



HAL
open science

Uncovering the socioeconomic structure of spatial and social interactions in cities

Maxime Lenormand, Horacio Samaniego

► **To cite this version:**

Maxime Lenormand, Horacio Samaniego. Uncovering the socioeconomic structure of spatial and social interactions in cities. *Urban Science*, 2023, 7 (1), pp.15. 10.3390/urbansci7010015 . hal-03964682

HAL Id: hal-03964682

<https://hal.inrae.fr/hal-03964682v1>

Submitted on 3 Jul 2024

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Article

Uncovering the Socioeconomic Structure of Spatial and Social Interactions in Cities

Maxime Lenormand ^{1,*}  and Horacio Samaniego ^{2,3,†} ¹ TETIS, University of Montpellier, AgroParisTech, Cirad, CNRS, INRAE, 34000 Montpellier, France² Laboratorio de Ecoinformática, Instituto de Conservación, Biodiversidad y Territorio, Universidad Austral de Chile, Campus Isla Teja s/n, Valdivia 5110290, Chile³ Instituto de Sistemas Complejos de Valparaíso, Valparaíso 7800003, Chile

* Correspondence: maxime.lenormand@inrae.fr

† These authors contributed equally to this work.

Abstract: The relationship between urban mobility, social networks, and socioeconomic status is complex and difficult to apprehend, notably due to the lack of data. Here we use mobile phone data to analyze the socioeconomic structure of spatial and social interaction in the Chilean urban system. Based on the concept of spatial and social events, we develop a methodology to assess the level of spatial and social interactions between locations according to their socioeconomic status. We demonstrate that people with the same socioeconomic status preferentially interact with locations and people with a similar socioeconomic status. We also show that this proximity varies similarly for both spatial and social interactions during the course of the week. Finally, we highlight that these preferential interactions appear to hold when considering city–city interactions.

Keywords: human mobility; socio-spatial networks; urban system; mobile phone data; urban computing; socio-informatics



Citation: Lenormand, M.; Samaniego, H. Uncovering the Socioeconomic Structure of Spatial and Social Interactions in Cities. *Urban Sci.* **2023**, *7*, 15. <https://doi.org/10.3390/urbansci7010015>

Academic Editor: Luis Hernández-Callejo

Received: 6 December 2022

Revised: 17 January 2023

Accepted: 26 January 2023

Published: 30 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Securing equal opportunities to access public infrastructure is a major challenge in urban planning [1], more so given the large concentration of wealth observed among the increasingly urban economies worldwide [2]. While these issues have largely been discussed in transportation [3], sociology [4], and physics [5] among other disciplines, the current deluge of spatially contextual information regarding the mobility and social interaction among humans is offering precise quantitative descriptions of emerging patterns of spatial socio-economic mixing across cities [5–9].

The analysis of trace information generated by mobile phones, credit cards, and transit cards, among others, has been shown to provide simple, synoptic, and near real-time descriptions of urban mobility that have expanded our understanding of mobility strategies towards fine-grained and contextual representations of how travel budgets are segmented across the different dimensions of human life [10–14]. Its adoption for urban planning and policy crafting, however, is still lagging, mostly due to the highly interdisciplinary endeavor involved in understanding the role, and impact, of mobility across the social, technological, and ecological fabric of urban life [15–17].

So far, different conclusions have emerged when describing the spatial context of social interactions, and, while important strides have been made, explaining how urban demographics and socioeconomic indicators relate to mobility still remains a challenge. Early work explicitly shows that existing correlations between mobile phone usage and wealth may be a starting point towards using information and communication technology (ICT) data for planning, where sensitive data are available for research [18,19]. When the spatial context is explicitly considered, mobility research, produced from different disciplines, seems to indicate that diversity of human trajectories across the city is a conserved

trait among social groups sharing similar status (social, economic, etc.) [5,8,20–23], albeit important differences exist across gender [20,24,25], income [21,26], residential location [16], and other aspects of human life [27].

What is now accepted, despite early predictions of a decline in the importance of space with the emergence of information and communications technologies in the sixties [28,29], is that ‘real’ social interactions connecting and exchanging wisdom, goods, and affection are highly relevant to explain the hierarchical patterns of mobility [9]. In fact, recent studies have shown the high predictive power of social ties to describe activities, interests, and locations in ego networks [30,31]. Hence, the notion that functional relationships between social networks and space are (strongly) mediated by the spatial opportunities available for human interaction seem to prevail across the literature [32,33]. We also know, that while social interactions are deeply associated with mobility, they only represent a limited fraction of movements across the city [34], hinting towards the existence of other components associated with mobility and social mixing. It is also becoming clear that multivariate analyses of the factors linked to travel schedules, while important, often provide only localized descriptions hampering generalizations of the phenomena compared to ICT traces that explicitly measure how individuals use urban spaces during their daily journey [7,35]. This has made ICT tracers great candidates to deepen our understanding of the dynamics of social mixing and the spatial environment in which they are embedded [36,37].

We here study the socioeconomic structure of spatial and social interactions using mobile phone records of a major provider in Chile. We begin by extracting the spatial and social networks of interactions. We then introduce an indicator, akin to an urban pulse [38], to assess the weekly mobility patterns of every urban location in Chile. We use this indicator to cluster the locations showing similar weekly mobility patterns. We obtained four spatial clusters strongly correlated with the socioeconomic status of its residents, which finally allow us to build and analyze coarse-grained spatial and social interaction matrices showing the emergence of a preferential association in terms of spatial and social interactions between people sharing similar socioeconomic status.

2. Materials and Methods

2.1. From Data to Networks

Our datasets are composed of Data Detail Records (XDR) and Call Detail Records (CDR) provided by Telefónica Chile, representing 37% share of the mobile phone market in Chile. The XDR dataset consists of billions of cellphone pings made by 4 million mobile phones during 3 weeks in March, May, and October 2015 in Chile. Each ping is characterized by its location (i.e., cellphone tower) and a timestamp. Each week has been divided into $T = 168$ h. We partitioned the country into $L = 3876$ locations following a Voronoi tessellation based on the cell phone towers’ position. Data processing started by identifying the mobile phone users’ home location for each week of observation [13]. We finally selected 2.5 million of reliable users with a *validated* home location for at least one of the three weeks. We removed users whose home locations were not possible to identify. We were thus able to identify 360 million of spatial events defined as the presence of a reliable user in a location at time $t \in [1, 168]$. This collection of spatial events has enabled us to build 168 spatial networks, one for each hour t . These networks are weighted and directed. A weight G_{ij}^t of a link between two locations was given by the number of users living in location i and that were present in location j at time t (i.e., all weeks combined). Similarly, we used the CDR dataset to identify 12.5 million of social events between reliable users. We defined a social event as a directed interaction (through a phone call) between two reliable users. In this case, we defined 168 social networks. The weight S_{ij}^t of a link corresponds to the number of social interactions made by users living in location i with users living in location j at time t for all weeks combined. More details regarding the data cleaning process are available in the section *Data preprocessing* (Table A1, Figures A1–A4) in Appendix A.

2.2. Pulse of a Location

We characterize the weekly mobility pattern of a location with a spatio-temporal indicator that we called the 'Pulse of a location'. We define such pulse P_i at location i as the time-evolution of the average distance between the location i and the position of its residents during a typical week. More specifically, the pulse P_i^t of location i at time $t \in [1, 168]$ corresponds to the average distance between the location i and the position of its residents at time t (Equation (1)).

$$P_i^t = \frac{1}{A_i} \frac{1}{G_i^t} \sum_{j=1}^L G_{ij}^t d_{ij} \quad (1)$$

where L is the total number of locations, d_{ij} the great circle distance between locations i and j , and $G_i^t = \sum_{j=1}^L G_{ij}^t$. The constant A_i is used as a normalization factor to ensure that $\sum_{t=1}^T P_i^t = 1$. Note that a large heterogeneity of location areas exists given the irregular location of antennas across the study area. This prompted us to only consider pulses associated with the 2294 locations having a surface area lower than 10 km² in order to compute pulses representative of the spatio-temporal status of the population.

2.3. Cluster Analysis

We rely on the ascending hierarchical clustering (AHC) method to identify different profiles of the pulse across locations. Ward's metric and Euclidean distances were taken as agglomeration method and dissimilarity metric, respectively [39]. The number of clusters was chosen by comparing the ratio between the within-group variance and the total variance. The purpose of this cluster analysis is to identify meaningful profiles of pulse that can be used as a proxy for the socio-economic structure of a location. Indeed, we make the assumption that differences in mobility behaviors and particularly between weekdays and weekends represent an important descriptor of the socioeconomic status of the location.

2.4. Measuring Spatial and Social Interactions

We construct two coarse-grained spatial and social interaction matrices λ and γ based on the aggregation of link weights, G_{ij}^t and S_{ij}^t , in space and time. More specifically, the fraction of spatial interaction from a cluster c to a cluster c' during a given time window Δ_t is defined as follows,

$$\lambda_{cc'} = \frac{1}{B_c} \sum_{i \in c} \sum_{\substack{j \in c' \\ j \neq i}} \sum_{t \in \Delta_t} G_{ij}^t \quad (2)$$

where Δ_t is the set of hours contained in the time window. The constant B_c is used as the normalization factor to ensure that the sum of interactions from a cluster c to the N clusters is equal to one, $\sum_{c'=1}^N \lambda_{c,c'} = 1$. The same formula is used to compute the social interactions $\gamma_{cc'}$ between and within clusters based on S_{ij}^t instead of G_{ij}^t .

