EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE

Open Access

CrossMark

Shopping mall attraction and social mixing at a city scale

Mariano G. Beiró[1*], Loreto Bravo[2,3], Diego Caro[2,3], Ciro Cattuto[4], Leo Ferres[2,3] and
Eduardo Graells-Garrido[2,3]

*Correspondence: mbeiro@fi.uba.ar
[1]Facultad de Ingeniería, Universidad de Buenos Aires, INTECIN (CONICET), Buenos Aires, Argentina
Full list of author information is available at the end of the article

Abstract

In Latin America, shopping malls seem to offer an open, safe and democratic version of the public space. However, it is often difficult to quantitatively measure whether they indeed foster, hinder, or are neutral with respect to social inclusion. In this work, we investigate if, and by how much, people from different social classes are attracted by the same malls. Using a dataset of mobile phone network records from 387,152 devices identified as customers of 16 malls in Santiago de Chile, we performed several analyses to study whether malls with higher social mixing attract more people. Our pipeline, which starts with the socio-economic characterization of mall visitors, includes the estimation of social mixing and diversity of malls, the application of the gravity model of mobility, and the definition of a co-visitation model. Results showed that people tend to choose a profile of malls more in line with their own socio-economic status and the distance from their home to the mall, and that higher mixing does positively contribute to the process of choosing a mall. We conclude that (a) there is social mixing in malls, and (b) that social mixing is a factor at the time of choosing which mall to go to. Thus, the potential for social mixing in malls could be capitalized by designing public policies regarding transportation and mobility to make some malls strong social inclusion hubs.

Keywords: Malls and retail; Human mobility; Socio-economic factors of mobility; Social mixing; Call/Data detail records; Urban informatics

1 Introduction

Shopping malls have a prominent place in the configuration of modern cities, affecting the daily activities, social relationships and mobility of their inhabitants. They arose in the U.S. in the 1950s, and have been idealized since then as democratic spaces to which all citizens have equality of access [1, 2].

Since then, the concept of "shopping mall" has been replicated throughout the world, and in many cities they are strong constituents of the urban space, specially in developing countries. In particular, shopping malls became popular in Chile in the early 80's, and were one of the first signs of the globalization and liberalization of consumption in Latin America. They are currently one of the main elements of Chilean popular culture, influencing people's mobility, purchase decisions, self-expression, and segregation processes [3]. Furthermore, in comparison with other emerging countries, Chile is the Latin-American country with the highest mall surface area per inhabitant [4].

However, the ideal of "democratic spaces" has been challenged in the literature. For example, in the U.S., most malls are located in suburban areas or areas without adequate public transportation, and some of them become unreachable for the average lower-middle class citizen (e.g., people without a car), thus promoting minority exclusion and segregation [2, 5–8], see also [9], for a discussion of how malls are in decline in the US, which is not the case in Chile. In the case of Santiago de Chile, as in many other large cities from developing countries, socio-economic class segregation has different characteristics than in the U.S.: first, the extensive public transportation system allows malls' close integration with pedestrian and city life; second, many low income people live in peripheral areas, and the arrival of malls to these zones has permitted the integration of middle- and low-income consumers [10]. These situations allow for cross-class encounters in malls.

Thus, the aim of our work is to assess the potential of shopping malls for social mixing and inclusion. We put this idea to test by employing a recent segregation framework [11] to measure if, and by how much, people from low and middle classes are attracted by malls that target high-income customers. We find that, indeed, there is social mixing—but not for everyone. Then we ask whether social mixing does play a role in mall selection; i.e., if people might choose a mall partly motivated by the opportunity of mixing with people of other socio-economical classes. We will focus our work in Chile, were we used Data Detail Records (XDRs) provided by Telefónica R&D to analyze the mobility patterns of people going to malls in its capital city, Santiago de Chile, in order to determine which factors influence mall choice and what types of social mixing can be found in malls. In order to understand the factors behind this selection, we fitted a gravity model of human urban mobility which describes the aggregated mobility flows of people from different city areas to malls in terms of population, mall size, and distance. We observed that this model can describe the socio-economic distribution of customers visiting each mall. After fitting the gravity model, we turned into a more individual model of mall selection in which we fitted the probability of a customer visiting some mall B given that they had visited mall A. From this model we observed that individual selection is not only conditioned by distance and mall size, but also by the similarity between malls; i.e., people visiting a high target mall will most probably visit another high target mall rather that a low-middle class mall. However, we observed that low and middle-class people do mix in shopping malls around the periphery of the city, but few of these people reach distant malls targeting the highest-income population.

To the best of our knowledge, this is one of the first studies of malls in Latin America using a large number of people by means of CDR (Call Detail Records) or XDR data. Our results suggest that social mixing and inclusion do happen in certain malls, and the results can be used to build models that will promote social inclusion by analyzing the factors that improve it.

2 Related work
General effects of malls in urban societies in the U.S. have been extensively studied in the literature of the social sciences. Among the most prominent works, L. Cohen studied their emergence in the context of suburbanization [12]; J. Goss depicted them as new civic spaces that manipulate behavior through the configuration of space, in order to provoke dispositions and facilitate consumption [6]; Wakefield and Baker have applied factor analysis to identify the determinants of customer excitement [13]. Reilly measured the effects

of mall size and distance on mall selection and proposed the "Law of Retail Gravitation" [14]: he suggested that larger retail centers present a higher attraction to customers, who would be willing to travel longer distances to arrive to them; Huff studied the concept of trading area and proposed a model for defining the catchment area of a retail center [15].

Currently, large shopping malls in the U.S. are suffering a decline since 2005, partly due to the increase of online retail, but also to the proliferation of convenience and neighborhood retail as part of the new Smart Growth movement that is reshaping many US cities [9]. Supporters of Smart Growth promote a fine-grained scale integration of housing, jobs and retail spaces, suggesting that mixing land uses and prioritizing open spaces in neighborhoods will promote inclusion and social mixing [16].

Several authors have studied how urban design for social encounters can contribute to reduce inequality. Among them, urbanist R. Sennett pointed out that cities should be designed in order to maximize encounters between people, and that social patterns of inequality might be altered by changing the physical landscape [17, 18]. In this line, Louail et al. suggested that it is possible to increase spatial equity in cities by redistributing shopping destinations, while conserving certain fundamental properties as travel distances at the same time [19]. Hristova et al. measured the socio-spatial diversity in London, observing that malls are among the top places bringing strangers together, supporting our hypothesis that malls are potential social mixers [20]. Finally, De Nadai et al. analyzed diversity in cities using mobile phone data, and they observed that cities that are more diverse tend to be more prosperous [21].

In a Latin-American context, A. Dávila analyzed the effects of the access to shopping malls for the new middle class [22]. In a case study in Bogotá (Colombia) she observed that, instead of reducing social differentiation, malls reproduce the cultural hegemony of a living traditional elite. Stillerman and Salcedo studied how people use and understand malls in Santiago de Chile, observing cross-class encounters in two malls in the periphery of the city [10]. He concluded that people use malls to satisfy different types of needs: from purchasing to self-expressing and keeping familiar, romantic and social relationships. Dinzey-Flores analyzed the antagonism between malls and poverty in Puerto Rico [23], which has also been seen (more violently) in Brazil [24].

Retail location has also been studied in relation to urban mobility: In [25], Cervero and Duncan have studied the impact of workplace proximity, housing, and retail accessibility on mobility, showing that job-housing proximity has a larger impact on total mobility than retail location. On the other side, Galati and Greenhalgh studied the inner mobility, contact duration and inter-contact time of mall customers [26]. In the context of big-data marketing, Chen et al. have used a large dataset for optimizing retail store location [27].

Different human mobility models have been proposed in the literature to describe the collective flows of people for commuting, migration or tourism: being the gravity model and the radiation model the most used. The gravity model of flow was proposed by W. Alonso [28] for modelling spatial flow phenomena between regions. It has been extensively used in the literature for modelling international air travel [29], traffic flow in highways [30], retailing [31, 32] and telecommunications across cities [33], among others.

The more recent radiation model [34] has proven to outperform the gravity model for commuting flows across cities. As an advantage it does not require real mobility data for fitting any parameters, as the flows are computed entirely from population distributional

data. However, its performance has been shown to be limited for intra-urban mobility [35, 36].

In this work we shall model the movements of people living in different areas of the city of Santiago to a set of large malls using the gravity model. Our selection is based on several reasons: first, the gravity model has a good performance at the intra-urban level; secondly, the same gravity law has been used for estimating the attractiveness of retail centers and the flow of economic goods between countries [37]. In particular, Reilly's law of retail gravitation models the attractiveness as directly proportional to the size of the retail center and inversely proportional to the square root of its distance [14]. This law has been generalized later in terms of the gravity model [38]. However, Reilly's model does not take cultural differences into account: it might be the case that people from certain social groups prefer certain types of malls instead of others. One of our hypothesis is that cultural differences might play a crucial role in the social mixing that emerges in malls. If that were the case, then we should see that social mixing cannot be exclusively described by distances, mall sizes, and Reilly's law. We will assess the performance of a generalized gravity model for describing the influx of people to malls, with and without considering a social attraction factor.

