



## Revue Sciences/Lettres

2 | 2014

Épistémologies digitales des sciences humaines et sociales

---

# Cartographie des pratiques du Vélo'v : le regard de physiciens et d'informaticiens

Jean-Baptiste Rouquier et Pierre Borgnat

---



### Édition électronique

URL : <http://journals.openedition.org/rsl/487>

DOI : [10.4000/rsl.487](https://doi.org/10.4000/rsl.487)

ISSN : 2271-6246

### Éditeur

Éditions Rue d'Ulm

### Référence électronique

Jean-Baptiste Rouquier et Pierre Borgnat, « Cartographie des pratiques du Vélo'v : le regard de physiciens et d'informaticiens », *Revue Sciences/Lettres* [En ligne], 2 | 2014, mis en ligne le 17 octobre 2012, consulté le 19 avril 2019. URL : <http://journals.openedition.org/rsl/487> ; DOI : [10.4000/rsl.487](https://doi.org/10.4000/rsl.487)

---

Ce document a été généré automatiquement le 19 avril 2019.

© Revue Sciences/Lettres

---

# Cartographie des pratiques du Vélo'v : le regard de physiciens et d'informaticiens

Jean-Baptiste Rouquier et Pierre Borgnat

---

## 1. Introduction

- 1 Les systèmes de vélos en libre service se révèlent aujourd'hui des réussites, leur nombre ne cesse de croître chaque année. Ils permettent de louer un vélo dans l'une des stations maillant la ville et de le rendre peu après dans n'importe quelle autre. Vélo'v est le premier système créé à l'échelle d'une grande ville : il a été développé depuis mai 2005, suivi par Paris et Barcelone en 2007. Ce principe suscite des questions de recherche sur l'usage et la transformation des transports, sur les méthodes scientifiques d'étude d'un système inédit et sur les problématiques de l'utilisateur ou de l'opérateur. Afin de répondre à ces questions, nous avons ressenti le besoin d'utiliser des cartes.
- 2 Cette démarche est donc intermédiaire entre la donnée (le point de départ, la raison de l'étude) et la théorie : théorie dans les buts de l'étude (comprendre le système Vélo'v, suggérer des faits généralisables à d'autres systèmes et des faits stylisés de sociologie), mais aussi théorie des méthodes utilisées et développées pour cette étude. Cela offre des perspectives épistémologiques :
  - parce que la méthode appelle la méthode (ici nous proposons quelques manières de faire des cartes), et une théorisation de son importance (comment le numérique transforme-t-il notre manipulation du texte et des données en général) ;
  - parce que cette étude permet un rapprochement entre SHS et sciences exactes (notamment informatique et physique).

## 2. Pourquoi faire des cartes ?

- 3 Ou encore, comment des physiciens et des informaticiens ont été amenés à s'intéresser aux sciences humaines et sociales et à passer des atomes aux humains.

### 2.1. L'environnement des systèmes complexes

- 4 Cette étude a été effectuée principalement au sein de l'Institut rhône-alpin des systèmes complexes ([www.ixxi.fr/](http://www.ixxi.fr/)). Selon le plan de la feuille de route (Chavalarias *et al.*, 2009), les systèmes complexes peuvent être étudiés soit sous l'angle des objets (par exemple il existe des organismes vivants, ce qui est l'objet de la biologie) soit sous l'angle des méthodes (beaucoup de systèmes évoluent selon des lois simples et déterministes, c'est le sujet des systèmes dynamiques). C'est une démarche par objet plutôt que par méthode qui nous a amenés à nous intéresser au système Vélo'v : appliquer nos méthodes et outils théoriques à cet objet, pour le mesurer, le comprendre, le prédire... Nos diagrammes statistiques n'ont donné qu'une vision partielle ; la carte est une représentation courante et utile, nous avons donc voulu en produire. La section 3 détaille les outils utilisés : essentiellement une construction *ad hoc* à l'aide des outils logiciels qui nous sont familiers, plutôt que l'utilisation de logiciels spécifiques de cartographie, puissants mais difficiles d'accès (il faut en sélectionner un, payer une licence, et apprendre à les utiliser : un investissement lourd qui ne servira que peu de fois). Après avoir produit ces cartes, nous les avons comparées, en grand nombre, ce qui mène facilement à souhaiter une animation, là aussi produite « à la main », solution qui permet plus de souplesse.
- 5 Les idées fusent au départ et il faut choisir les cartes qui sont à la fois réalisables en temps raisonnable, instructives et acceptables pour un journal. Réaliser une carte, la présenter à d'autres chercheurs, donne de nouvelles idées. Par exemple, un détail tel qu'une station vide au milieu d'une zone de stations pleines attire notre attention. Pour mieux comprendre, on peut réaliser une autre carte spécialisée, mais on peut aussi faire des statistiques, extraire les chiffres concernant ce détail : ce n'est pas toujours en une carte que se raffine une carte.

### 2.2. Des masses de données à la disponibilité croissante

- 6 Le début de ce siècle a vu l'émergence progressive de grandes masses de données : la plupart des études reposaient avant sur une collecte manuelle, précise mais nécessairement limitée en nombre. Aujourd'hui, de nombreux jeux de données à grande échelle deviennent accessibles (Science Staff, 2011) : publications et liens sur des réseaux sociaux, traces de communications mobiles ou d'achats, encyclopédies, corpus littéraires, biologie... « Accessible » signifie réunir plusieurs conditions :
- exister sous format numérique, au sein d'un organisme privé ou public ;
  - recevoir une certaine attention : être enregistré avec des conventions stables, dans un format de fichier durable (*i.e.* lisible plusieurs années plus tard malgré des changements de logiciels), sur un support renouvelé régulièrement.
- Ce point est loin d'être le cas systématiquement (ne serait-ce que la conservation plus d'une année !). Les entreprises, essentiellement sur internet, dont les données sont la raison d'être (Twitter, Flickr, Youtube, IMDB...) sont naturellement à la pointe, mais si les entreprises

traditionnelles ont conscience de l'intérêt stratégique de leurs données, elles allouent peu de moyens à leur gestion quand elles n'ont pas un besoin immédiat tel que la facturation ou les obligations juridiques.