To rigorously quantify the structure of these interactions, we use the index Φ proposed in [40] to measure the hotspots' hierarchical structure of cities. In our case, this index allows us to quantify the importance of interactions between close clusters (i.e., $|c - c'| \leq 1$) among all interactions as the index relies on the tridiagonal trace of the matrix λ (Equation (3)), where $\delta_{cc'}$ is the Kronecker delta). Such an approach provides a succinct representation of the preferential relationships between locations across the study area. The same procedure is used to compute the index, now associated with the social interactions, by using γ instead of λ in the formula.

$$\Phi = \frac{\sum_{c,c'=1}^N \lambda_{cc'} (\delta_{cc'} + \delta_{c(c'-1)} + \delta_{(c-1)c'})}{\sum_{c,c'=1}^N \lambda_{cc'}} \quad (3)$$

The values of Φ range from 0 to 1. A value of 1 means that all the elements of the matrix that are not on the tridiagonal are equal to 0. In other words, all the interactions occur within the same cluster or with the closest cluster. A value of 0 means that the tridiagonal trace of the matrix is null, implying an absence of interactions within the same cluster or with the closest cluster. However, this specific case is clearly unrealistic, so to rescale the value of Φ in a relevant order of magnitude we proposed the following min-max normalization to obtain the metric $\bar{\Phi}$ (Equation (4)).

$$\bar{\Phi} = \frac{\Phi_h - \Phi}{\Phi_h - 1} \quad (4)$$

where Φ_h is the index obtained with a null model based on Equation (2), in which a cluster is randomly assigned to every location (preserving the total number of locations per cluster). The value of Φ_h is then averaged over 100 random reassignments. $\bar{\Phi}$ varies from 0, when the proximity between clusters is equivalent to the one obtained with the null model, to 1, when only interactions between nearby clusters occur. More details regarding the impact of the number of random reassignments used to compute Φ_h on $\bar{\Phi}$ are available in Figure A10 in the section *Null model* in Appendix D.

3. Results

3.1. Pulse of a Location and Socioeconomic Structure

Based on the ratio between the within-group variance and the total variance (Figure A5 in Appendix C), 18 clusters were found. As can be seen in Figure A6 in Appendix C, 92 percent of the location is covered by four main clusters. The four average pulses associated with these clusters are displayed in Figure A7 in Appendix C. The rest of the locations are gathered into 3 small clusters (Figure A8 in Appendix C) and 11 outliers (category *Others* in Figure A6) that we decided to discard because they contain too few locations (or even one location for the outliers) to allow for a rigorous analysis.

Thus, we obtain four main pulse profiles, gathering 92% of the locations. Figure 1 shows a profile of the average pulse activity for each of these four clusters. Not surprisingly, each profile exhibits a typical day–night temporal activity pattern where individuals are moving, on average, further away from their residence during the day compared to night hours. Some differences can nevertheless be observed between the different days of the week. The average distance from home tends to increase from Wednesday to Saturday and then decrease from Sunday to Tuesday. The difference between day and night is also more pronounced on weekdays compared to weekends. The main difference between profiles is mostly based on the difference in mobility behaviors between weekdays and weekends. This difference is very pronounced for the locations belonging to cluster 1 (representing 25% of the locations). Indeed, people living in locations belonging to cluster 1 tend to roam farther away from home during weekends compared to weekdays. This difference slightly decreases for the 26% and 34% locations belonging to cluster 2 and 3, respectively. The opposite behavior is observed for people living in cluster 4 (7% of locations) that tend to be more or less at the same distance from their home irrespective of the day of the week. This pattern is congruent with descriptions of individual mobility journeys in which working-class groups tend to exhibit longer journeys to work compared with more affluent sectors in Chile [24,41].

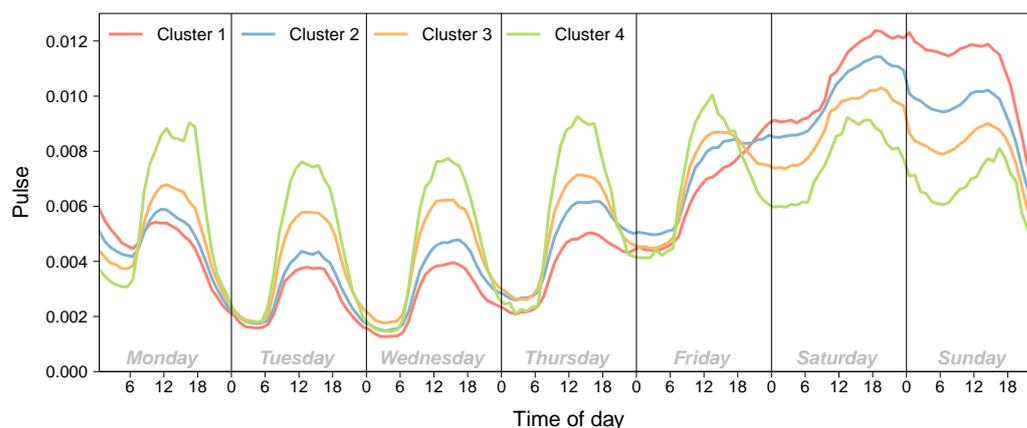


Figure 1. Average pulse associated with the four clusters. Plots displaying the standard deviations are available in Figures A7 and A8. It is worth noting that the fraction of reliable users (i.e., mobile phone users with a validated home location) is stable between the different clusters (Figure A9 in Appendix C).

In order to understand the origin of the observed differences in mobility behavior between weekdays and weekends, we investigated the relationship between the pulse of a location and its socioeconomic status. To do so, we attach to each location the socioeconomic structure of its residents (when the information was available). The indicator used is divided into five relevant socioeconomic categories labeled ABC1, C2, C3, D, and E, with ABC1 as the most wealthy group and E as the group with the lowest income and educational level. The socioeconomic structure of a location is based on the surface area dedicated to the socioeconomic category of each census track intersecting the location (see the section *Socioeconomic structure of the locations* in Appendix B for further details). The relationship between these four clusters and the five socioeconomic categories is plotted in Figure 2. We observe in Figure 2A how the fraction of the surface area of locations belonging to a given cluster is distributed among the socioeconomic categories for the whole country. It is worth noting that a socioeconomic gradient exists from cluster 1, characterized by an over-representation of wealthy neighborhoods (i.e., comparatively larger red bars for the categories ABC1 and C2), to cluster 4, which shows an over-representation of neighborhoods with low incomes and educational levels (i.e., larger green bars for categories D and E). The comparison of the spatial distribution of clusters (Figure 2B) and socioeconomic categories (Figure 2C) in Gran Santiago (the largest city) confirms these results. Indeed, this particular spatial pattern of socioeconomic distribution has been described in detail in the literature, with a concentration of more affluent neighborhoods projected in a cone-shaped area that starts at the center of the city and opens towards the east and northeast outskirts of Santiago. This particular spatial segregation pattern has recently been corroborated by newer research [7,42]. This pattern is particularly apparent while looking at the spatial distribution of clusters 1 and 2 in Figure 2B.

