Toward Finding Latent Cities with Non-Negative Matrix Factorization

Eduardo Graells-Garrido
Data Science Institute
Universidad del Desarrollo
Telefonica R&D
Santiago, Chile
egraells@udd.cl

Diego Caro
Data Science Institute
Universidad del Desarrollo
Telefonica R&D
Santiago, Chile
dcaro@udd.cl

Denis Parra
Dept. of Computer Science
Faculty of Engineering
Pontificia Universidad Catolica
Santiago, Chile
dparra@ing.puc.cl

ABSTRACT
In the last decade, digital footprints have been used to cluster population activity into functional areas of cities. However, a key aspect has been overlooked: we experience our cities not only by performing activities at specific destinations, but also by moving from one place to another. In this paper, we propose to analyze and cluster the city based on how people move *through* it. Particularly, we introduce *Mobilicities*, automatically generated travel patterns inferred from mobile phone network data using NMF, a matrix factorization model. We evaluate our method in a large city and we find that *mobilicities* reveal latent but at the same time interpretable mobility structures of the city. Our results provide evidence on how clustering and visualization of aggregated phone logs could be used in planning systems to interactively analyze city structure and population activity.

Author Keywords
Mobile Phone Networks; Urban Informatics; Urban Mobility; Non-Negative Matrix Factorization.

INTRODUCTION
The increasing availability of digital footprints, such as Web/App access logs, user-generated content, and mobile phone network data, has allowed to characterize the city at spatio-temporal granularities never seen before. This means that the different functional areas of the city can be estimated, based not only on how planners thought that the city would be lived, but on how people actually used the different spaces available to them [12]. However, this is not enough to understand the city. As Charles Montgomery says in his book, Happy City: *"When we talk about cities, we usually end up talking about how various places look and perhaps how it feels to be there in those places. But to stop there misses half the story, because they way we experience most parts of cities is at velocity: we glide past on the way to somewhere else. City life is as much about moving* through *landscapes as it is about being* in *them"* [32].

Since people may spend a considerable amount of time while moving through the city, and the quality of that time has a strong influence on mood, health, and productivity [39], it is important to understand city structure with respect to mobility. Given that the growth of cities is faster than the capability of traditional methods to understand the city, it is important to have cost-effective ways to analyze the city at scale [46].

In this paper, inspired by Montogomery's ideas, we estimated a characterization of the city defined by the collective experience of its several areas. Particularly, we analyzed intra-city transportation inferred from mobile phone network records, which we represented in a *Waypoints Matrix*. This matrix, similar to document-term matrices used in Information Retrieval, was decomposed using Non-Negative Matrix Factorization (NMF) [11]. We interpreted and labeled the obtained components, which we denoted *Mobilicities*. We evaluated our pipeline by performing a case study in Santiago, Chile, using mobile phone network data from the biggest telecommunications company in the country. Our pipeline delivered interpretable results, in contrast with a well-established method. We concluded that *mobilicities* can be used within an intelligent user interface aimed at mobility and transportation-analysis tasks.

BACKGROUND AND RELATED WORK
There has been a flurry of research of mobile phone network data known as *eXtended and Call Detail Records* (X/CDR), as evidenced in recent surveys on the area [4, 8]. Some examples

UISTDA '18, March 11, 2018, Tokyo, Japan

include: understanding socio-economic factors on the population [41], understanding family and social relations [13], characterizing response to emergencies and critical events [33], crime detection [5], credit scoring [40], test of urban theories [14]; and the provision of a cost-effective way of understanding population dynamics and behavior in developing countries [21].

The mobile phone network events of a given device depict a spatio-temporal trajectory that can be processed to infer trips, by using geometric approaches based on transportation rules [19], or by clustering events in the trajectory [7]. When individual trips are known, it is possible to aggregate them into Origin-Destination (OD) matrices. This analysis is common in the literature from X/CDR [2, 23, 16, 7], and shows that inferring individual mobility is a relevant problem.

Other important aspects are the characterization of land use (*e.g.*, residential areas, business areas, *etc.*) and functional areas (*i.e.*, delimited areas that serve specific or multiple land uses). Since it is crucial to understand the dynamics of these aspects, functional areas have been measured, monitored and categorized using digital footprints [45, 43] and X/CDR [34, 26, 42, 1, 18]. A similar work to ours has applied NMF to understand trip purpose, and build functional areas based on the spatial distribution of such purposes [36].

