Crowdsourcing the Robin Hood effect in cities

Thomas Louail, Maxime Lenormand, Juan Murillo Arias, Jose Javier Ramasco

Crowdsourcing the Robin Hood effect in cities

Thomas Louail,[1,*] Maxime Lenormand,[2,*] Juan Murillo Arias,[3] and José J. Ramasco[1]

[1]*Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), Campus UIB, 07122 Palma de Mallorca, Spain*

[2]*Irstea, UMR TETIS, 500 rue François Breton, FR-34093 Montpellier, France*

[3]*BBVA Data & Analytics, Avenida de Burgos 16D, E-28036 Madrid, Spain*

Socioeconomic inequalities in cities are embedded in space and result in neighborhood effects, whose harmful consequences have proved very hard to counterbalance efficiently by planning policies alone. Considering redistribution of money flows as a first step toward improved spatial equity, we study a bottom-up approach that would rely on a slight evolution of shopping mobility practices. Building on a database of anonymized credit card transactions in Madrid and Barcelona, we quantify the mobility effort required to reach a reference situation where commercial income is evenly shared among neighborhoods. The redirections of shopping trips preserve key properties of human mobility, including travel distances. Surprisingly, for both cities only a small fraction (~ 5%) of trips need to be altered to reach equity situations, improving even other sustainability indicators. The method could be implemented in mobile applications that would assist individuals in reshaping their shopping practices, to promote the spatial redistribution of opportunities in the city.

The growth of economic inequality has raised concern and attention in recent years [1, 2]. In cities, entangled processes such as location choices of households and businesses, daily mobility, segregation and closure attitudes, central planning, or global economic restructuring contribute to these inequalities becoming embedded in space. Over the course of several decades their joint actions have given rise to spatially segregated cities, characterized by uneven distributions of capital among their neighborhoods. While the intensity of socioeconomic inequalities vary from one city to another, the general observation that "some neighborhoods are poorer than others" has been made for cities with different age, in every continent, and for different periods in urban history [3–6]. An abundant literature has long depicted the neighborhood effect [7], and highlighted its societal costs and enduring consequences [8–12].

Over the last decade, increasing volumes of digital geographic footprints have been passively produced by individuals using mobile ICT devices, and these footprints have been increasingly analyzed by scientists as well. These data are not free of biases [13] or privacy concerns [14], but they undeniably constitute an important asset for understanding social phenomena in detailed spatio-temporal contexts [15–20]. They also have the potential to reveal the information required to coordinate individuals' actions, so that large groups of people can tackle issues which are distributed and spatial by nature. This is particularly true in the case of mobility networks, which already integrate such footprints in feedback mechanisms: people produce data when moving, and their travel decisions are partly guided by the data produced by others. Examples include GPS navigation using real-time traffic data, local search and discovery of new places, or location-based dating applications. So far these footprints have been mainly used in applications intended to enhance individual satisfaction (time savings, discovery of a location, encounter of a partner), but they have also fostered spontaneous and large-scale solidarity movements during disasters (e.g. Facebook's safety check, or the use of dedicated Twitter hashtags). An important question thus concerns the possibility to scale up the use of such pervasive data, in order to address complex social issues with distributed and coordinated approaches, issues for which improvements would necessarily occur on longer timescales. There is a need to relate smart technology with sustainability and spatial justice in cities [21], and this implies building upon the existing practices of individuals. Here we develop further this idea by focusing on a complex problem: the reduction of spatial inequality in large cities.

The "Robin-Hood effect" refers to a process through which capital is redistributed to reduce inequality. A spatial and city-scale implementation would then consist in taking from the rich neighborhoods to give to the poor. This role is normally played by the city's governance, and is essential to mitigate spatial inequality. However, studies in cities worldwide have demonstrated that top-down planning and fiscal policies alone are inefficient in significantly counterbalancing the consequences of the neighborhood effect [22, 23]. It has also been long emphasized that developing economic activity in disadvantaged regions indirectly benefits the surrounding populations, by fostering job opportunities, transport facilities and increased safety [24]. Here we study an original approach to rebalance economic activity among a city's neighborhoods, that would not incur any additional environmental or monetary costs, but would instead require a slight evolution of shopping mobility practices. According to surveys, shopping and leisure trips account for 15% to 20% of the individuals' daily trips [25]. Such trips virtually move money from one part of the city to another, and directly contribute to shape

* Corresponding authors (thomas.louail@ifisc.uib-csic.es, maxime.lenormand@irstea.fr) who contributed equally to this work.

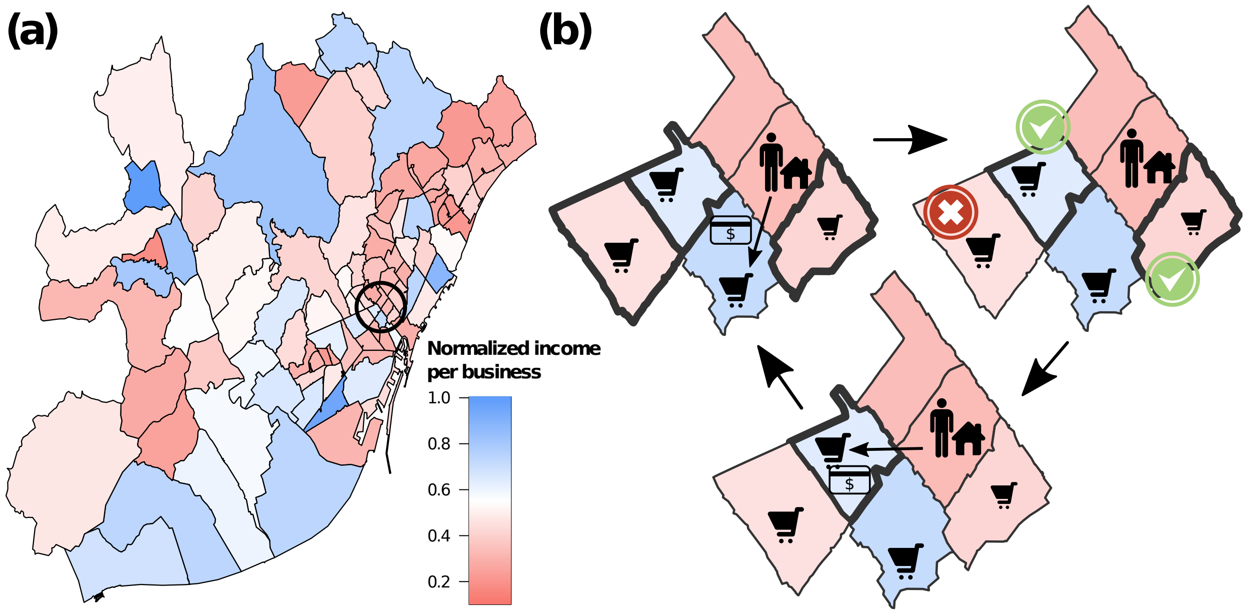

Figure 1. Rewiring urban shopping trips. (a) Average income per business in the neighborhoods of Barcelona resulting from individual transactions. The average income has been normalized by the maximum value among neighborhoods. The data correspond to 2011 and is displayed by zip code in the metropolitan area. From this perspective, some neighborhoods are five times richer than others. (b) The general principle common to the iterative rewiring methods. At each step a transaction is randomly selected, along with the possible alternative businesses (highlighted in bold). If rewiring the transaction to one of them (randomly selected) decreases inequality between neighborhoods and matches the other constraints, then rewiring is performed.

the distribution of wealth across neighborhoods. By connecting areas, shopping trips foster metropolitan integration and social cohesion [26], and the resulting money flows are a key component of the development of territories [27]. In large cities there are businesses of various types in most of the neighborhoods, and every time one has to buy usual products such as food, gas or clothes, one can actually choose among several stores and several neighborhoods, even without increasing the distance traveled.

In the following we demonstrate that more evenly balanced cities are reachable by the cumulative addition of small and reasonable changes in a limited fraction of individuals' shopping destinations. While there exists various spatial indicators for quantifying territorial inequalities, static indicators fail to provide a clear picture of the collective effort required to reach a certain level of redistribution. This in mind, we quantify the proportion of individual shopping trips that should be redirected to evenly share the commercial income between neighbourhoods, a conceivable path toward the spatial redistribution of opportunities in the city. We show that alternative shopping mobility scenarios not only allow to distribute money more evenly in space, but also to enhance the spatial mixing of city residents through their shopping mobility, without increasing the total distance traveled, nor changing the effective purchases and the mobility routines of individuals.

RESULTS

Data

We use a dataset containing one year of credit card transactions for more than 150 000 anonymous users in over 95 000 businesses of Barcelona and Madrid. Each transaction is time-stamped and contains the information collected by the bank on both the cardholder and the business. It also includes the customer's age and residence neighborhood, the business category and its geographical coordinates (see the Appendix for details). From these data there are two obvious ways to estimate inequality among neighborhoods: first, in measuring the income of their residents – indirectly estimated through the total amount of money spent during the year; second, in measuring the total income resulting from the commercial activity of businesses located in these neighborhoods. The map on Figure 1a shows the latter for Barcelona, and reveals that some neighborhoods are indeed five times "richer" than others. This measure is particularly interesting because it results from the spatial organization of shopping trips, which may be much easier to alter than any other type of our daily trips, notably commuting.

