Factorisation matricielle sous contraintes pour l'analyse des usages du métro parisien

Mickael Poussevin[1,2], Nicolas Baskiotis[1,2] et Vincent Guigue[1,2]

[1]Université Pierre et Marie Curie, PRES Sorbonne-Universités
[2]Laboratoire d'Informatique de Paris 6, UMR 7606, CNRS

Résumé

Le but de cet article est de présenter une approche robuste pour catégoriser les usages du métro parisien. En collaboration avec le STIF (Syndicat des Transports en Ile de France), nous avons analysé l'ensemble des données billétiques (validation de titre de transport) sur le réseau métro (14 lignes) durant 3 mois pour les usagers disposant d'un forfait mensuel. Cela représente environ 80 millions de déplacements effectués par 600 000 usagers. Pour étudier les usages, nous avons décrit un usager en fonction des stations qu'il utilise plus ou moins fréquemment et nous nous sommes focalisés sur ses habitudes journalières et hebdomadaires ce qui correspond à plusieurs vues d'un même utilisateur. Nous avons ensuite appris un dictionnaire des usages par factorisation matricielle. Les contraintes de parcimonie nous ont permis d'isoler des comportements usuels significatifs. Chaque utilisateur et station du réseau peut ensuite être re-exprimé comme une combinaison de ces usages, ce qui permet de grouper des populations et des zones géographiques sur un critère pertinent.

Mots-clef : Factorisation matricielle, clustering, comportement

1 Introduction

La mobilité urbaine est une préoccupation majeure aujourd'hui : les transports en communs sont au coeur des politiques d'aménagement du territoire et l'ANR a créé un défi spécifique sur la thématique. C'est dans ce contexte que nous proposons d'analyser les traces des usagers dans le métro parisien.

Les données billétiques, fournie par le STIF (Syndicat des Transports en Ile de France), correspondent aux traces anonymisées des utilisateurs lorsqu'ils valident leur titre de transport sur les bornes du métro. Ces données sont volumineuses (80 millions de validations et 600 000 usagers sur 91 jours) et bruitées (seule l'entrée dans le réseau est référencée) ce qui expliquent qu'elles n'aient pas été exploitées jusqu'à maintenant. Nous proposons ici une étude démontrant le potentiel des données billétiques pour analyser les usages du métro parisien.

La littérature sur la mobilité urbaine est vaste et hétérogène, de l'évaluation des politiques d'aménagement du territoire [BPS02] à la caractérisation des déplacements en véhicules personnels [BHG06], taxis [PJW+12], transports en commun [FKR+13], vélos en libre service [RCOG13] ou même piétons [LXMW12]. Depuis une dizaine d'années, les études quantitatives se multiplient parallèlement au déploiement de moyens technologiques permettant de suivre les usagers : les réseaux cellulaires ont permis d'analyser les échelles des déplacements en fonction de leur fréquences [BHG06] et plusieurs études démontrent même qu'il est possible de prévoir une grande majorité des déplacements que nous faisons quotidiennement [SQBB10]. Concernant les transports en commun, certaines études se focalisent sur des données de sondages pour caractériser les changements de comportements liés à la mise en service de nouvelles lignes [Gol02]. Les données quantitatives sont récentes et liées à l'adoption de systèmes d'identification RFID des usagers à Londres, Lisbonne ou Paris. Elles ont jusqu'ici été exploitées pour la mise en évidence des goulets d'étranglement spatio-temporels des réseaux [CSC12] ou la prédiction de certains usages comme le fait d'utiliser une ligne de bus un jour donné [FKR+13]. Cependant à ce jour, aucun article n'étudie les habitudes et usages des population dans le temps et l'espace. De telles approches ont été tentées sur les taxis à Shanghai [PJW+12] ou les Vélib' parisien [RCOG13] mais sans pouvoir suivre un même utili-

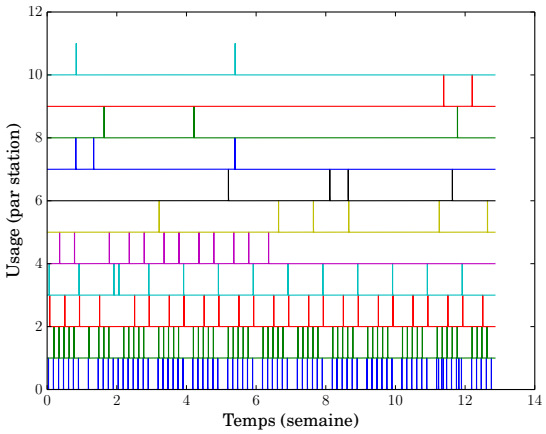

FIGURE 1 – Représentation d'un usager du réseau sur 91 jours en triant les logs par station. Pour cet utilisateur, 10 stations sont utilisées, plus ou moins fréquemment.

sateur sur plusieurs trajets. Les données dont nous disposons maintenant nous permettent donc d'aborder cette tâche sous un nouveau jour en centrant notre étude sur l'usager lui-même.