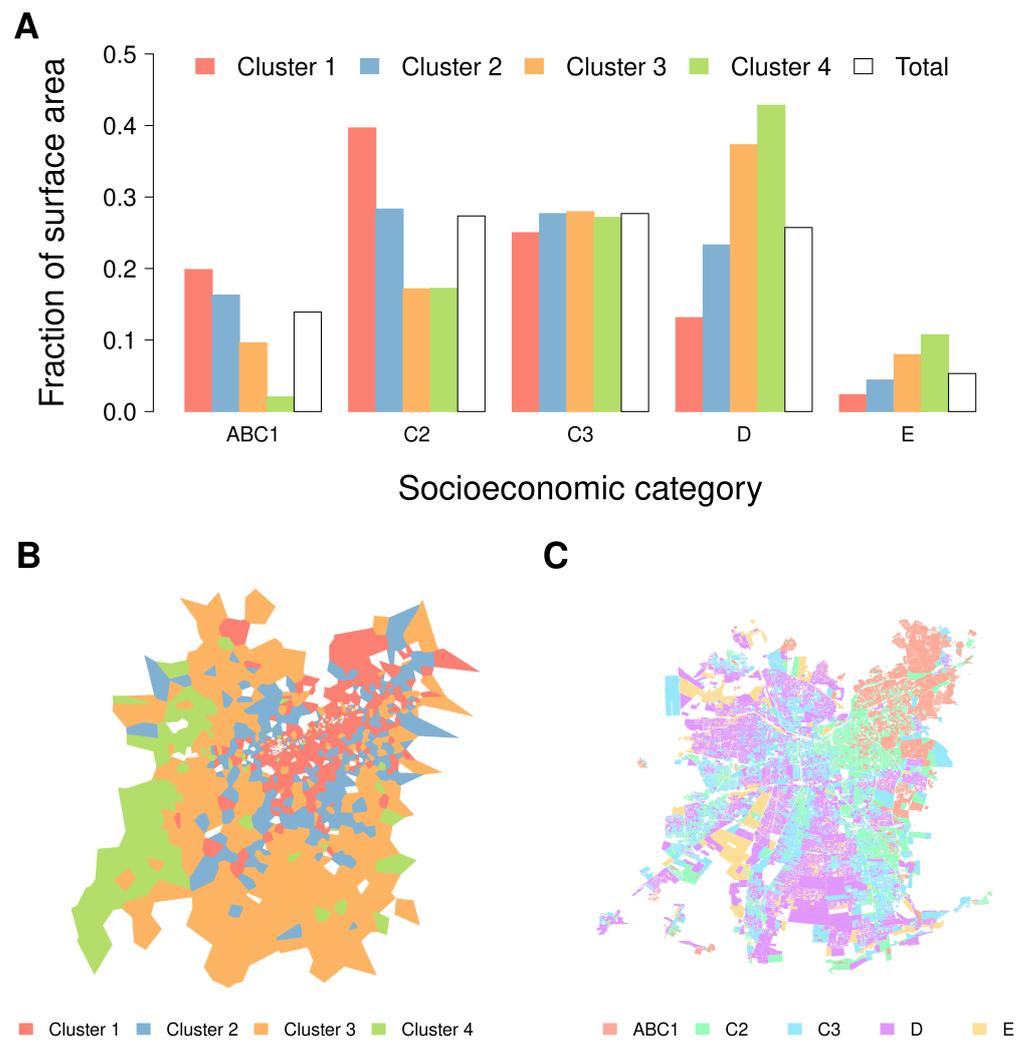


Figure 2. Socioeconomic characteristic of the clusters. (A) Fraction of surface area dedicated to each socioeconomic category according to the cluster (colored bars) and in total (white bar) for the whole country. (B) Maps of the four clusters in Gran Santiago (the largest city). (C) Spatial distribution of socioeconomic categories in Gran Santiago (the largest city).

3.2. Socio-Spatial Interactions Analysis

The results obtained for a week window period (i.e., $\Delta_t = |[1, 168]|$) are presented in Figure 3A,B. Each bar represents an element of the interaction matrices and can be interpreted as the probability of spatial and social interactions between two clusters during a typical week. The figures indicate that locations belonging to the same cluster—or similar clusters—tend to mostly interact with each other compared to their interaction with other locations, both spatially and socially. We also observe that these preferential interactions are less marked for social interactions (Figure 3B) than for spatial ones (Figure 3B).

Our results show a $\bar{\Phi}$ value of 0.52 for the spatial interaction matrix (Figure 3A) and 0.44 for the social interaction matrices (Figure 3B). These values demonstrate that a clear proximity exists in terms of spatial and social interactions between locations sharing similar socioeconomic features. It also shows that such a pattern is not just driven by spatial constraints. In other words, these results clearly show that people living in locations of a given socioeconomic status tend to move in, and socially interact with, people living in locations of the same, or similar, socioeconomic status. While slightly higher for the spatial interactions than for the social ones, it is particularly remarkable that both $\bar{\Phi}$ values are quite high, as this metric intrinsically considers a random model of interactions that

effectively considers spatial autocorrelation. That is, it explicitly considers what could happen in a random situation.

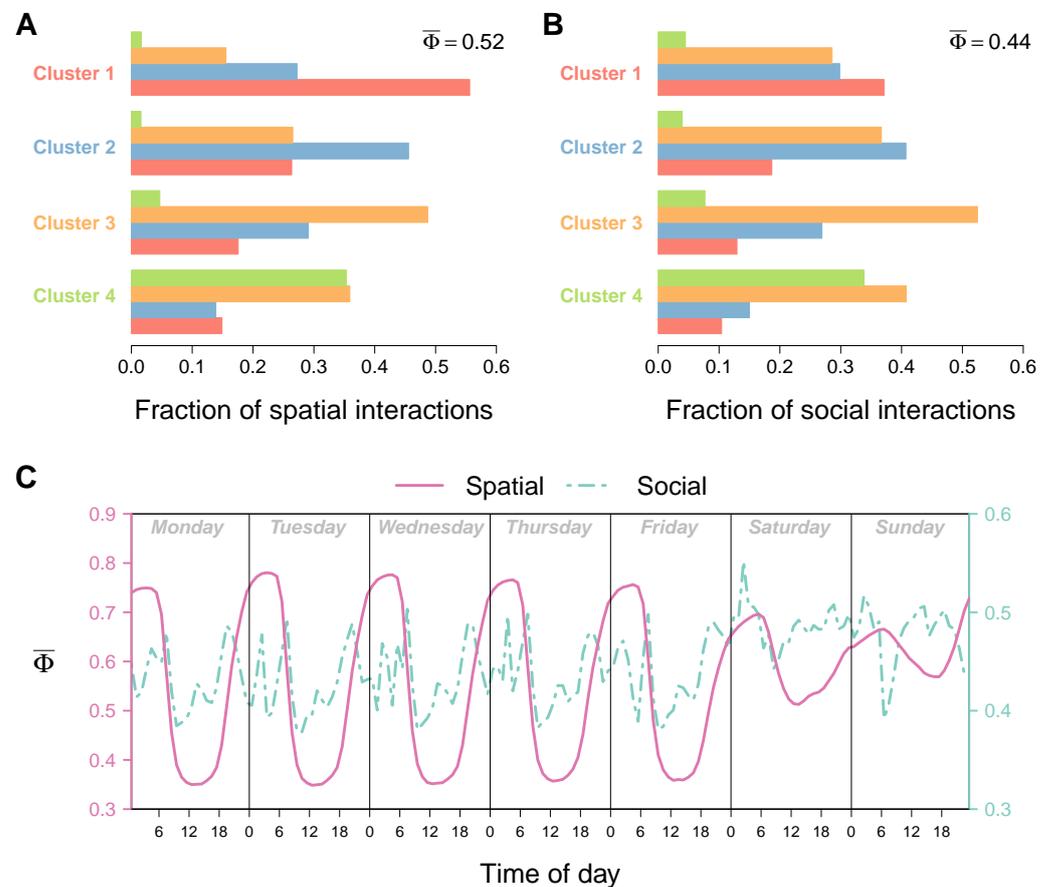


Figure 3. Socio-spatial interactions analysis. (A,B) The fraction of spatial (A) and social (B) interaction within and between clusters. The values of $\bar{\Phi}$ obtained with both matrices are displayed. (C) Temporal evolution of $\bar{\Phi}$ across week hours for the spatial interactions (in pink) and social interactions (in green).

In order to deepen the socio-spatial interaction analysis, we plot in Figure 3C the temporal evolution of $\bar{\Phi}$ during a typical week using a time window of one hour (see Equation (2)). As expected, the value of $\bar{\Phi}$ varies greatly according to the day of the week and the hour of the day. A greater variation is observed for the spatial interactions compared to the social ones. During weekdays, the spatial proximity between clusters is higher during the night with a $\bar{\Phi}$ value going from 0.75 to 0.8 compared to the 0.35 observed during the 11–19 h span. The variations decrease during weekend days with less proximity during night hours ($\bar{\Phi} = 0.7$) and more during the day ($\bar{\Phi} = 0.6$). This result also suggests that structural dependence between clusters, revealed by $\bar{\Phi}$, is more relevant when everybody is at home, confined to their individual socio-economic groups. Social interactions, in turn, show more nuanced and noisy results, presumably due to the comparatively lower number of social events during the night (see Figure A4). Nevertheless, we also observe that $\bar{\Phi}$ decreases during weekdays. In such a time span, particularly during the morning, people tend to interact less with people living in a similar cluster, a pattern that increases during the evening hours. However, it is interesting to note that this increase in social interactions with people living in a similar cluster starts earlier over the course of the day. It is also characterized by two peaks, one halfway through the day and another one around 6 pm. During weekend days, $\bar{\Phi}$ is quite stable with a value fluctuating around 0.5.

Finally, Figure 4 shows the $\bar{\Phi}$ index for intra- and inter-city interactions. In this case, an additional constraint is added in Equation (2) to only consider interactions between

locations that belong to cluster c in one city with locations belonging to cluster c' in the same or in another city, hence highlighting same-cluster interaction. We focus here on the three largest cities in terms of population in Chile. As it can be observed in Figure 4, the values of $\bar{\Phi}$ capturing spatial (Figure 4A) and social interactions (Figure 4B) between locations in the same city are in line with $\bar{\Phi}$ values obtained for the whole country. We also note that these preferential spatial and social interactions hold for several pairs of cities such as the people living in Concepción interacting with locations and people living in Santiago.