3 Methods

In this section we will introduce the three methods we will use in this article to measure social mixing, study the attraction to malls and cluster them according to their costumer profiles.

3.1 Quantifying social mixing

Our motivation is that malls promote social mixing in terms of co-locations of people from different socioeconomic status in the same spaces. This should be reflected in the choice of malls, traditionally understood as distance-based. If social mixing is an important factor when choosing which mall to visit, then, if two malls are at equivalent distance, then the one with highest social mixing should be preferred. To be able to test this, first, we need to quantify social mixing. We do so by analyzing segregation.

Several models of segregation exist, yet, recently Louf and Barthelemy [11] proposed a new model that considers a null model of spatial segregation, where the exposure of a population α to a population β is defined as:

$$E_{\alpha\beta} = \frac{1}{N_\alpha} \sum_{m=1}^{M} n_\alpha(m) r_\beta(m) \tag{1}$$

with

$$r_\beta(m) = \frac{n_\beta(m)/N_\beta}{n(m)/N}, \tag{2}$$

where m is the mall index, N_α is the total number of people belonging to category α, $n_\alpha(m)$ is the total number of visitors to mall m belonging to category α, $r_\alpha(m)$ is the *representation* of category α in mall m as computed by equation (2), $n(m)$ is the total number of visitors to mall m, and N is the total number of people. The exposure metric is interpreted as

follows: if $E_{\alpha\beta} > 1$, then social mixing happens between subpopulations α and β within malls. Conversely, if $E_{\alpha\beta} < 1$, then both categories are segregated.

To define visitor categories, we bin users into percentiles according to their socio-economic characteristics. These percentiles are based on the Human Development Index, or HDI. This index, when not directly available, can be estimated with a method proposed by the United Nations [39]. Its formula includes income distribution, life expectancy and education. As such, the value of E depends on the distributions of the percentiles of the HDI of its visitors, according to their home place.

While E is a measure of segregation/mixing, we still need a measure of the social diversity within a specific mall. To this end, we use the Shannon Entropy S_m with respect to the percentiles of HDI of its visitors:

$$S_m = \sum_{q=1}^{Q} p_q \log p_q, \tag{3}$$

where S_m is the entropy of mall m, and p_q is the fraction of visitors to m that belong to HDI percentile q.

Note that, by definition, the representation term in the model compares the relative population that visit a mall to the expected value in an unsegregated city [11]. This implies random interactions with respect to social status. As found in earlier work, mall visits are strongly influenced by distance, and thus, interactions may not follow a random pattern. Then, we shall compare the observed social mixing against a null model in which visitors always choose their nearest mall.

3.2 A gravity mobility model for mall visits

The gravity model of flow has been extensively used to model human mobility in different contexts [28, 29, 33], and in particular for retailing [31, 32]. It considers the flow between two nodes (i, j) as directly proportional to some powers of their populations, and inversely proportional to some power of the distance between them:

$$F_{ij} = G \frac{M_i^{\alpha} M_j^{\beta}}{D_{ij}^{\gamma}}, \tag{4}$$

where G is a proportionality constant, M_i is the population of a square grid in the city, computed from census data, M_j is mall size in terms of total rental space, and D_{ij} is the distance between the center of the square grid and the mall.

The traditional approach for fitting this model consists of applying a logarithmic transformation, leading to a linear model on the logarithms of the variables:

$$\log(F_{ij}) = \log(G) + \alpha \log(M_i) + \beta \log(M_j) - \gamma \log(D_{ij}) + \epsilon_{ij}, \tag{5}$$

where ϵ_{ij} represents an additive, independent error term. This linearized model can be fitted through OLS (ordinary least squares) [40]. However, this approach has several limitations: it cannot model the zero observations (which must be thrown away), and the estimated coefficients can have significant biases under heteroskedasticity [41]. As an alternative, we replace the linear regression by a Generalized Linear Model (GLM) [42])

with a Negative Binomial distribution for count data:

$$\mathbf{E}[F_{ij}] = \exp\big[\log(G) + \alpha \log(M_i) + \beta \log(M_j) - \gamma \log(D_{ij})\big]. \tag{6}$$

This Negative Binomial GLM is fitted by maximizing the log-likelihood function. The maximization of this function does not have a closed analytical solution, but as the function is convex convergence is guaranteed by applying standard optimization techniques such as gradient descent or iteratively reweighted least squares (IWLS).

To account for the social mixing factor, in addition to the baseline gravity model we shall consider a distance-modulated model, where the distance to a mall varies according to its social diversity: malls with higher entropy appear as closer. This model is specified as:

$$\mathbf{E}[F_{ij}] = \exp\left[\log(G) + \alpha \log(M_i) + \beta \log(M_j) - \gamma \frac{\log(D_{ij})}{S_j}\right], \tag{7}$$

where S_j is the social diversity (entropy) of mall j according to its distribution of HDI percentiles (Eq. (3)). This modulated distance would allow to differentiate malls that, with respect to a visitor, are within the same distance, but with different social mixing properties.

3.3 Clustering malls according to customer profiles

In order to better understand the motivations behind mall selection, we built a co-visitation network representing common mall customers. This is a weighted directed network whose nodes v_i are the 16 malls, while the weighted edges (v_i, v_j) between them represent the conditional probability of visiting mall v_j given that someone visited mall v_i. We built a similarity matrix S between malls using the Kolmogorov–Smirnov S_{ij} distance between the customer profile distributions of malls. We then built a Logit model for describing the conditional probability that a customer visits mall v_j given that they also visit mall v_i. We fitted this model using a logistic regression:

$$\mathbf{E}[p_{j|i}] = \big(1 + \exp\big[-\log(G) - \beta \log(M_j) - \lambda \log(S_{ij}) + \gamma \log(D_{ij})\big]\big)^{-1}, \tag{8}$$

where G is a parameter representing the *odds* of the event in the logistic regression.
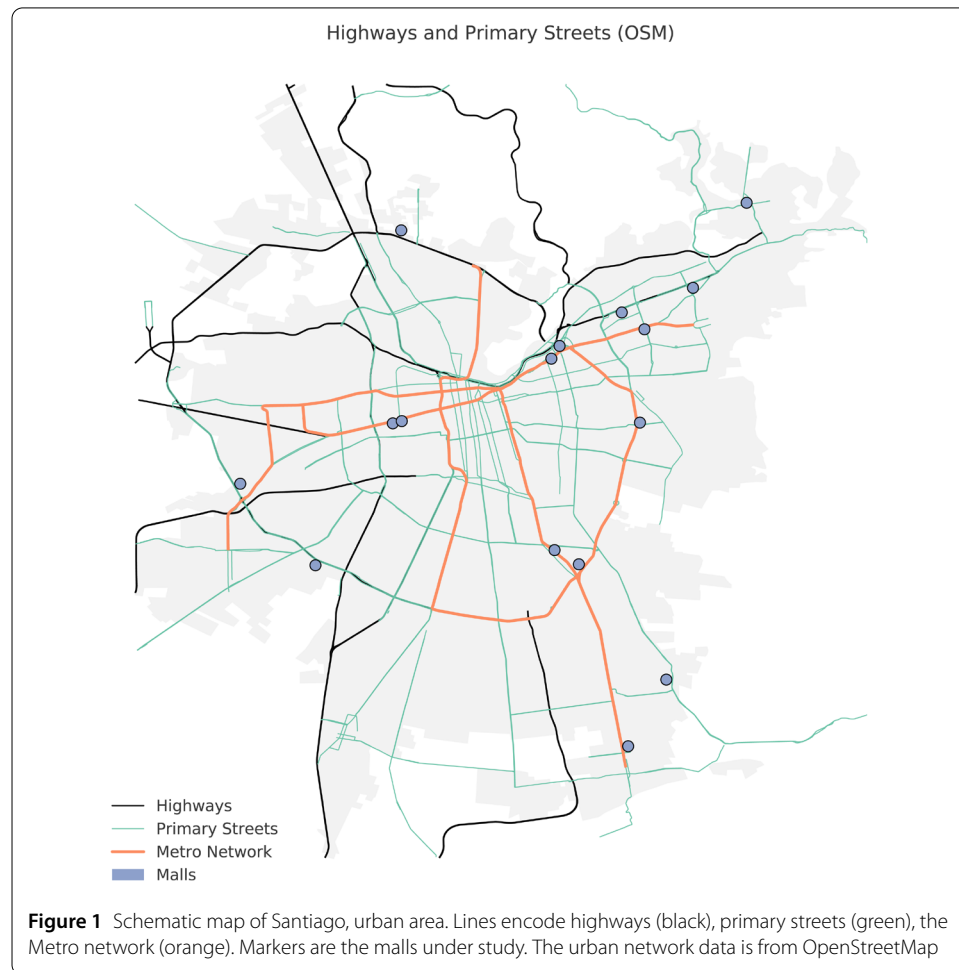
4 Datasets and data preprocessing