Comme le montre la figure 1, nos informations sont des triplets (usager, station, temps). Les données sont riches mais difficile à exploiter : nous proposons d'agréger les données sur les trois échelles de fréquences et deux échelles temporelles. Nous distinguons donc d'une part les stations fréquemment utilisées de celles plus marginale et nous exprimons le temps de validation en fonction de l'heure de la journée et du jour de la semaine. Une fois chaque utilisateur ramené à une forme vectorielle, nous effectuons une factorisation matricielle à la manière de [PJW+12]. Nous ajoutons des contraintes de parcimonie et de forme pour obtenir une décomposition explicite : nous apprenons *in fine* un dictionnaire de fonctions temporelles correspondant à des usages. Un des atomes du dictionnaire représente par exemple le fait de valider à 8h45, 5 jours par semaine, ce que nous interprétons comme un départ au travail. En parallèle, chaque utilisateur est décomposé comme un sous-ensemble pondéré des atomes du dictionnaire. Cette technique, parfois aussi appelée *apprentissage de dictionnaire* est courante en traitement du signal et particulièrement en analyse musicale pour déterminer les principales composantes d'un morceau et séparer différentes sources (instruments, chanteurs...) [WP05].

2 Etat de l'art

Nous proposons un état de l'art en deux parties qui aborde d'une part le problème de la mobilité urbaine et ensuite les techniques de factorisations matricielles et leur mode de fonctionnement.

2.1 Mobilité urbaine

Dans la littérature, le problème de la mobilité urbaine est étudié à différent niveau. D'abord au niveau des politiques d'aménagement du territoire [BPS02] pour promouvoir des indicateurs de performances mesurable puis au niveau des déplacements eux-mêmes ensuite. Les trajets en voiture font l'objet de plusieurs études [BHG06, GHB08] : les auteurs caractérisent essentiellement l'échelle des déplacement au cours du temps en utilisant les réseaux de téléphonie mobile pour localiser et suivre une population cible. L'équipe de A. Barabàsi insiste sur la prédicabilité de nos déplacements qui sont très récurrents. Dans leur étude [SQBB10], ils montrent que plus de 90% de nos mouvements sont théoriquement prédictibles. Dans la même logique, leur article suivant [WPS+11] va plus loin en mettant en parallèle les comportements de déplacements concrets et les profils des usagers dans les réseaux sociaux. Pour ce faire, les auteurs travaillent sur des données complètes de téléphonie mobile (CDR, Call Detail Record) contenant la localisation ainsi que les données échangées. Ils mettent en évidence le parallèle entre les profils virtuels et de déplacement pour conclure que ces derniers sont caractéristiques des usagers. L'étude très récente [LLC+14] propose une caractérisation des déplacements (toujours basé sur les traces collectées dans les réseaux de téléphonie mobile) pour les jours de semaine dans les 31 principales agglomérations espagnoles. Les auteurs mettent en évidence les points chauds de chaque ville, c'est à dire les centres névralgiques rassemblant le plus de population. Ils montre que la dynamique et la répartition de ces points est caractéristique de chaque ville.

Les données de téléphonie mobile sont aussi utilisées pour l'analyse du trafic routier et la détection d'anomalie [Her10]. [LZC+11] propose même une analyse sur la causalité des anomalies : une première description sous forme d'arbre permet d'identifier les comportements typiques (via les branches fréquentes), puis les anomalies sont détectées comme des écarts aux chemins fréquents et l'article se concentre sur les effets de causalité entre anomalies pour isoler la source des problèmes. Les données viennent des GPS de 33000 taxis pékinois sur 6 mois et représentent 800 millions de kilomètres.

[PJW+12] travaille également sur les traces GPS de 2000 taxis de Shanghai mais ils utilisent la factorisation matricielle pour mettre en évidence les habitudes collectives des chauffeurs.