The key difference between the aforementioned work and our proposal is the focus. Other work focuses on the destinations of trips, as well as activities performed *within* places. As such, their definition of functional area is limited by those places that, in transportation terms, *attract* people [20]; yet, as mentioned by urbanists, the city is experienced in sequence by moving from one place to another [32]. Each citizen has an unique version of the city, built upon the sequence of nodes, landmarks, and paths traversed [28]. In this paper we show that, by using mobility inferred from X/CDR, and using NMF to decompose/cluster the different cities experienced by mobile phone users, we are able to identify the different *Mobilicities* that comprise a big urban city.Even though we have centered the discussion around mobile phone network data, it is possible to infer transportation and urban patterns from other sources, particularly social media. Twitter has been shown to be a good predictor of commuter flows [31] at several scales [27]. Now, these approaches have the same limitation as previous approaches: a focus on the origins and destinations of trips, mainly due to their way of modelling mobility: using gravity and radiation models (see [29] for a comparison). Twitter data, while massive and longitudinal, does not allow to infer within-trip behavior.

METHODS

Our methods can be summarized in a pipeline of three steps:

1. *Trip detection* from X/CDR data, which, for each device in the dataset, identifies its corresponding daily trips, with origin, intermediate, and destination towers.
2. The construction of a *Waypoint Matrix* W that aggregates the intermediate towers of trips, of a given period of time, into a device-antenna matrix.
3. The decomposition of W using NMF into the product of two matrices, U and T, according to a number of k components, which we denote as *mobilicities*.

Trip Detection. To detect trips from X/CDR traces, we resort to an algorithm based on transportation rules and trajectory simplification [19]. The algorithm builds a space-time trajectory from daily X/CDR events, where space is the cumulative distance between consecutive connected towers, starting from zero at the first connection of the day. This trajectory is simplified using a line-simplification algorithm. Then, each segment from the simplified trajectory is categorized according to transportation rules, such as the relationship between the approximated trip distance and time, which is visually inspected through the slope of the segment. In other words, the trip detection allows us to separate X/CDR events into the following: *stationary* events (the user was performing an activity), *trip start* events (denoting the origin), *trip end* events (denoting the destination), and *within-trip* events (denoting mobility).

Building the Waypoints Matrix. Ideally, characterization of within-trips events does not need to be done using aggregation. For instance, GPS data allows to do rich clustering over specific trajectories [47]. However, due to the billing purpose of X/CDR data, it is possible that trips have few within-trip events because of the billing cycle. Since we will focus on within-trip events, we need to aggregate such events from an extended period of time. Furthermore, some trips do not have within-trip events, such as those with duration near to the billing cycle time, and those within zones with low tower density. Hence, by aggregating all within-trip events for a user in a period of time, the likelihood of identifying the towers that characterize a specific user's mobility increases. We use this schema to define a *Waypoints Matrix* W, defined as:

$$w_{i,j} = \frac{\#\text{ of } \textit{within-trip} \text{ events of user } u_i \text{ at tower } t_j}{\#\text{ of } \textit{within-trip} \text{ events of user } u_i}$$

This schema is equivalent to the L1-normalized document-term matrices found in Information Retrieval, but without weighting with Tf-Idf [44]. We do not apply Tf-Idf because its purpose is to identify discriminative features; conversely, we want to extract collective features. Additionally, note that this matrix is different to those used in related work with NMF decompositions [36]: there is a semantic difference between within-trips and trip start/end events. To avoid this polysemic behavior, we focus only on within-trips events.

Applying Non-Negative Matrix Factorization. To represent how users interact with towers, we propose to decompose this matrix into two: $W = U \times T$, where U is a $|u| \times k$ matrix that encodes k user latent features for $|u|$ users, and T is a $k \times |t|$ matrix that encode k latent tower features for $|t|$ different cell phone towers. Note that, by definition, all $w_{i,j} \geq 0$. NMF allows us to decompose the matrix W into two non-negative matrices, which gives a lower rank approximation for W, such that $W \approx U \times T$ [24]. This can be formalized as the following optimization problem: $\min_{U,T} ||W - U \times T||_F$ subject to U and T be non-negative, where number of rows in U and the

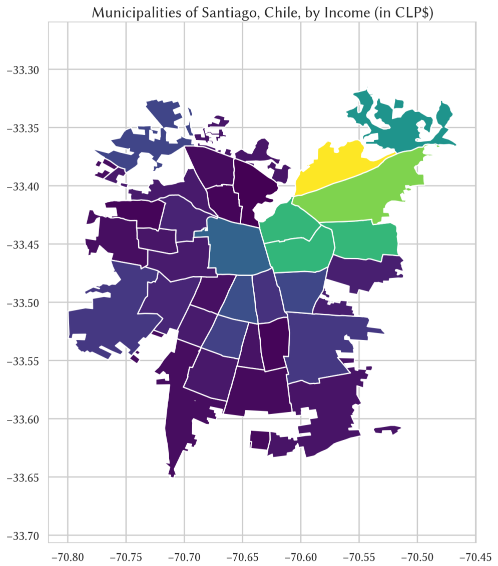

Figure 1: Choropleth map of the urban area of Santiago, Chile. Each municipality is colored according to their average income (in CLP$).

number of columns in T correspond to the desired lower-rank approximation k.