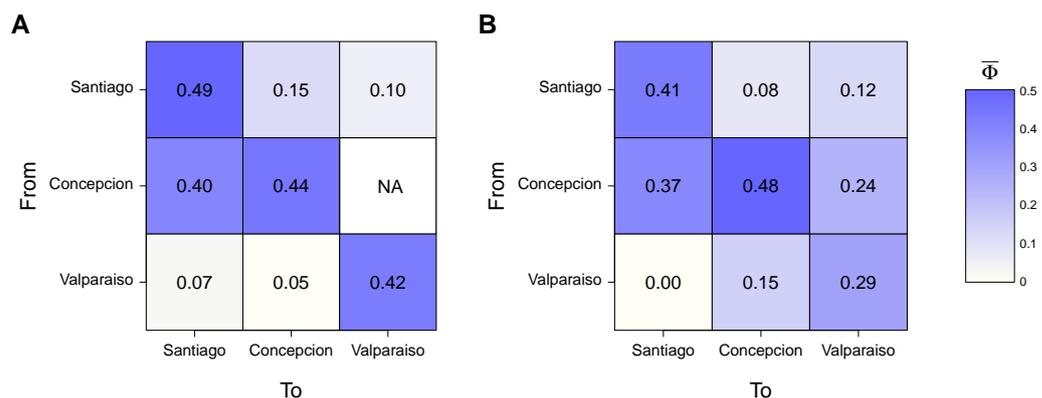


Figure 4. Intra-city and inter-cities socio-spatial interactions analysis. The index values are based on spatial (A) and social (B) interactions between locations in the same city (diagonal) or from one city to another. There were not enough data available (NA) to measure the spatial interactions between Concepción and Valparaíso.

4. Discussion

This study not only concurs with other studies showing how mobile phone data may aid in shaping a better understanding of the socioeconomic structure of spatial and social interactions in urban systems, but it also proposes a methodological approach to assess the hierarchical structure of spatiotemporal interactions across the city. By defining two temporal networks representing interactions stemming from highly resolved spatial and social events, we are able to describe how people ascribed to a particular socioeconomic level within the city interact with their environment and with people living in other locations across weekly hours. Similarly to [43], the net result here shows that people living in locations of a given socioeconomic status preferentially interact with locations and people sharing similar socioeconomic levels. Additionally, while this proximity varies similarly for both spatial and social interactions during the course of the week, social interactions measured by the voice calls between users exhibit a more nuanced association between socioeconomic status, much like what has recently been described in the literature [44]. This may be the product of a combination of factors, including the fact that the events captured by our voice call dataset are composed of a combination of professional, personal, and leisure interactions that may increase social mixing.

Our study sheds new light on the understanding of social mixing using large datasets. In fact, the mounting availability of such types of information is contributing to making large strides to describe the effect of segregation on the various realms of our society [5,25,45,46]. While our results contribute to the analysis and understanding of the relationship between urban mobility, social networks, and socioeconomic status, they also raise a number of new questions with regard to their generalization. In this regard, we will argue that the large sample used in this analysis, in terms of spatial and social events attached to a substantial number of mobile phones users in Chile (see Table A1 in Appendix A.2) provides an empirical description of how socioeconomic status relates to spatial and social interactions at several levels.

For instance, the role of space has been a central topic in understanding social tie formation. At the local scale, space has been described to determine interactions through distance, urban configuration [47,48], and specific locations fostering social interactions [49]. These studies highlight not only the importance of space-mediated interactions but also the relevance of social interactions such as relationship maintenance among friends [36]. While these conclusions go beyond this particular work, we envisage that ongoing improvements in the identification of residences [50] and transportation modes [51], among others, will clearly foster more granular descriptions of urban dynamics. In fact, they may even shift the focus to more localized descriptions of social and transportation behavior, as has recently been seen by the analysis of the ongoing COVID-19 pandemic in Chile [45,52]. At broader scales, spatial limitations (e.g., the modifiable areal unit problem) have recently been invoked to highlight the difficulties describing spatial aspects of segregation [42]. While this old geographic issue may certainly hamper the possibility to inform social mixing from mobile phone datasets, other equally important aspects of areal distributions may acquire relevance in the spatially explicit descriptions of cities. For instance, the definition of urban entities may also concur with the MAUP to describe the correct functional extension to which urban descriptions should be attached [8,53]. In spite of this, it is interesting to note that the preferential interactions among socioeconomic status in Chile, as described here, appear to hold even when considering interactions between cities hinting towards an intrinsic property of social systems as opposed to a particular constraint (e.g., spatial) imposed on the interaction network [54].

Finally, it is also worth noting that the usage of the hierarchy index proposed in [40] used here provides a simple conceptual means to compare both social and spatial networks across the whole country that is independent of urban shape, while still capturing the spatial hierarchy of mobility within and between cities [55].

Author Contributions: Conceptualization, M.L. and H.S.; methodology, M.L. and H.S.; software, M.L. and H.S.; validation, M.L. and H.S.; formal analysis, M.L. and H.S.; investigation, M.L. and H.S.; resources, M.L. and H.S.; data curation, M.L. and H.S.; writing—original draft preparation, M.L. and H.S.; writing—review and editing, M.L. and H.S.; visualization, M.L. and H.S.; supervision, M.L. and H.S.; project administration, M.L. and H.S.; funding acquisition, M.L. and H.S. All authors have read and agreed to the published version of the manuscript.

Funding: The work of M.L. was supported by a grant from the French National Research Agency (project NetCost, ANR-17-CE03-0003 grant). H.S. was supported by the Chilean Agency of Research and Development ANID (FONDECYT Regular grant #1211490).

Data Availability Statement: The mobile phone datasets used in this study are available on request from the corresponding author.

Acknowledgments: Thanks to Isidro Puig from the OCUC for his help on census data.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Data Preprocessing

Appendix A.1. Call and Location History

The data used in this study consists of Call Detail Records (CDR) and Data Detail Records (XDR) provided by Telefónica Chile representing 37% share of the mobile phone market in Chile.

Our first dataset is composed of billions of cellphone pings made by 4 million mobile phones during 3 weeks in March, May, and October 2015 in Chile. Each ping is characterized by its location (i.e., Voronoi cell) and a timestamp informing us of the hour, the day of the week, and the week when the ping has occurred. We structured the dataset in a four-column location history table (week, hour, user, location). Each line represents a *spatial event* informing us of the presence of a user in a location during a given week at a given time. If the presence of a user was detected in several locations during the same hour, we

chose the location with the highest number of events. In the event of a tie, one of them was drawn at random.

In addition, we relied on a second dataset to compute the history of calls between mobile phone users. This dataset was structured into a four-column call history table (week, hour, caller, callee). Each line represents a *social event*. A social event is characterized by a phone call made by a caller to a callee during a given hour and a given week. This means that if a caller called several times the same callee during a given hour, only one social event has been considered.

Appendix A.2. Identification of the Users' Place of Residence

The first step consisted in identifying the users' place of residence to filter out users with a low number of spatial events and/or exhibiting irregular mobility patterns. For each of the three weeks periods and for each user, we applied the following procedure to extract the home locations:

- First, we focused on the user's spatial events occurring during nighttime hours (between 9 pm and 8 am included). Only days of the week from Monday to Thursday were considered ($N = 48$ h in total). We note that N_u is the number of events occurring during nighttime hours.
- We applied here a **first filter** by considering only users with a number of spatial events higher than a fraction $\delta_A = N_u/N$ of the total number of nighttime hours.
- We identified the location in which the user has localized the highest number of spatial events during nighttime hours. We define this location as her or his home location.
- A **second filter** was also implemented to select only users whose fraction of events occurring at their home location during nighttime is larger than a fraction δ_R of the total number of events during nighttime.

As explained in [13], the first filter δ_A is applied to discard users having too low a number of spatial events. The last filter allowed us to adjust the degree of confidence in the identification of the home location. We chose to fix δ_A to 0.3 and δ_R to 0.3 which seems to be a good interplay, allowing us to remove users not active enough and/or exhibiting irregular mobility patterns during the time period (Figure A1) while preserving the spatial distribution of inhabitants observed in Chile (Figure A2). The number of reliable users (i.e., with a validated home location) is available in Table A1.

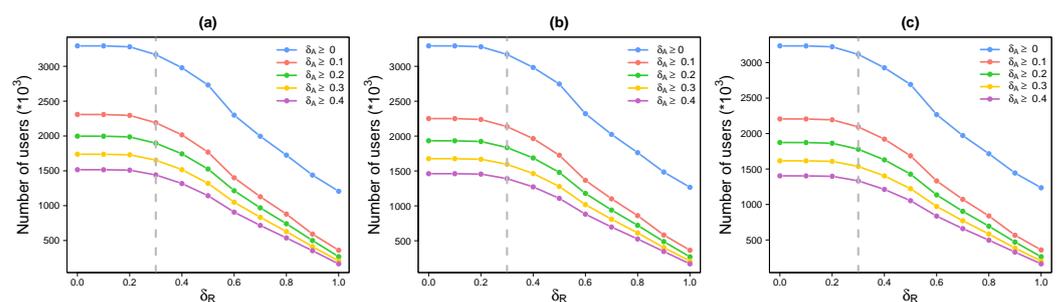


Figure A1. Influence of the parameters. Number of reliable users during the first (a), second (b), and third (c) week as a function of δ_R and for different values of δ_A . The vertical bars indicate the value $\delta_R = 0.3$.

Table A1. Number of users (all) and reliable users according to the week of observation and in total.

