)$$

Cette approximation permet d'obtenir des transitions progressives lorsque la focale est déplacée par l'utilisateur. Les flux peuvent être exprimés en déplacements / heure / kilomètres carrés et sont estimés par une fonction $f : R^2 \times R^2 \rightarrow R^+$. Pour les visualiser, nous proposons de permettre à l'utilisateur de fixer x (ou y) et de représenter $f(y|x)$ grâce à une heatmap en appliquant une simple échelle de couleur adaptée pour faire ressortir les zones où le flux est important. La Figure 3 présente le résultat obtenu lorsque la focale (représentée à l'aide d'un marqueur) est positionnée sur le quartier de la goûte d'Or à Paris, suivant que l'on s'intéresse aux destinations ou aux origines des trajets.

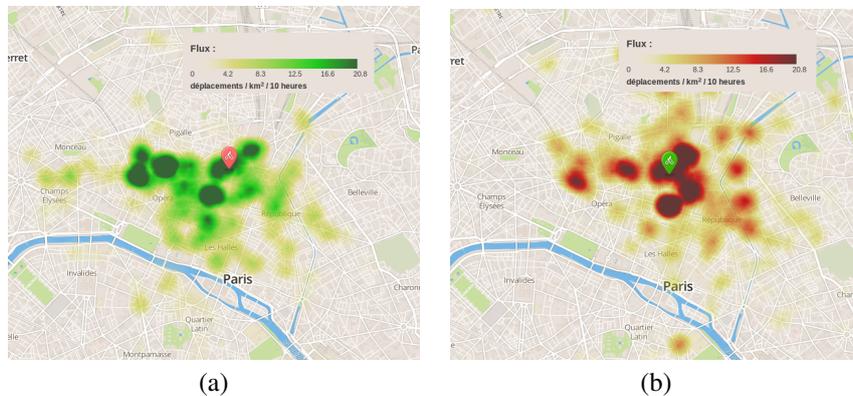


Figure 3. (a) Représentation des destinations des trajets partant du quartier de la goûte d'Or à Paris, (b) Représentation des origines des trajets s'achevant dans le quartier de la goûte d'Or à Paris.

Jusqu'à maintenant nous n'avons considéré qu'une matrice O/D statique or les origines et destinations des vélos en libre-service ne sont pas stationnaires : on observe, en effet, différentes phases dans une journée rythmé en particulier par nos déplacements domicile-travail (Côme *et al.*, 2014).

3. Estimation et non-stationnarité

Si on suppose que le processus est stationnaire, il est possible d'estimer tout simplement le flux par simple comptage (en utilisant les données billettiques dans le cas des VLS).

$$od_{ij} = \frac{\#dep(i, j)}{\#heures}$$

L'hypothèse de stationnarité n'est cependant pas adaptée au système qui nous intéresse. Il est par contre assez naturel de définir des plages horaires pendant lesquelles le système est supposé être stationnaire. Différentes matrices peuvent ainsi être estimées en modifiant l'intervalle pendant lequel le système est considéré comme stationnaire. On peut par exemple s'intéresser à la matrice correspondant à la plage horaire 8h-9h des jours de semaine ou à celle de 16h-19h des week-ends. Cela revient à supposer que le système est cyclo-stationnaire de période 1 semaine et stationnaire par morceaux. L'estimation peut dans ce cas de figure également être aisément réalisée par simple comptage après filtrage des données n'appartenant pas à la tranche horaire spécifiée. Pour permettre à l'utilisateur de passer d'une matrice O/D à l'autre nous avons simplement ajouté dans l'interface un menu déroulant permettant de changer de tranche horaire.

4. Implémentation et résultats

Nous avons réalisé un prototype (présenté sur la Figure 4 et accessible ici : <http://vlsstats.ifsttar.fr/od/>) qui permet d'explorer les flux enregistrés à Paris, Londres et New York. Le découpage en 5 plages de fonctionnement supposées stationnaires a été fait en s'inspirant de nos travaux précédents sur Paris (Côme *et al.*, 2014). Enfin, six mois de données billettiques ont été utilisées pour estimer les matrices Origine / Destination, pour chacun des systèmes. Les données de Paris datent de 2011, celles de Londres de 2012 et celles de New-York de 2014. Les paramètres des deux noyaux ne pouvant être estimés à partir des données, l'interface permet à l'utilisateur de jouer avec leurs valeurs dans un intervalle a priori raisonnable (50m - 500m). L'utilisateur peut également changer de ville et de tranche horaire et bien sûr modifier la position du marqueur représentant l'origine ou la destination de référence. Ce prototype a été développé grâce à des technologies Web et en particulier grâce aux bibliothèques Leaflet et D3. L'utilisation d'un rendu de la heatmap sur l'élément HTML 5 Canvas et d'un quad-tree pour le stockage des points a permis d'obtenir des transitions assez fluides lors du changement de focus.