Even though there is a variety of methods to decompose matrices, we choose NMF, which has been applied in similar contexts with interpretable results [36]. Then, we define a mobilicity m_c as the weighted set of towers within the c component of the decomposition, *i.e.*, the c-th column of matrix T. The parameter k must be chosen manually, and its value should be ideally decided jointly between data scientists and domain experts according to the context. Previous strategies for choosing k have focused in measuring the stability of the components [6] and in the variation of the residual sum of squares curve between the original matrix and its decomposition [17, 22]. However, we prefer to manually choose the number of components as these methods do not allow us to incorporate external information such as the socio-economic distribution of the city.

CASE STUDY: SANTIAGO, CHILE

We performed a case study on Santiago, the capital of Chile, with almost 8 million inhabitants. Its urban area covers a surface of 867.75 square kilometers, and is composed of 35 independent administrative units called municipalities (*c.f.* Fig. 1). Because this city has experienced accelerated growth, and it is expected to keep growing at least until 2045 [38], understanding its structure at scale is an important and timely task.

Datasets

Mobile Phone Network Data. We studied an anonymized X/CDR dataset from Telefónica Chile, the biggest telco. in Chile, with a market share of 33% in 2016. The dataset contained records between July 27th and August 10th from 2016. In total, we analyzed 124,414 users, who had enough connections to the cell towers under analysis to estimate their daily

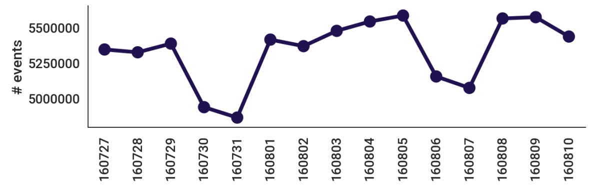

Figure 2: Number of events in the dataset per day. The effect of weekends in the number of events is easily identifiable.

Metric	
# of Trips	4,213,400
# of Users	124,415
# of Users with Within-Trip events	95,027
Mean trips per user	33.87
Std. Dev.	19.62
Min	1
Percentile 25%	18
Percentile 50%	34
Percentile 75%	48
Max	140

Table 1: Statistics with respect to the number of inferred trips.

trips between 6AM and midnight, and had either a pre-paid or contract subscription. They generated an average of 5.33 million billing records per day (*c.f.* Fig. 2). The average inter-event times for users range within 14.71 and 30.96 minutes, which shows a billing cycle between fifteen and thirty minutes.

Telefónica has 1,464 cell phone towers in the municipalities under consideration. We discarded towers that were installed in in-door contexts (*e.g.*, malls, hospitals, *etc.*). This is possible because tower meta-data includes their geographical position and their name. For out-door towers, the name usually contains the nearest crossing, while in-door towers contain the name of the place they lie in. In total, there were $|t| = 1{,}082$ out-door towers (see Fig. 3, Towers). The only in-door towers that we kept were those installed within underground metro stations.

OpenStreetMap. OSM (http://openstreetmap.org) is a crowd-sourced maps platform. We downloaded a dump of its data for Chile, and then identified the highways within Santiago. We used this information to contextualize the different *mobilicities* identified by the NMF. We did so by finding the out-door towers that lie within 250 meters of each highway, as shown in Fig. 3 (Labeled Towers).

Trip Detection and Waypoints Matrix

Using the trip inference algorithm we detected 4,213,400 trips for 124,415 users (*c.f.* Table 1 for descriptive statistics). Fig. 4 shows the departure time distribution of all trips. One can see that business days exhibit expected peak-times related to work hours, and that weekends exhibit different patterns, such as a higher density at lunch time.

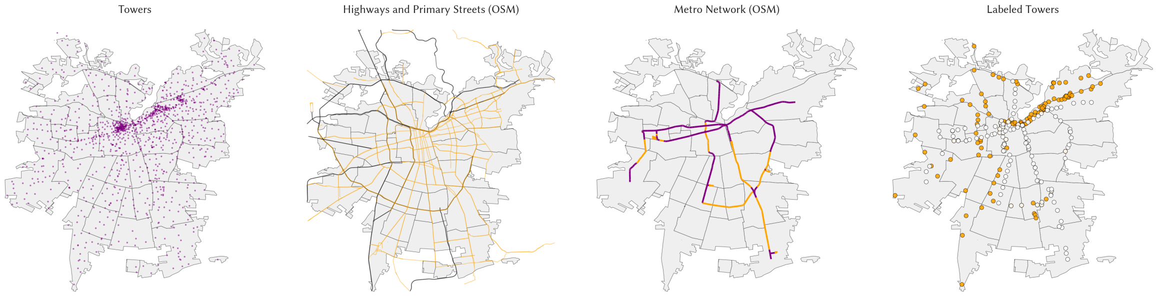

Figure 3: Maps of Santiago: cell phone tower network, highway and primary streets, the metro network, and the set of labeled towers according to their distance to highways or metro lines.

