ESTEBAN MORO

A city is not a static tree: understanding urban areas through the lens of real-time behavioral data

La ciudad no es un árbol estático: comprender las áreas urbanas a través de la óptica de los datos de comportamiento en tiempo real

Abstract

Cities are the main ground on which our society and culture develop today and will evolve in the future. Against the traditional understanding of cities as physical spaces mostly around our neighborhoods, recent use of large-scale mobility datasets has enabled the study of our behavior at unprecedented spatial and temporal scales, much beyond our static residential spaces. Here we show how it is possible to use these datasets to investigate the role that human behavior plays in traditional urban problems like segregation, public health, or epidemics. Apart from measuring or monitoring such problems in a more comprehensive way, the analysis of those large datasets using modern machine learning techniques or causality detection permits to unveil of the behavioral roots behind them. As a result, only by incorporating real-time behavioral data can we design more efficient policies or interventions to improve such critical societal issues in our urban areas.

Keywords

Human behavior, mobility data, segregation, public health

A dynamic tree

Most understanding of our cities comes from the static mapping of data to the geographical layout of urban areas. The traditional census is a remarkable example of how data can help policymakers, citizens, and communities understand their communities better. The census is used to collect demographic data about the population, which is then used to inform decision-making at all levels of government. The data collected through the census helps to determine how federal funds are distributed to states and localities, how electoral districts are drawn, and how businesses locate themselves. The census is also used to track social and economic trends over time. The basic idea behind the census data is that “we are where we sleep,” that is, that our residential spaces determine most of our behavior and sometimes most of our life outcomes. That is, unfortunately, partially true. In the US, up to 50% of upward social mobility is determined by where we were

A city is not a static tree: understanding urban areas through the lens of real-time behavioral data

La ciudad no es un árbol estático: comprender las áreas urbanas a través de la óptica de los datos de comportamiento en tiempo real

ESTEBAN MORO

Esteban Moro, "A city is not a static tree: understanding urban areas through the lens of real-time behavioral data.", *ZARCH* 19 (diciembre 2022): 28-39. ISSN versión impresa: 2341-0531 / ISSN versión digital: 2387-0346.
https://doi.org/10.26754/ojs_zarch/zarch.2022197407
Orcid: 0000-0003-2894-1024

born[1]. The most predictive trait of getting a higher education is the zip code we spend our early years[2] in, and even our life expectancy depends strongly on where we live. Because of that, most of our current understanding of cities is still based on census or survey information.

However, increasing access to different and more efficient transportation modes, the new spatial organization of home-work-shopping urban areas, and more access to leisure time suggest that the areas covered by our activity spaces go beyond our residential neighborhoods. In fact, our research shows that in major US cities, most of the people we encounter or interact with live more than 15km away from us[3]. Not unexpectedly, we only spend 3% of our time with people living in our block. Even in an ideal "20-minute" city (where access to most services is within 20 minutes), the places where people meet and interact would be far away from their neighborhoods. A large part of our lives happen in those activity spaces, where most of our working hours, leisure, or social interactions happen. This suggests that the census may not be capturing the complete picture of our lives and our behavior. More interestingly, those activity spaces are more prone to favor highly diverse interactions with people of different socio-demographic groups. The soul of the city, understood as a big social machine, lies mainly in those interconnected activity spaces. .

As a consequence, those traditional static ways of understanding cities based on our residence fall short of anticipating, monitoring, or forecasting the rapid and complex evolution that our society

1 Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. "Is the United States still a land of opportunity? Recent trends in intergenerational mobility." *American Economic Review* 104, no. 5 (2014): 141-47.

2 Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment." *American Economic Review* 106, no. 4 (2016): 855-902.

3 Moro, Esteban, Dan Calacci, Xiaowen Dong, and Alex Pentland. "Mobility patterns are associated with experienced income segregation in large US cities." *Nature communications* 12, no. 1 (2021): 1-10.

is experiencing. While our residence might be fixed for years, activity spaces constantly change, even during the day. Moreover, most censuses or surveys are collected every ten years, leaving us explaining what happened ten years ago rather than nowcasting or predicting how our cities are evolving. The recent pandemic has highlighted the shortcomings of using outdated, non-integrated, and slowly processed data to manage and anticipate the spreading of COVID-19 and the particular relevance of real-time, more granular, and high-frequency data[4].