- être mis à la disposition de chercheurs (éventuellement sous clause de confidentialité). Là aussi les mentalités évoluent, et de plus en plus d'entreprises voient dans la mise à disposition de leur données l'opportunité d'en retirer des enseignements, voire de favoriser l'émergence de services innovants créés par des tiers.

- 7 Malgré cette disponibilité croissante, un gros jeu de données est toujours considéré comme rare, et un collègue qui découvre un gisement de données facile à exploiter ressent encore une certaine excitation. Ainsi, un travail commence parfois par « Voici de belles données, quelles questions peut-on se poser dessus ? ». Cet état d'esprit fut présent quand nous avons identifié le jeu de données Vélo'v (décrit plus en détail à la section 3.1.). Il constitue en effet un jeu de données rare, car il n'est pas seulement à grande échelle, il est exhaustif. Une collecte manuelle ne peut porter sur un très grand nombre d'individus, une méthode pour pallier cette limitation est donc de soigneusement choisir un échantillonnage. Cette étape est ici balayée : la base contient tous les trajets enregistrés pas l'opérateur. De plus, les frontières du système sont claires : l'influence d'un extérieur (d'un flux de passagers n'habitant pas la zone de couverture de Vélo'v) ne se remarque qu'à l'emplacement des gares ; ce qui représente un problème de moins par rapport à d'autres systèmes.
- 8 Une autre particularité de ces données est leur intérêt stratégique. L'opérateur avait notamment fait fermer le site (O'Brien) qui présente une belle visualisation de ses données en temps réel. On peut regretter que les accords entre municipalités et opérateurs privés ne prévoient pas encore le sort des données. Comme elles sont gérées par l'opérateur, elles sont par défaut sa propriété, mais on peut envisager des contrats qui spécifient que ces données doivent être rendues publiques, notamment car elles sont financées en partie par l'argent public et générées par les utilisateurs.

### Combiner les sources

- 9 Depuis que les données de diverses sources sont faciles à obtenir, sous une forme manipulable automatiquement, il apparaît des sites les combinant, nommés *mashups*. Un exemple de *mashup* consiste à afficher une carte provenant de Google Maps, à y représenter un certain nombre de point d'intérêt cliquables, chaque point d'intérêt étant un lien soit vers une photo du lieu, soit vers un article de Wikipédia sur ce lieu. Ce mouvement a aussi inspiré notre façon d'imaginer une carte, de l'augmenter par des données variées. Nous avons ainsi réalisé une carte combinant :
- le relief et notamment les deux collines de Lyon ;
  - les stations ;
  - les disponibilités : cliquer sur une station affiche ses disponibilités instantanées (vélos et places), ainsi que ses disponibilités historiques (par exemple, « le lundi 8h, il y a 80 % de chances de trouver une place »).
- 10 L'opérateur a souhaité que cette carte ne soit pas publiée.
- 11 On peut noter que notre étude statistique combine entre autres des données Vélo'v, des données démographiques spatialisées et des données météorologiques.

### 2.3. Un intérêt en tant qu'utilisateur

- 12 Lyon avait été choisie pour implanter le premier système de vélo en libre service de grande ampleur. Comme utilisateur de la première heure, avec l'excitation et l'esprit pionnier d'explorer un nouveau système où les usages restent à définir, nous voulions l'utiliser au mieux et éventuellement suggérer des outils utiles aux utilisateurs. À l'intérêt du scientifique s'est donc superposé celui de l'utilisateur, qui voulait optimiser son propre usage des données Vélo'v.
- 13 La culture d'informaticien, pratiquant des TIC, nous fait questionner chaque système, un peu comme les enfants curieux face à un réveil que l'on donnait à démonter. Avoir utilisé de nombreux outils technologiques permet aussi de savoir ce qui est possible ailleurs, et donc d'envisager des sites internet (on penserait aujourd'hui plutôt à des applications mobiles) informant l'utilisateur.
- 14 Pour ces deux raisons, nous avons voulu disposer de cartes indiquant la disponibilité des stations. En effet, si l'on connaît bien la station au pied de chez soi (par exemple à côté d'une université, donc pleine la journée, vidée le soir par les étudiants), on voudrait, en allant faire une course, savoir si l'on déposera facilement son vélo dans tel quartier dans vingt minutes, et si une heure après on trouvera facilement un vélo pour repartir.
- 15 L'opérateur a mis en place un système indiquant ces disponibilités (vélos et places) en temps réel, ce qui est déjà un outil très utile. Pour aller plus loin, nous voulions une prédiction à court terme, une visualisation de l'évolution des disponibilités. C'est ainsi que nous avons commencé avec une carte montrant historiquement comment ces disponibilités ont évolué, en moyenne.

### 2.4. Un outil de communication

- 16 Une carte animée a un attrait certain sur la personne qui visite un site web ou qui passe devant une série de posters (sans doute *via* un réflexe aussi vieux que celui du chasseur ou de la proie) ; en faisant voir battre le cœur de la ville, elle facilite l'accroche du spectateur. Même si elle nécessite des explications, elle est un outil pour capter l'attention et susciter l'intérêt, pour « vendre » ses recherches, *i.e.* pour que le destinataire reparte avec une image claire, positive et durable de l'étude présentée. De plus, elle rend le message plus facile à comprendre et à retenir, selon le vieil adage « Un dessin vaut mieux qu'un long discours ». L'animer est une façon naturelle de prendre en compte le temps et l'évolution du système.
- 17 Ce point participe à un glissement de la façon d'envisager une étude : on prend en compte dès le début le potentiel des résultats à être communiqués facilement, à faire rêver les organismes de financement et le grand public et donc à les faire adhérer au projet. Ces cartes ont d'ailleurs été appréciées par des décideurs.