Our study focuses in 16 malls located in the urban area of Santiago, Chile [3]. Table 1 shows the rental space of these malls, and Fig. 1 shows them in the urban context, in terms of the urban network of streets and transportation services.
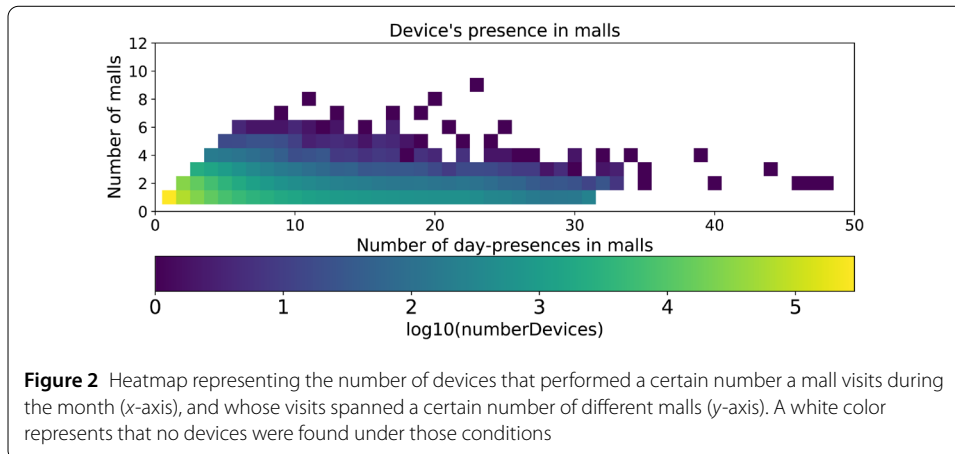
The Santiago urban area is comprised by 35 *"comunas"*,[a] containing 13,239 cellphone antennas, distributed in 1377 towers. Some antennas are located indoors, such as in underground metro stations, important public and private buildings, and malls, thus reflecting the importance of the latter in the city culture. As indoor mall antennas are usually low power, connections to them are almost surely established from inside the mall. It is very improbable that a passerby can connect to an inner antenna from outside, as it is also improbable that a mall visitor will establish a connection to some antenna outdoors. Antennas inside malls were identified in two different ways: (i) by means of a description

Table 1 Malls under study in Santiago, with their size in square meters (according to Chile's Chamber of Commerce), and number of unique visitors in our dataset during August, 2016

Mall	Rental Space (sqm)	Visitors
Alto las Condes	115,000	122,061
Apumanque	20,100	36,154
Mall Arauco Maipú	54,000	94,788
Mall Costanera Center	173,000	163,234
Espacio Urbano Puente Alto	25,500	77,153
Mall Florida Center	110,000	145,436
Mall Vivo Panorámico	7000	69,646
Mall Parque Arauco	111,000	170,494
Mall Paseo Estación	46,700	59,056
Mall Plaza Alameda	60,000	26,478
Mall Plaza Egaña	80,000	58,239
Mall Plaza Norte	74,900	127,768
Mall Plaza Oeste	144,000	148,716
Mall Plaza Tobalaba	54,200	2070
Mall Plaza Vespucio	116,000	96,523
Portal La Dehesa	79,000	29,268

Figure 1 Schematic map of Santiago, urban area. Lines encode highways (black), primary streets (green), the Metro network (orange). Markers are the malls under study. The urban network data is from OpenStreetMap

field in the dataset; (ii) by drawing each mall's polygon and using the towers' coordinates to determine which ones were inside the polygons.[b] Both methods turned out to be highly consistent. Our selection process yielded 481 indoor cellphone antennas placed *inside* the

Figure 2 Heatmap representing the number of devices that performed a certain number a mall visits during the month (*x*-axis), and whose visits spanned a certain number of different malls (*y*-axis). A white color represents that no devices were found under those conditions

malls of interest, which represent 29% of all indoor antennas in the Santiago urban area, highlighting the important role of shopping malls in Santiago.

The XDR dataset consists of 1,023,118 unique anonymized mobile devices (containing a Movistar/Telefónica SIM card) triggering events with the indoor mall antennas and interacting with the Telefónica network in the course of one single month (August 2016). XDRs (rather than the more common CDRs or Call Detail Records) store interactions for, for example, accesses to websites, applications that communicate over the Internet [43]. As such, XDRs only have one communicating antenna, (CDRs have two instead: one for the caller and one for the callee), and the cost is measured in down/upload units (bytes) since the last registered event rather than minutes consumed. Given the prevalence of data rather than voice-driven communication nowadays (interacting with the internet is more common than calling someone), XDRs allow much more time granularity, at around 15–30 minutes between logged events. Calls are more sparse. Each XDR record in our dataset contains information logged by Telefónica in Santiago during that month, including latitude/longitude pairs, hashed phone number, time-stamp, and bytes downloaded [44].

Figure 2 shows a heatmap of the number of devices that were found as visiting a certain number of different malls during the month (*y*-axis), and totalizing a certain number of visits (*x*-axis), a visit being defined as the presence of the device inside a mall during a specific day. For the 81,027 devices with more than 10 day-presences in malls, we conservatively determined that they were not mall customers and discarded them from the dataset (the are probably related to SIM cards associated to card readers, and to people who are not visitors, such as employees or providers). We identified 942,091 devices (of the original 1,023,118) as "real visitors." The remaining ones might be associated with staff, mall providers or utility devices such as credit card readers, points-of-purchase, *etc.* The total number of visitors per mall is shown on Table 1.

For those devices identified as visitors we determined the area where they live (the "home antenna") by finding the first (before 8AM) and last (after 10PM) cell tower they were connected to each day of the month. We then computed, for each user: (i) The number of times in which we detected a first/last tower connection, and (ii) the relative frequency of the most observed tower (i.e., the number of times in which the user's first or last tower was that tower, divided by the previous quantity). Results show that more than 70% of the devices were connected at least 80% of days. About half the users repeated
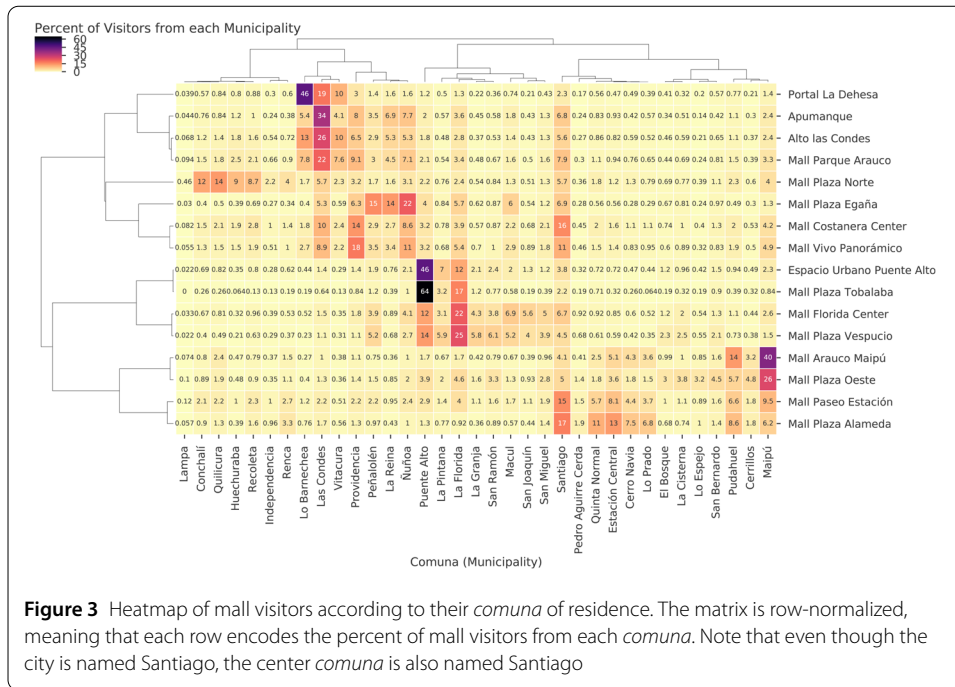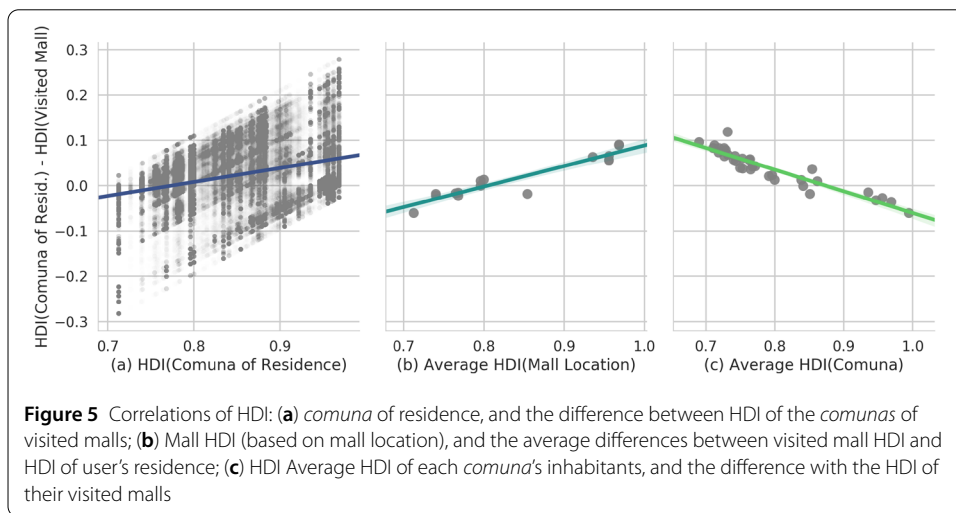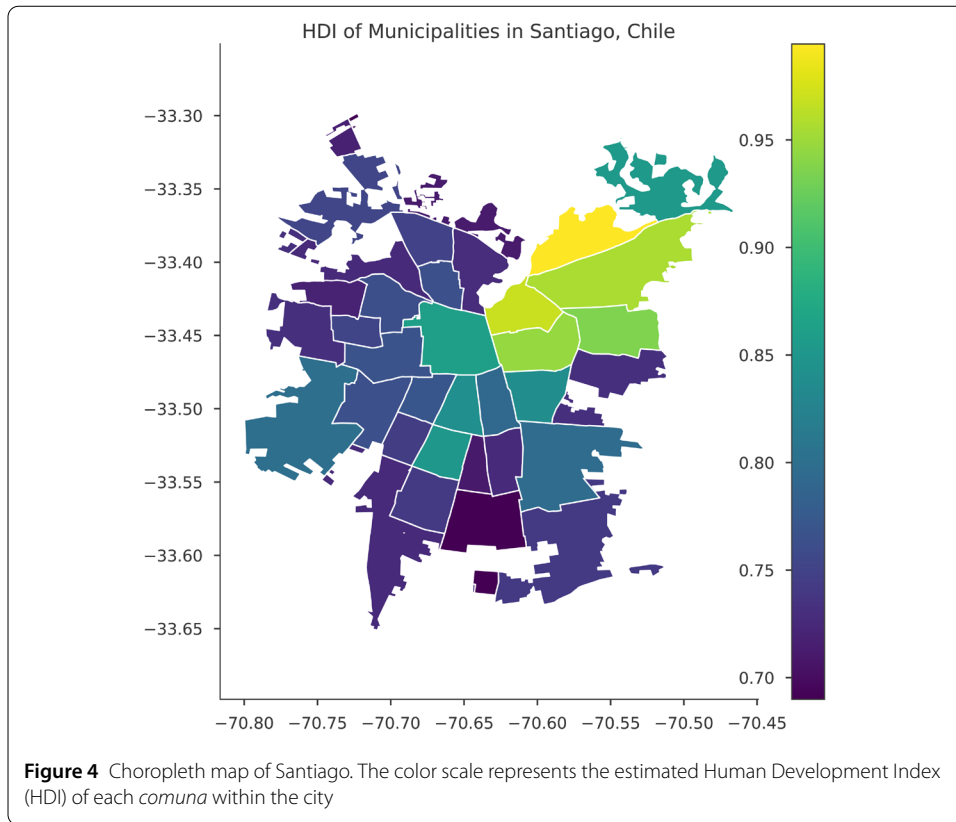
Figure 3 Heatmap of mall visitors according to their *comuna* of residence. The matrix is row-normalized, meaning that each row encodes the percent of mall visitors from each *comuna*. Note that even though the city is named Santiago, the center *comuna* is also named Santiago