Rewiring the shopping trips networks

From the data for both cities we construct the bipartite spatial network whose nodes are individuals

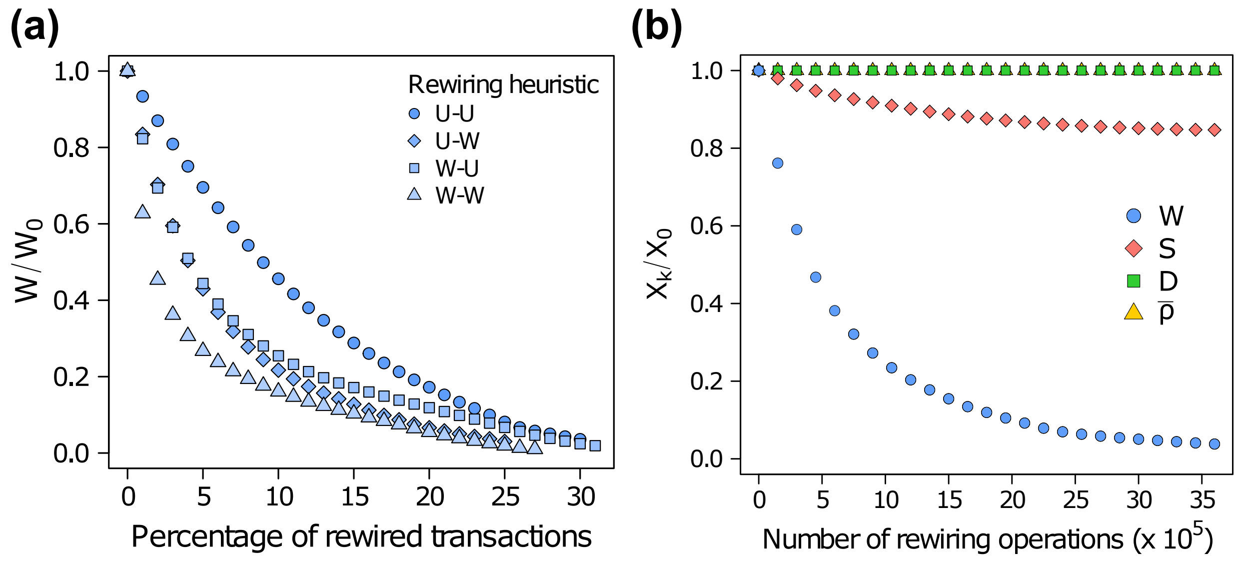

Figure 2. **Decreasing spatial inequality in the city by adapting daily shopping destinations.** (a) Decrease of wealth inequality among neighborhoods as a function of the fraction of transactions rewired, for various rewiring methods. Four combinations of choice heuristics are considered, "Uniform-Uniform" (U-U in the legend), "Uniform-Weighted" (U-W), "Weighted-Uniform" (W-U) and "Weighted-Weighted" (W-W). (b) Decrease of wealth inequality (W/W_0) while preserving the spatial mixing index (S/S_0), the total distance traveled (D/D_0) and the exploration rate ($\bar{\rho}/\bar{\rho}_0$), as a function of the number of rewiring operations. Values have been averaged over hundreds of replications. The bars represent the minimum and the maximum values obtained but in most cases are too close to the average to be seen (see Figure S9-S10 in Appendix for Madrid).

and businesses, and whose edges stand for transactions (see Figure S2 in Appendix). We then perform rewiring experiments, in which randomly selected transactions are redirected toward alternative businesses of the same category, but located elsewhere in the city. The purpose of each rewiring operation is to incrementally decrease inequality in the spatial distribution of business income W (see Figure1a). Furthermore, we consider three key properties of urban mobility: the distance traveled D, the individual mobility routines $\bar{\rho}$ (measured as the tendency to return to already visited businesses), and finally the spatial mixing of residents of different neighborhoods S. W and S are defined as distances to homogeneous spatial configurations. At each step, an individual transaction is selected along with all the candidate businesses toward which it could be reasonably redirected. One of them is randomly chosen, and if rewiring the transaction results in distributing money more evenly in space while matching the other constraints, then the change is accepted. The purchases and the amount of expenses of each individual are preserved, and this iterative process is run until the rewiring rate falls below a given threshold. Since the rewiring process is stochastic, all the results have been averaged over hundreds of replications (see Methods for more details about the four metrics and constraints considered in the rewiring process).

Numerous rewiring methods fulfilling these conditions could be proposed. In particular, besides the random selection of transactions ('Uniform' sampling), we can choose them according to a probability proportional to their amount and/or inversely proportional to the income of the target neighborhood ('Weighted' sampling). Even more informed methods might be proposed, but for the sake of simplicity, only basic random procedures are considered in the following.

Reachability of even spatial distributions

We first investigate the reachability of an even spatial distribution of the commercial income resulting from individual purchases, while the variables S, D and $\bar{\rho}$ remain in the range of their empirical values. To address this question, we apply the method with the four constraints of Equation 5 such as $\alpha_W = 0$, $\alpha_S = 1$, $\alpha_D = 1$ and $\alpha_\rho = 1$. This constitutes the *Reference* experiment. Figure 2a shows the evolution of inequality in the urban area of Barcelona as a function of the fraction of rewired transactions, according to various methods. Surprisingly, even with basic random methods, it is possible to reduce spatial inequality by more than 80% while reassigning only 20% of individual transactions. All the methods produce the same qualitative behavior – an early regime of very fast decay, followed by a regime of slower decay. Weighted methods are naturally more efficient, and allow to reach spatial equity by redirecting a smaller fraction of transactions. In particular, a reduction of 80% in W/W_0 can be obtained by rewiring only 5%

of the transactions if the method is double weighted.

The state of the other variables S, D and ρ is also monitored along the process, as shown in Figure 2b for a Uniform-Uniform method. What makes the previous results remarkable is in fact that income redistribution is achieved without increasing the distance traveled by individuals (D), nor changing their mobility routines ($\bar{\rho}$). Moreover, a positive side-effect is to increase the frequency of encounters of individuals living in different parts of the city – as indicated by the decrease of S/S_0 –, an effect that could not be derived from the rewiring constraints alone ($\alpha_S = 1$). The increase of spatial mixing is the consequence of individual shopping trips more evenly distributed in the city space, required to homogenize the income among neighborhoods. The behavior of S is non trivial, notably because one could imagine unrealistic solutions to this problem that would simultaneously even the business income distribution among neighborhoods and decrease the total distance traveled, by rewiring most of the shopping trips to the closest neighborhood containing businesses of the relevant category. The spatial mixing of individuals in this case would decrease dramatically, and S the distance to an homogeneous spatial mixing would increase. Here the decrease of S/S_0 guarantees that it is not the case.

Preservation of human mobility properties

We wish to control further the likelihood of the rewired shopping mobility networks, and ensure that they preserve the spatial properties of individual human mobility. A small set of indicators are useful to describe the statistical and spatial properties of human mobility [20]. These include the jump length between consecutive locations Δ_r, the radius of gyration r_g and the tendency to return to already visited places ρ. In our case, for each individual ρ is simply defined as the ratio between the number of unique businesses visited and the total number of transactions.

Figure 3 shows their empirical and simulated values, plus the average distance $\bar{d}$ traveled by each individual for each shopping trip (see the Methods section for details on the calculation of shopping trips distances). On each panel both curves overlap almost perfectly, indicating that the rewiring has no significant effect on the key mobility properties. Simulated distributions of $\bar{d}$ and r_g are slightly more peaked than the empirical ones. Finally, we showed in a previous study that young adults tend to spend their money further from their neighborhood of residence [28], and coherently their shopping trips are those that are the most affected in the simulated scenarios (see Figure S5 in Appendix).

Multi-objective improvement

We now perform multi-criteria rewiring experiments, to measure to what extent redistribution can be achieved while improving simultaneously other important aspects of urban mobility. To this end we perform the series of experiments summarized in Table 1. The objective is to even the wealth distribution among neighborhoods ($\alpha_W = 0$) and simultaneously improve either S, D or $\bar{\rho}$ without worsening the other two. Figure 4 gives the relative gains and losses upon the four indicators, and the last two columns of Table 1 contain the asymptotic values obtained for the reduction rate of wealth inequality, for Barcelona (B) and Madrid (M).

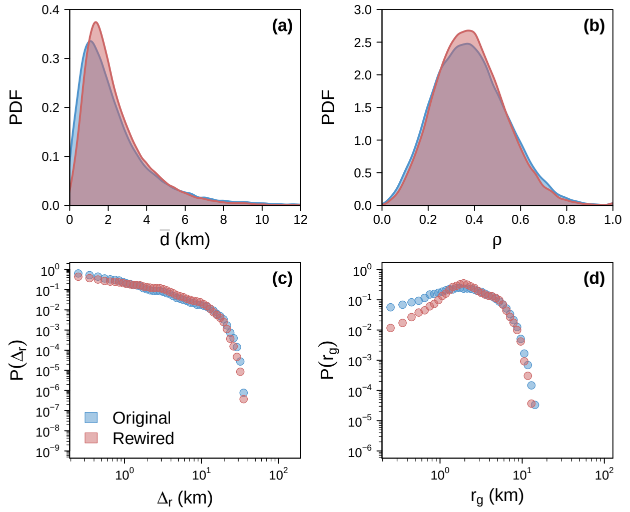

Figure 3. Observed and simulated distributions of human mobility indicators. The distribution of jump lengths Δ_r, the radius of gyration r_g, the tendency to return to already visited places (ρ) and the individual average distance traveled ($\bar{d}$) are considered. Values measured on the empirical data are in blue, while those obtained after rewiring are in red. The calculation of Δ_r and r_g is based on the business' exact geographical coordinates. The simulated distributions plotted here correspond to one particular replication, see the Figure S4 for the robustness of the results and Figure S8 in Appendix for the same curves for Madrid.