#### Parenthèse sur la vidéo

- 18 La vidéo a cependant le défaut d'imposer son rythme au lecteur. Le débit d'information devrait être calibré pour un lectorat précis, en fonction de ce qu'il sait déjà du sujet, de son éducation, de l'attention qu'il porte à la vidéo, etc. Nous sommes habitués à sauter des paragraphes, à lire en diagonale, moins à sauter un passage d'une vidéo. Les outils

sont d'ailleurs encore insuffisants et les gros hébergeurs de vidéo en ligne ne proposent même pas la lecture en accéléré.

- 19 Notons cependant que du temps de la dominance de la télévision l'utilisateur de vidéo était passif, aujourd'hui la consultation de vidéo devient active : on choisit ce que l'on veut regarder, un bouton permet (sur certains lecteurs récents) de revenir en arrière pour bien comprendre un passage ou une action sportive. Ces nouvelles fonctionnalités participent à une consultation plus efficace de la vidéo, de même que des projets ambitieux visent à créer automatiquement des métadonnées sur les vidéos : transcription du texte prononcé, détection des changements de scène, identification des personnes voire des objets présents, etc. Il s'agit donc d'une convergence entre texte et image, ou plus précisément une adaptation à l'image et la vidéo des méthodes d'archivage et de consultation éprouvées pour le texte.

### 3. Outils

- 20 Les outils utilisés pour produire nos cartes sont en majorité *ad hoc*, en partie par méconnaissance du métier de cartographe et de ses outils, en partie pour la souplesse que permet une telle approche, enfin car le travail de chercheur est jugé sur les publications et non sur les outils utilisés. En effet, un article de recherche n'inclut pas les petits programmes écrits spécifiquement pour une étude. Rendre un outil utilisable et le publier double environ la quantité de travail nécessaire à le développer pour soi, sans compter la maintenance à assurer ensuite, cela est donc rarement effectué. Cependant, ce type de travail est progressivement reconnu comme faisant partie de la mission du chercheur.
- 21 Les briques utilisées sont classiques : base de données (MySQL), langage de programmation (OCaml, Python), un outil d'analyse de données (Matlab), et quelques utilitaires tels que Mencoder pour assembler en une vidéo les cartes générées une par une pour chaque heure. Nous écrivons par exemple un petit programme qui interroge la base de données, effectue quelques traitements et génère une image SVG (de façon rudimentaire, il nous a par exemple fallu aligner les systèmes de coordonnées du fond de carte et des stations). Nous apprenons en même temps les rudiments de cartographie : nécessité d'une légende, représentation par couleur ou par taille pour une proportion ou une quantité, etc.

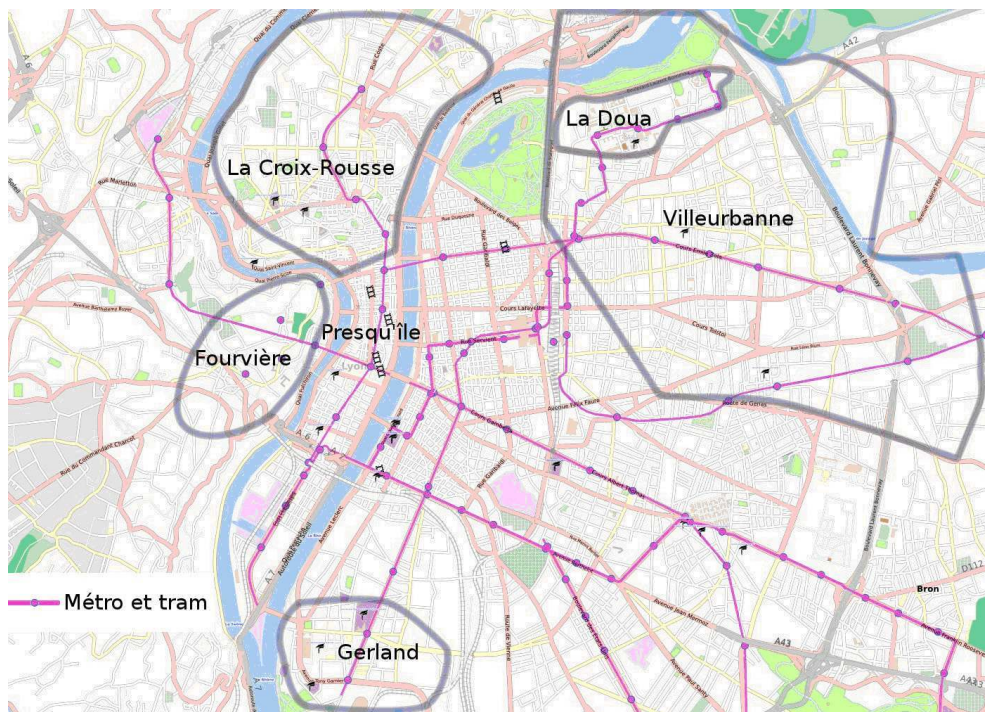
#### 3.1. Les données Vélo'v

- 22 Le système Vélo'v est constitué de 343 stations espacées en moyenne de 250 m, 4 000 vélos utilisés chacun en moyenne cinq fois par jour, pour des trajets moyens de 2.5 km et 15 min. Les données que nous avons pu traiter sont :
- la liste des 13 millions de trajets effectués entre mai 2005 et décembre 2012 ;
  - le nombre de vélos et places libres dans chaque station, heure par heure, pendant 15 mois ;
  - la météo observée (précipitation, température, vent), heure par heure ;
  - des données démographiques, par zone d'environ 2 000 habitants ;
  - les coordonnées, altitude et taille des stations, les jours fériés, le nombre d'abonnés et quelques autres données mineures.
- 23 Sans documentation accompagnant la base, un travail de compréhension a été nécessaire ainsi que le nettoyage des données.