5. Conclusion

Cet article traite du problème de la visualisation des matrices O/D, rendue complexe par la masse de données que celles-ci contiennent. À cet effet, nous proposons une

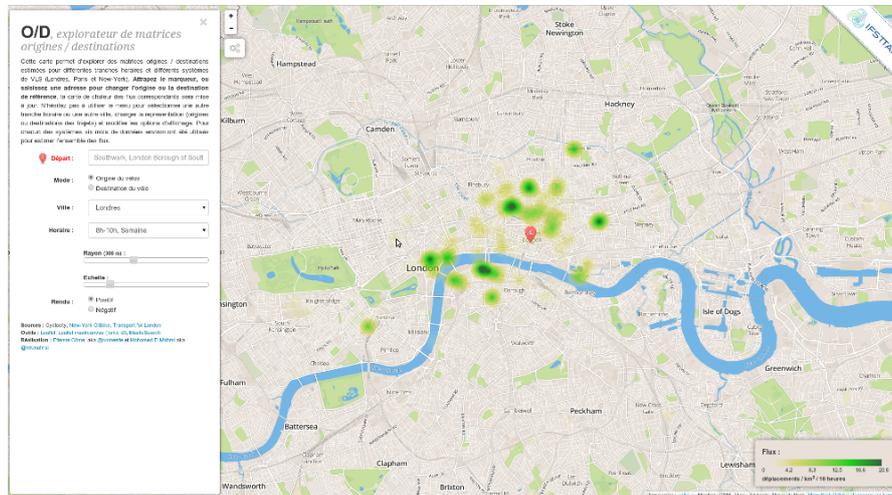


Figure 4. Interface d'exploration proposée pour Londres.

cartographie interactive qui permet de cartographier ce type de matrices. Cette méthode offre la possibilité de changer la focale (l'origine ou la destination de référence) de la carte ce qui permet de simplifier la vue et la rendre plus intuitive en ne présentant à l'utilisateur que les informations pertinentes vis à vis du lieu qui l'intéresse. De plus, afin de tenir compte de la non-stationnarité éventuelle des données, nous proposons de visualiser les matrices O/D sur des petites périodes sur lesquelles le système est supposé être stationnaire. La méthode proposée a été illustrée sur l'exemple du VLS dans trois grandes villes : Paris, Londres et New York. Les cartographies obtenues permettent de positionner les données dans leur contexte géographique, contribuant ainsi à simplifier l'analyse des flux de déplacement.

Bibliographie

- Bertin J. (1967). *Sémiologie graphique*. Paris: Editions Gauthier Villars.
- Cui W., Zhou H., Qu H., Wong P. C., Lin X. (2008). Geometry-based edge clustering for graph visualization. *Transactions on Visualization and Computer Graphics*, vol. 14, n° 6, p. 1227–1284.
- Côme E., Randriamanamihaga A., Oukhellou L., Aknin P. (2014). Spatio-temporal analysis of dynamic origin-destination data using latent dirichlet allocation. application to the vélib' bike sharing system of paris. In *Proceedings of 93rd annual meeting of the transportation research board*.
- Haynes K., Fotheringham A. (1984). *Gravity and spatial interaction models*. Sage.
- Holten D., Van Wijk J. (2009). Force-directed edge bundling for graph visualization. *Computer Graphics Forum*, vol. 28, n° 3, p. 983–990.

- Phan D., Ling X., Yeh R., Hanrahan P. (2005). Flow map layout. In *Ieee symposium on information visualization infovis 2005*, p. 219–224.
- Rae A. (2009). From spatial interaction data to spatial interaction information? geovisualisation and spatial structures of migration from the 2001 uk census. *Computers, Environment and Urban Systems*, vol. 33, p. 161–178.
- Tobler W. (1976). Spatial interaction patterns. *Journal of Environmental Systems*, vol. 6, n° 4, p. 271–301.
- Tobler W. (1987). Experiments in migration mapping by computer. *The American Cartographer*, vol. 14, n° 2, p. 155–163.
- Ullman E. L. (1980). Geography as spatial interaction. In R. R. Boyce (Ed.), p. 13-27. University of Washington Press.
- Wilkinson L. (1979). Permuting a matrix to a simple pattern. In *the statistical computing section of the american statistical association*, p. 409–412.
- Wilkinson L., Friendly M. (2009). The history of the cluster heat map. *The American Statistician*, vol. 63, n° 2, p. 179–184.