Date	# Users (All)	# Reliable Users
15 to 21 March 2015	3,292,923	1,657,048
10 to 16 May 2015	3,292,647	1,598,571
2 to 8 August 2015	3,236,122	1,539,621
Total	4,064,476	2,565,365

As mentioned in the previous section, there are some holes in the user history location with hours with no events. Nevertheless, we observe in Figure A3 that, during each week of observation, 75% of the reliable users have at least 100 spatial events (60% of the maximum value).

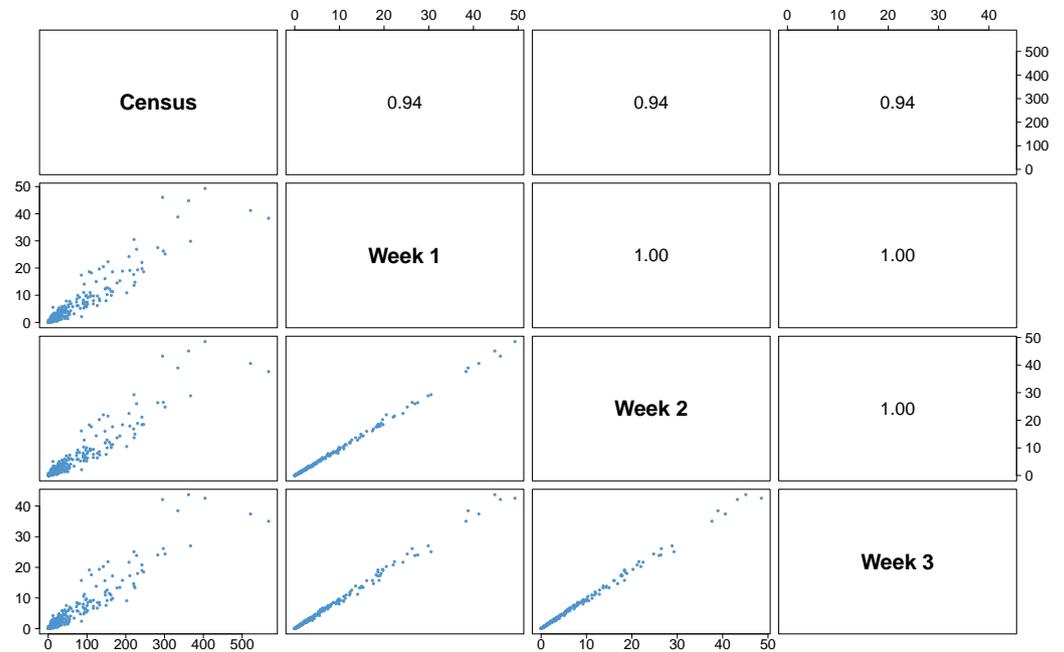


Figure A2. Comparison between census and XDR data. Each scatter plot and its associated Pearson correlation coefficient represents a comparison between the number of inhabitants (expressed in thousands of individuals) in the census and the number of inhabitants (expressed in thousands of individuals) estimated with XDR data (i.e., reliable users) during the three weeks of observation. Each point represents one municipality of Chile.

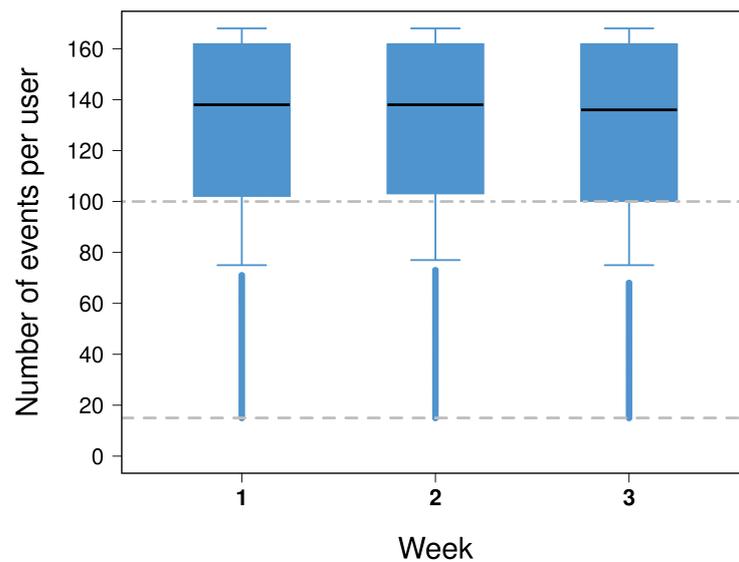


Figure A3. Boxplots of the number of events per reliable user according to the week of observation. The dashed grey line represents the minimum value (15 is the minimum value required to pass the first filter in the home identification). The dash-dotted line represents the limit of 100 events. The maximum value is 168 (number of hours in the week). Each boxplot is composed of the first decile, the lower hinge, the median, the upper hinge, and the last decile. The blue dots represent the outliers.

Appendix A.3. From Events to Networks

A table summarizing the number of reliable users and their associated numbers of spatial and social events is available in Table A2. The associated temporal evolution is available in Figure A4. Finally, the collections of spatial and social events have enabled us to construct 168 spatial networks and 168 social networks. The weight G_{ij}^t of a link between two locations i and j at time t is equal to the number of users living in location i that were present in location j at time t (all weeks combined). Similarly, the link weight S_{ij}^t of a social network is equal to the number of social interactions made by users living in location i with users living in location j at time t (all weeks combined).

Table A2. Number of reliable users, spatial and social events per week and in total.

Date	#Reliable Users	#Spatial Events	#Social Events
15 to 21 March 2015	1,657,048	129,760,887	4,433,505
10 to 16 May 2015	1,598,571	126,359,359	4,207,538
2 to 8 August 2015	1,539,621	120,960,807	3,905,935
Total	3,023,946	377,081,053	12,546,978

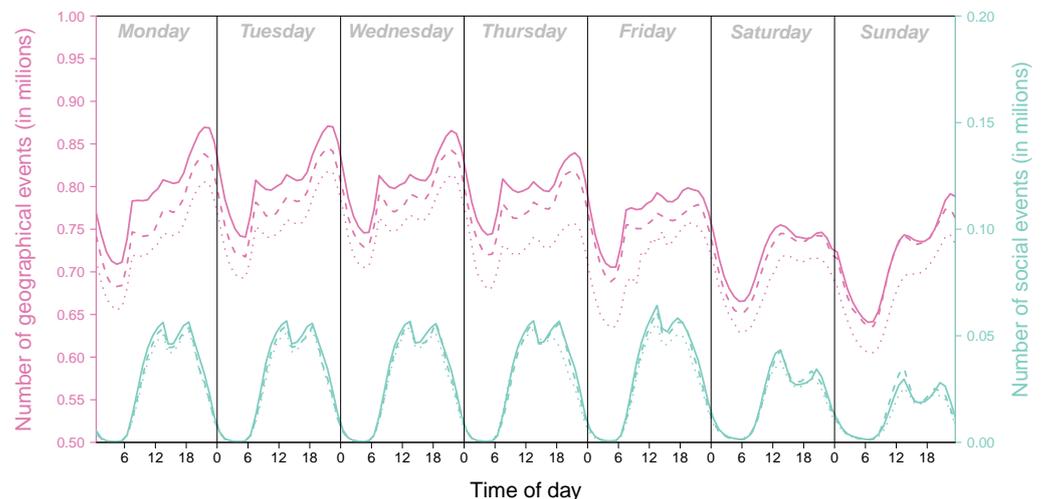


Figure A4. Number of spatial events (in pink) and social events (in green) according to the hour of the day. Each line represents a week of observation.

Appendix B. Socioeconomic Structure of the Locations

As mentioned in the main text, we attached to each of the 3876 locations some information regarding the socioeconomic level of their residents when the information was available. To do so, we relied on the socioeconomic map of Chile proposed by Adimark [7,56]. These maps are available from the *Observatorio de Ciudades UC* (OCUC) website (<https://ideocuc-ocuc.hub.arcgis.com/>, last accessed 6 December 2022) in Shapefile format for five major Chilean cities.

- Antofagasta in 2002 available at https://ideocuc-ocuc.hub.arcgis.com/datasets/fbde68b6c3d547c8adfcc17d196e1e88_0, last accessed 6 December 2022.
- Coquimbo y La Serena in 2002 available at <https://ideocuc-ocuc.hub.arcgis.com/>, last accessed 6 December 2022.
- Gran Concepción in 2002 available at https://ideocuc-ocuc.hub.arcgis.com/datasets/f62f12fae97548fd8c71cb405d40e5f2_0, last accessed 6 December 2022.
- Gran Santiago in 2012 available at https://ideocuc-ocuc.hub.arcgis.com/datasets/c264bc8bca7f45bc8ae74329557628b2_0, last accessed 6 December 2022.
- Puerto Montt and Puerto Varas in 2002 available at https://ideocuc-ocuc.hub.arcgis.com/datasets/91deae3707ff447f961b4e2a5cf2300d_0, last accessed 6 December 2022.

- Valparaíso in 2002 available at https://ideocuc-ocuc.hub.arcgis.com/datasets/b9458dbbc94343e58ea5fc9c5def03f9_0, last accessed 6 December 2022.