### 3.2. Fond de carte : Open Street Map

- 24 Pour obtenir facilement le droit d'utiliser un fond de carte dans une publication il est préférable qu'il soit libre de droit, ce qui a naturellement orienté notre choix vers Open Street Map (OSM). Pour le définir en référence à des sites connus, OSM est Google Maps en Wikipédia : une carte que chacun peut éditer. Précisément, c'est une galaxie centrée sur une grosse base de données cartographique (25 Go compressée) autour de laquelle gravitent :
- un moteur de rendu, qui transforme ces données vectorielles en images, en une carte consultable sur OpenStreetMap.org et ailleurs ;
  - une communauté d'enthousiastes, qui complètent voire corrigent ces cartes ;
  - des outils pour faciliter les contributions, soit directement en ligne (bouton « éditer » comme sur Wikipédia), soit localement pour plus de performance ;
  - des outils pour tirer partie de ces cartes : plan d'un parc ou d'un quartier utilisé par la municipalité, plan des pistes de ski sur son smartphone, guidage GPS en vacances sans connexion, plan vélo faisant ressortir pistes cyclables, points d'eau et arceaux de stationnement, etc. ;
  - des réalisations variées : comparaison de la qualité des données avec d'autres cartes, statistiques (on y apprend que « Church street » est un nom trois fois plus répandu que « School Street », alors que pour les équivalents allemands, « Kirchstraße » est deux fois moins utilisé que « Schulstraße »), etc. OSM est par définition ouvert, ce qui permet d'en faire beaucoup de choses, et beaucoup de choses en ont été faites, par opposition à d'autres outils propriétaires, mieux finis, mais ne permettant pas de publier ou de réutiliser leurs cartes.
- 25 Les données ont la qualité d'un wiki : détaillé, précis, sans garantie, partiel, sans obligation de suivre les conventions. Elles sont plus complètes que Google Maps en zones denses et développées, notamment pour les piétons (elles mentionnent par exemple les toilettes, escaliers, boîtes aux lettres, etc.) mais lacunaires en zones peu peuplées. OSM a disposé d'une cartographie d'Haïti (créée dans un élan d'aide après le séisme) ou de la Corée du Nord avant Google Maps.

Figure 1 – Quartiers de Lyon mentionnés dans le texte.



- 26 Nous avons utilisé l'un de ces outils pour créer notre propre fond de carte (cf. figure 1), mettant en avant les éléments structurants pertinents pour Vélo'v, notamment les transports en commun, les universités et les cinémas. Quelques retouches manuelles ont été faites avec un logiciel de traitement d'image (Gimp). Regarder un tel fond de carte donne des idées de corrélations et d'effets à tester, par exemple : les trajets se font-ils le long des principales lignes de transport en communs ou bien de façon orthogonale, en complément ?

## 4. Cartes et commentaires

### 4.1. Lignes de métro

- 27 Nous pouvons d'ores et déjà répondre à la question précédente : les trajets se font le long des principales lignes de transport en communs. Sur la carte publiée dans Jensen *et al.* (2010) et reproduite à la figure 2, on retrouve nettement les lignes de métro, déduction faite des trajets impliquant un dénivelé<sup>1</sup>. Une interprétation est que les lignes de métro correspondent bien aux besoins en déplacement (et aux grands boulevards), mais qu'elles ne les absorbent pas entièrement : sur certains trajets le Vélo'v serait préféré pour sa vitesse sur courtes distances, pour son coût moindre, pour sa souplesse, etc.

### 4.2. Probabilité de trouver une place (ou un vélo)

- 28 Avec le point de vue utilisateur mentionné à la section 2.3., nous avons voulu répondre à la question suivante : en arrivant à la station, quelle est la probabilité de pouvoir poser



son vélo ? C'est-à-dire, en moyenne, sur les mois précédents, la station a-t-elle souvent été pleine ou bien y avait-il au moins une place ? Qu'il y ait une seule ou dix places importe peu à l'utilisateur au moment où il arrive. Il ne faut donc pas confondre cette statistique avec le taux de remplissage des stations, dont l'intérêt concerne par exemple les flux de vélos, et partant, de personnes.

- 29 Cette probabilité de pouvoir poser son vélo dépend bien sûr de l'heure et du jour de la semaine, la statistique est donc calculée pour 24h et 7 jours, soit 168 « heures de la semaine » au total.
- 30 Nous avons fait quelques simplifications :
- on ne prend pas en compte le remplissage actuel de la station ;
  - on ne considère que la possibilité de poser son vélo immédiatement. Certaines stations (notamment à côté de la gare Part-Dieu) ont un gros débit et on a de bonnes chances qu'une place se libère en attendant quelques minutes. Inversement, quand une station est pleine, les données ne précisent pas si elle contient effectivement 0 place ou si 2 personnes attendent pour poser leur vélo et le nombre de places est alors mieux modélisé par « -2 » ;
  - on s'intéresse de même à la probabilité de trouver un vélo, et les données ne disent pas si le dernier vélo disponible a un pneu crevé.
- 31 Pour une heure de la semaine donnée, on place alors chaque station sur la carte, colorée en rouge d'autant plus foncé qu'elle a de chances d'être pleine, en bleu d'autant plus foncé qu'elle a de chances d'être vide. On pourrait représenter ces deux informations à la fois. Mais il s'avère qu'il y a très peu de stations qui soient, à une heure donnée, à la fois souvent pleines et souvent vides. On simplifie donc et l'on représente chaque station soit dans un ton de rouge, soit dans un ton de bleu, le plus foncé des deux.
- 32 On crée alors une carte pour chaque heure de la semaine, que l'on enchaîne en une vidéo. Quelques extraits sont sur la figure 3. Les effets les plus saillants, qui se répètent les jours ouvrés, sont les suivants :
- les quartiers résidentiels, par exemple Villeurbanne, se vident le matin et se remplissent le soir ;
  - les universités périphériques, La Doua et Gerland, se vident le soir et se remplissent le matin ;
  - les collines (la Croix-Rousse et Fourvière) sont toujours vides ;
  - la Presqu'île, centre animé, se remplit le soir et se vide brutalement juste après les derniers métros. Vélo'v semble donc un bon complément des transports en commun.
- 33 Cette vidéo est disponible sur [rouquier.org/jb/research/velov](http://rouquier.org/jb/research/velov)

Figure 2 – Stations Vélo'v et flux quotidien moyen dans les rues de Lyon (Jensen *et al.*, 2010). Les trajets sont calculés à partir des flux entre paires de stations, en supposant qu'un cycliste emprunte le même chemin qu'un piéton. Une telle carte peut notamment aider à choisir les endroits où construire des pistes cyclables sera le plus utile.