All of this, while companies like Google or Uber have real-time access to how people move and interact in urban areas! Today we have the potential to produce high-resolution updates of how people purchase, get to school, move, get a job, or interact by leveraging new digital sources of information from mobile data, social media, WiFi networks, phone apps, and credit cards[5]. Companies are using this wealth of data to micro-segment clients based on their demographic but also their behavioral traits. But cities are still using primary segments of census groups (residential areas, gender, age, unemployment) to map and intervene in problems such as inequality, gentrification, or transportation. In the last years, we have witnessed an enormous amount of scientific studies and projects focused on using these new digital data sources to estimate an abundance of demographic indicators. Satellite images and machine learning techniques are used to map poverty at high resolution[6], Facebook Ads are used to measure the gender digital gap in the whole world[7], and the analysis of social media posts allows us to understand the dynamics of unemployment[8]. These works show the potential of Big Data to complement official statistics, especially in situations where the spatial or temporal resolution of official statistics is not enough. Thus, cities are a particular instance in which Big Data can help most in policy-making. City halls are aware of this situation and have led this change by creating open data portals, collaboration with companies, or Smart Cities initiatives. Those projects mainly focus on measuring the efficient allocation of resources in problems like energy, pollution, or transportation, leaving behind important societal issues like inequality, segregation, gentrification, or public health.

Large-scale, high-resolution, longitudinal, and dynamical information from Big Data sources enables us to ask a much broader set of questions in the city. Specifically, it allows us to understand human behaviors in the activity space and their relationship with urban environment, and social or cultural context something which is not possible through traditional lab experiments, censuses or surveys. Yet, the increasing volume, complex structures, and dynamics of behavioral data, as well as the research questions on the complicated human interactions, stretch the limit of conventional methods. The emerging field of data-driven "Computational Social Science" is sparked by the massive amounts of digital records of human behaviors[9]. New technologies offer at least three advantages in urban contexts.

First, the data provides much higher resolution information both at a temporal and spatial scale, including granular information about user behaviors and individual connections. That resolution enables us to understand not only the "when" and "where" (where most Smart Cities initiatives end), but more importantly, the "why", i.e., the underlying mechanism for the human decision-making process and thus design interventions accordingly. Cities are living labs where millions of individuals perform a myriad of experiments every day. Choosing a new restaurant, testing a new route to the workplace, trying a new gym, or starting to exercise. For example, the same individual is

4 Aleta, Alberto, David Martin-Corral, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E. Dean et al. "Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19." *Nature Human Behaviour* 4, no. 9 (2020): 964-971.

5 Kandt, Jens and Michael Batty. Smart cities, big data and urban policy: Towards urban analytics for the long run. *Cities* 109 (2021): 102992.

6 Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. "Combining satellite imagery and machine learning to predict poverty." *Science*353, no. 6301 (2016): 790-794.

7 Garcia, David, Yonas Mitike Kassa, Angel Cuevas, Manuel Cebrian, Esteban Moro, Iyad Rahwan, and Ruben Cuevas. "Analyzing gender inequality through large-scale Facebook advertising data." *Proceedings of the National Academy of Sciences* 115, no. 27 (2018): 6958-6963.

8 Llorente, Alejandro, Manuel Garcia-Herranz, Manuel Cebrian, and Esteban Moro. "Social media fingerprints of unemployment." *PloS one* 10, no. 5 (2015): e0128692.

9 Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barabasi, Devon Brewer, Nicholas Christakis et al. "Social science. Computational social science." *Science (New York, NY)* 323, no. 5915 (2009): 721-723.

constantly exposed to different urban environments, which we could use as "treatments," mimicking the holy grail of random control trials in behavioral experiments. At the same time, those behavioral changes can be monitored to assess the impact of urban interventions more rapidly and more deeply, uncovering their actual effect on our cities. This is an essential aspect of having extensive, comprehensive data about human behaviors in the city: the level of detail of digital records allows us to disentangle the direct effect of interventions and separate them from other confounding variables. Most evaluations of policies are based on correlations or percentage changes e.g., in how people use a transportation route. We measure the success of a new route by the relative change in people using that bus. But many other confounders can explain that increase, like, for example, the rise in gas prices. Furthermore, a simple correlation cannot disentangle the effect of the intervention on different socio-demographic groups or how that intervention in a bus route impacted other transportation services in the city. Or any other process. In our previous example, a new route can give some people chances to work in a different environment that can foster more diverse interactions, business opportunities, etc.