their most common first-last tower at least 60% of the days in which they are seen. In order to ensure a good confidence for the home location determination, we only kept these last consumers (387,152 unique devices). A Pearson correlation coefficient was computed to asses the relationship between our home-antenna identification approach and the actual population of those *comunas*, given by the 2017 census information [45], obtaining a positive correlation of 0.84.
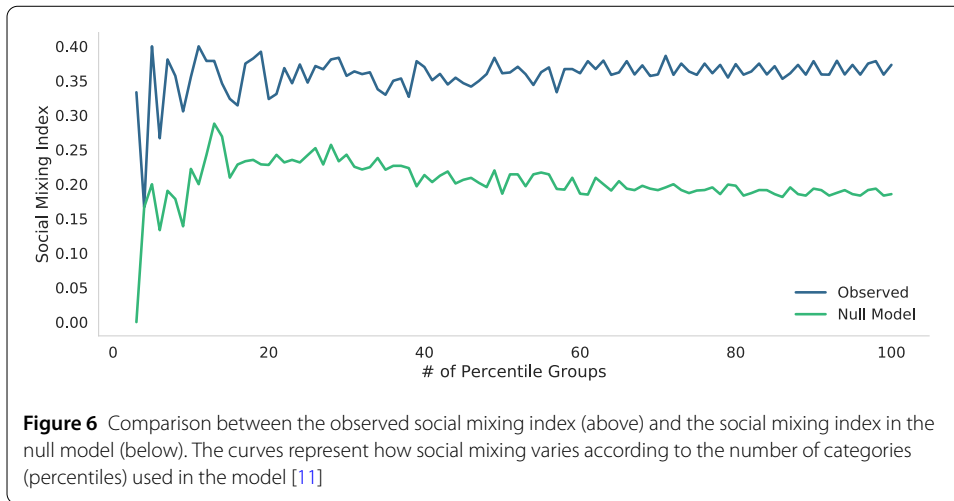
The heatmap in Fig. 3 shows the distribution of the *comunas* of residence of mall visitors. One can see that several malls are tailored at specific *comunas* due to the high fraction of visitors from one or two (*e.g.*, Mall Arauco Maipú exhibits a high fraction of visitors from Maipú; Mall Plaza Tobalaba exhibits a high fraction of visitors from Puente Alto), while others exhibit high diversity (*e.g.*, Mall Plaza Norte, Costanera Center, Panoramico, Mall Plaza Vespucio, *etc.*). This implies that, in terms of the *comuna* of residence, there may be social mixing in some malls.

A further dataset needed for the purposes of this work was a measure of the Human Development Index, or HDI. Unfortunately, in Chile, the last formal/official calculation of the HDI at the *comuna* level was done in 2005 [46]. To make it more up to date we computed the HDI for Chile for the period 2013-2015, following the guidelines in [39].[c] We correlated our results against those of 2005, obtaining a Pearson correlation of 0.90. We use our HDI calculation in the rest of this paper. Figure 4 shows the geographical HDI distribution, putting into evidence the socio-economic segregation of the city.

One direct way to explore social mixing is to aggregate mall visits per user, estimate the difference between HDI of visited mall (in terms of a mall's location) and HDI of *comuna* of residence, and then calculate the correlation of those differences with the HDI of each *comuna*. Figure 5 shows three correlations of such differences: (a) individual differences (Pearson's $\rho = -0.49$, $p < 0.001$); (b) aggregated per visited mall ($\rho = 0.65$, $p < 0.01$); and

Figure 4 Choropleth map of Santiago. The color scale represents the estimated Human Development Index (HDI) of each *comuna* within the city

Figure 5 Correlations of HDI: (**a**) *comuna* of residence, and the difference between HDI of the *comunas* of visited malls; (**b**) Mall HDI (based on mall location), and the average differences between visited mall HDI and HDI of user's residence; (**c**) HDI Average HDI of each *comuna*'s inhabitants, and the difference with the HDI of their visited malls

(c) aggregated per *comuna* of residence ($\rho = -0.9$, $p < 0.001$). In all cases, the difference was weighted by the number of days each user visited a mall. These results support our motivation: on the one hand, users have a tendency to visit malls in areas with higher HDI than their residence's; on the other hand, malls tend to receive visitors from areas with lower HDI. However, the *comuna* level is too coarse to bring conclusions in this matter. In the next section we evaluate exposure to people with different HDI using the theory described in the methods section, at a finer granularity.

Figure 6 Comparison between the observed social mixing index (above) and the social mixing index in the null model (below). The curves represent how social mixing varies according to the number of categories (percentiles) used in the model [11]