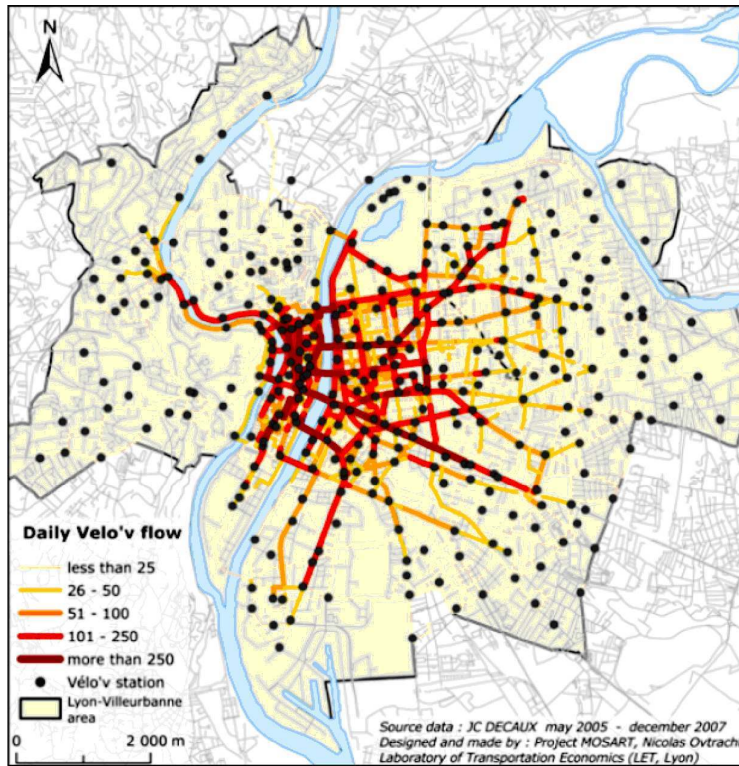
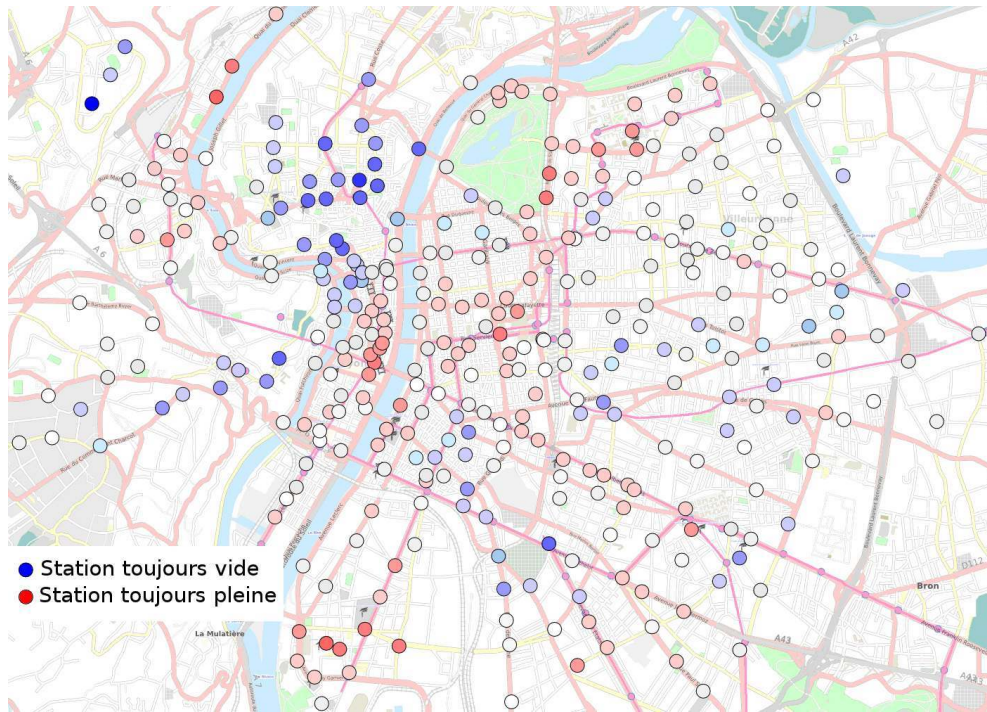
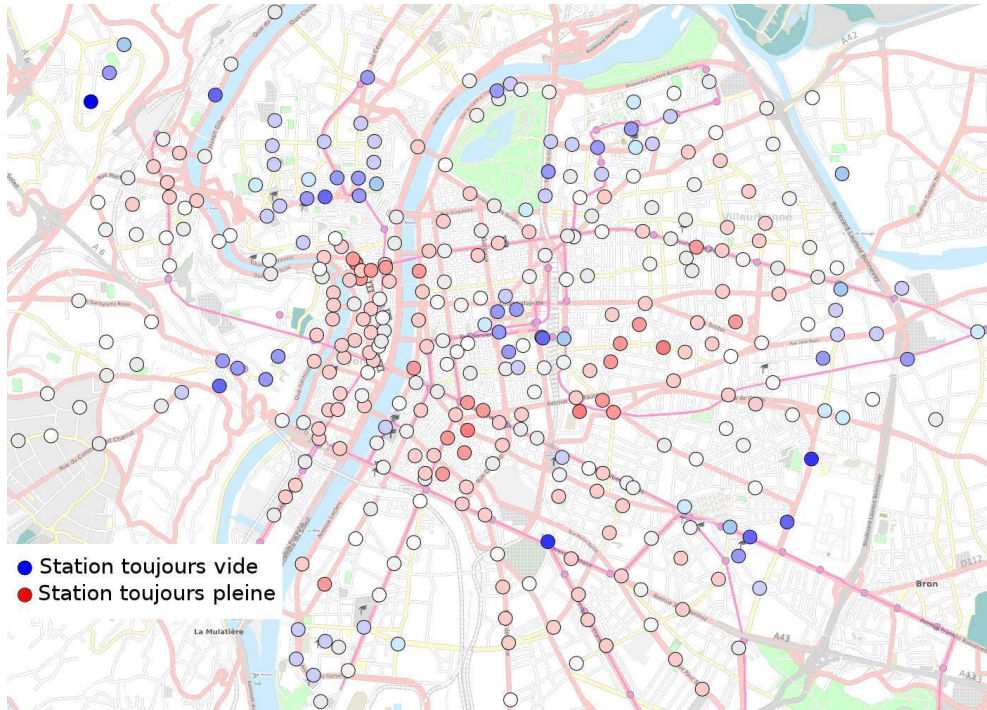