As we know, the city is not a tree but rather a complex system of intertwined connections between different systems, areas, or people[10]. Traditional census data or statically defined surveys cannot incorporate that level of granularity to describe or anticipate unexpected relationships and consequences. Advances in machine learning, propensity matching, and causal techniques that are traditionally used by companies in their A/B experiments are starting to be incorporated into urban studies, giving us a much more controlled way to detect and monitor the relationship between people's behavior and urban structure and interventions. Only by having extensive and real-time data about people's behavior can we see the whole dynamics of the tree in the city rather than how a single branch sways in the wind.

Secondly, digital technologies enable us to have a more comprehensive picture of even small human behaviors at a city scale for an extended period. Small changes can result in profound evolutions in urban areas. For example, predicting gentrification has been an elusive task because it starts with minor changes in the behavior of small groups in particular areas. Only by having comprehensive data about how people move, visit and consume can we detect potential demographic and behavioral changes that lead to the gentrification of urban areas[11].

Similarly, it is impossible to prepare for extreme events using only traditional data based on surveys done five or ten years ago under everyday conditions. The recent pandemic, but also natural disasters like hurricanes, earthquakes, extreme heat waves, or flooding, have put our cities under unique stress situations which are hard to anticipate, monitor, or manage. But digital records of human behaviors can help us complement the action of humanitarian and relief agencies. For example, mobile phone data are now commonly used to monitor evacuation routes taken by individuals[12], or even social media posts are used to evaluate the damage of hurricanes in different areas[13]. Geolocated behavioral data from location-based apps on mobile phones was heavily used during the pandemic to investigate the effect of lockdowns, the number of people staying home, investigating the places where COVID-19 transmission events were happening, or even evaluate the potential impact of reopening policies in urban areas[14]. The pandemic is a clear example of the value of digital traces in understanding behavior in urban areas since the propagation of the virus, and its containment were mediated by sudden changes in the behavior of individuals. Without incorporating in real-time those changes of behavior, epidemiologists and public health officials cannot efficiently manage and act quickly.

Finally, digital traces can capture small but pervasive changes in the city for a long time, enabling us to understand how society evolves. Companies have been collecting these datasets for decades

10 Alexander, Christopher. *A city is not a tree*. Sustasis Press/Off The Common Books, 2017.

11 Cook, Cody. "Heterogeneous Preferences for Neighborhood Amenities: Evidence from GPS data." *Available at SSRN 4212524* (2022).

12 Deng, Hengfang, Daniel P. Aldrich, Michael M. Danziger, Jianxi Gao, Nolan E. Phillips, Sean P. Cornelius, and Qi Ryan Wang. "High-resolution human mobility data reveal race and wealth disparities in disaster evacuation patterns." *Humanities and Social Sciences Communications* 8, no. 1 (2021): 1-8.

13 Kryvasheyeu, Yury, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. "Rapid assessment of disaster damage using social media activity." *Science advances* 2, no. 3 (2016): e1500779.

14 Aleta, Alberto, David Martín-Corral, Michiel A. Bakker, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi et al. "Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas." *Proceedings of the National Academy of Sciences* 119, no. 26 (2022): e2112182119.

now, allowing us to study our cities longitudinally in time. For example, we can study the permanent impact of COVID-19 on our society. While the development of effective vaccines has successfully suppressed the mortality rates of COVID-19, the new behavioral habits and social norms that we have acquired during the pandemic, such as higher rates of work from home, and dramatic changes in physical activity, sleep, time use, and mental health, could have a long-lasting impact on society. For example, we found that our activity spaces have become less diverse (in terms of income diversity) than before the pandemic[15]. As a result, we are around 15% more segregated in our cities than in 2019. This is probably the most significant change in segregation in our cities in decades, adding to already substantial segregation in some of them. Without behavioral data from digital traces, it would take years to quantify it using traditional residence-based measures.