Table 2 Results of quantization of HDI values of users into quantiles

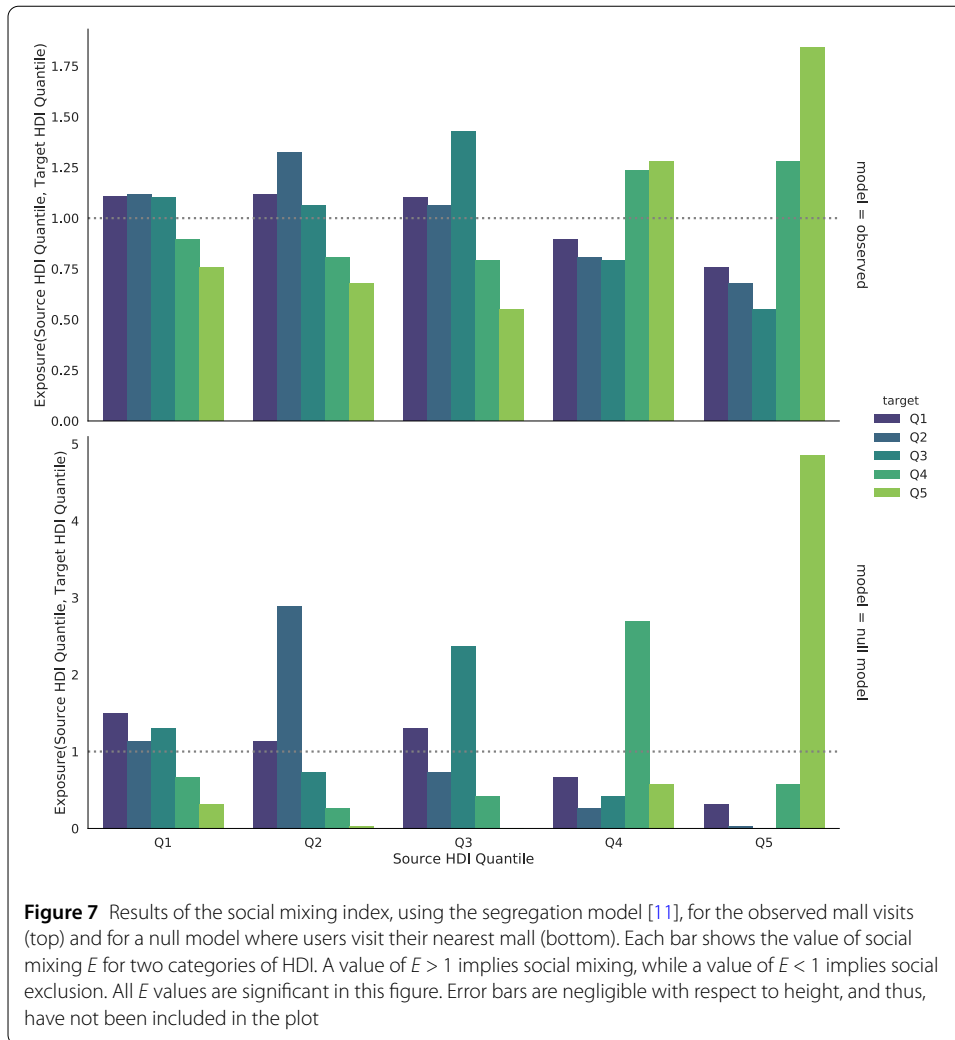| Quantile	HDI Range	# of Users
Q1	(0.689, 0.732]	56,863
Q2	(0.732, 0.764]	50,335
Q3	(0.764, 0.8]	52,921
Q4	(0.8, 0.946]	57,585
Q5	(0.946, 0.995]	48,164

5 Results

5.1 Social mixing seen through a segregation theory

Since the model by [11] requires the definition of categories, first we defined the number of percentiles to include in the analysis. We evaluated the social mixing model for all percentiles in [3, 100]. For each evaluation, we estimated a Social Mixing Index (SMI), defined as the fraction of pairwise exposures that were significant according to the model. In formal terms:

$$\text{SMI}(i_p) = \frac{|(\alpha, \beta)\ \forall E_{\alpha\beta} > 1, \alpha \neq \beta, E \text{ is significant}|}{|(\alpha, \beta)\ \forall \alpha \neq \beta|},$$

where i_p is the number of percentiles, α and β are two different percentiles, and E is the social mixing model (Eq. (1)). Figure 6 shows the results, where it can be seen that the model is fairly stable with respect to the number of percentiles. The mean value of SMI is 0.35, with a standard deviation of 0.03. As result, we resort to five percentiles, or quantiles (SMI(5) = 0.4). Quantiles are commonly used in the socio-economic literature, allowing for comparison of our results to similar data. The Figure also shows the behavior of the null model considering that users visit their nearest mall. This null model shows less social mixing than observed (in average, the observed SMI is 57% $\pm$ 10% higher than that of the null model).

To estimate quantiles, we binned users according to their HDI distribution, assigning to each user the HDI of their *comuna*. Table 2 shows the results. Then, we evaluated $E_{\alpha\beta}$, where α and β were the quantiles. In Fig. 7, Top, which displays the several values of $E_{\alpha\beta}$ for all quantile pairs, one can see that there is varying behavior in the value of E (all E values are significant with 99% confidence according to the variance estimation in [11]). We observe

Figure 7 Results of the social mixing index, using the segregation model [11], for the observed mall visits (top) and for a null model where users visit their nearest mall (bottom). Each bar shows the value of social mixing E for two categories of HDI. A value of $E > 1$ implies social mixing, while a value of $E < 1$ implies social exclusion. All E values are significant in this figure. Error bars are negligible with respect to height, and thus, have not been included in the plot

that the first three quantiles (Q1, Q2, and Q3) tend to have a value of $E > 1$ between them, and $E < 1$ with the other two quantiles (Q4 and Q5). However, the opposite trend happens for quantiles Q4 and Q5. When $\alpha = \beta$, the E metric is called the *isolation* of α: I_α. A value $I_\alpha > 1$ implies that the α groups tends to co-locate with people from the same group. Three groups show high values of isolation: Q2, Q3 and Q5. Q5 is the most interesting case, as it presents the highest isolation of all, low values (*e.g.*, exclusion) with the lower three quantiles. These results imply that social mixing happens, but with restrictions—people co-locate with others who are in "nearby" quantiles of HDI, indicating that there is social mixing but that the economic segregation of the city plays an important role.

Figure 7, bottom, shows the behavior of the null model, where it can be seen that all quantiles are over-exposed to themselves, and, in case of inclusion, it is infrequent. This motivates the hypothesis that it is not only distance that drives mall visits, and, given that the observed social mixing exhibits a prominent positive difference with the null model, it could be an important factor in the gravity model.

5.2 Fitting the gravity model

We fitted the gravity model of flow in equations (6) and (7) between areas of the city and malls. Since there is social mixing, and our hypothesis states that people look for this co-location, we hypothesized that introducing the malls' entropy into the model should produce a significant effect.

The gravity model will describe aggregated trips between source nodes (the user's residence) and destination nodes (the malls). Even though the HDI of the mall visitors was defined at the *comuna* level, in order to fit and evaluate the model we needed a finer spatial granularity than that of *comunas*. To do so, we divided the city into a set of squares confirming a regular grid. We assigned the antennas into the squares by clipping their latitude and longitude to two decimals. Thus, each square grid has a ratio of 1.1 km and might aggregate many antennas. We finally represent each square by an "abstract" antenna located in its center, which will be the source point for all trips starting in its real antennas. In turn, the trip destination points are the centroids of the mall polygons mentioned in Sect. 4.
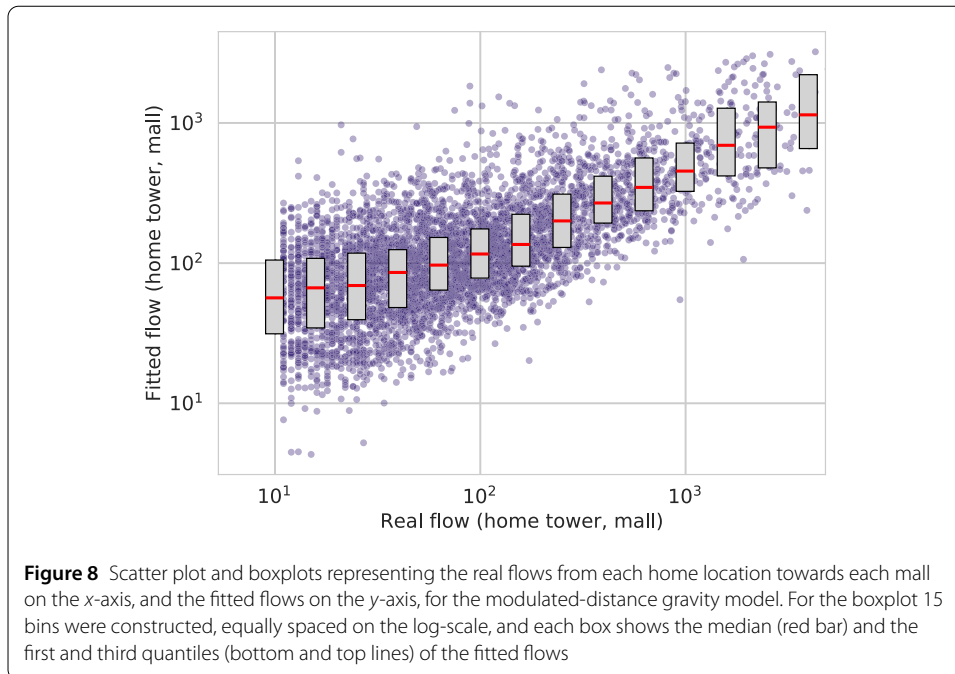
The gravity model requires the definition of masses for all source and destination points. Regarding the former, the population of each square in the grid is estimated using data from the 2012 Census at the level of census zones (census zones are aggregations of typically a few blocks). On the other hand, the "mass" of a given mall was defined as its total rental space (data taken from the Chamber of Commerce of Chile [47]), according to the ideas in Reilly's law [14] and Huff's model [15].

Using the XDR dataset, the model was calibrated by extracting the number people living at each abstract antenna C and visiting each mall M. As discussed above, a person is said to visit mall M if that person has been there at least once during the day, with the provisos made above to rule out non-visitor events. Likewise, we assume that a visitor lives in an abstract antenna C when that person's home antenna is within the boundaries of the square represented by C.