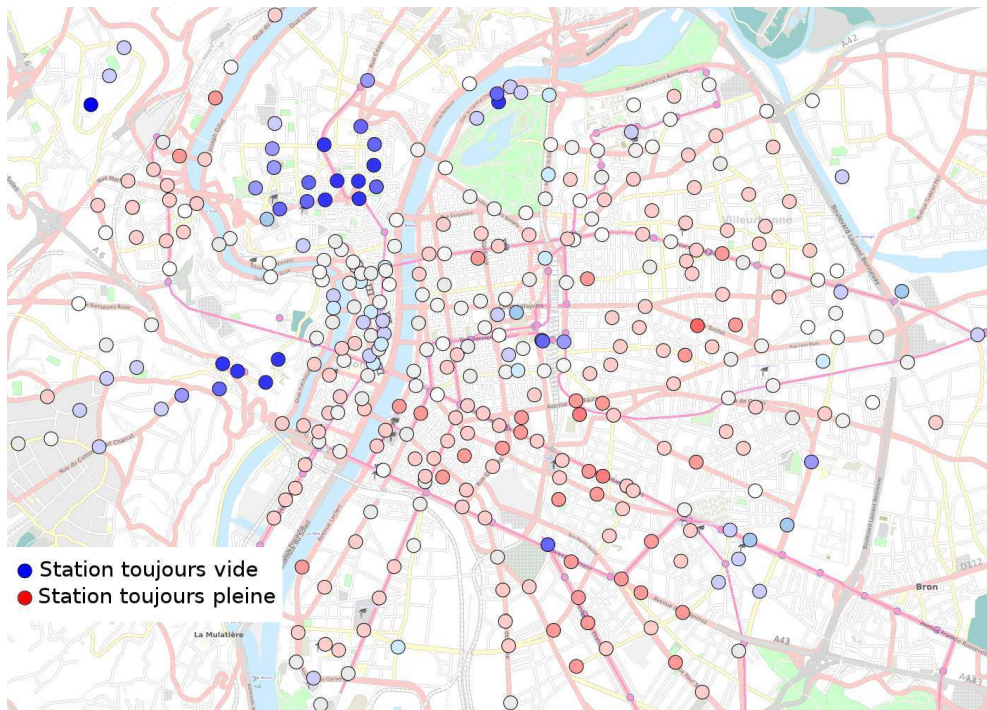
Figure 3 – Probabilité qu'une station soit vide ou pleine.



**MARDI 10H**



**MARDI 20H**



**MARDI 1H**

### 4.3. Résumé des résultats

- 34 Les travaux autour de Vélo'v dans notre équipe ont ainsi conduit à la publication de plusieurs articles (cf. *infra* en bibliographie) ; à chaque fois la discipline de la revue correspond à la discipline de l'un des auteurs.
- 35 Les conclusions principales de ces publications sont :
- Le fonctionnement des stations diffère grandement selon leur emplacement. Les effets dominants sont la descente des collines et les trajets pendulaires (entre quartiers résidentiels et quartiers d'emplois et d'études), avec un pic d'utilisation supplémentaire à midi. Puis viennent des effets cycliques assez attendus : dans un quartier de commerces et d'affaires, le remplissage des stations est bien moindre lors des mois d'hiver ; dans un quartier universitaire l'activité est dictée par le calendrier scolaire ; les parcs attirent les vélos le week-end après-midi, les zones commerciales, le samedi mais pas le dimanche.
  - On peut expliquer le nombre de locations d'une heure donnée à l'aide d'un modèle linéaire contenant la saison, le jour et l'heure (qui est un effet cyclostationnaire), la météo (température et pluie) et le nombre d'abonnés (*i.e.* la taille du système, qui est une évolution lente non stationnaire), avec une erreur quadratique moyenne de 10 %.
  - La régulation (transport de vélo sur remorque) se fait depuis les stations des gares, centre commerciaux et campus vers le haut des collines.
  - Regrouper les stations selon le nombre de vélos échangés entre elles conduit à des groupes bien déterminés géographiquement, car la plupart des trajets sont locaux. La moyenne des distances est en effet de 2.5 km, celle des durées de 1/4h. On ne peut en déduire directement la vitesse moyenne car les trajets plus longs sont plus rapides. La vitesse moyenne varie ainsi de 14.5 km/h tôt en semaine, à 10 km/h les après-midi de week-end. On observe aussi une vitesse moyenne plus élevée entre 12h et 14h, ou bien en hiver, ou encore les mercredis.
  - Concernant la distance parcourue et l'itinéraire emprunté, les vélos ont un comportement plus proche de celui des piétons que de celui des voitures.
- 36 L'intérêt de l'étude est :
- pour le physicien : tester les méthodes d'analyse de données et de traitement du signal, en développer de nouvelles ;
  - pour le sociologue : comprendre comment les lyonnais se sont appropriés ce mode de transport ;
  - pour l'opérateur : optimiser l'exploitation ;
  - pour les élus : contrôler la bonne réalisation de l'opérateur, évaluer l'impact du système, présenter un bilan.