Good behavioral data

New technologies also come with new problems. Mobility data have a complex structure and content that can only be analyzed by data scientists with specialized training. Many researchers prepared to ask and answer crucial questions in urban research that require mobility data lack the technical skills to do so. Most public agencies working in the urban space are unprepared technically to deal with these datasets or with the sophisticated machine learning algorithms to process them.. Additionally, such data are highly vulnerable to invasions of privacy, as highlighted by previous scandals. Consequently, access to mobility is even further limited to a small set of researchers with close relationships with the companies that gather and organize these data. However, recent advances in differential privacy, aggregation of the data, or private-public alliances like “Data for Good” Initiatives show a growing interest and possibility to access and analyze non-traditional data through key partnerships and collaborations[16], or by sharing the data in aggregated formats. In this sense, machine learning techniques can be used to train generative models to generate synthetic datasets or low-embedding representations of the original data that allow sharing it without privacy concerns[17].

Another primary concern is the potential biases in the data. It is well-established that technology use is biased in multiple ways. Individuals of specific demographic backgrounds are less likely to use cell phones or use them less regularly, creating population-level biases in who is visible in the data. Furthermore, individuals may use their cell phones at different times, potentially overrepresenting certain activities and places and underrepresenting others. Current aggregation techniques, however, treat the underlying data as entirely accurate, and all subsequent analyses suffer from this assumption, inviting the “big data paradox” of systematically inaccurate and highly precise insights[18,19,20]. Since those biases are typically over-represented in more vulnerable populations that do not have access to digital technology, the blind use of these data can propagate and exacerbate those inequalities. Once again, the high level of detail of the mobility data from digital traces can help reduce the potential biases in the data by adapting traditional pre- or post- stratification techniques used in survey design to the mobility data. Pre-stratification techniques like defining socio-demographic panels from the company’s user base that have a better representation of each census area have been successfully used to study segregation behavior[21] and behavior during the pandemic[22]. Post-stratification of the results to the whole population is

15 Yabe, Takahiro, Bernardo Garcia Bulle Bueno, Xiaowen Dong, AlexSandy Pentland, and Esteban Moro. “Behavioral changes during the pandemic worsened income diversity of urban encounters.” *arXiv preprint arXiv:2207.06895* (2022).

16 See examples at https://datacollaboratives.org/.

17 Berke, Alex, Ronan Doorley, Kent Larson, and Esteban Moro. “Generating synthetic mobility data for a realistic population with RNNs to improve utility and privacy.” In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*, pp. 964-967. 2022.

18 Coston, Amanda, Neel Guha, Derek Ouyang, Lisa Lu, Alexandra Chouldechova, and Daniel E. Ho. “Leveraging administrative data for bias audits: Assessing disparate coverage with mobility data for COVID-19 policy.” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, (2021): pp. 173-184. 2021.

19 Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. “The parable of Google Flu: traps in big data analysis.” *Science* 343, no. 6176 (2014): 1203-1205.

20 Moro, Esteban, Dan Calacci, Xiaowen Dong, and Alex Pentland. “Mobility patterns are associated with experienced income segregation in large US cities.” *Nature communications* 12, no. 1 (2021): 1-10.

21 Moro, Esteban, Dan Calacci, Xiaowen Dong, and Alex Pentland. “Mobility patterns are associated with experienced income segregation in large US cities.” *Nature communications* 12, no. 1 (2021): 1-10.

22 Aleta, Alberto, David Martín-Corral, Michiel A. Bakker, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi et al. “Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas.” *Proceedings of the National Academy of Sciences* 119, no. 26 (2022): e2112182119.

also possible by employing traditional weighting techniques[23], as well as more sophisticated methods[24]. However, in most cases, mobility data from digital traces lacks information about the socio-demographic features of the users. In those cases, small surveys or controlled experiments in which opted-in users share their mobility data can be an invaluable source for understanding and correcting the potential biases in the data. Fortunately, a number of "data cooperatives" and experimental platforms exist to recruit and even engage individuals in participatory experiments of "citizen science"[25]. Complementing large datasets collected passively from millions of users with well-defined small surveys would create better behavioral data that portray everybody in the city justly.