These data inform us of the dominant socioeconomic categories of the resident of each ‘manzana’ (i.e., census block). There are five categories labeled ABC1, C2, C3, D, and E with ABC1 as the most wealthy group and E as the group with the lowest income and educational level. For each location, we computed the area of the intersection between the Voronoi cell and the census blocks (if any) for each category. To identify the socioeconomic structure of each cluster, we computed the fraction of surface area (of the locations composing this cluster) dedicated to each socioeconomic category.

Appendix C. Clustering Analysis

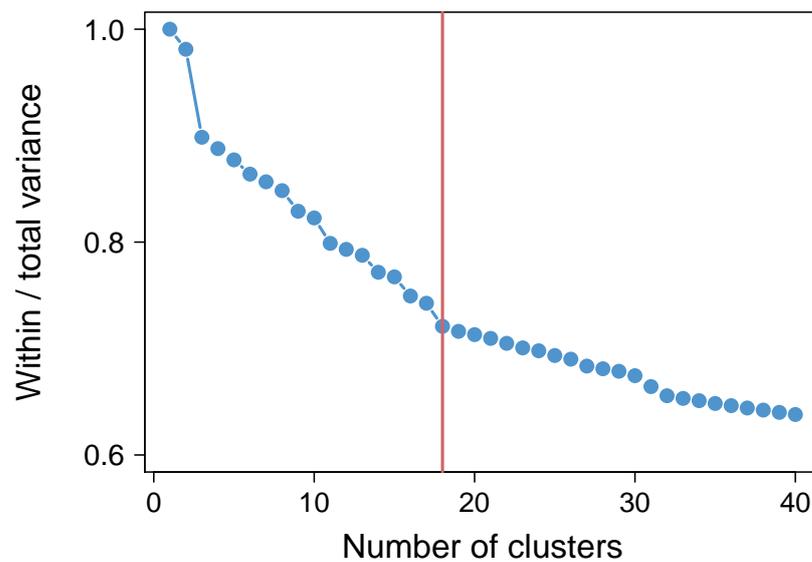


Figure A5. Ratio between the within-group variance and the total variance as a function of the number of clusters. The red line represents the selected number of clusters.

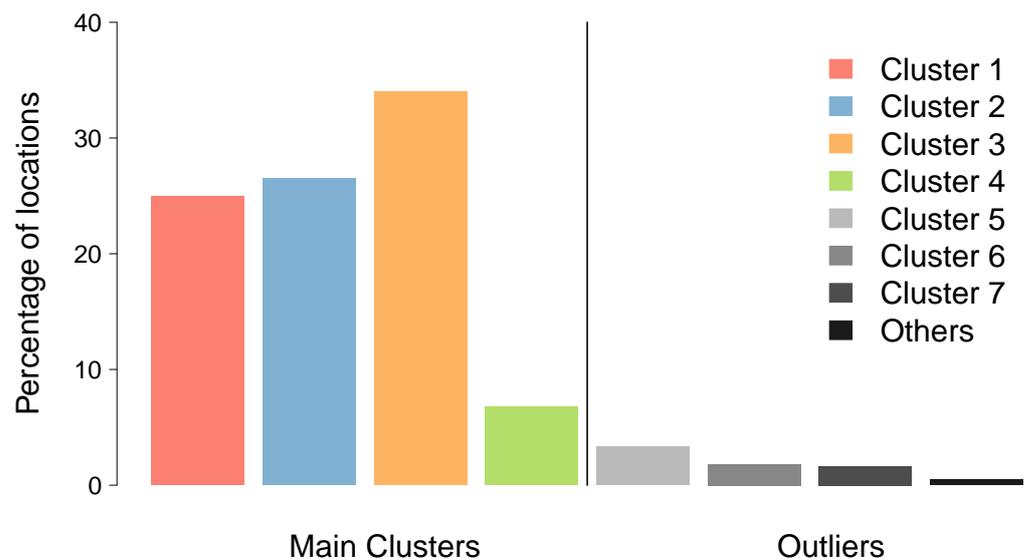


Figure A6. Percentage of locations by cluster.

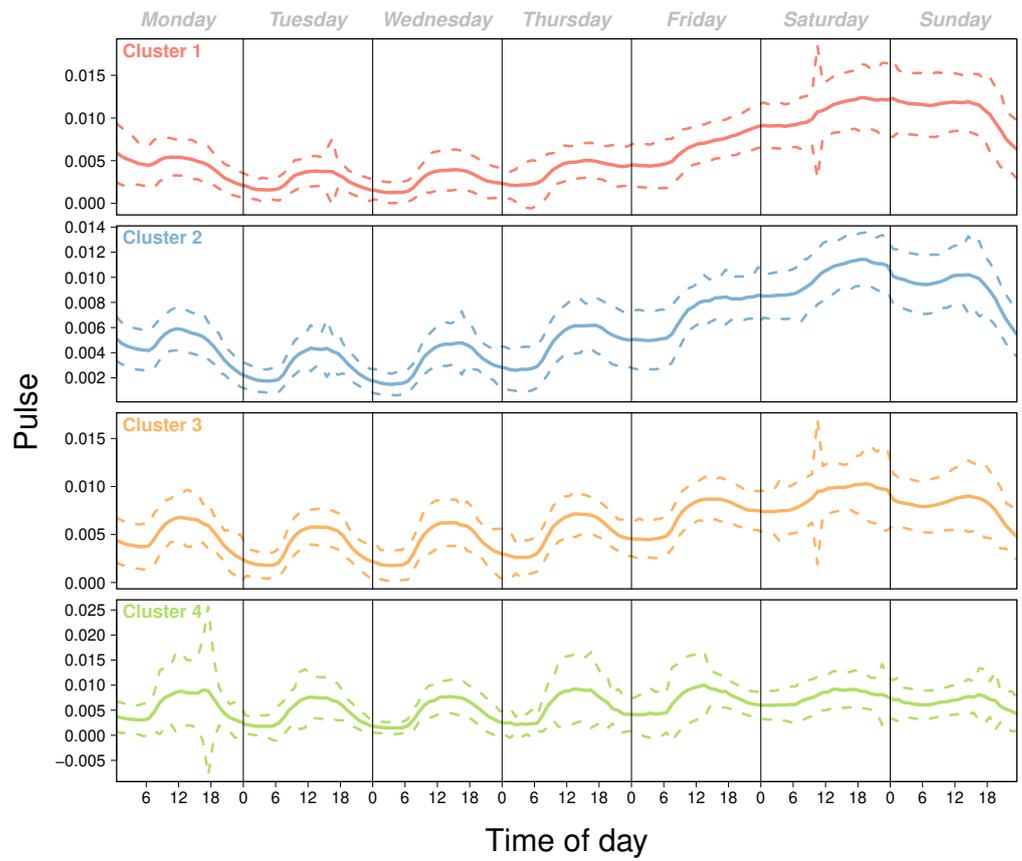


Figure A7. Pulses associated with the four main clusters. The solid lines represent the average pulse, while the dashed lines represent one standard deviation.

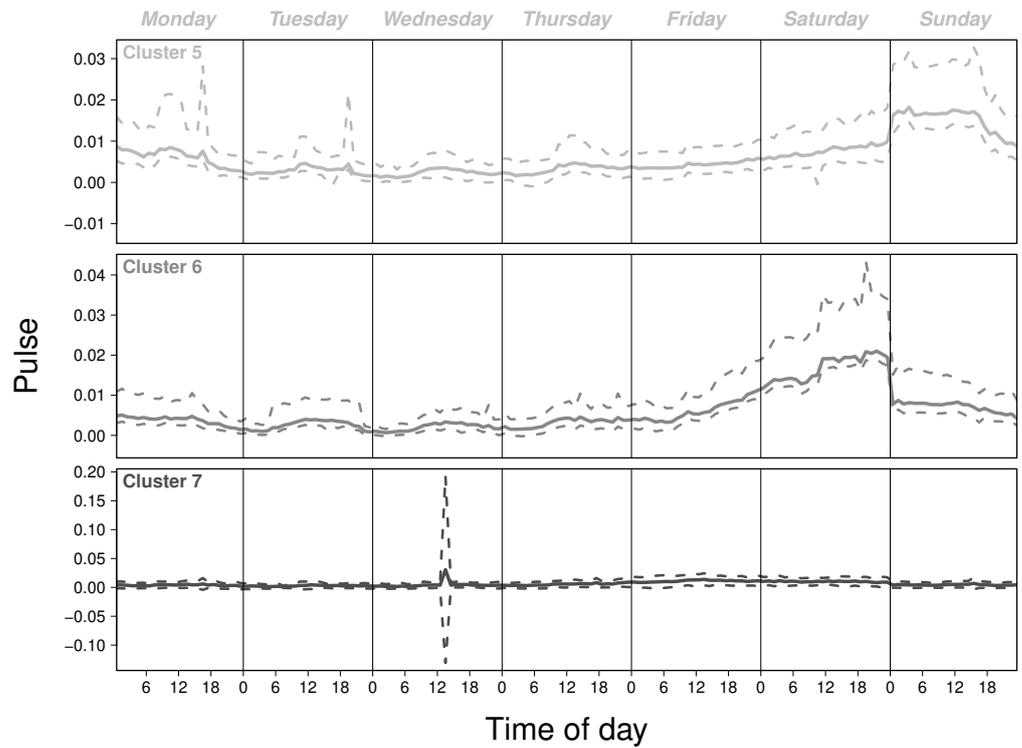


Figure A8. Pulses associated with the three additional clusters. The solid lines represent the average pulse, while the dashed lines represent one standard deviation.