## 5. Bilan épistémologique

- 37 Nous avons développé ces cartes pour avoir une vision globale du système Vélo'v et rapprocher ces analyses des quartiers que nous connaissons pour les avoir parcourus. Nous sommes conscients qu'elles n'ont pas le fini que l'on peut trouver sur des cartes faites par des géographes. Elles ont été développées en même temps que d'autres analyses, dans une même étude du système. Cela nous a permis d'interroger les rapprochements et différences entre SHS et sciences dures, dont voici les principaux enseignements.

## 5.1. Réflexion sur nos pratiques

- 38 Notre notion de carte a quelques différences, et est sans doute moins mature que celle des cartographes. C'est une représentation des données parmi d'autres, adaptée pour visualiser la composante spatiale des données. Un symptôme de cette approche est que certaines figures que nous avons produites superposent à la carte nos schémas habituels tels qu'un champ de vecteurs : nos cartes pourraient être plus simples. L'un de nos buts est de condenser l'information pour que celui qui a produit la carte voit apparaître une organisation globale. Si cette information ressort, la carte peut être publiée sans grande modification, à destination de collègues de la même discipline. Il nous semble qu'en cartographie l'objet d'étude est d'abord le territoire, et la carte en étant sa représentation mentale constitue un passage obligé. Le but y est plus souvent de simplifier l'information pour faire passer un message clair.
- 39 Notre pratique des sciences dures apporte une expérience qui aide à identifier ce qui est réalisable ou non parmi différentes manières de transformer les données en cartes : ce que l'on peut attendre des statistiques, les visualisations (éventuellement animées) existantes et à créer, etc. Cette expérience influe ensuite sur le choix des analyses que l'on va réaliser. Manier les outils lors de ces analyses enrichit l'expérience acquise, constituant ainsi une boucle de rétroaction.
- 40 Dans notre étude, il n'y a pas eu comme on pourrait le penser d'allers-retours nombreux entre une idée, une carte, une nouvelle idée d'étude à mener, une nouvelle carte, qui génère à son tour des questions. Le cheminement est plutôt descendant : les idées de cartes à réaliser foisonnent, nous réalisons celles que nous estimons à la fois instructives et raisonnablement facile à produire, les leçons apprises lors de cette production permettent de mieux filtrer les idées selon ces deux critères (intérêt et coût). Les idées de nouvelles cartes à produire sont plus souvent générées à partir des idées précédentes qu'à partir des cartes réalisées.
- 41 L'outil est créé en même temps que le résultat. D'ailleurs beaucoup de publications en sciences dures mettent en avant l'outil ou la méthode comme contribution principale de l'article, plus que les résultats qu'il a permis d'obtenir dans un cas particulier.

## 5.2. Impact de l'informatique et des masses de données

- 42 Nous assistons peut-être à l'avènement du « macroscopie », selon l'expression de Börner (2011) :
- Just as the microscope made it possible for the naked human eye to see cells, microbes, and viruses, thereby advancing biology and medicine, and just as the telescope opened the human mind to the immensity of the cosmos and the conquest of space - the macroscopie promises to help make sense of yet another dimension - the infinitely complex. Macroscopes provide a « vision of the whole », « helping us synthesize » the related elements and detect patterns, trends, and outliers while granting access to myriad details.
- 43 Les cartes sont un outil de représentation et d'écriture du monde. Puisque l'informatique et l'internet renouvellent l'écriture, ils renouvellent l'épistémologie des SHS. On peut prendre l'image d'un morceau d'argile : un ordinateur est un outil aux possibilités innombrables, que l'on modèle selon ses besoins ; l'informatique apporte une grande souplesse, rendant beaucoup de traitements possibles, rapides (le délai entre une idée et

sa réalisation est raccourci) voire automatisés. On peut dès lors imaginer de nombreuses représentations, ce qui fait que, ironiquement, l'imagination redevient l'une des limites de la production : si beaucoup de cartes sont possibles, encore faut-il que l'idée naisse dans le cerveau d'un chercheur en géographie ou en informatique. L'informatique favorise la quantité : puisque la production est facile, il est possible de réaliser beaucoup de cartes « juste pour voir ». Bien pensée, cette quantité peut devenir qualité, par exemple par une exploration systématique, ou bien en stimulant la créativité par de nombreuses propositions.

- 44 Bien maniée, la taille des données augmente aussi la confiance dans les résultats. Mais ce n'est pas là la seule évolution dans les pratiques. En effet, avoir quantitativement plus de données produit du qualitatif : plus que se restreindre à un quartier, on peut étudier la ville entière qui coûtait avant trop cher en collecte de données ; plus que regarder des mouvements d'ensemble, on peut étudier des détails fins qui étaient avant perdus dans le bruit. On peut pratiquer une sociologie qui étudie une grande population et pourtant prend en compte chaque individu dans ses particularités, sans simplifier chacun en un individu type. Aussi, on fait de plus en plus d'analyse :
- visuelle : une représentation graphique permet parfois de découvrir une pépite qui ne ressemble pas à des résultats précédents ;
  - exploratoire : on ne cherche plus à valider une hypothèse formulée *a priori*, mais à déceler des phénomènes intéressants dans les données.
- 45 Le second point est une nouvelle approche majeure. La connaissance est désormais autant produite en commençant par les données que par des théories.

#### Remerciements

Éric Guichard pour m'avoir relancé jusqu'à ce que cet article prenne corps, Pablo Jensen, Patrice Abry, Patrick Flandrin, Éric Fleury, Antoine Scherrer, Céline Robardet, Luc Merchez pour avoir participé à l'étude de Vélo'v, JC Decaux pour nous avoir confié les données.