Ces dernières références illustrent une tendance récente : la prise en compte de différents capteurs pour localiser les utilisateurs dans leur activités. [LXMW12] décrit l'utilisation de scanners bluetooth pour la reconstruction des trajectoires piétonnes des visiteurs du zoo de Duisburg. Les systèmes de vélo en libre service permettent également de suivre les trajets des utilisateurs du service. Les études [BC11, RCOG13] analysent respectivement les parcours à Londres et Paris. La première référence se positionne plus sur l'analyse de graphe et la visualisation d'une grande masse de données connectées et met en avant les notions de connectivité et de centralité pour exprimer les clusters comportementaux. La seconde étude concerne 2.5 millions de trajets parisiens et permet aux auteurs de catégoriser les trajets et leur dynamique : ils analysent quelques clusters spatio-temporels centrés sur les horaires de bureau, les week-end au parc et les transports nocturnes une fois les transports en commun fermés. Cependant, les données de cet article ne sont contiennent pas d'identifiant utilisateur et il n'est donc pas possible de suivre les acteurs du système.

Concernant les transports en commun, plusieurs études se basent sur les analyses qualitatives et les données de sondage qui se focalisent en général sur une portion de trajet ou un usage en particulier [Gol02]. Depuis une dizaine d'année, l'adoption de systèmes d'identification RFID des usagers à Londres, Lisbonne ou Paris ouvre la voie à des études quantitatives. [CSC12] propose une étude sur la répartition spatio-temporelle des usagers du métro londonien et isole les goulets d'étranglements du réseau dans le but de trouver des solutions pour fluidifier le trafic. L'étude porte sur une période d'un mois et les auteurs se sont focalisés sur trois stations particulières qui sont prises comme références pour modéliser les zones résidentielles, de travail et les hubs (gares). Les données sont agrégées et un seuil est défini à partir duquel une station est déclarée en état de surcharge. Une étude du réseau lisboète centrée sur l'utilisateur propose de personnaliser l'accès à l'information sur les problèmes de trafic dans le réseau en prédisant les usages d'une personne dans les transport [FKR+13]. Les données rassemblent 24 millions de trajets et 800 000 usagers sur 61 jours et permettent d'extraire des profils d'usage pour les jours de semaine et le week-end mais les auteurs se sont concentrés sur la tâche de prédiction à l'échelle de la journée et des lignes de bus : ils cherchent à anticiper le fait qu'une personne emprunte une ligne un jour donné.

Notre approche est basé sur les usagers : les abonnements étant nominatifs, il devient possible de suivre les personnes et de décrire les habitudes pour extraire des usages classiques à l'échelle de la journée et de la semaine. Nous utilisons une approche à base de factorisation matricielle similaire à [PJW+12, RCOG13] mais le suivi des usagers nous permet d'ajouter la notion de périodicité dans la caractérisation des trajets.

2.2 Décomposition des sources composant un signal

En changeant complètement de domaine applicatif, notre approche se rapproche des travaux sur la séparation de sources (type Independant Component Analysis) musicale [WP05] ou de l'identification des notes dans un morceau [VBB08]. Certaines approches récentes permettent de modéliser directement les signaux composés d'impulsions [HB11], cependant la variabilité intrinsèque de nos données nous pousse plutôt vers une description lissée des données : nous considérons que la mesure de temps associée aux validations n'est pas significative à la minute près, de nombreux facteurs pouvant expliquer des variations légères.

Les algorithmes de factorisation matricielle sont couramment utilisés en analyse de données de puis longtemps et parfaitement décrits dans [GVL96]. Dans notre cas, il s'agit de décomposer la matrice des logs des usagers deux parties : d'une part un dictionnaire, composé de distributions de probabilités de logs dans le temps et d'autre part un code, exprimant chaque utilisateur comme une combinaison d'atomes du dictionnaire. Ce problème est parfois renommé *dictionary learning* [KDMR+03] pour insister sur l'optimisation de la forme des atomes du dictionnaire. Le problème de décomposition d'un signal sur une famille de fonction est quand à lui classique depuis longtemps : la littérature sur le sujet est très vaste et comporte des références fameuse comme [Tib96].