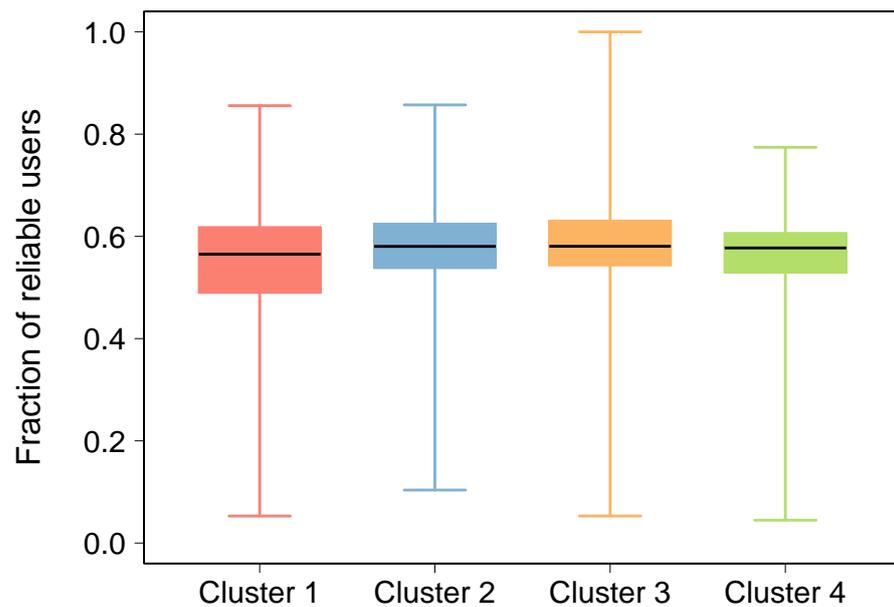


Figure A9. Boxplots of the fraction of reliable users per cluster. Each boxplot is composed of the minimum value, the first quartile, the median, the third quartile, and the maximal value.

Appendix D. Null Model

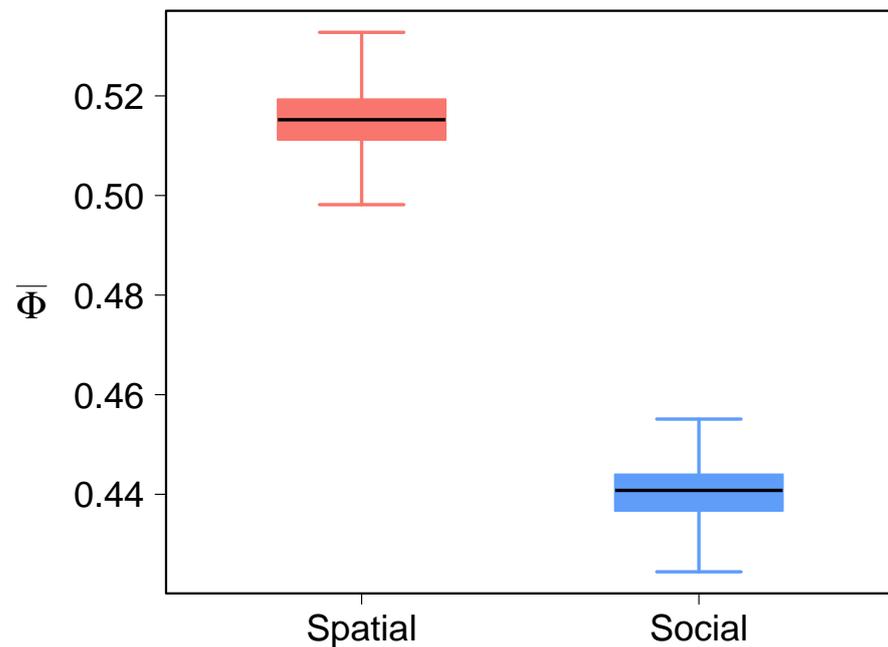


Figure A10. Boxplots of $\bar{\Phi}$ for the spatial and social interaction matrices. Each boxplot is composed of 100 $\bar{\Phi}$ values, each of them obtained with a Φ_h value based on one random assignment. Each boxplot is composed of the minimum value, the first quartile, the median, the third quartile, and the maximal value.

References

1. Hall, P.; Tewdwr-Jones, M. *Urban and Regional Planning*; Routledge: Oxfordshire, UK, 2019.
2. Alvaredo, F.; Chancel, L.; Piketty, T.; Saez, E.; Zucman, G. *World Inequality Report 2018*; Belknap Press: Cambridge, MA, USA, 2018.
3. El-Geneidy, A.; Levinson, D.; Diab, E.; Boisjoly, G.; Verbich, D.; Loong, C. The cost of equity: Assessing transit accessibility and social disparity using total travel cost. *Transp. Res. Part A Policy Pract.* **2016**, *91*, 302–316. [[CrossRef](#)]

4. Jones, M.; Pebley, A.R. Redefining neighborhoods using common destinations: Social characteristics of activity spaces and home census tracts compared. *Demography* **2014**, *51*, 727–752. [[CrossRef](#)] [[PubMed](#)]
5. Lenormand, M.; Samaniego, H.; Chaves, J.; da Fonseca Vieira, V.; Silva, M.; Evsukoff, A. Entropy as a Measure of Attractiveness and Socioeconomic Complexity in Rio de Janeiro Metropolitan Area. *Entropy* **2020**, *22*, 368. [[CrossRef](#)]
6. Steele, J.E.; Sundsøy, P.R.; Pezzulo, C.; Alegana, V.A.; Bird, T.J.; Blumenstock, J.; Bjelland, J.; Engø-Monsen, K.; de Montjoye, Y.A.; Iqbal, A.M.; et al. Mapping poverty using mobile phone and satellite data. *J. R. Soc. Interface* **2017**, *14*, 20160690. [[CrossRef](#)] [[PubMed](#)]
7. Dannemann, T.; Sotomayor-Gómez, B.; Samaniego, H. The Time Geography of Segregation during Working Hours. *R. Soc. Open Sci.* **2018**, *5*, 180749. [[CrossRef](#)]
8. Cottineau, C.; Vanhoof, M. Mobile phone indicators and their relation to the socioeconomic organisation of cities. *ISPRS Int. J. -Geo-Inf.* **2019**, *8*, 19. [[CrossRef](#)]
9. Alessandretti, L.; Aslak, U.; Lehmann, S. The scales of human mobility. *Nature* **2020**, *587*, 402–407. [[CrossRef](#)]
10. De Montjoye, Y.A.; Hidalgo, C.A.; Verleysen, M.; Blondel, V.D. Unique in the Crowd: The privacy bounds of human mobility. *Sci. Rep.* **2013**, *3*, 1376. [[CrossRef](#)]
11. Blondel, V.D.; Decuyper, A.; Krings, G. A survey of results on mobile phone datasets analysis. *EPJ Data Sci.* **2015**, *4*, 10. [[CrossRef](#)]
12. Barbosa, H.; Barthelemy, M.; Ghoshal, G.; James, C.R.; Lenormand, M.; Louail, T.; Menezes, R.; Ramasco, J.J.; Simini, F.; Tomasini, M. Human mobility: Models and applications. *Phys. Rep.* **2018**, *734*, 1–74. [[CrossRef](#)]
13. Lenormand, M.; Louail, T.; Barthelemy, M.; Ramasco, J.J. Is spatial information in ICT data reliable? *arXiv* **2016**, arXiv:1609.03375.
14. Lenormand, M.; Murillo Arias, J.; San Miguel, M.; Ramasco, J.J. On the importance of trip destination for modeling individual human mobility patterns. *J. R. Soc. Interface* **2020**, *17*, 20200673. [[CrossRef](#)] [[PubMed](#)]
15. Entwisle, B. Putting people into place. *Demography* **2007**, *44*, 687–703. [[CrossRef](#)] [[PubMed](#)]
16. Shelton, T.; Poorthuis, A.; Zook, M. Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landsc. Urban Plan.* **2015**, *142*, 198–211. [[CrossRef](#)]
17. Salganik, M.J. *Bit by Bit: Social Research in the Digital Age*; Princeton University Press: Princeton, NJ, USA, 2018.
18. Blumenstock, J.; Cadamuro, G.; On, R. Predicting poverty and wealth from mobile phone metadata. *Science* **2015**, *350*, 1073–1076. [[CrossRef](#)]
19. Frias-Martinez, V.; Virseda, J. On the relationship between socio-economic factors and cell phone usage. In Proceedings of the Fifth International Conference on Information and Communication Technologies and Development, ICTD '12, Atlanta, GA, USA, 12–15 March 2012; pp. 76–84.
20. Lenormand, M.; Picornell, M.; Garcia Cantú, O.; Tugores, A.; Louail, T.; Herranz, R.; Barthelemy, M.; Frías-Martínez, E.; Ramasco, J.J. Comparing and modeling land use organization in cities. *R. Soc. Open Sci.* **2015**, *2*, 150459. [[CrossRef](#)]
21. Pappalardo, L.; Vanhoof, M.; Gabrielli, L.; Smoreda, Z.; Pedreschi, D.; Giannotti, F. An analytical framework to nowcast well-being using mobile phone data. *Int. J. Data Sci. Anal.* **2016**, *2*, 75–92. [[CrossRef](#)]
22. Alessandretti, L.; Sapiezynski, P.; Sekara, V.; Lehmann, S.; Baronchelli, A. Evidence for a conserved quantity in human mobility. *Nat. Hum. Behav.* **2018**, *2*, 485–491. [[CrossRef](#)]
23. Barbosa, H.; Hazarie, S.; Dickinson, B.; Bassolas, A.; Frank, A.; Kautz, H.; Sadilek, A.; Ramasco, J.J.; Ghoshal, G. Uncovering the socioeconomic facets of human mobility. *arXiv* **2020**, arXiv:2012.00838.
24. Jiron, P. Mobility on the Move: Examining Urban Daily Mobility Practices in Santiago de Chile. Ph.D. Thesis, London School of Economics and Political Science, London, UK, 2009.
25. Gauvin, L.; Tizzoni, M.; Piaggese, S.; Young, A.; Adler, N.; Verhulst, S.; Ferres, L.; Cattuto, C. Gender gaps in urban mobility. *Humanit. Soc. Sci. Commun.* **2020**, *7*, 11. [[CrossRef](#)]
26. Lotero, L.; Hurtado, R.G.; Floría, L.M.; Gómez-Gardeñes, J. Rich do not rise early: Spatio-temporal patterns in the mobility networks of different socio-economic classes. *R. Soc. Open Sci.* **2016**, *3*, 150654. [[CrossRef](#)] [[PubMed](#)]
27. Palmer, J.R.B.; Espenshade, T.J.; Bartumeus, F.; Chung, C.Y.; Ozgencil, N.E.; Li, K. New Approaches to Human Mobility: Using Mobile Phones for Demographic Research. *Demography* **2013**, *50*, 1105–1128. [[CrossRef](#)] [[PubMed](#)]
28. McLuhan, M. *The Medium Is the Message*; Routledge: Oxfordshire, UK, 1964.
29. Yeung, H.W.C. Capital, State and Space: Contesting the Borderless World. *Trans. Inst. Br. Geogr.* **1998**, *23*, 291–309. [[CrossRef](#)]
30. Bagrow, J.P.; Liu, X.; Mitchell, L. Information flow reveals prediction limits in online social activity. *Nat. Hum. Behav.* **2019**, *3*, 122–128. [[CrossRef](#)]
31. Song, C.; Qu, Z.; Blumm, N.; Barabasi, A.L. Limits of Predictability in Human Mobility. *Science* **2010**, *327*, 1018–1021. [[CrossRef](#)] [[PubMed](#)]
32. Urry, J. Social networks, travel and talk. *Br. J. Sociol.* **2003**, *54*, 155–175. [[CrossRef](#)]
33. Netto, V.M.; Meirelles, J.; Ribeiro, F.L.; Federal, U.; Uff, F. Social Interaction and the City: The Effect of Space on the Reduction of Entropy. *Complexity* **2017**, *2017*, 6182503. [[CrossRef](#)]
34. Cho, E.; Myers, S.A.; Leskovec, J. Friendship and mobility: User movement in location-based social networks. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11, San Diego, CA, USA, 21–24 August 2011; pp. 1082–1090.
35. Le Roux, G.; Vallée, J.; Commenges, H. Social segregation around the clock in the Paris region (France). *J. Transp. Geogr.* **2017**, *59*, 134–145. [[CrossRef](#)]