The results of the Negative Binomial regression for our baseline gravity model (Eq. (6)) showed that all the regressors are significant according to their t-values. The multiplicative constant is the least significant one ($t(G) = 6.00$, $p < 0.001$), while distance is the most significant one ($t(d_{ij}) = 54.84$, $p < 0.001$). In particular, from the fact that $\gamma/\beta = 3.4$ we observe that our results do not obey Reilly's law of retail gravitation, which suggested that $1.5 \leq \gamma/\beta \leq 2.5$. In the modulated-distance gravity model including the malls' entropy (Eq. (7)) all regressors are significant too, with $t(d_{ij}) = 57.45$, $p < 0.001$. All regression coefficients are positive, implying that there is a negative effect of distance and a positive effect of mall social diversity on mall election.

We compared both models' performance and found that the modulated-distance gravity model produced a significant improvement with respect to the baseline model in terms of a lower BIC (Bayesian Information Criterion) value (Δ BIC = $BIC_{md} - BIC_{base} = -36{,}992 - (-36{,}880) = -112$). This implies that people are more attracted by malls with higher social mixing customer profiles, and we can measure this social attraction in terms of the entropy of the mall. The plot in Fig. 8 shows a comparison between the real and fitted flows of customers from different squares to the malls in the distance-modulated gravity model.

Finally, we are interested in comparing the differences in the profile of customers visiting each mall in the real dataset and according to the flows in the fitted model. We show these differences in Fig. 9 by plotting a Kernel Density Estimation (KDE) of the HDI indexes of the mall customers in the real data and in the fitted data.

Figure 8 Scatter plot and boxplots representing the real flows from each home location towards each mall on the *x*-axis, and the fitted flows on the *y*-axis, for the modulated-distance gravity model. For the boxplot 15 bins were constructed, equally spaced on the log-scale, and each box shows the median (red bar) and the first and third quantiles (bottom and top lines) of the fitted flows

5.3 Co-visitation probability model

Figure 10 shows the network of co-visitations between malls, where the edge-width represents the weight w (the graph is thresholded for $w \geq 0.10$). The network shows that co-visitation is strongly regulated by distance, but we wonder whether people tend to choose similar malls in terms of profile too. In order to run the co-visitation Logit model we computed the similarity between malls as explained in Sect. 3. This similarity matrix is visualized in Fig. 11. The malls in this Figure were sorted according to the 3 groups that we found by applying a spectral clustering algorithm to the similarity matrix. The cluster (Alto las Condes, Apumanque, Portal La Dehesa, Mall Parque Arauco) is composed by high-target malls, only accessible to people from areas with high HDI. The cluster (Mall Costanera Center, Mall Plaza Egaña, Mall Vivo Panorámico) is composed by three malls in downtown with high mixing level. Finally, the remaining large cluster is composed by 9 malls receiving customers from low and middle classes.

Fitting the co-visitation probability regression model in Eq. (8) we observed that this probability diminishes with distance, as well as it increases with the similarity between malls. A comparison between the real and fitted co-visitation probabilities is shown in Fig. 12.

Figure 13 shows a KDE of the joint distribution of customer HDI and mall HDI (calculated with our method above) for each mall visit. The figure shows a clear effect of visitors of middle-lower classes visiting middle-lower class malls (represented by the high peak in the lower-left corner). Instead, people from places with the highest HDI's visit high-target malls (peak in the top-right) but also intermediate malls.

6 Discussion

The power of malls to *attract* people from diverse social classes may be explained by several factors: every mall, for example, may have particular target demographics; the time it takes to get to the mall may be such that mall-goers usually choose malls that are easier

Figure 9 Kernel Density Estimation of the HDI index distribution of customers visiting each mall, according the XDR data (purple) and to the fitted gravity model in terms of the influx of people from different *comunas* (orange). KDE bandwidth was set to 0.15
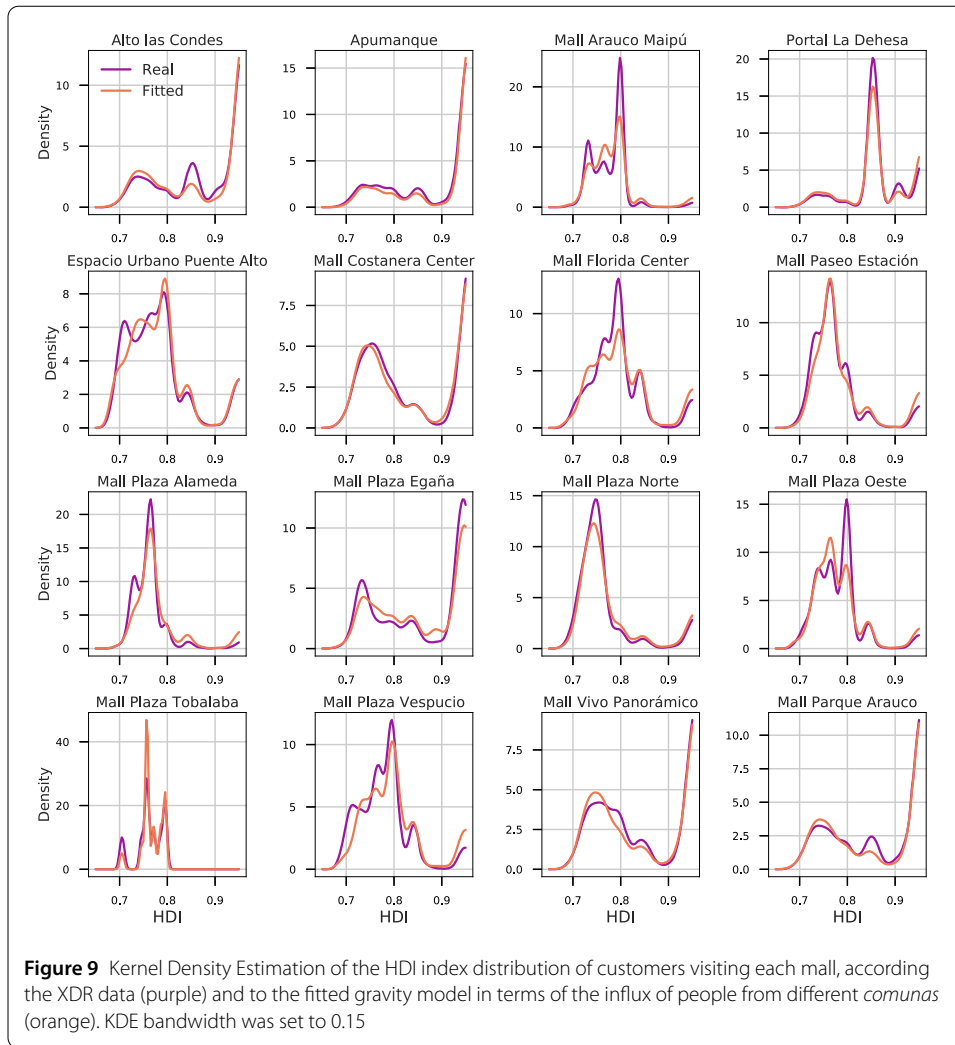
to access; or, simply, larger malls (in terms of, for instance, leased space) will attract more people.

With cellphone data, and given our adjusted gravity model of mall visitation, we have found that the number of people visiting malls can be described in terms of the size and distance to the mall, and that malls with higher mixing attract more people.

Our model cannot determine whether this attraction itself is derived from such diversity. One potential explanation is that diversity is a consequence of accessibility, as the transportation network in Santiago makes some malls easier to reach than others, adding another factor that may modulate distances from a visitor's perspective. The co-visitation probability model (discussed in Sect. 5.3) provided some evidence that the issue is more complex: given the visits to a specific mall, visits to other malls are influenced by the socio-economic target of the former. This is also expressed in the clustering of malls by customer profile, and lends some evidence to the idea that people have social preferences in their mall choices.