Behavior, slow and fast

The behavior of the city is the result of millions of actions of its citizens, trying to optimize in real-time their day, resources, work, leisure, or social connections. The urban infrastructure constrains their chosen actions, forcing them to behave similarly. However, it still is hard to imagine how public transportation agencies, businesses, or city officials can predict so well traffic, the number of people buying groceries or using the public library. The underlying reason explaining this conundrum is that human behavior is very repetitive and, thus, very predictable. Studies of human mobility using GPS or mobile phones have discovered that we spend 75% of our time only in 5 places in the city[26]. As a result, 90% of our movements are predictable[27]. Our behavior, especially on working days, is organized around home, work, children, groceries, and the gym. And we constantly repeat the time and order of those actions throughout the year, instinctively, even without thinking about it.

On the contrary, the restaurant we choose for friends or family dinner or how we spend Sunday morning can take us hours to decide, resulting in a highly unpredictable decision. A decision based on dozens of contextual pieces of information, time constraints, and social restrictions. As a result, recommendation algorithms have a hard time predicting precisely the restaurant you would like to go to for dinner. Although they are very good at forecasting that dinner will happen with a high probability. This slower, deliberative, and more logical behavior is hard to predict and typically happens during weekends or holidays.

The dichotomy between these two modes of behavior or mobility is similar to Kahneman's "fast" and "slow" modes of thinking[28]. Human behavior in the city is the result of the combination of "fast" actions that are highly predictable, regular in time, and almost instinctive, and "slow" deliberative and more logical actions that are harder to anticipate. Slow behavior typically results in the exploration of new places, while fast behavior takes us to return to already visited places[29]. Recent studies have found that slow and fast mobility is not equally distributed across the population. Some individuals perform more routine repetitive movements (returners), while others have a more explorative behavior (explorers). High-income, more educated individuals tend to have more explorative behavior than low-income, less educated and living in minority neighborhoods[30]. Although the results vary significantly by geography, probably signaling the effect of the urban environment and cultural or economic characteristics of different cities.

As a result of highly predictable and regular movements, our lives in the city tend to be similar. Despite the considerable complexity of the mobility of millions of users visiting millions of places

23 Salganik, Matthew J. *Bit by bit: Social research in the digital age*. Princeton University Press, 2019.

24 Zagheni, Emilio, and Ingmar Weber. "Demographic research with non-representative internet data." *International Journal of Manpower* (2015).

25 Fritz, Steffen, Linda See, Tyler Carlson, Mordechai Muki Haklay, Jessie L. Oliver, Dilek Fraisl, Rosy Mondardini et al. "Citizen science and the United Nations sustainable development goals." *Nature Sustainability* 2, no. 10 (2019): 922-930.

26 Schläpfer, Markus, Lei Dong, Kevin O'Keeffe, Paolo Santi, Michael Szell, Hadrien Salat, Samuel Anklesaria, Mohammad Vazifeh, Carlo Ratti, and Geoffrey B. West. "The universal visitation law of human mobility." *Nature* 593, no. 7860 (2021): 522-527.

27 Song, Chaoming, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. "Limits of predictability in human mobility." *Science* 327, no. 5968 (2010): 1018-1021.

28 Kahneman, Daniel. *Thinking, fast and slow*. Macmillan, 2011.

29 Pappalardo, Luca, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. "Returners and explorers dichotomy in human mobility." *Nature communications* 6, no. 1 (2015): 1-8.

30 Moro, Esteban, Dan Calacci, Xiaowen Dong, and Alex Pentland. "Mobility patterns are associated with experienced income segregation in large US cities." *Nature communications* 12, no. 1 (2021): 1-10.

every day, we have found that they can be described by only twelve latent interpretable activity behaviors on how people combine shopping, eating, working, or using their free time[31]. The whole lifestyle of people is made up of the weighted combination of those "genes" of behavior that all of us have but in different combinations. For example, we find that people that make trips to errands also visit fast food outlets frequently, while heavy users of public transportation (bus) also spend much time in the neighborhood and entertainment. Working life is also related to nightlife. Some people have more working life behaviors, some others more local trips, and most of us go shopping during the weekend. But all of us have those twelve "genes" of behavior.