Table 1. Experiments performed. Column W indicates the relative gain of $(W_0 - W)/W_0$. The first value is for Barcelona (B) and the second for Madrid (M).

Experiment	α_W	α_S	α_D	$\alpha_{\bar{\rho}}$	W (B/M)
(a) Reference	0	1	1	1	96.4%/99.5%
(b) Spatial mixing ↑	0	0.75	1	1	85.9%/78.1%
(c) 50% energy savings	0	1	0.5	1	87.4%/84.8%
(d) 25% energy savings	0	1	0.75	1	94.7%/98.8%
(e) Exploration rate ↑	0	1	1	1.25	96.8%/99.9%
(f) Exploration rate ↑↑	0	1	1	1.5	97.3%/100%

Experiments prove that it is possible to combine significant improvements on several dimensions simultaneously. This is not an issue with the method but rather with the objectives set, which are somewhat opposite. Most individuals perform their shopping trips

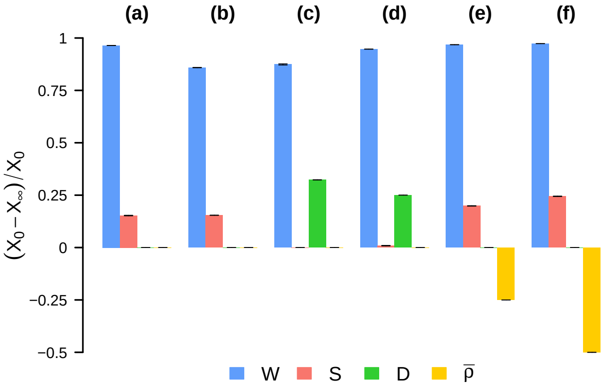

Figure 4. **Multi-criteria improvement of shopping mobility.** Each group of bars gives the relative gains or losses for the four indicators W, S, D and $\bar{\rho}$. Experiments are described in Table 1. See Figure S11 in Appendix for Madrid.

near their residence – as highlighted by the empirical distributions in Figure 3 – and consequently it is not feasible to simultaneously diversify the neighborhoods where an individual regularly travels to – in order to improve spatial mixing S – and at the same time decrease the total travel distance D. More surprisingly, experiment (b) indicates that it is also not possible to simultaneously improve the wealth redistribution and the spatial mixing of individuals. The two indicators are based on different metrics (the amount of money spent per business for W and the number of trips for S), which imply different reference egalitarian situations (see Figure S6 in Appendix for more details). Optimization is thus a trade-off between the various consequences of shopping mobility at the city scale. However, experiments (c) and (d) prove that it is possible to significantly decrease the total distance traveled and in the same time to strongly reduce wealth inequality among neighborhoods, but not as much as in the reference experiment. Still, it is remarkable that experiment (d) results in an alternative mobility network such that the spatial inequality of the average business income is reduced by 95%, while the total distance associated to shopping mobility is reduced by 25%, the level of spatial mixing is preserved, as well as the individual mobility routines. Finally experiments (e) and (f) show that even if residents deeply restructured their mobility routines, and typically started going to a new business each time they perform a new shopping trip, keeping control of the total distance traveled in the city would prevent from increasing the mixing of individuals coming from different neighborhoods beyond 25%. The gains in terms of wealth redistribution would not be significant when compared to the reference experiment (a).

DISCUSSION

Reducing urban segregation and increasing spatial justice are some of the major challenges faced by cities worldwide, and the digital footprints passively produced by their residents constitute a promising resource to help addressing these issues from the bottom. This study is a first attempt to quantify the relation between shopping mobility and the spatial distribution of economic activity in the city. The alternative shopping trips resulting from our experiments offer an interesting trade-off between the preservation of essential aspects – the effective purchases of individuals and households, and their mobility properties – and some reasonable changes in the places where they spend their money. The addition of small changes in the shopping destinations of individuals can dramatically impact the spatial distribution of money flows in the city, and the frequency of encounters between residents of different neighborhoods, even if the total number of changes remains small. These results have important consequences, and they lead in particular to the decisive question of the effective implementation of alternative shopping travels, like those drawn by our experiments. While the decision process behind each individual redirection may appear intricate for a single person, one could easily imagine dedicated mobile applications, querying databases very similar to the one used in this paper. Their purpose would be to assist their users in a transition toward a more socially and spatially concerned shopping mobility.

Limitations of the study

However one should keep in mind that individuals do not guide most of their travel decisions by philanthropy, but instead by balancing accessibility, price and business characteristics. Individuals first choose their casual shopping destinations with regard of transport facilities and travel time budget [29]. Here as accessibility information we considered the Euclidean distance between neighborhoods. However in urban environments the Euclidean distance is rarely a direct proxy for travel time [30], and some of the rewired shopping trips are unlikely to be performed in the real world. We also assumed that every shopping trip follows the simple pattern $A \to B \to A$, and we did not consider the more complicated case of chained trips (e.g. $A \to B \to C \to A$) during which individuals join several trips associated with different purposes [31]. People also tend to choose the places where they spend money according to several other key factors, the price of products in the first place, but also according to some more personal appreciations, such as the "ambiance" of neighborhoods and the feeling of well-being they provide. In large cities, the neighborhoods strongly differ in the quality of their planning and architecture, in their public spaces, in their amenities and leisure opportunities, commercial fabric, in their safety. Additionally, the changes might be considered as problematic, since a profound spatial reorganization of shopping mobility in the city could have consequences on the spatial structure of employment in the first place (see Figure S7 in Appendix), and then on residences. For the sake of simplicity we considered none of these factors, but they could be implemented in more involved frameworks derived from our work.

Concluding remarks

There are recent, encouraging examples of fast and wide adoptions of new daily practices, whose benefits are essentially collective. Examples include garbage differentiation [32], the increasing role of bicycle in urban transport and the development of bicycle sharing systems [33], or the open-source movement and the dedication of a growing number of individuals to collectively build free knowledge databases (e.g. Wikipedia, StackExchange, the free software movement). In these cases, the remuneration of participants, if any, is essentially symbolic. The success of a spatial counterpart to these altruistic behaviors could rejuvenate the meaning of the so-called sharing economy [21]. As citizens produce the data that document their location and activity patterns, in return these data could serve not only the specific interest of the institution collecting them, but also support fair socioeconomic initiatives. This study brings evidence that these geographical footprints we passively produce can support bottom-up responses to big societal issues, an expected feature of truly smart cities.

METHODS

Estimation of shopping trips distances.

We made the assumption that to each transaction is implicitly associated a trip originating from either the main activity neighborhood during working time, or the neighborhood of residence, depending on the hour of the day and day of the week (see the Appendix for more details). The shopping trip distance is then the Euclidean distance between the centroids of the origin and destination neighborhoods.

Variables considered when redirecting the shopping trips

The primary objective when rewiring the shopping mobility networks is to even the distribution of commercial income among the neighborhoods (here zipcodes), while respecting a set of constraints to ensure that the alternative networks produced are reasonable, and possess interesting properties.

We denote by W the *wealth inequality among a city's neighborhoods*. It is defined as the distance to a reference egalitarian situation, where the commercial income resulting from individual purchases would be equally shared among all neighborhoods in average. More specifically, the level W_k of *wealth inequality* after k rewiring operations is

$$W_k = \sum_{i=1}^{N} (\overline{w}_k^i - w^*)^2, \tag{1}$$

where N is the number of neighborhoods, $\overline{w}_k^i$ is the average income of businesses located in neighborhood i after k rewiring operations, and w^* represents the wealth per neighborhood in the reference configuration where commercial income is evenly distributed across neighborhoods, such that

$$w^* = \frac{1}{N} \sum_{i=1}^{N} \overline{w}_k^i. \tag{2}$$

An important aspect of individuals mobility in the city is to prevent some neighborhoods from ghettoization. To measure to what extent individuals residing in various neighborhoods mix in the city space as a result of their displacements, for each neighborhood i we count the number of times $(s_k^{i1}, ..., s_k^{iN})$ the residents of i traveled to each of the N neighborhoods (i included), after k rewiring operations. Then, by averaging the vector of trips over all the neighborhoods, we compute the geographical diversity index S_k obtained after k rewiring operations,

$$S_k = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} (s_k^{ij} - s_i^*)^2, \tag{3}$$

where s_i^* represents the homogeneous distribution of visits originating from i and in direction to all neighborhoods,

$$s_i^* = \frac{1}{N} \sum_{j=1}^{N} s_k^{ij}. \tag{4}$$

The third aspect we want to control is the distance traveled by individuals. For obvious reasons it is desirable that the total travel distance does not increase. Summing the distances traveled by individuals for all their shopping trips, we can compute D_k the total distance measured on the network after rewiring k transactions.