36. Carrasco, J.A.; Hogan, B.; Wellman, B.; Miller, E.J. Agency in social activity interactions: The role of social networks in time and space. *Tijdschr. Econ. Soc. Geogr.* **2008**, *99*, 562–583. [[CrossRef](#)]
37. Gonzalez, M.C.; Hidalgo, C.A.; Barabási, A.L. Understanding individual human mobility patterns. *Nature* **2008**, *453*, 779. [[CrossRef](#)]
38. Miranda, F.; Doraiswamy, H.; Lage, M.; Zhao, K.; Gonçalves, B.; Wilson, L.; Hsieh, M.; Silva, C.T. Urban Pulse: Capturing the Rhythm of Cities. *IEEE Trans. Vis. Comput. Graph.* **2017**, *23*, 791–800. [[CrossRef](#)] [[PubMed](#)]
39. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 5th ed.; Springer: New York, NY, USA, 2009.
40. Bassolas, A.; Barbosa-Filho, H.; Dickinson, B.; Dotiwalla, X.; Eastham, P.; Gallotti, R.; Ghoshal, G.; Gipson, B.; Hazarie, S.A.; Kautz, H.; et al. Hierarchical Organization of Urban Mobility and Its Connection with City Livability. *Nat. Commun.* **2019**, *10*, 4817. [[CrossRef](#)] [[PubMed](#)]
41. Jiron, P.; Carrasco, J.A. Understanding Daily Mobility Strategies through Ethnographic, Time Use, and Social Network Lenses. *Sustainability* **2020**, *12*, 312. [[CrossRef](#)]
42. Garreton, M.; Basauri, A.; Valenzuela, L. Exploring the correlation between city size and residential segregation: Comparing Chilean cities with spatially unbiased indexes. *Environ. Urban.* **2020**, *32*, 569–588. [[CrossRef](#)]
43. Wang, Q.; Phillips, N.E.; Small, M.L.; Sampson, R.J. Urban mobility and neighborhood isolation in America's 50 largest cities. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 7735–7740. [[CrossRef](#)]
44. Xu, Y.; Santi, P.; Ratti, C. Beyond Distance Decay: Discover Homophily in Spatially Embedded Social Networks. *Ann. Am. Assoc. Geogr.* **2021**, *112*, 505–521. [[CrossRef](#)]
45. Mena, G.E.; Martinez, P.P.; Mahmud, A.S.; Marquet, P.A.; Buckee, C.O.; Santillana, M. Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile. *Science* **2021**, *372*, eabg5298. [[CrossRef](#)]
46. Beiró, M.G.; Bravo, L.; Caro, D.; Cattuto, C.; Ferres, L.; Graells-Garrido, E. Shopping Mall Attraction and Social Mixing at a City Scale. *EPJ Data Sci. Data* **2018**, *7*, 49. [[CrossRef](#)]
47. Wang, P.; Zhang, J.; Liu, G.; Fu, Y.; Aggarwal, C. Ensemble-Spotting: Ranking Urban Vibrancy via POI Embedding with Multi-view Spatial Graphs. In Proceedings of the 2018 SIAM International Conference on Data Mining (SDM), San Diego, CA, USA, 3–5 May 2018; pp. 351–359.
48. Liu, H.; Guo, Q.; Zhu, H.; Fu, Y.; Zhuang, F.; Ma, X.; Xiong, H. Characterizing and Forecasting Urban Vibrancy Evolution: A Multi-View Graph Mining Perspective. *ACM Trans. Knowl. Discov. Data* **2022**, *accepted*. [[CrossRef](#)]
49. Small, M.L.; Adler, L. The Role of Space in the Formation of Social Ties. *Annu. Rev. Sociol.* **2019**, *45*, 111–132. [[CrossRef](#)]
50. Pappalardo, L.; Ferres, L.; Sacasa, M.; Cattuto, C.; Bravo, L. Evaluation of home detection algorithms on mobile phone data using individual-level ground truth. *EPJ Data Sci.* **2021**, *10*, 29. [[CrossRef](#)] [[PubMed](#)]
51. Graells-Garrido, E.; Caro, D.; Parra, D. Inferring modes of transportation using mobile phone data. *EPJ Data Sci.* **2018**, *7*, 49. [[CrossRef](#)]
52. Gozzi, N.; Tizzoni, M.; Chinazzi, M.; Ferres, L.; Vespignani, A.; Perra, N. Estimating the effect of social inequalities on the mitigation of COVID-19 across communities in Santiago de Chile. *Nat. Commun.* **2021**, *12*, 2429. [[CrossRef](#)] [[PubMed](#)]
53. Sotomayor-Gómez, B.; Samaniego, H. City limits in the age of smartphones and urban scaling. *Comput. Environ. Urban Syst.* **2020**, *79*, 101423. [[CrossRef](#)]
54. Onnela, J.P.; Arbesman, S.; González, M.C.; Barabási, A.L.; Christakis, N.A. Geographic constraints on social network groups. *PLoS ONE* **2011**, *6*, e16939. [[CrossRef](#)] [[PubMed](#)]
55. Pumain, D. An evolutionary theory of urban systems. In *International and Transnational Perspectives on Urban Systems*; Springer: Singapore, 2018; pp. 3–18.
56. Adimark. Mapa Socioeconómico de Chile. 2009. Available online: <https://ideocuc-ocuc.hub.arcgis.com/> (accessed on 6 December 2022).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.