Those latent behaviors represent combined aspects of our life that occur concurrently and which could be used to devise successful holistic interventions to change people's lifestyles. For example, people that run many errands might choose fast food because they are time-poor or because errands take place around specific food environments (food swamps). Our results can help design public health interventions that incorporate those distinct lifestyles to identify those routines and habits that are most risky for health. Finally, those latent behaviors are independent of the area people live or their socio-demographics and could enrich the current Census by including the composition of the different latent behaviors to study urban areas. In fact, we found that they are much associated with dynamics like experience segregation, transportation, or healthy behaviors in cities, even after controlling for demographic features.

Segregated behavior

Rising economic inequality and its spatial counterpart, segregation, are one of the most critical problems worldwide, and combating inequality has moved to the forefront of the policy debate, the different agendas, and sustainable development goals for 2030. Despite this, our current understanding of inequality or segregation and its relationship with other urban problems like transportation, gentrification, or even social participation is still based on residential census or survey information.

In the last few years, different research groups have started to get a deeper understanding of the multidimensional nature of inequality and segregation in our cities. Using massive data sets of geolocalized mobile phone data calls or social media posts, different researchers have shown the intertwined nature of segregation and mobility in urban areas. For example, even though residents of disadvantaged neighborhoods travel far and wide, they mostly visit poor or underprivileged communities, and thus, their relative isolation and segregation persist[32,33]. Segregation in mobility is also seen at the level of gender or race: women's travel patterns are different from men's, and access and use of public transportation mediate this difference[34]. Segregation happens even at different times of the day due to the different mobility rhythms in the city of different ethnic groups[35]. Segregation does not only happen in the spatial domain but also in the social-network domain. For example, social media and credit card data[28] and mobile phone calls[36] were used to show that income segregation has the same fingerprint in social media as the one found in spatial mobility. Not only do people move to areas that are similar in income, they even use social media or mobile phone calls to communicate with similar people, even if they are in the other part of the city. These findings challenge our traditional understanding of segregation and inequality in the city. Segregation is not only reflected in the economic or racial sorting within neighborhoods; it is encoded into our behavior

31 Yanni Yang, Alex Pentland, Esteban Moro. "Identifying latent activity behaviors and lifestyles using mobility data to describe urban dynamics." *arXiv preprint arXiv:2209.12095* (2022).

32 Wang, Q., Phillips, N. E., Small, M. L., & Sampson, R. J. (2018). Urban mobility and neighborhood isolation in America's 50 largest cities. *Proceedings of the National Academy of Sciences*, *115*(30), 7735-7740.

33 Dong, Xiaowen, Alfredo J. Morales, Eaman Jahani, Esteban Moro, Bruno Lepri, Burcin Bozkaya, Carlos Sarraute, Yaneer Bar-Yam, and Alex Pentland. "Segregated interactions in urban and online space." *EPJ Data Science* 9, no. 1 (2020): 20.

34 Gauvin, Laetitia, Michele Tizzoni, Simone Piaggesi, Andrew Young, Natalia Adler, Stefaan Verhulst, Leo Ferres, and Ciro Cattuto. "Gender gaps in urban mobility." *Humanities and Social Sciences Communications* 7, no. 1 (2020): 1-13.

35 Silm, Siiri, and Rein Ahas. "The temporal variation of ethnic segregation in a city: Evidence from a mobile phone use dataset." *Social Science Research* 47 (2014): 30-43.

36 Xu, Yang, Alexander Belyi, Paolo Santi, and Carlo Ratti. "Quantifying segregation in an integrated urban physical-social space." *Journal of the Royal Society Interface* 16, no. 160 (2019): 20190536.