La factorisation matricielle résout un double problème (apprentissage du dictionnaire et décomposition sur ce dictionnaire) comportant beaucoup de paramètres. Afin d'obtenir des résultats sensés, il est nécessaire d'ajouter des contraintes sur la décomposition. Les plus classiques sont la non-négativité et la parcimonie [Hoy04]. La première signifie que les éléments à décomposer (les profils d'usager dans notre problème) sont exprimés comme une somme pondérée positivement des éléments du dictionnaire (aucune soustraction n'est admise). La

seconde contrainte est une forme de régularisation : le but est de reconstruire les éléments en utilisant un minimum d'atomes du dictionnaire. Nous utiliserons la régularisation $\mathcal{L}_1$ qui est reconnue pour son efficacité dans les formulations du LASSO [Tib96] et du compressive sensing [Bar07]. Pour la résolution, nous avons opté pour une variante de la formulation multiplicative de [LS00].

3 Modèle

Nous commentçons par décrire la mise en forme des données que nous avons retenue. Nous étudierons ensuite en détail l'algorithme de factorisation matricielle implémenté puis nous verrons la contrainte spécifique que nous avons ajouté pour donner plus de sens aux atomes du dictionnaire.

3.1 Description des données

Pour chaque utilisateur (cf fig. 1), nous avons distingué les stations utilisées fréquemment ($\geq$ 24 fois dans la base, c'est à dire 2 fois par semaine), modérément ($\geq$ 8 fois) et rarement. Nous avons fusionné toutes les informations dans chacune des trois catégories sur deux échelles de temps : chaque log est référencé au niveau d'une journée (sur des intervalles de 15 minutes) et au niveau d'une semaine (sur des intervalles de 2 heures). Chaque utilisateur i est donc composé de 3 échelles fréquentielles $\mathbf{x}_i = \{\mathbf{x}_i^{freq}, \mathbf{x}_i^{mod}, \mathbf{x}_i^{rare}\}$ avec $\mathbf{x}_i^j = \{\mathbf{x}_i^{j,jour}, \mathbf{x}_i^{j,semaine}\}$.

La taille des intervalles de temps considérés donne : $\mathbf{x}_i^{j,jour} \in \mathbb{R}^{96}$ et $\mathbf{x}_i^{j,semaine} \in \mathbb{R}^{84}$. Les données sont normalisées de façon que : $\sum_t x_{it}^{j,jour} = 1$ et $\sum_t x_{it}^{j,semaine} = 1$. Les deux échelles temporelles sont concaténées mais les échelles fréquentielles sont décrites distinctement : nous avons donc 3 vues des données $\{X^{freq}, X^{mod}, X^{rare}\} \in \mathbb{R}^{3\times N\times 180}$. TODO valeur de N

i	indice des usagers	t	indice temporel
j	indice de l'échelle (fréquent, modéré, rare)		
k	indice des fonctions (ou atomes) du dictionnaire		
n	indice des stations de métro		

TABLE 1 – Tableau récapitulatif des significations des indices

3.2 Factorisation matricielle

Nous traitons les décompositions de X^{freq}, X^{mod} et X^{rare} de manière indépendante : ces trois matrices correspondent pour nous à différents usages et il semble donc naturel de construire des dictionnaires spécifiques pour les différentes échelles de fréquences. Chaque matrice X est approximée par un code U sur un dictionnaire V en résolvant le problème suivant [1] :

$$\underset{U,V}{\operatorname{argmin}} \left((X - UV)^2 + \lambda \|U\|_1 \right) \quad \text{avec} : \forall k, \ \sum_t v_{kt} = 1 \tag{1}$$

Le problème (1) est résolu en utilisant l'algorithme hybride proposé dans [Hoy04].

1. Initialisation aléatoire de :
   U (code) et V (dictionnaire)
2. Boucle d'optimisation
   (a) Gradient : $V = V + \mu U^T(X - UV)$
   (b) Projection : $V = \max(V, 0)$
   (c) Décomposition parcimonieuse :
   $U = U. * \left((XV^T)./(UVV^T + \lambda)\right)$