These results show that malls have potential as places for social mixing, and urban planners could take this into account at the time of designing common spaces to promote diversity in the city, with the positive effect evidenced in the literature [19, 21]. However,
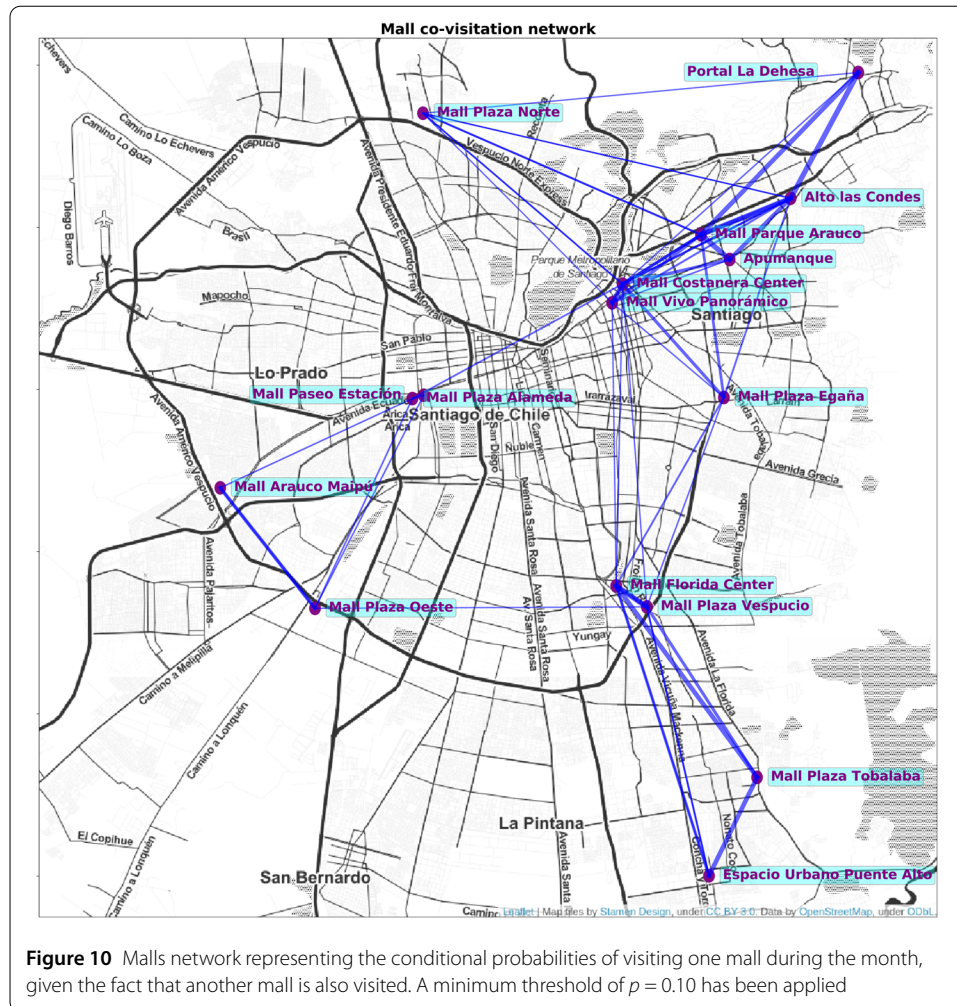
Figure 10 Malls network representing the conditional probabilities of visiting one mall during the month, given the fact that another mall is also visited. A minimum threshold of $p = 0.10$ has been applied
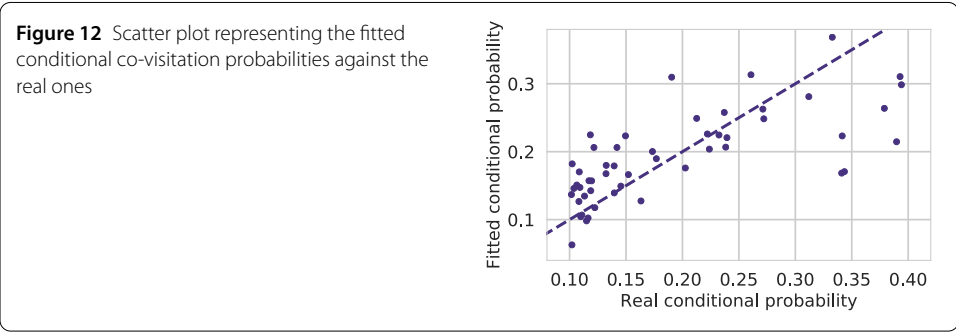
while people might be motivated by the idea of social interaction, urban mobility will likely be a limiting factor. Places may designed with accessibility in mind, and policies could be enacted to regulate where malls are located, what they offer and how they can be accessed in order to maximize its potential as a democratic space.

6.1 Scope and limitations

As with most C/XDR-based research, there are several limitations that should be acknowledged. Perhaps one of the most important ones has to do with the nature of the dataset itself. Telefónica owns about a third of the Chilean cellphone marketshare. Although the largest, with the second one Entel at about 20%, this introduces some biases that are hard to identify and quantify. For instance, in its recent history, Telefónica invested heavily on 4G technology, biasing the population of clients towards a higher socio-economic segment. This might be biasing the results obtained and discussed above, in terms of geographic localization, with higher numbers of interactions coming from richer *comunas*.
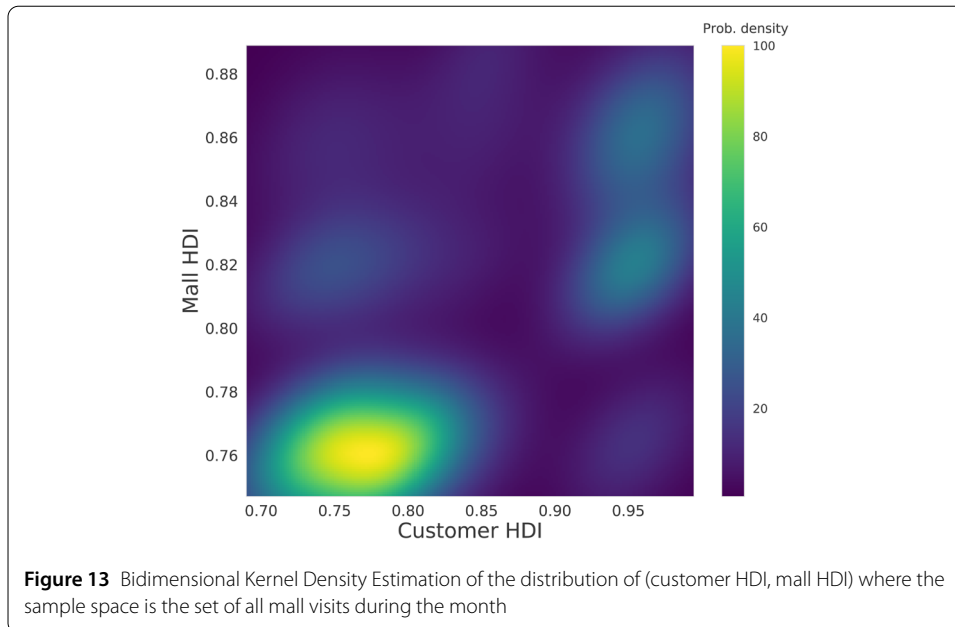
The calculation of HDI used in this work comprises an integration of several data sources. Most importantly, however, some of these sources belong to different periods of time, the oldest being from 2013 while the most recent being 2015. Thus, the HDI here is a time-range rather than a snapshot of one particular year. We do not consider this to

Figure 11 Similarity matrix between malls. The similarity between a pair of malls was computed as the Kolmogorov–Smirnov distance between the customer profile distributions of the malls. Malls were clustered into 3 groups using spectral clustering

Figure 12 Scatter plot representing the fitted conditional co-visitation probabilities against the real ones

be a major issue here because there has not been a sizeable shift in any of the dimensions measured by the HDI (like education, income or mortality rates) in the *comunas* of Santiago (or all of Chile for that matter) in the period under study. It is, nonetheless, important to mention it here.

Another important consideration is that we analyzed co-locations within the same space. Co-location does not necessarily imply face-to-face interaction, yet, it provides a venue for *potential* face-to-face interactions [11]. Critics may rightly say that this is a limitation of the study, however, recent findings acknowledging that co-location can be used as proxy for face-to-face interactions for certain network metrics [48]. Hence, more than a limitation of our work, we believe that future work should analyze what elements may transform *potential* interactions into *actual* ones.

Figure 13 Bidimensional Kernel Density Estimation of the distribution of (customer HDI, mall HDI) where the sample space is the set of all mall visits during the month

Lastly, a comparison between malls and commercial neighborhoods/streets should be made. That is, there are commercial neighbourhoods in Santiago where social classes might mingle more straightforwardly (such as Patronato [49] and others). As discussed, malls have a considerable proportion of indoor antennas, which allows us to identify mall visits with a higher degree of certainty, while, conversely, commercial streets may have false positives, *e.g.*, people that pass through but that do not *visit* the place as destination. This is a much more difficult problem that falls out of the scope of this paper.

7 Conclusion

At least in Latin America, and in Chile in particular, shopping malls have an important cultural place. Many people visit these places not only for the goods they carry, but to eat, for leisure and in general to interact with other people as well. Usually, poorer people go to malls because they are free to enter and open to all, quite safe, and ultimately a nice, comfortable place to spend time in.

In this work we looked into the nature of malls as "social mixers", analyzing whether malls with higher mixing attract more people. We have done so using XDR data to measure mobility patterns of the people of Santiago, the capital and largest metropolitan area of Chile, where many social classes co-exist. We conclude that there is social mixing in malls, that social mixing affects mall visitation patterns, and that high-end malls can attract low-income visitors.

This work can be extended in several dimensions. First of all, it would be interesting to have a more fine-grained description of what kind of mixing happens in those malls where mixing is found (like the Costanera Center). For example, we would like to analyze whether people make a point about going to these malls as day trips during weekends or social mixing may happen only on off-business hours (on leisure time). Another dimension has to do with the role of mall workers, since they may also promote social mixing. A third dimension for future study is people behavior inside malls: since malls have (sometimes more than 100) indoor antennas, we would like to know whether people coming from

lower-income *comunas* behave differently inside malls. It will be important to find out whether they visit high-end stores, or simply go to the food-court, for example. Knowing this will help personalize the buying experience of non-target customers, perhaps being more fine-tuned to customers of different socio-economic segments and their likes and dislikes.

Using the results of this study, we expect public policies can be enacted by governments or public organizations to push social mixing, for example by improving mall accessibility or by carefully deciding where new malls should be opened. All in all, malls are an important landmark of cities, just their sheer volume has a strong impact on the mobility patterns of the people in the cities they're in. We ultimately expect to take advantage of their *attraction* properties to promote social mixing.