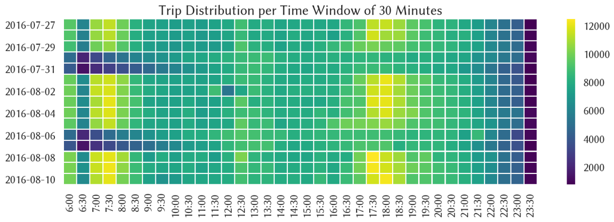

Figure 4: Trip departure time distribution per day.

After detecting trips, we built the W matrix, of dimensions $|u| \times |t|$. Note that $|u| = 95{,}027$, because not all users had within-trip events.

Factorization of the Waypoints Matrix

To perform the NMF decomposition, we chose $k = 8$, because the city is usually divided into six big areas (north, south, east, west, south-east, center) and, since we expect that the results exhibit relationship with modes of transportation, we wanted to see the effect of private and public transportation. Thus, $k = 8$ is an arguably reasonable choice (note that we discuss the choice of k at the next section).

Tower-Component Matrix. Figure 5 shows the results of the factorization, with one map for each component-tower column from the matrix. One can see that there is a strong geographical clustering of towers, which may be explained as W is essentially a co-occurrence matrix.

Fig. 6 show how the sets of labeled towers (those near highways, near surface metro and within underground metro stations) relate to each component. This allows to see that some components tend to be more associated than others to some modes of transportation: C1, C2, C3 and C7 are more associated to metro than highways, while C4 exhibits the opposite behavior. Having both figures into account, the following is an interpretation of each *mobilicity*:

- C0 : the east side of the city, including part of the center, next to the yellow metro line and one important highway of the city. This area is characterized for its business districts and high income residential areas (*c.f.* Fig. 3). As such, it is likely that its residents do not use public transportation, nor visit other *mobilicities*. Note how metro towers have lower association with this component in Fig. 6.
- C1 : people that live in the southern part of the city, mostly between two metro lines. Since this area is characterized by low income, this means that they need to take a bus to reach the metro.
- C2 : the southeast area of the city, which is characterized by their dependency of two metro lines. This component contains mixed-income municipalities.
- C3 : In contrast with the previous components, this one is completely focused on public transportation: it fully contains two metro lines in full, and partially other two. It also contains bus corridors that tend to connect to metro lines.
- C4 : the northern part of the city, which is mostly residential and of low income. The component also has a main street of the city as a kind of tentacle, showing that people who lives/work in this area, but who work/lives in another, uses this street as a way to get into the component.
- C5 : the south-west part of the city, which is connected to downtown primarily through a highway and a metro line that is parallel to the highway.
- C6 : the western area of the city. This area contains one of the most populated municipalities in the city.
- C7 : similar to C6, but extending its reach to center areas of the city through a metro line a bus corridors. This makes this component dependent on public transport, and thus, the routes followed by its inhabitants tend to cluster, in contrast to what happens in C6.

In summary, latent cities seem to be comprised by three kinds of clusters of towers: those where people lives and moves, enclosed by specific limits (C0, C1, C6), those where people lives and work, but in different areas of the city, connected through the transportation network (C2, C4, C5, C7), and transportation infrastructure (C3).

User-Component Matrix. The user-component matrix may suggest that users pass through different *mobilicities* in their daily lives. Fig. 7 explores this potential behavior, by displaying how a sample of 25,000 users cluster around the corresponding components. One can see that, indeed, users tend to have a primary component, but they still belong to others. This would be the equivalent to, for instance, living in the suburbs, and having to travel long distances to go to work. Note that many users from C0, the wealthiest part of the city, have negligible association to other components – something

Figure 5: The eight *mobilicities* of the tower-component matrix obtained by performing NMF.

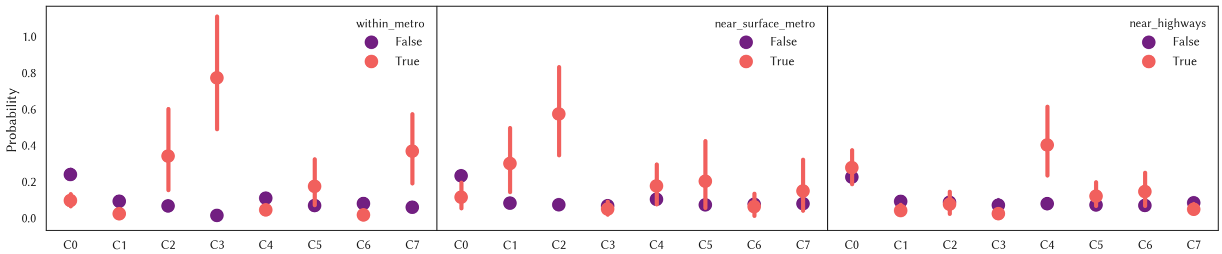

Figure 6: Point-plot of the average association of the labeled sets of towers into the different *mobilicities*.