De base, un tel système aboutit à la constitution d'un dictionnaire contenant des atomes parcimonieux mais multi-modaux comme le montre la figure 2 (haut). Le premier atome représente à la fois un départ au travail autour de 9h30 et un retour vers 19h, 5 jours par semaine, les deux atomes suivants montrent des comportements plus variables. Nous souhaitons nous focaliser sur des comportements plus ciblé et notre hypothèse est que ces comportements correspondent à un seul horaire dans la journée, nous avons donc ajouté une contrainte gaussienne sur la forme des atomes journaliers, en centrant la gaussienne sur le pic le plus haut :

$$\forall i, \quad \mathbf{v}_k^{jour} \leftarrow \mathbf{v}_k^{jour}. * \exp(-\frac{(\mathbf{t} - t_{max})^2}{2\sigma^2}) \tag{2}$$

où $.*$ désigne la multiplication terme à terme, t est le vecteur temps, t_{max} le temps auquel $\mathbf{v}_k^{jour}$ atteint son maximum et σ est l'écart type temporel de l'atome $\mathbf{v}_k^{jour}$. En intégrant cette contrainte toutes les 20 itérations, nous obtenons des atomes mono-modaux. Cette contrainte de mono-modalité joue pleinement son rôle comme le montre les atomes de la figure 2 (bas).

3.3 Clustering

L'identification des usages du métro est un objectif intermédiaire de cet article, nous souhaitons *in fine* proposer un clustering sur les populations d'usager et les stations en exploitant la base des usages.

1. Les trois échelles fréquentielles étant traitées de manière indépendante, nous avons retirée les indices j pour gagner en clarté.

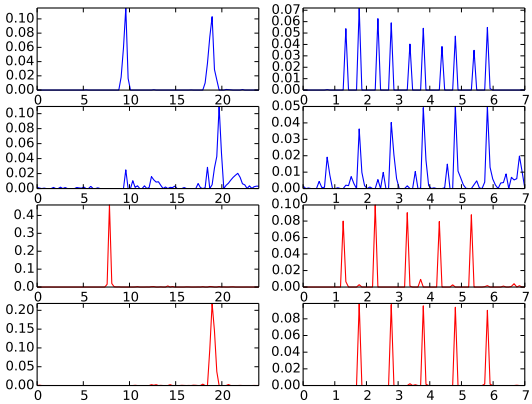

FIGURE 2 – Exemples d'atomes issu du dictionnaire appris par factorisation matricielle non-négative parcimonieuse. Chaque ligne correspond à un atome du dictionnaire et chaque atome est constitué d'une vision jour (à gauche) et d'une vision semaine (à droite). Les atomes du haut (bleus) sont ceux issus du modèle de base. Les atomes du bas (rouge) sont ceux appris avec une contrainte de mono-modalité.

3.3.1 Projection des usagers sur les stations

A l'issue de la factorisation matricielle, les usagers sont projetés sur les atomes du dictionnaire pour chaque échelle fréquentielle : $\mathbf{x}_i^j = \sum_k u_{ik}^j \mathbf{v}_k^j$. Or dans la base de données originale, nous connaissons le lien entre utilisateurs et stations (ainsi que la fréquence d'usage de ladite station) cf figure 1. Nous proposons de projeter simplement les utilisateurs sur le réseau de station : à chaque fois qu'un utilisateur utilise une station, le profil de l'utilisateur (pour l'échelle de fréquence correspondante) est ajouté à la station.

Nous obtenons donc un profil des usages de la station de la forme : $\mathbf{s}_n^j = \sum_k \alpha_{nk}^j \mathbf{v}_k^j$.

3.3.2 Algorithme de clustering multi-instances

Que nous nous intéressions aux stations ou aux utilisateurs, nous manipulons des objets complexes formés d'une série de fonctions temporelles pondérées. Nous proposons de traiter ces objets comme des sacs d'instances, nous utiliserons une métrique adaptée pour comparer différents éléments comme dans [GFKS02] :

$$\langle \mathbf{x}, \mathbf{x}' \rangle = \sum_{j,k,k'} u_k^j u_{k'}'^j \langle \mathbf{v}_k^j \mathbf{v}_{k'}^j \rangle = \sum_j \mathbf{u}^j \Sigma^j \mathbf{u}'^{jT} \qquad (3)$$

en pré-calculant les similarités entre atomes des dictionnaires (par échelle) dans des matrices $\Sigma^j = V^j V^{jT}$.

En se basant sur cette métrique robuste, nous avons développé une variante de l'algorithme des k-moyennes [ZZ09]. Chaque prototype est naturellement lui-même un sac d'instances pondérées correspondant aux éléments qui lui sont affectés.

4 Résultats expérimentaux

5 Conclusion

Références

[Bar07] Richard Baraniuk. Compressive sensing. *IEEE signal processing magazine*, 24(4), 2007.