Finally, in order to preserve individual mobility routines and the tendency of individuals to return to already visited places, for each individual we calculated the exploration rate ρ_k. It is defined as the number of unique businesses he/she has visited divided by his/her total number of transactions, after k rewiring operations. Considering the empirical peaked distribution of ρ_k among the population of customers, in the following we consider the mean value $\bar{\rho}_k$.

Rewiring constraints

The rewiring experiments can be formalized as a constraint satisfaction problem (CSP) over a spatial bipartite network. We consider four constraints, each of them concerns one of the variables mentionned in the previous section, and quantifying an important aspect of shopping mobility:

- Constraint C_W applies on the *wealth distribution*; it ensures that each destination change contributes to homogenize the distribution of commercial income across neighborhoods.
- C_S, a constraint on the *spatial mixing* of individuals, resulting from their shopping trips. To be accepted, a rewiring operation has to preserve the diversity of neighborhoods visited, hence the degree of spatial mixing of individuals residing in different neighborhoods.
- C_D, a constraint on the *total distance traveled*, to guarantee that each destination change does not result in increasing the total distance traveled. The distance associated to each individual transaction is calculated with regard to the individual's main activity place at this moment of the day (see section above for details).
- $C_{\bar{\rho}}$, a constraint on the spatial *exploration rate* of individuals, to preserve the behavioral mobility routines measured in the population.

All constrains have the same form: constraint C_X is satisfied if the following condition holds

$$X_{k+1} \leq \begin{cases} X_k & \text{if } X_{k+1} \geq \alpha X_0, \\ \alpha X_0 & \text{otherwise,} \end{cases} \tag{5}$$

where k denotes the number of rewiring operation, and α is a parameter positive or equal to zero. Equation 5 allows to fix the objective upper bound for each variable X with respect to its empirical value X_0. While X is greater than αX_0, each rewiring operation must decrease X. Once X is smaller than αX_0, Equation 5 ensures that none of the following rewiring operations will increase X above this value. An experiment is then defined by a set of four values $(\alpha_W, \alpha_S, \alpha_D, \alpha_{\bar{\rho}})$ that specify the desired maximal value for each of the four dimensions of interest.

We favored a numerical approach because of the large number of transactions (~ 10^7) and because of the constraints imposed to guarantee that the simulated networks possess realistic and interesting properties.

ACKNOWLEDGEMENTS

We wish to thank Julie Vallée and Suma Desu for providing feedback on an early version of the manuscript. We also thank Jean-Loup Guillaume for fruitful discussions on the rewiring problem. Partial financial support has been received from the Spanish Ministry of Economy (MINECO) and FEDER (EU) under the project INTENSE@COSYP (FIS2012-30634), and from the EU Commission through projects EUNOIA and INSIGHT. The work of TL has been funded under the PD/004/2015, from the Conselleria de Educación, Cultura y Universidades of the Government of the Balearic Islands and from the European Social Fund through the Balearic Islands ESF operational program for 2013-2017. JJR acknowledges funding from the Ramón y Cajal program of MINECO.

[1] T. Piketty. *Capital in the Twenty-First Century*. Harvard University Press, Cambridge, MA, USA, 2014.
[2] J. E. Stiglitz. *The Great Divide: Unequal Societies and What We Can Do About Them*. W. W. Norton & Company, New York, 1 edition edition, 2015.
[3] J.-L. Pinol. *Histoire de l'Europe urbaine*. Seuil, Paris, France, 2003.
[4] W. Goldsmith, E. J. Blakely, and B. Clinton. *Separate Societies: Poverty and Inequality in U.S. Cities, 2nd Edition*. Temple University Press, Philadelphia, Pa, 2nd edition edition, 2010.
[5] T. Cassiers and C. Kesteloot. Socio-spatial Inequalities and Social Cohesion in European Cities. *Urban Studies*, 49(9):1909–1924, 2012.
[6] UN-HABITAT. *State of African Cities Re-imagining sustainable urban transitions*. State of Cities - Regional Reports. UN-Habitat, 2014.
[7] J. Friedrichs, G. Galster, and S. Musterd. Neighbourhood effects on social opportunities: the European and American research and policy context. *Housing Studies*, 18(6):797–806, 2003.
[8] Judith R. Blau and Peter M. Blau. Cost of Inequality - Metropolitan Structure and Violent Crime. *American Sociological Review*, 47:114–129, 1982.
[9] Jeanne Brooks-Gunn, Greg Duncan, and J. Lawrence Aber, editors. *Neighborhood Poverty, Volume 1: Context and Consequences for Children*. Russell Sage Foundation, New York, 1997.
[10] J. Vallée, E. Cadot, F. Grillo, I. Parizot, and P. Chauvin. The combined effects of activity space and neighbourhood of residence on participation in preventive health-care activities: The case of cervical screening in the Paris metropolitan area (France). *Health & Place*, 16(5):838–852, 2010.
[11] B. Womack. *Across 110th street*. 1972.
[12] R. Chetty and N. Hendren. The Impacts of Neighborhoods on Intergenerational Mobility: Childhood Exposure Effects and County-Level Estimates. Technical report, Harvard Univ. Dept. of Economics, 2015.
[13] K. Lewis. Three fallacies of digital footprints. *Big Data & Society*, 2(2):2053951715602496, 2015.
[14] Y.-A. de Montjoye, L. Radaelli, V. K. Singh, and A. Pentland. Unique in the shopping mall: On the

reidentifiability of credit card metadata. *Science*, 347(6221):536–539, 2015.

[15] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. V. Alstyne. Computational Social Science. *Science*, 323(5915):721–723, 2009.

[16] N. Eagle, M. Macy, and R. Claxton. Network Diversity and Economic Development. *Science*, 328(5981):1029–1031, 2010.

[17] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007.

[18] X. Lu, L. Bengtsson, and P. Holme. Predictability of population displacement after the 2010 Haiti earthquake. *Proceedings of the National Academy of Sciences*, 109(29):11576–11581, 2012.

[19] L. Sun, K. W. Axhausen, D.-H. Lee, and X. Huang. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences*, 110(34):13774–13779, 2013.

[20] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[21] D. McLaren and J. Agyeman. *Sharing Cities: A Case for Truly Smart and Sustainable Cities*. Urban and Industrial Environments. MIT Press, Cambridge, MA, USA, 2015.

[22] W. Ostendorf, S. Musterd, and S. De Vos. Social Mix and the Neighbourhood Effect. Policy Ambitions and Empirical Evidence. *Housing Studies*, 16(3):371–380, 2001.

[23] D. Satterthwaite and D. Mitlin. *Reducing Urban Poverty in the Global South*. Routledge, London ; New York, 2013.

[24] R. Kanbur and A. J. Venables. *Spatial Inequality and Development*. Oxford University Press, 2005.

[25] The American Association of State Highways and Transportation Officials (AASHTO). Commuting in America 2013, the national report on commuting patterns and trends. Technical report.

[26] K. Miciukiewicz and G. Vigar. Mobility and Social Cohesion in the Splintered City: Challenging Technocentric Transport Research and Policy-making Practices. *Urban Studies*, 49(9):1941–1957, 2012.

[27] J.-F. Ruault. *L'effet de la consommation de passage sur le développement et l'intégration métropolitaine des territoires en Ile-de-France*. Paris Est, 2014.

[28] M. Lenormand, T. Louail, O. G. Cantú-Ros, M. Picornell, R. Herranz, J. Murillo Arias, M. Barthelemy, M. San Miguel, and J. J. Ramasco. Influence of sociodemographics on human mobility. *Scientific Reports*, 5:10075, 2015.

[29] W. W. Recker and L. P. Kostyniuk. Factors influencing destination choice for the urban grocery shopping trip. *Transportation*, 7(1):19–33, 1978.

[30] R. Gallotti, A. Bazzani, S. Rambaldi, and M. Barthelemy. A stochastic model of randomly accelerated walkers for human mobility. *arXiv:1509.03752 [cond-mat, physics:physics]*, 2015.

[31] C. M. Schneider, V. Belik, T. Couronné, Z. Smoreda, and M. C. González. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84):20130246, 2013.

[32] G. M. Robinson and A. D. Read. Recycling behaviour in a London Borough: Results from large-scale household surveys. *Resources, Conservation and Recycling*, 45(1):70–83, 2005.

[33] J. Dill and T. Carr. Bicycle Commuting and Facilities in Major U.S. Cities: If You Build Them, Commuters Will Use Them. *Transportation Research Record: Journal of the Transportation Research Board*, 1828:116–123, 2003.

[34] Instituto Nacional de Estadstica (National Institute for Statistics). Available: http://www.ine.es. Accessed 2016 April 26.

APPENDIX

Data preprocessing

The dataset contains information about 14 million bank card transactions made by customers of the Banco Bilbao Vizcaya Argentaria (BBVA) in the metropolitan areas of Barcelona and Madrid in 2011. For both case studies, we only consider the credit card payments whose amount was inferior to 1000 euros, and which were made inside the metropolitan areas, by bank customers that lived and worked in the metropolitan area in 2011. Each transaction is characterized by its amount (in euro currency) and the time when the transaction has occurred. Each transaction is also linked to a customer and a business. Customers are identified with an anonymized customer ID, connected with sociodemographic characteristics (gender, age and occupation) and their postcode of residence. In the same way, businesses are identified through an anonymized business ID, a business category id, and the geographical coordinates of the credit card terminal. Since we are primarily interested in daily shopping mobility, we chose to consider the business categories that account for the top 90% of the daily shopping trips (see Figure S1). The proportions of shopping trips associated to each of the 20 business categories we selected are available in Table S1.