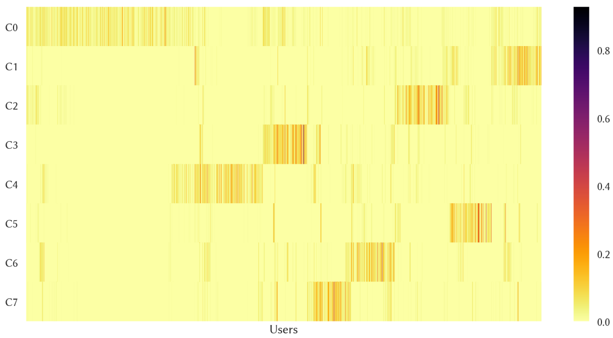

Figure 7: Heatmap of a random sample of 25K users and their corresponding component associations.

expected due to the economical segregation of the city (*c.f.* Fig. 3).

Understanding k. We showed the eight *mobilicities* (*c.f.* Fig. 5) to domain experts, who gave informal feedback – it made sense to split the city in this way, as we have interpreted earlier. Even though a formal evaluation with domain experts is left for future work, here we discuss the patterns that emerge when varying the parameter k which is the rank of the factorized matrix. We do so by estimating mobilicities with $k' = 4$ (*c.f.* Fig. 8) and $k'' = 12$ (*c.f.* Fig. 9).

With four mobilicities, the clustering is mostly geographic: three components split the city. However, the fourth component is related to public transportation: it reconstructs several metro lines and bus corridors. In this aspect, it seems that using $k' = 4$ allows to obtain a similar result to $k = 8$. Then, if one would like to differentiate cell towers with respect to general transportation patterns, this could be a reasonable choice.

With twelve mobilicities, the geographical clustering is still present, but the routes that connect distinct parts of the city become more evident – meaning that a mobilicity is comprised by one or two sectors of close towers (for instance, home and work locations), plus “bridges” that connect one mobilicity to another, based on the common routes followed by people. This behavior is expected, due to the co-occurrence property of the Waypoints Matrix.

In summary, several values of k allow to infer soft-partitions of the city, as well as the way its inhabitants move between those partitions. A mobilicity may be a soft-partition, a soft-partition with bridges to other mobilicities, or a network of those bridges – namely, a transportation network.

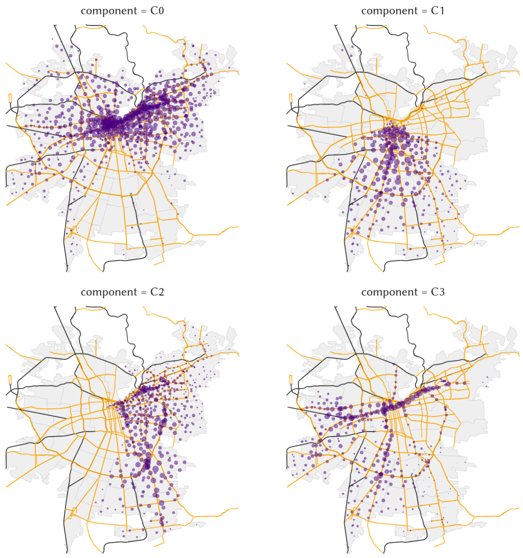

Figure 8: Mobilicities obtained with k =4.

Comparing Interpretability with PCA/Truncated SVD. To discuss further whether NMF is a good choice of model in terms of interpretability, we estimated a Truncated SVD decomposition (equivalent to PCA) with $k = 8$. Fig. 10 shows that, in contrast to NMF, there is no geographical clustering nor correspondence to any infrastructure available in the city. Thus, even though PCA is a widely used dimensionality reduction technique, it does not allow the interpretation nor clustering of the city in the same way as NMF does.

CONCLUSIONS

We proposed the concept of *mobilicities*, which denotes the different cities experienced by the inhabitants of a big city, and depict its dynamics with respect to mobility and usage of modes of transportation. The suitability of NMF to this kind of spatial data could be related to the fact that NMF is equivalent to spectral clustering [15], which has performed well when grouping trip destination data [12]. However, as we have noted in our motivation, our input is not destination nor origin data; instead, it is *spatial location while moving*. This focus was inspired by the book Happy City [32], reflecting that our purpose was to help domain experts and policy designers to make better, happier cities. Such purpose implies collaboration between the emerging field of data science and the corresponding disciplines – transportation and urban planning. However, evidence-based policy in those areas requires transparency and interpretability, and many state of the art machine learning techniques do not offer both qualities [9]. In this paper, we have shown that NMF does offer both qualities when applied to mobility data, and thus, is a promising technique to apply in the field of Urban Computing [46].