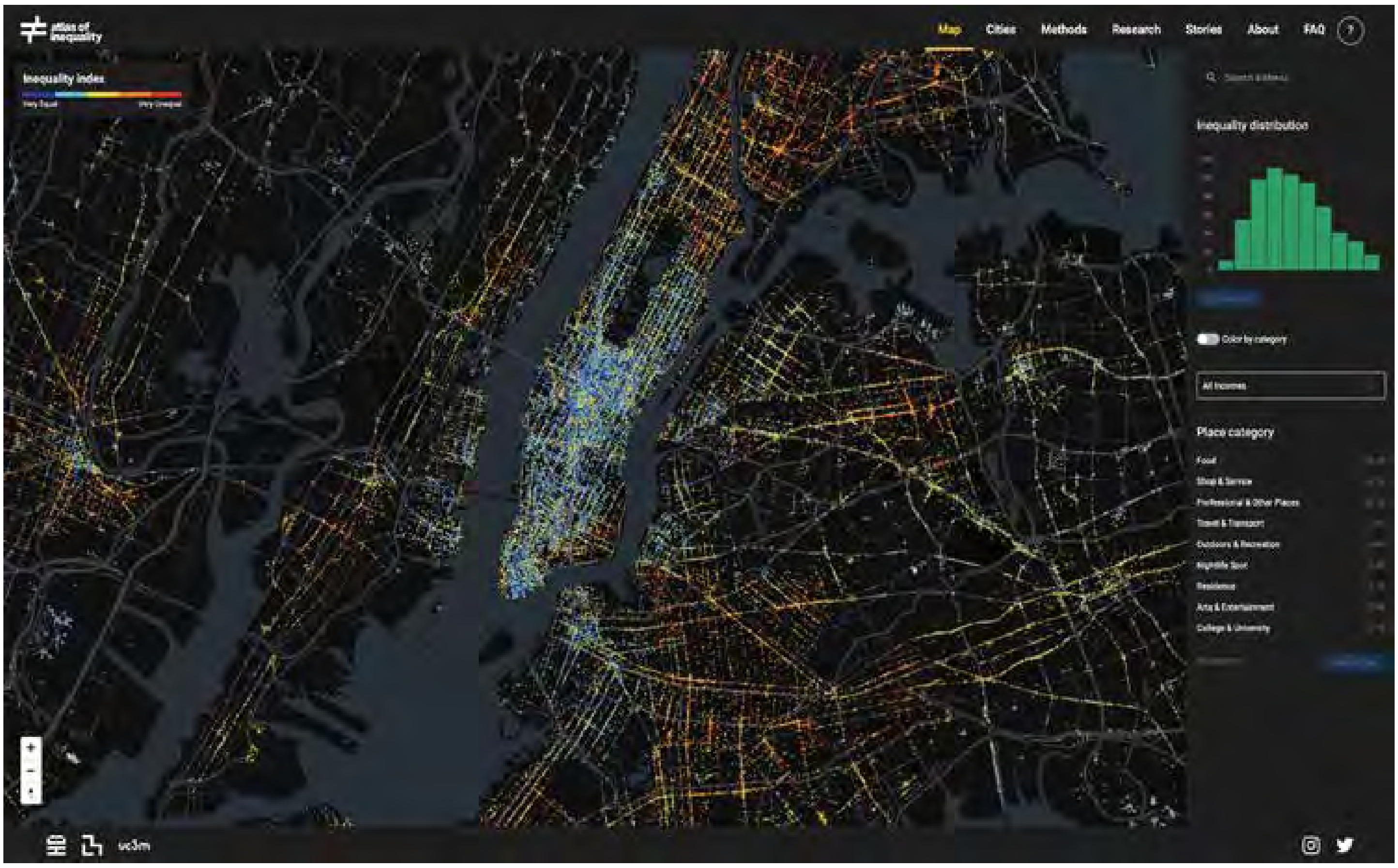

Figure 1. Snapshot of the "Atlas of Inequality" (for the city of New York) by MIT & UC3M, an interactive platform where users can navigate the inequality in the city by place (restaurant, shop, park, etc.). The Highly integrated places, visited by many different income groups, are colored in blue, while highly segregated are colored in red.

as we move, interact, or even communicate with the rest of the city. However, most of the traditional policies to improve social integration are place-based, such as affordable housing and easy access to transportation. The initial results from mobility data show that those policies misalign with how people experience segregation and are not sufficient per se to achieve better integration in our cities.