Table S1. Proportion of shopping trips associated to each of the 20 business categories selected.

Category	Barcelona	Madrid
Supermarket	17.71	14.84
Hypermarket	10.25	12.09
Gas Stations	8.41	9.49
Restaurants	8.20	6.73
Retail store	5.84	2.82
Clothing store chain	5.48	4.67
Clothing store	5.16	7.33
Pharmacy, optical and orthopedics	4.52	3.81
Department store	3.14	5.73
Hair and beauty	2.88	2.72
Electronics, computers and appliances	1.97	1.44
Bars and café	1.85	1.60
Shoe store	1.71	1.43
Toys and sports articles	1.43	1.33
Bookshop, music shop and stationery	1.42	1.04
Fast food restaurants and chains	1.13	2.38
Car dealership and garage	1.02	1.01
Bazaar	1.01	1.06
DIY store	0.99	1.08
Hospitals, clinics and doctors	0.91	0.88

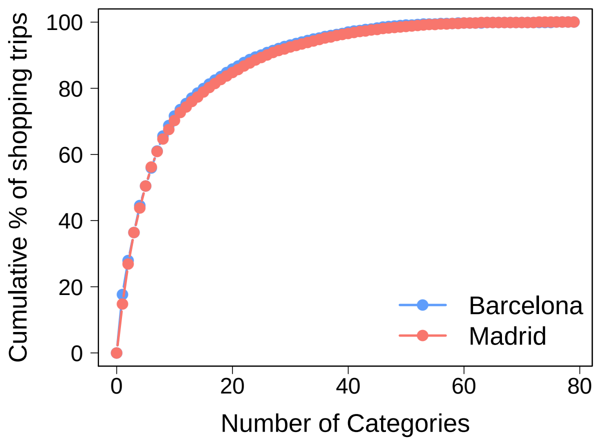

Figure S1. **Cumulative proportion of shopping trips as a function of the number of categories.** In blue the metropolitan area of Barcelona; In red the one of Madrid

Formal description of the rewiring process

From the data we extract $G(R, B, T)$ the bipartite network of all credit card transactions performed by the city residents in businesses located in the city, during the entire year. R is the set of residents, B the set of businesses and T the set of transactions. Table S2 contains the characteristic attributes of the network in the two cities studied. Each city is partitioned in N spatial units/neighborhoods (here the units correspond to zip codes) and the network G is spatial: each resident and each business is located in one neighborhood. We denote R_i (resp. B_i) the set of residents (resp. businesses) located in the neighborhood i. The sets are disjointed and we have $R = \cup R_i$ and $B = \cup B_i$, with $i \in 1..N$. Additionally the businesses are also partitioned in C categories according to the products they sell, and we have $B = \cup B_c, c \in 1..C$. The edges of the network represent the card transactions, hence implicitly the shopping trips. We note $t^k_{r,b}$ the k-th transaction performed by resident r in business b, and by $w(t^k_{r,b})$ its amount.

Table S2. **Summary statistics of the two metropolitan areas and of the two transactions networks.**

Statistics	Barcelona	Madrid
Number of neighborhoods	97	123
Number of inhabitants (2009)	3,218,071	5,512,495
Area (km^2)	634	1,935
Number of customers	42,023	118,447
Number of businesses	40,618	55,148
Number of transactions	3,640,961	10,025,642

The rewiring methods we implemented operate directly at the micro scale of the individual transactions, and each rewiring operation $t_{r,b} \to t_{r,b'}$ consists in se-

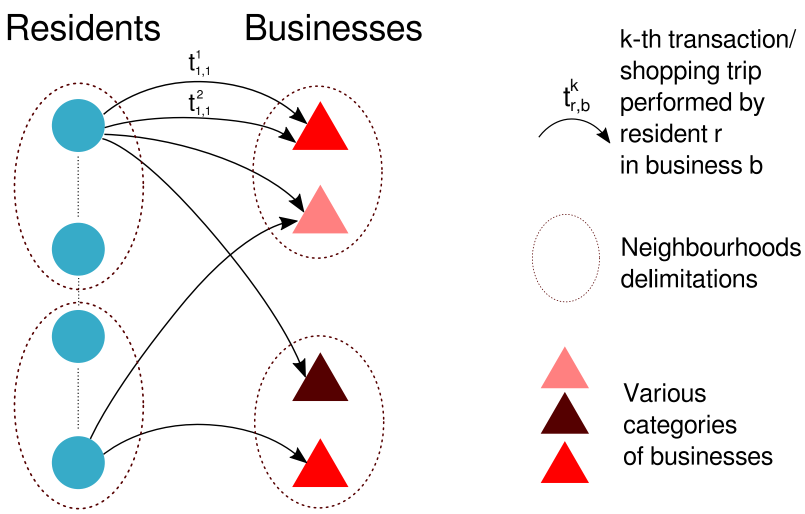

Figure S2. The bipartite network of transactions.

lecting a business $b' \neq b$, such than b' and b are of the same category c but are located in different neighborhoods. The rewiring occurs only if b' fulfills a number of additional constraints (namely C_W, C_D, C_S and C_ρ) which are expressed at the macro-scale of the entire city.

The network is rewired iteratively, i.e. transaction per transaction. A transaction $t_{r,b}$ is picked up randomly (uniform or weighted sampling, as described in the main text). A neighborhood is chosen among the set of all neighborhoods that contain some businesses b' of the same category than b. Both the transaction and the candidate business can be picked up through a uniform or weighted random sampling. Finally, if the neighborhood change $j \to j'$ matches the four constraints, then the transaction/edge is rewired. The process stops when the rewiring rate falls below 0.001.

Estimation of shopping trips distances and identification of the users' main daytime activity location

We made the assumption that to each transaction is implicitly associated a trip originating from either the main activity neighborhood during working time, or the neighborhood of residence, depending on the hour of the day and day of the week. The shopping trip distance is then defined as the Euclidean distance between the centroids of the origin and destination neighborhoods.

We already know the neighborhood of residence that we can assign as the place of main activity during night time (i.e between 7pm and 8am) on week days and Saturday and Sunday. In addition to the neighborhood of residence, for each individual we can determine the neighborhood in which he/she was the most frequently located during the typical working hours of working days, i.e. from 8pm to 7am, from Monday to Friday.

To do so, for each individual we count the number of unique couples $(day, hour)$ during which he/she was located in each neighborhood. For our study we keep only the individuals for which credit card is a casual mode of payment, and for which we can then reasonably assume that their card purchases and corresponding shopping trips are representative of their shopping mobility in general. Regarding the available statistics for Spain on the share of credit card payments among all payments, we decided to keep individuals whose data displayed at least $N = 20$ unique couples $(day, hour)$ during the entire year. For each of these individuals, we then determine the neighborhood in which they were the most frequently located during typical working hours. If this neighborhood accounted for less than one third of the time $\delta = 1/3$ in his/her entire set of locations, then the individual is discarded. As it can be seen in Figure S3a and Figure S3c, the couple of value $(N, \delta) = (20, 1/3)$ allow us to keep enough users and discard the users not showing enough regularity to estimate their main daytime activity location.

Finally, we can estimate the commuting flows between neighborhoods and assess the accuracy of the results by comparing these flows with those obtained from the 2011 Spanish census in Barcelona and Madrid [34]. The census data is at the municipal level, which implies that the neighborhoods must be aggregated at the municipality scale to be able to perform the comparative analysis. Figure S3b and Figure S3d show a scattered plot with the comparison between the flows obtained with the two matrices. A good agreement between the two ODs is obtained.

The source code of this method is available at `https://github.com/maximelenormand/Most-frequented-locations`.