---

## BIBLIOGRAPHIE

- Borgnat, P., Abry, P. et Flandrin, P., « Modélisation statistique cyclique des locations Vélo'v à Lyon », *22<sup>e</sup> Colloque sur le traitement du signal et des images. GRETSI*, 2009a.
- Borgnat, P., Abry, P., Flandrin, P. et Rouquier, J.-B., « Studying Lyon's Vélo'v : a statistical cyclic model », *European Conference on Complex Systems, ECCS'09*, 2009b.
- Borgnat, P., Fleury, É., Robardet, C. et Scherrer, A., « Spatial analysis of dynamic movements of Vélo'v, Lyon's shared bicycle program », *European Conference on Complex Systems, ECCS'09*, 2009c.
- Borgnat, P., Robardet, C., Rouquier, J.-B., Abry, P., Fleury, E. et Flandrin, P., « Shared bicycles in a city : A signal processing and data analysis perspective », *Advances in Complex Systems*, vol. 14, n° 3, 2011, p. 415-438.
- Börner, K., « Plug-and-play macroscopes », *Communications of the ACM*, vol. 54(3), 2011, p. 60-69.
- Chavalarias, D. et al., *French roadmap for complex systems 2008-2009*, 2009.

Jensen, P., Rouquier, J.-B., Ovtracht, N. et Robardet, C., « Characterizing the speed and paths of shared bicycles in Lyon », *Transportation Research Part D : Transport and Environment*, vol. 15(8), 2010, p. 522-524.

Merchez, L. et Rouquier, J.-B., « L'usage des vélos en libre service (VLS) comme révélateur des rythmes urbains : le cas des stations de Vélo'v à Lyon », *Données Urbaines*, n° 6, 2011.

Michau, G., Robardet, C., Merchez, L., Jensen, P., Abry, P., Flandrin, P. et Borgnat, P., « Peut-on attraper les utilisateurs de Vélo'v au Lasso ? », *23<sup>e</sup> Colloque sur le traitement du signal et des images. GRETSI*, 2011.

O'Brien, O. G., Bike share maps.

Science Staff, « Dealing with Data », *Science*, vol. 331(6018), 2011.

## NOTES

1. Les lignes A et B sont ainsi bien dessinées par les trajets en Vélo'v, la ligne C est absente car elle monte la colline de la Croix-Rousse, la ligne D est aussi nette, sauf lorsqu'elle traverse la colline de Fourvière pendant que les cyclistes la contournent.

---

## RÉSUMÉS

L'étude des données de location du système de vélos en libre service Vélo'v, installé dans Lyon et Villeurbanne, amène à poser des questions de méthodes quand il s'agit, en tant qu'informaticien ou physicien et non de géographe ou de cartographe, de trouver comment faire des cartes représentant ces données. Partant d'une discussion sur les masses de données numérisées accessibles aux scientifiques, tout particulièrement pour des études en sciences sociales, des outils utiles pour manipuler sont évoqués. L'exemple de l'analyse des déplacements en Vélo'v permet d'illustrer pourquoi et comment réaliser des cartes qui décrivent des résultats sur les types de déplacements effectués ou sur la disponibilité des vélos ou des places aux stations. Une conclusion traite des pratiques, en partie nouvelles, liées aux masses de données.

From the study of the data of Vélo'v, the Bicycle Sharing System of Lyon and Villeurbanne, several issues are discussed about how researchers from computer science and physics, and not from geography or cartography, were lead to draw maps representing these data. First, the diversity of available digital data is shown, especially for studies related to social or human studies. Some tools to manipulate them are discussed. The example of the Vélo'v data of trips illustrates why and how to propose maps describing features about the movements made with these bikes, or about the availability of bikes or free stands at stations. This practice gives way to a conclusion that suggests a whole new meaning related to the mass of data.



## INDEX

**Mots-clés :** Vélo'v, Vélib, VLS (Vélos en libre service), carte, cartographie, open street map, système complexe, informatique, physique, analyse de données, épistémologie

**Keywords :** Vélo'v, vélib, BSS (Bicycle Sharing System), map, cartography, open street map, complex systems, computer science, physics, data analysis, epistemology

## AUTEURS

### JEAN-BAPTISTE ROUQUIER

Enseignant à l'ENS de Lyon et ingénieur chercheur à Eonos.

Parmi les publications :

Avec C. R. Shalizi, R. Haslinger, K. L. Klinkner et C. Moore, « Automatic filters for the detection of coherent structure in spatiotemporal systems », *Physical Review E*, vol. 73, n° 3, 2005, p. 036104.

Avec P. Jensen, P. Kreimer et Y. Croissant. « Scientists who engage with society perform better academically », *Science and Public Policy*, vol. 35, n° 7, 2008, p. 527-541.

Avec X. R. Yan, Y. J. Zhu et C. Moore, « Active learning for hidden attributes in networks », in E. Airoldi, J. Kleinberg, J. Leskovec et J. Tenenbaum (éd.), *22nd Annual Conference on Neural Information Processing Systems, Workshop on Analyzing Networks and Learning with Graphs*, décembre 2009.

Avec P. Jensen, N. Ovtracht, et C. Robardet, « Characterizing the speed and paths of shared bicycles in Lyon », *Transportation Research Part D : Transport and Environment*, vol. 15, n° 8, 2010, p. 522-524.

### PIERRE BORGNAT

Chargé de recherche CNRS au Laboratoire de physique de l'ENS de Lyon.

Parmi les publications :

Avec G. Dewaele, K. Fukuda, P. Abry et K. Cho, « Seven Years and One Day: Sketching the Evolution of Internet Traffic », *Proceedings of the 28th IEEE INFOCOM 2009*, Rio de Janeiro, 2009, p. 711-719.

Avec P. Abry, P. Flandrin, C. Robardet, J.-B. Rouquier, et E. Fleury, « Shared Bicycles in a City : A Signal Processing and Data Analysis Perspective », *Advances in Complex Systems*, vol. 14, n° 3, 2011, p. 415-438.

Avec N. Tremblay, A. Barrat, C. Forest, M. Nornberg et J.-F. Pinton, « Bootstrapping under constraint for the assessment of group behavior in human contact networks », *Arxiv :1212.3524*, décembre 2012.

Avec C. Robardet, P. Abry, P. Flandrin, J.-B. Rouquier, et N. Tremblay, « A Dynamical Network View of Lyon's Vélo'v Shared Bicycle System », in N. Ganguly, A. Mukherjee, B. Mitra, F. Peruani et M. Choudhury (éd.), *Dynamics of Time-Varying Networks*, Springer, 2013.