List of abbreviations
CDR, Call Detail Records; HDI, Human Development Index; KDE, Kernel Density Estimation; XDR, Data Detail Records.

Availability of data and materials
The Telefónica Movistar mobile phone records have been obtained directly from the mobile phone operator through an agreement between the Data Science Institute and Telefónica R&D. This mobile phone operator retains ownership of these data and imposes standard provisions to their sharing and access which guarantee privacy. Anonymized datasets are available from Telefónica R&D Chile (http://www.tidchile.cl) for researchers who meet the criteria for access to confidential data. Other datasets extracted from public sources have also been made available at the dedicated git repository https://zenodo.org/badge/latestdoi/106930683 together with the source code.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All authors contributed to the design of the study. MB, LF and EG designed the experiments. MB, EG, LF and DC performed data analysis. All authors participated in manuscript preparation. All authors read and approved the final manuscript.

Author details
[1]Facultad de Ingeniería, Universidad de Buenos Aires, INTECIN (CONICET), Buenos Aires, Argentina. [2]Data Science Institute, Faculty of Engineering, Universidad del Desarrollo, Santiago, Chile. [3]Telefónica R&D, Santiago, Chile. [4]ISI Foundation, Torino, Italy.

Endnotes
[a] We have decided to maintain the Spanish word for the administrative unit of Chile: "county" or "commune" being close translations to English, but rather artificial.
[b] The mall polygons are available at https://zenodo.org/badge/latestdoi/106930683.
[c] Data from CASEN: https://goo.gl/LuE6s3, source code at https://zenodo.org/badge/latestdoi/106930683.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Graham S, Aurigi A (1997) Virtual cities, social polarization, and the crisis in urban public space. J Urban Technol 4(1):19–52
2. Staeheli LA, Mitchell D (2006) USA's destiny? Regulating space and creating community in American shopping malls. Urban Stud 43(5–6):977–992
3. Salcedo R, De Simone L (2013) Los Malls En Chile: 30 Años. Cámara Chilena de Centros Comerciales, Santiago

4. Pradel D (2017) Chile se ubica como el país de la región con más metros cuadrados de malls por habitante. http://www.economiaynegocios.cl/noticias/noticias.asp?id=337568. [Online; accessed January 31, 2018]

5. Austin R (1997) Not just for the fun of it: governmental restraints on black leisure, social inequality, and the privatization of public space. S Cal L Rev 71:667

6. Goss J (1993) The "magic of the mall": an analysis of form, function, and meaning in the contemporary retail built environment. Ann Assoc Am Geogr 83(1):18–47

7. Wilson E (1995) The rhetoric of urban space. New Left Rev 209:146

8. Lofland LH (2017) The public realm: exploring the city's quintessential social territory. Routledge, London

9. Chen O, Kernan JBJ (2017) Equity research. retail's disruption yields opportunities-store wars! Cowen and Company

10. Stillerman J, Salcedo R (2012) Transposing the urban to the mall: routes, relationships, and resistance in two Santiago, Chile, shopping centers. J Contemp Ethnogr 41(3):309–336

11. Louf R, Barthelemy M (2016) Patterns of residential segregation. PLoS ONE 11(6):0157476

12. Cohen L (1996) From town center to shopping center: the reconfiguration of community marketplaces in postwar America. Am Hist Rev 101(4):1050–1081

13. Wakefield KL, Baker J (1998) Excitement at the mall: determinants and effects on shopping response. J Retail 74(4):515–539

14. Reilly WJ (1931) The law of retail gravitation. WJ Reilly

15. Huff DL (1964) Defining and estimating a trading area. J Mark 28(3):34–38

16. Duany A, Speck J, Lydon M, Goffman E (2011) The smart growth manual. SSPP 7(2):89–90. https://doi.org/10.1080/15487733.2011.11908078

17. Sennett R (1992) The uses of disorder: personal identity and city life. WW Norton & Company, New York

18. Sennett R (2018) Building and dwelling: ethics for the city. Farrar, Straus and Giroux, London

19. Louail T, Lenormand M, Arias JM, Ramasco JJ (2017) Crowdsourcing the Robin hood effect in cities. Appl Netw Sci 2(1):11

20. Hristova D, Williams MJ, Musolesi M, Panzarasa P, Mascolo C (2016) Measuring urban social diversity using interconnected geo-social networks. In: Proceedings of the 25th international conference on World Wide Web, pp 21–30. International World Wide Web Conferences Steering Committee

21. De Nadai M, Staiano J, Larcher R, Sebe N, Quercia D, Lepri B (2016) The death and life of great Italian cities: a mobile phone data perspective. In: Proceedings of the 25th international conference on World Wide Web, pp 413–423. International World Wide Web Conferences Steering Committee

22. Dávila A (2016) El mall: the spatial and class politics of shopping malls in Latin America. 1st edn. University of California Press, California

23. Dinzey-Flores ZZ (2017) Spatially polarized landscapes and a new approach to urban inequality. Lat Am Res Rev 52(2)

24. Goncalves AA (2014) Conflicting frames: the dispute over the meaning of rolezinhos in brazilian media. PhD thesis, Massachusetts Institute of Technology

25. Cervero R, Duncan M (2006) Which reduces vehicle travel more: jobs-housing balance or retail-housing mixing? J Am Plan Assoc 72(4):475–490. https://doi.org/10.1080/01944360608976767

26. Galati A, Greenhalgh C (2010) Human mobility in shopping mall environments. In: Proceedings of the second international workshop on mobile opportunistic networking. MobiOpp '10. ACM, New York, pp 1–7

27. Chen X, Xu F, Wang W, Du Y, Li M (2018) Geographic big data's applications in retailing business market. In: Big data support of urban planning and management. Springer, Cham, pp 157–176

28. Alonso W (1976) A Theory of Movements: I, Introduction. Institute of Urban & Regional Development, University of California Berkeley, CA. Working paper No. 266

29. Grosche T, Rothlauf F, Heinzl A (2007) Gravity models for airline passenger volume estimation. J Air Transp Manag 13(4):175–183

30. Jung W, Wang F, Stanley HE (2008) Gravity model in the Korean highway. Europhys Lett 81(4):48005

31. Lovelace R, Birkin M, Cross P, Clarke M (2016) From big noise to big data: toward the verification of large data sets for understanding regional retail flows. Geogr Anal 48(1):59–81

32. Piovani D, Molinero C, Wilson A (2017) Urban retail location: insights from percolation theory and spatial interaction modeling. PLoS ONE 12(10):0185787

33. Krings G, Calabrese F, Ratti C, Blondel VD (2009) Urban gravity: a model for inter-city telecommunication flows. J Stat Mech Theory Exp 2009(07):07003

34. Simini F, González MC, Maritan A, Barabási A-L (2012) A universal model for mobility and migration patterns. Nature 484(7392):96–100

35. Liang X, Zhao J, Dong L, Xu K (2013) Unraveling the origin of exponential law in intra-urban human mobility. Sci Rep 3

36. Palchykov V, Mitrović M, Jo H-H, Saramäki J, Pan RK (2014) Inferring human mobility using communication patterns. Sci Rep 4

37. Tinbergen J (1962) Shaping the world economy; suggestions for an international economic policy. Books (Jan Tinbergen)

38. Batty M (1978) Reilly's challenge: new laws of retail gravitation which define systems of central places. Environ Plan A 10(2):185–219

39. Klugman J (2010) Human development report 2010–20th anniversary edition. The real wealth of nations: pathways to human development

40. Goldberger AS (1968) The interpretation and estimation of Cobb–Douglas functions. Econometrica: 464–472

41. Silva JS, Tenreyro S (2006) The log of gravity. Rev Econ Stat 88(4):641–658

42. McCullagh P (1984) Generalized linear models. Eur J Oper Res 16(3):285–292

43. Calabrese F, Ferrari L, Blondel VD (2014) Urban sensing using mobile phone network data: a survey of research. ACM Comput Surv 47(2):25–12520

44. Graells-Garrido E, Ferres L, Caro D, Bravo L (2017) The effect of Pokémon go on the pulse of the city: a natural experiment. EPJ Data Sci 6(1):23

45. Chile, 2017 Census: Population by *Comuna* separated by sex and age. http://www.censo2017.cl/descargue-aqui-resultados-de-comunas/. [Online; accessed November 20, 2017]

46. Programa de las Naciones Unidas para el Desarrollo: Las trayectorias del desarrollo humano en las comunas de Chile (1994–2003). Temas de Desarrollo Humano Sustentable. PNUD: MIDEPLAN, Santiago, Chile (s.f)
47. Cámara Chilena de Centros Comerciales: Catastro de Centros Comerciales. http://www.camaracentroscomerciales.cl/centros-comerciales-de-chile/. [Online; accessed November 20, 2017]
48. Génois M, Barrat A (2018) Can co-location be used as a proxy for face-to-face contacts? EPJ Data Sci 7(1):11
49. Patronato B (2010) Barrio Patronato—Wikipedia, The Free Encyclopedia. [Online; accessed January 31, 2018] https://en.wikipedia.org/wiki/Barrio_Patronato