Limitations and Future Work. Critics may rightly say that we need a well-defined criteria to choose k. Future work should tackle this limitation using intelligent user interfaces aimed at domain experts. This opens two lines of research within the IUI: on the one hand, we could try other factorization methods for positive-only data such as SLIM, which has shown promising results in the past [25], and would allow to understand how the choice of k influences the output and its interpretability. On the other hand, visualization and exploratory interfaces are tools valued by domain experts [10], and mobility has been a recurring topic in visual analytics [3]. Finally, our work did not consider the temporal aspects of transportation. Hence, future work should consider how to incorporate that dimension into the definition of Mobilicities.

Acknowledgements. The analysis was performed using Jupyter Notebooks [37], jointly with the *scikit-learn* [35], *pandas* [30], and *geopandas* libraries. The maps on this paper include data from ©OpenStreetMap contributors. We also thank Telefónica R&D in Santiago for facilitating the data for this study, in particular Pablo García Briosso. The author Denis Parra has been funded by Conicyt, Fondecyt grant 11150783, as well as Fondef grant id16i10222 and the BRT+ Centre of Excellence funded by VREF. Finally, we thank the anonymous reviewers for the insightful comments that helped to improve this paper.

REFERENCES

1. Rein Ahas, Anto Aasa, Siiri Silm, and Margus Tiru. Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: case study with mobile positioning data. *Transportation Research Part C: Emerging Technologies*, 18(1):45–54, 2010.
2. Lauren Alexander, Shan Jiang, Mikel Murga, and Marta C González. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 2015.
3. Natalia Andrienko and Gennady Andrienko. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3–24, 2013.
4. Vincent D Blondel, Adeline Decuyper, and Gautier Krings. A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(1):1, 2015.
5. Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Emmanuel Letouzé, Nuria Oliver, Fabio Pianesi, and Alex Pentland. Moves on the street: Classifying crime hotspots using aggregated anonymized data on people dynamics. *Big Data*, 3(3):148–158, 2015.
6. Jean-Philippe Brunet, Pablo Tamayo, Todd R. Golub, and Jill P. Mesirov. Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences*, 101(12):4164–4169, 2004.
7. Francesco Calabrese, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Computing*, 10(4):0036–44, 2011.
8. Francesco Calabrese, Laura Ferrari, and Vincent D Blondel. Urban sensing using mobile phone network data: a survey of research. *ACM Computing Surveys (CSUR)*, 47(2):25, 2015.
9. Davide Castelvecchi. Can we open the black box of AI? *Nature News*, 538(7623):20, 2016.
10. Tao Cheng, Garavig Tanaksaranond, Chris Brunsdon, and James Haworth. Exploratory visualisation of congestion evolutions on urban transport networks. *Transportation Research Part C: Emerging Technologies*, 36:296–306, 2013.
11. Andrzej Cichocki and Anh-Huy Phan. Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 92(3):708–721, 2009.
12. Justin Cranshaw, Raz Schwartz, Jason Hong, and Norman Sadeh. The Livehoods Project: Utilizing social media to understand the dynamics of a city. In *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.