[BC11] Anil Bawa-Cavia. Statistical analysis of dynamic urban networks. Technical report, UCL, 2011.

[BHG06] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. The scaling laws of human travel. *Nature*, 439(7075) :462–465, 2006.

[BPS02] John A Black, Antonio Paez, and Putu A Suthanaya. Sustainable urban transportation : performance indicators and some analytical approaches. *Journal of urban planning and development*, 128(4) :184–209, 2002.

[CSC12] Irina Ceapa, Chris Smith, and Licia Capra. Avoiding the crowds : understanding tube station congestion patterns from trip data. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, pages 134–141. ACM, 2012.

[FKR+13] Stefan Foell, Gerd Kortuem, Reza Rawassizadeh, Santi Phithakkitnukoon, Marco Veloso, and Carlos Bento. Mining temporal patterns of transport behaviour for predicting future transport usage. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 1239–1248. ACM, 2013.

[GFKS02] Thomas Gärtner, Peter A Flach, Adam Kowalczyk, and Alex J Smola. Multi-instance kernels. In *ICML*, volume 2, pages 179–186, 2002.

[GHB08] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding

individual human mobility patterns. *Nature*, 453(7196) :779–782, 2008.

[Gol02] John C Golias. Analysis of traffic corridor impacts from the introduction of the new athens metro system. *Journal of Transport Geography*, 10(2) :91–97, 2002.

[GVL96] Gene H. Golub and Charles F. Van Loan. *Matrix Computations (3rd Ed.)*. Johns Hopkins University Press, 1996.

[HB11] Chinmay Hegde and Richard G Baraniuk. Sampling and recovery of pulse streams. *Signal Processing, IEEE Transactions on*, 59(4) :1505–1517, 2011.

[Her10] Ryan Jay Herring. *Real-Time Traffic Modeling and Estimation with Streaming Probe Data using Machine Learning*. PhD thesis, University of California, Berkeley, 2010.

[Hoy04] Patrik O Hoyer. Non-negative matrix factorization with sparseness constraints. *The Journal of Machine Learning Research*, 5 :1457–1469, 2004.

[KDMR+03] Kenneth Kreutz-Delgado, Joseph F Murray, Bhaskar D Rao, Kjersti Engan, Te-Won Lee, and Terrence J Sejnowski. Dictionary learning algorithms for sparse representation. *Neural computation*, 15(2) :349–396, 2003.

[LLC+14] Thomas Louail, Maxime Lenormand, Oliva García Cantú, Miguel Picornell, Ricardo Herranz, Enrique Frias-Martinez, José J Ramasco, and Marc Barthelemy. From mobile phone data to the spatial structure of cities. *arXiv preprint arXiv :1401.4540*, 2014.

[LS00] Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2000.

[LXMW12] Thomas Liebig, Zhao Xu, Michael May, and Stefan Wrobel. Pedestrian quantity estimation with trajectory patterns. In *Machine Learning and Knowledge Discovery in Databases*, pages 629–643. Springer, 2012.

[LZC+11] Wei Liu, Yu Zheng, Sanjay Chawla, Jing Yuan, and Xie Xing. Discovering spatio-temporal causal interactions in traffic data streams. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1010–1018. ACM, 2011.

[PJW+12] Chengbin Peng, Xiaogang Jin, Ka-Chun Wong, Meixia Shi, and Pietro Liò. Collective human mobility pattern from taxi trips in urban area. *PLoS ONE*, 7, 04 2012.

[RCOG13] Andry Randriamanamihaga, Etienne Côme, Latifa Oukhellou, and Gérard Govaert. Clustering the vélib origin-destinations flows by means of poisson mixture models. In *22th European Symposium on Artificial Neural Networks (ESANN 2013)*, 2013.

[SQBB10] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968) :1018–1021, 2010.

[Tib96] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[VBB08] Emmanuel Vincent, Nancy Bertin, and Roland Badeau. Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 109–112. IEEE, 2008.

[WP05] Beiming Wang and Mark D Plumbley. Musical audio stream separation by non-negative matrix factorization. In *Proc. DMRN summer conf*, pages 23–24, 2005.

[WPS+11] Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-László Barabási. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1100–1108. ACM, 2011.

[ZZ09] Min-Ling Zhang and Zhi-Hua Zhou. Multi-instance clustering with applications to multi-instance prediction. *Applied Intelligence*, 31(1) :47–68, 2009.