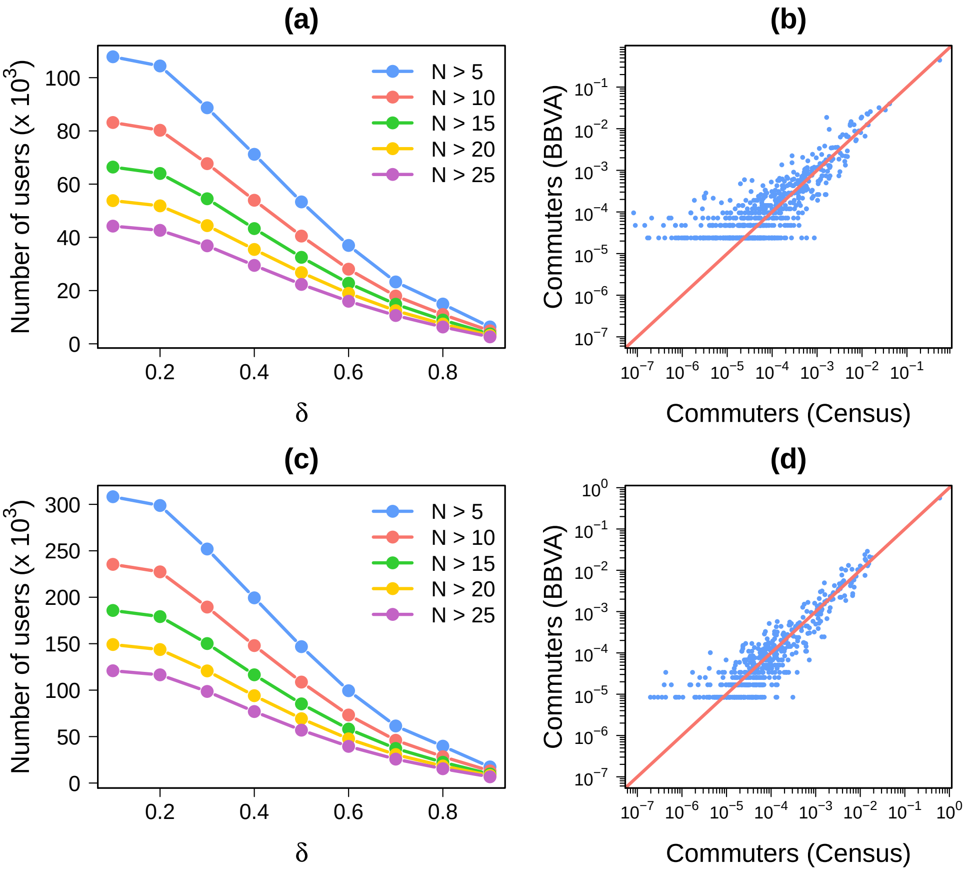

Figure S3. **Identification of the users' main daytime activity location in Barcelona ((a)-(b)) and Madrid ((c)-(d)).** (a) and (c) Number of users according to N and δ. (b) and (d) Comparison between the non-zero flows obtained with the credit card dataset ($(N, \delta) = (20, 1/3)$) and the census data. The values have been aggregated at the municipality scale. The values have been normalized by the total number of commuters for both OD tables. Blue points are scatter plot for each pair of municipalities. The red line represents the x = y line.

SUPPLEMENTARY FIGURES

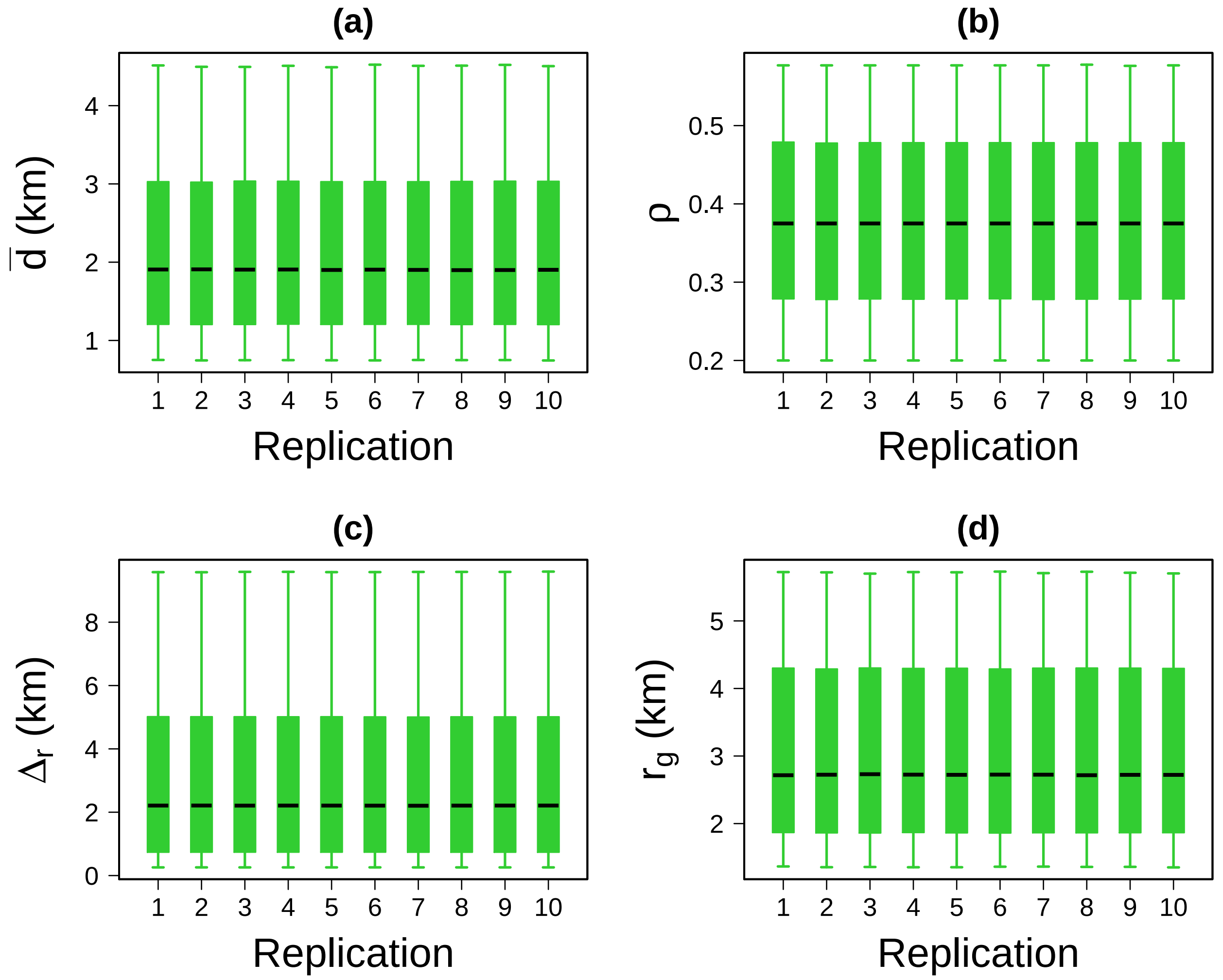

Figure S4. **Individual human mobility indicators' distributions obtained with ten replications of the algorithm.** (a) Individual average distance traveled $\bar{d}$. (b) Exploration rate ρ. (c) Jump length distribution Δ_r. (d) Radius of gyration r_g. The boxplot is composed of the first decile, the lower hinge, the median, the upper hinge and the 9th decile.

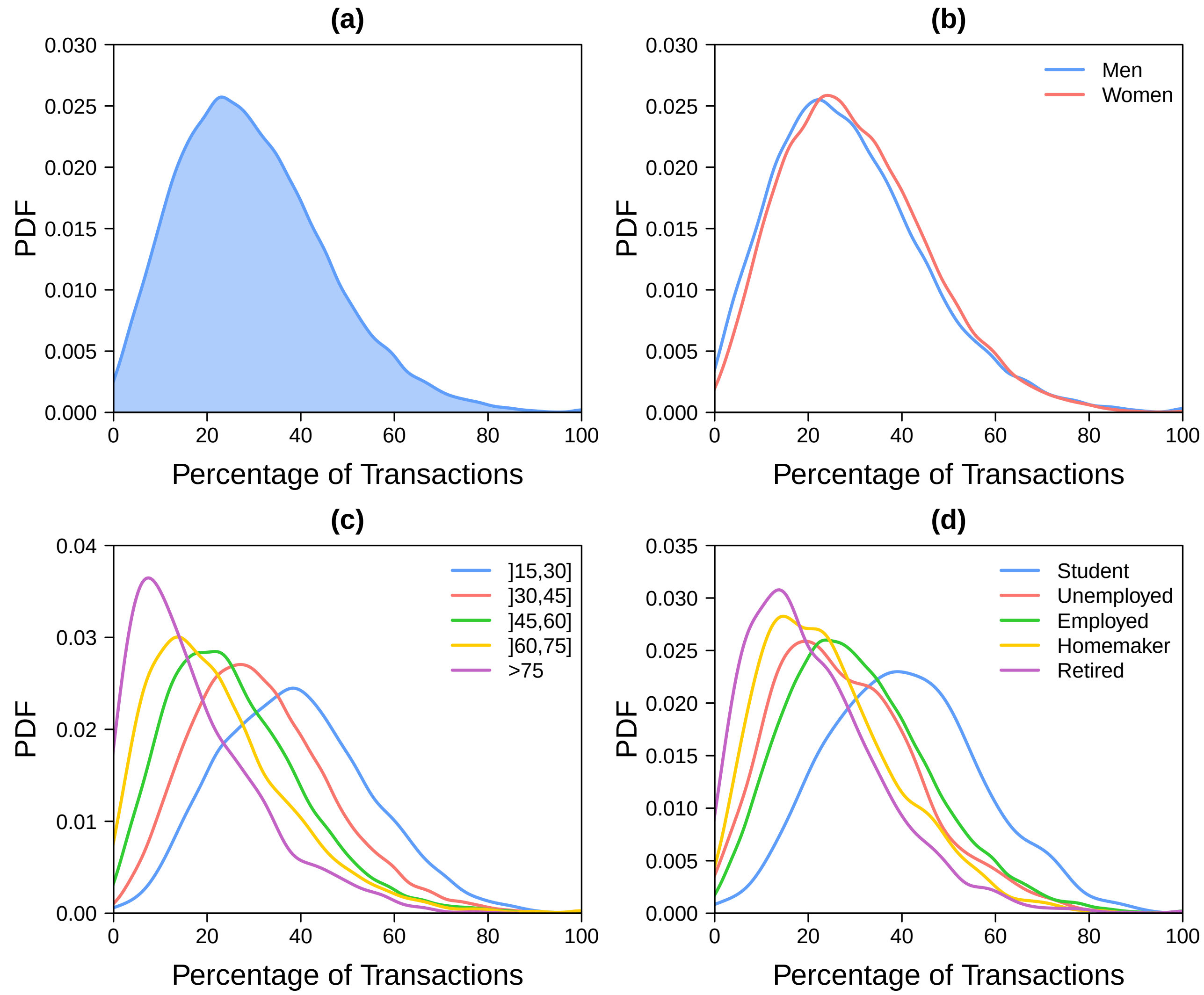

Figure S5. **Probability density functions of the individual percentage of rewired transactions.** (a) Total. (b) By Gender. (c) By age. (d) By occupation.