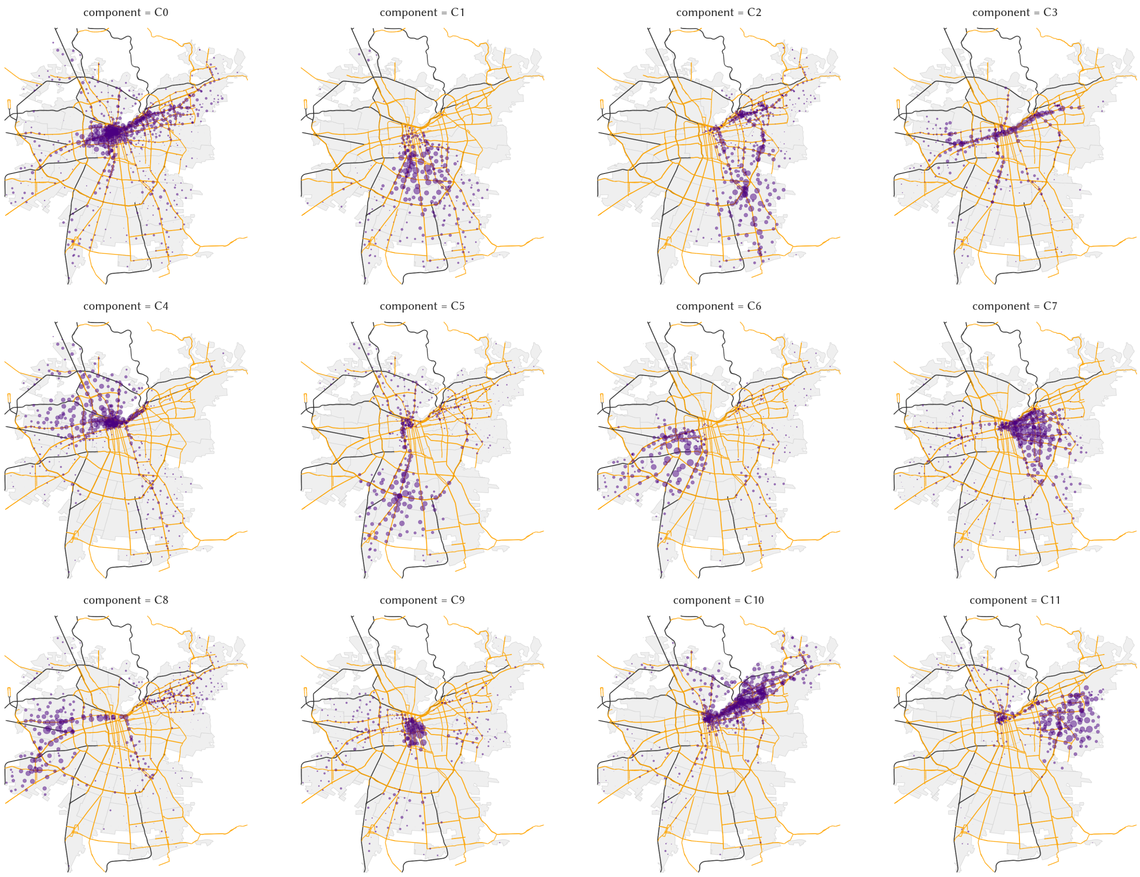

Figure 9: Mobilicities obtained with k =12.

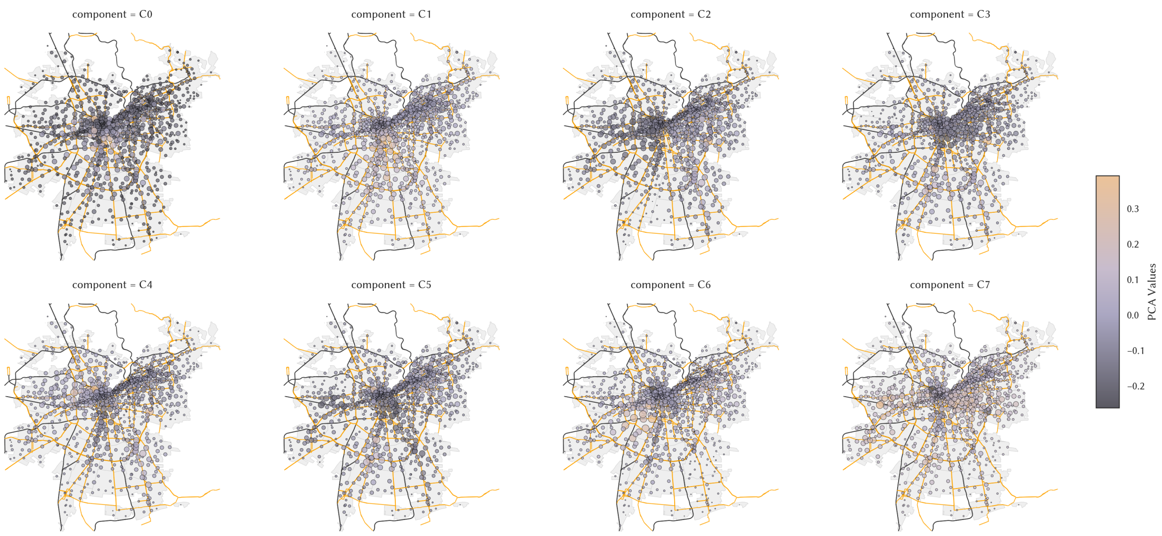

Figure 10: Tower-component results of the application of Truncated SVD/PCA to the Waypoints Matrix, with k = 8.

13. Tamas David-Barrett, Janos Kertesz, Anna Rotkirch, Asim Ghosh, Kunal Bhattacharya, Daniel Monsivais, and Kimmo Kaski. Communication with family and friends across the life course. *PloS one*, 11(11):e0165687, 2016.

14. Marco De Nadai, Jacopo Staiano, Roberto Larcher, Nicu Sebe, Daniele Quercia, and Bruno Lepri. The death and life of great Italian cities: a mobile phone data perspective. In *Proceedings of the 25th International Conference on World Wide Web*, pages 413–423. International World Wide Web Conferences Steering Committee, 2016.

15. Chris Ding, Xiaofeng He, and Horst D Simon. On the equivalence of nonnegative matrix factorization and spectral clustering. In *Proceedings of the 2005 SIAM International Conference on Data Mining*, pages 606–610. SIAM, 2005.

16. Vanessa Frias-Martinez, Cristina Soguero, and Enrique Frias-Martinez. Estimation of urban commuting patterns using cellphone network data. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, pages 9–16. ACM, 2012.

17. Attila Frigyesi and Mattias Höglund. Non-negative matrix factorization for the analysis of complex gene expression data: identification of clinically relevant tumor subtypes. *Cancer informatics*, 6(2003):275–92, jan 2008.