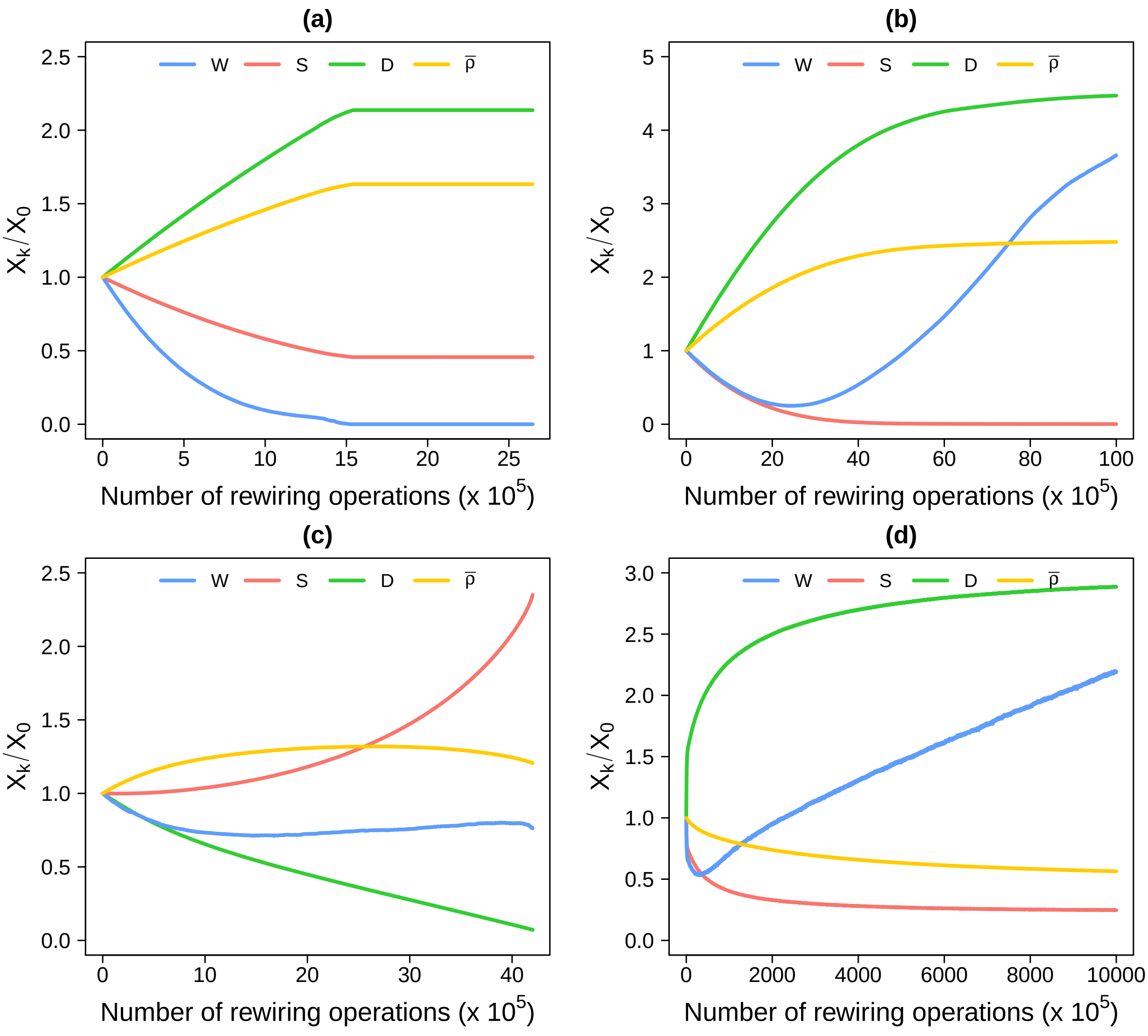

Figure S6. **Evolution of W, S, D and $\bar{\rho}$ as a function of the number of rewiring transactions.** (a) $(\alpha_W, \alpha_S, \alpha_D, \alpha_{\bar{\rho}}) = (0, +\infty, +\infty, +\infty)$. (b) $(\alpha_W, \alpha_S, \alpha_D, \alpha_{\bar{\rho}}) = (+\infty, 0, +\infty, +\infty)$. (c) $(\alpha_W, \alpha_S, \alpha_D, \alpha_{\bar{\rho}}) = (+\infty, +\infty, 0, +\infty)$. (d) $(\alpha_W, \alpha_S, \alpha_D, \alpha_{\bar{\rho}}) = (+\infty, +\infty, +\infty, 0)$.

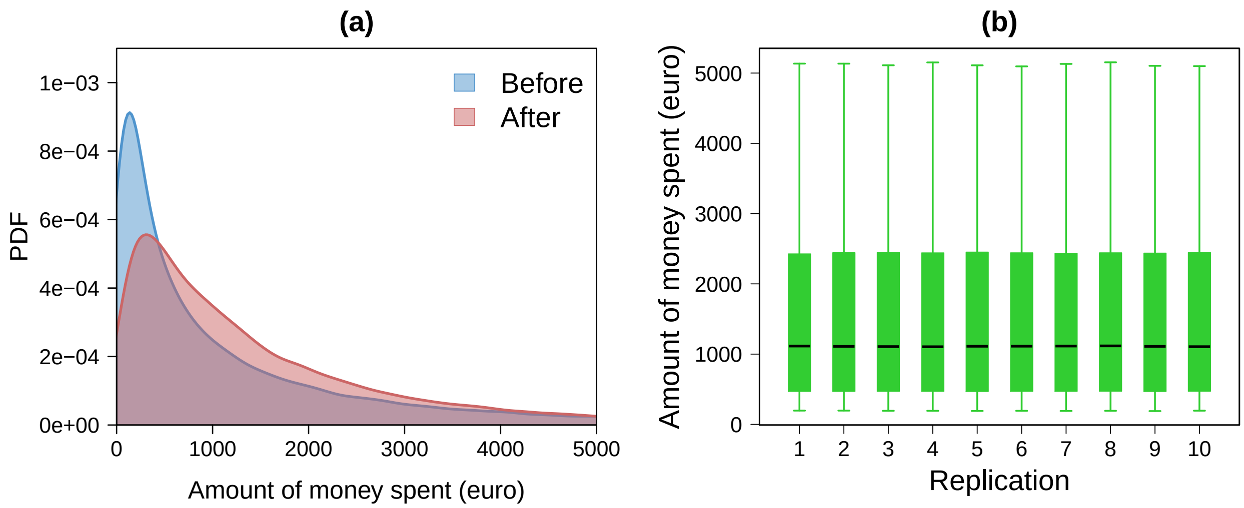

Figure S7. Probability density functions of the total amount of money spent by business in 2011, in Barcelona. (a) Comparison between the original distribution and the one obtained after applying the rewiring algorithm. (b) Distributions obtained with ten replications of the algorithm. The boxplot is composed of the first decile, the lower hinge, the median, the upper hinge and the 9^{th} decile.