18. Eduardo Graells-Garrido, Oscar Peredo, and José García. Sensing urban patterns with antenna mappings: the case of Santiago, Chile. *Sensors*, 16(7):1098, 2016.

19. Eduardo Graells-Garrido and Diego Saez-Trumper. A day of your days: estimating individual daily journeys using mobile data to understand urban flow. In *Proceedings of the Second International Conference on IoT in Urban Space*, pages 1–7. ACM, 2016.

20. Randolph Hall. *Handbook of transportation science*, volume 23. Springer Science & Business Media, 2012.

21. Martin Hilbert. Big data for development: A review of promises and challenges. *Development Policy Review*, 34(1):135–174, 2016.

22. Lucie N. Hutchins, Sean M. Murphy, Priyam Singh, and Joel H. Graber. Position-dependent motif characterization using non-negative matrix factorization. *Bioinformatics*, 24(23):2684–2690, 2008.

23. Md Shahadat Iqbal, Charisma F Choudhury, Pu Wang, and Marta C González. Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40:63–74, 2014.

24. Da Kuang, Chris Ding, and Haesun Park. *Symmetric Nonnegative Matrix Factorization for Graph Clustering*, pages 106–117. 2012.

25. Santiago Larraín, Denis Parra, and Alvaro Soto. Towards improving top-N recommendation by generalization of SLIM. In *RecSys Posters*, 2015.

26. Maxime Lenormand, Miguel Picornell, Oliva G Cantú-Ros, Thomas Louail, Ricardo Herranz, Marc Barthelemy, Enrique Frías-Martínez, Maxi San Miguel, and José J Ramasco. Comparing and modelling land use organization in cities. *Royal Society Open Science*, 2(12):150449, 2015.

27. Jiajun Liu, Kun Zhao, Saeed Khan, Mark Cameron, and Raja Jurdak. Multi-scale population and mobility estimation with geo-tagged tweets. In *Data Engineering Workshops (ICDEW), 2015 31st IEEE International Conference on*, pages 83–86. IEEE, 2015.

28. Kevin Lynch. *The image of the city*, volume 11. MIT press, 1960.

29. A Paolo Masucci, Joan Serras, Anders Johansson, and Michael Batty. Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Physical Review E*, 88(2):022812, 2013.

30. Wes McKinney. Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56, 2010.

31. Graham McNeill, Jonathan Bright, and Scott A Hale. Estimating local commuting patterns from geolocated Twitter data. *EPJ Data Science*, 6(1):24, 2017.

32. Charles Montgomery. *Happy city: transforming our lives through urban design*. Macmillan, 2013.

33. Benyounes Moumni, Vanessa Frias-Martinez, and Enrique Frias-Martinez. Characterizing social response to urban earthquakes using cell-phone network data: the 2012 Oaxaca earthquake. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 1199–1208. ACM, 2013.

34. Anastasios Noulas, Cecilia Mascolo, and Enrique Frias-Martinez. Exploiting Foursquare and cellular data to infer user activity in urban environments. In *2013 IEEE 14th International Conference on Mobile Data Management*, pages 167–176. IEEE, 2013.

35. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

36. 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(4):e34487, 2012.

37. Fernando Pérez and Brian E Granger. IPython: a system for interactive scientific computing. *Computing in Science & Engineering*, 9(3):21–29, 2007.

38. Olga Lucia Puertas, Cristian Henríquez, and Francisco Javier Meza. Assessing spatial dynamics of urban growth using an integrated land use model. application in Santiago metropolitan area, 2010–2045. *Land Use Policy*, 38:415–425, 2014.

39. Heiko Rüger, Simon Pfaff, Heide Weishaar, and Brenton M Wiernik. Does perceived stress mediate the relationship between commuting and health-related quality of life? *Transportation research part F: traffic psychology and behaviour*, 50:100–108, 2017.

40. Jose San Pedro, Davide Proserpio, and Nuria Oliver. Mobiscore: towards universal credit scoring from mobile phone data. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 195–207. Springer, 2015.

41. Victor Soto, Vanessa Frias-Martinez, Jesus Virseda, and Enrique Frias-Martinez. Prediction of socioeconomic levels using cell phone records. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 377–388. Springer, 2011.

42. Jameson L Toole, Michael Ulm, Marta C González, and Dietmar Bauer. Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD international workshop on urban computing*, pages 1–8. ACM, 2012.

43. Carmen Karina Vaca, Daniele Quercia, Francesco Bonchi, and Piero Fraternali. Taxonomy-based discovery and annotation of functional areas in the city. In *Ninth International AAAI Conference on Web and Social Media*, 2015.

44. R Baeza Yates and B Ribeiro Neto. Modern information retrieval: the concepts and technology behind search. *Addison-Wesley Professional*, 2011.

45. Jing Yuan, Yu Zheng, and Xing Xie. Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 186–194. ACM, 2012.

46. Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):38, 2014.

47. Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th international conference on World wide web*, pages 791–800. ACM, 2009.