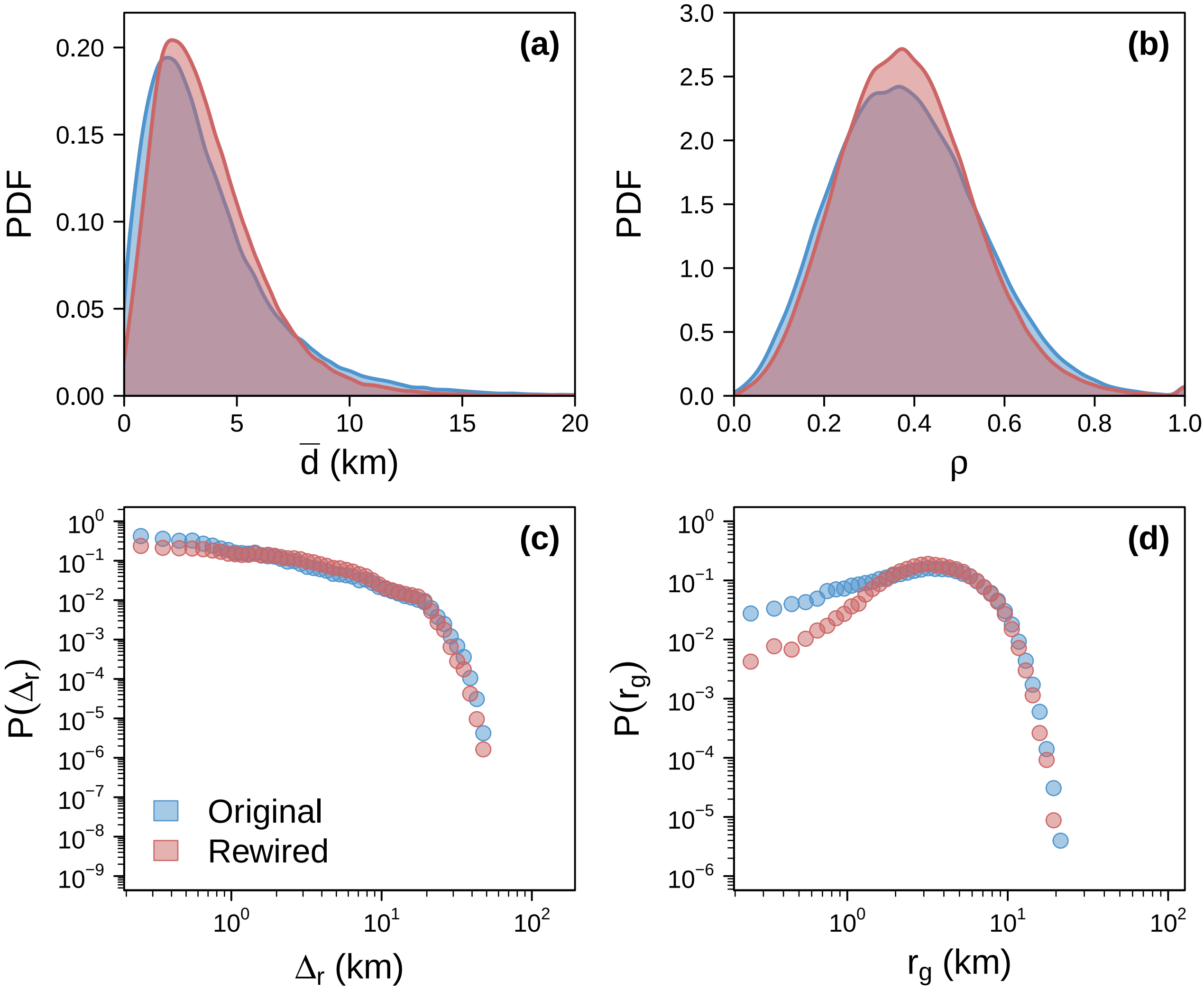

Figure S8. **Observed and simulated distributions of human mobility indicators in Madrid.** The distribution of jump lengths Δ_r, the radius of gyration r_g, the tendency to return to already visited places (ρ) and the individual average distance traveled ($\bar{d}$) are considered. Values measured on the empirical data are in blue, while those obtained after rewiring are in red. The calculation of Δ_r and r_g is based on the business' exact geographical coordinates.

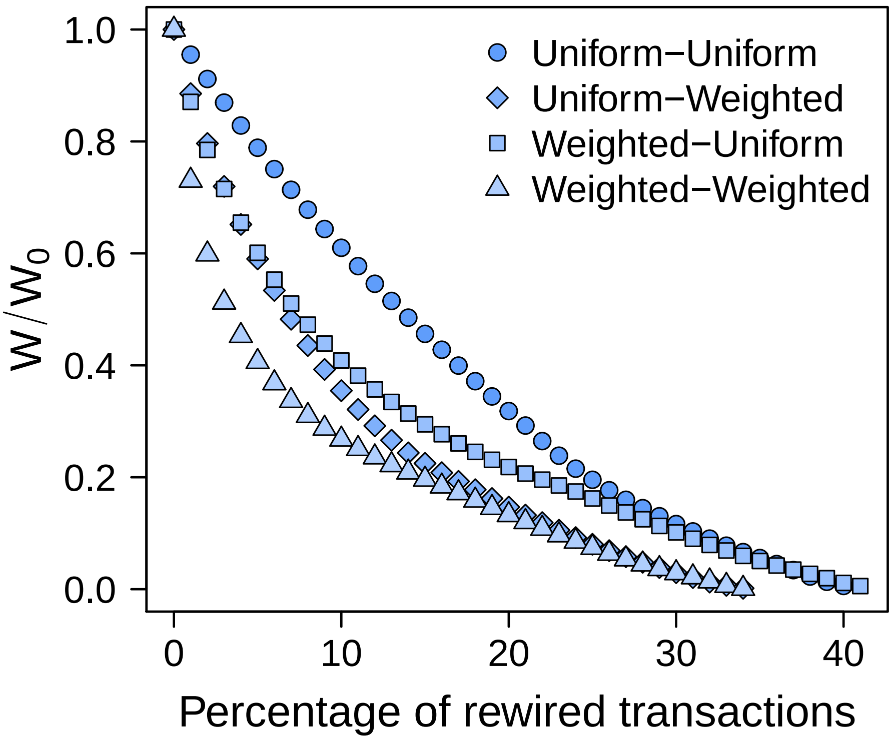

Figure S9. Decrease of wealth inequality among neighborhoods as a function of the fraction of transactions rewired, for various heuristics (Madrid). Four heuristics are considered, "Uniform-Uniform", "Uniform-Weighted", "Weighted-Uniform" and "Weighted-Weighted".

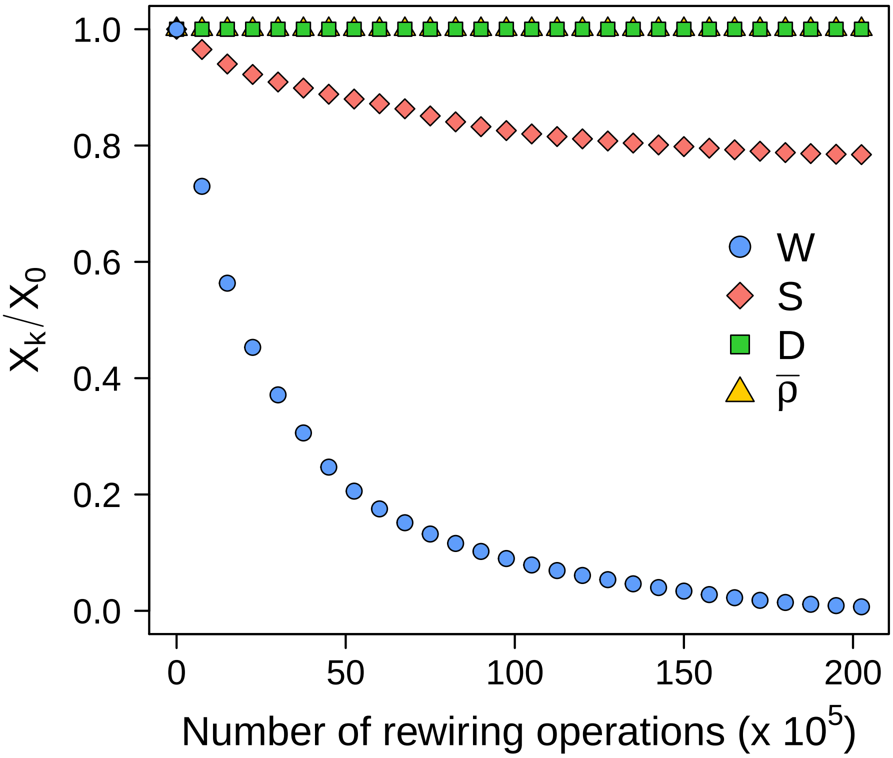

Figure S10. **Decrease of wealth inequality (W/W_0) while preserving the spatial mixing index (S/S_0), the distance traveled (D/D_0) and the exploration rate ($\bar{\rho}/\bar{\rho}_0$) as a function of the number of rewiring operations (Madrid case).** Values have been averaged over hundreds of replications. The bars represent the minimum and the maximum values obtained but in most cases they are too close to the average to be seen.

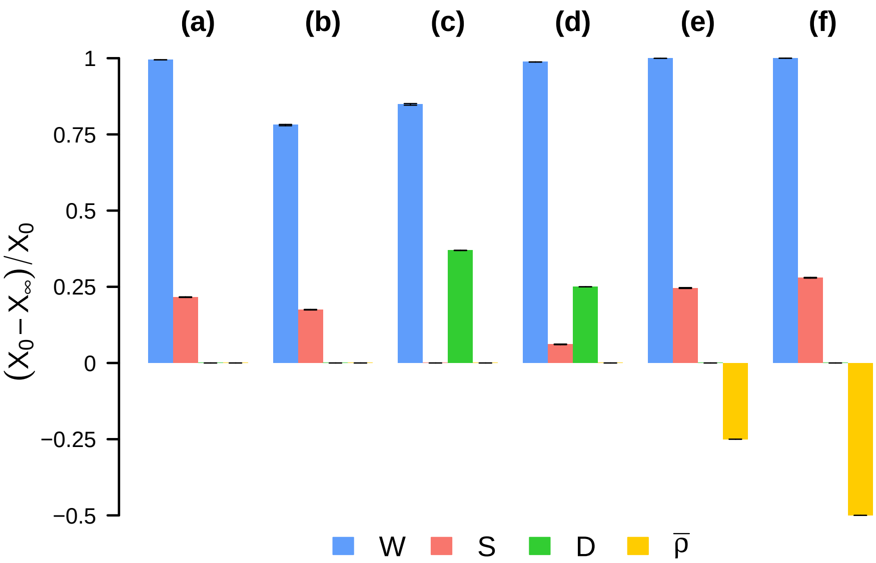

Figure S11. **Multi-criteria improvement of shopping mobility in the city of Madrid.** Each group of bars gives the relative gains or losses for the four indicators W, S, D and $\bar{\rho}$.