Since segregation is encoded into our behavior, we need a better understanding of people's movements, choices, and opportunities to mix with other social classes. However, the spatial resolution of data from mobile phone calls or social media used in previous studies does not allow us to investigate which places, activities, or opportunities are the main drivers (or barriers) of social integration in our cities. Leveraging high-resolution data from mobile phone apps and other sources, we have built the "Atlas of Inequality"[37] (see Figure 1) a platform and research project to understand the "where" and the "why" of how people experience segregation as they move, work, shop in the city[38]. Using data from location-based services apps in millions of anonymous mobile phones, we investigate the segregation where it happens, that is, at places that are visited by people.

The platform shows two crucial new dimensions of segregation in our cities. The first one is that segregation happens even at the street level. Places located in the very same block can have a very different composition of visitors: mixed places can be only a few dozen meters away from those that are highly segregated, even just across the street. Where we get coffee, where we buy groceries, and where we grab take-out often reflect our choices and in turn, determine the kinds of people we interact with every day. It's important to note that these choices that people make are usually constrained by things like affordability, location, and social groups. But as a result, socioeconomic inequality in exposure to other groups is encoded in part by these choices, not just where people live. The second

37 http://inequality.media.mit.edu

38 Moro, Esteban, Dan Calacci, Xiaowen Dong, and Alex Pentland. "Mobility patterns are associated with experienced income segregation in large US cities." *Nature communications* 12, no. 1 (2021): 1-10.

important result is that the type of place matters for their segregation. In the Atlas of Inequality, we can see that some types of restaurants (fast food) are much more segregated than healthy choices (e.g., vegetarian or Asian). Or that some workplaces like factories or warehouses are more segregated than offices or co-working spaces. Also, most cultural and entertainment places do not work towards integrating people: most of the museums, classical music, or theaters are highly segregated, with the only exception of Science Museums which are also one of the most integrated places in all the cities studied. Still, the most segregated places in cities are schools and worship places. These results hold even if we controlled for the uneven distribution of different types of places in the cities.

The fine-grained structure of place segregation and the fact that individuals move much larger distances than home census areas challenge the notion that segregation in cities is driven by stationary locations (e.g., home, work, school environments). Official statistics show that the average person's commute is 40 miles per day in the US and 50 km in the EU. Thus, geographical accessibility within our neighborhoods to a particular type of place is not a limitation for a person to be segregated or not. To understand the reason why people experience segregation, in the second part of the project, we analyzed the mobility patterns of 4.5 million people to study how their choices of places determine their segregation patterns. Results demonstrated that experienced inequality depends much more on where people spend their time than where they live or work. Specifically, around 45% of experienced inequality depends on the home neighborhood demographic characteristics. In comparison, 55% of segregation is attributable to our lifestyles, i.e., the set of choices and opportunities we make when we move around the city and visit places. This significant result shows that segregation is an emergent behavioral process that appears at the level of places rather than a static attribute of regions or neighborhoods.

Finally, we found that explorers, those that have more "slow" behaviors in the city, are the ones more integrated. However, returners that mostly have repetitive, habitual behaviors experience a more segregated life in the city. Once again, those different kinds of behavior are related to different aspects of our life. The findings in this research show potential paths to identify proper interventions to make public spaces, transportation, and community businesses more equitable. Although "place-based" interventions might still be relevant, interventions that help people change their lifestyles or behavior might translate into more significant effects on the segregation of people in urban areas. Beyond individual or single-place interventions, our research shows that potential changes in behavior could happen in "exploratory" aspects of our lives rather than possible immutable regular patterns. And that if we want to deploy interventions to alleviate the experience segregation of urban dwellers, we should act upon whole latent behaviors configuring our lifestyles, rather than slightly changing a particular aspect of our mobility.

Healthy behaviors

Segregation is not the only significant problem in our city. The pandemic has revealed the profoundly rooted health problems in our society in general and urban areas in particular. Large cities have been hit harder by COVID-19 with more cases and, sadly, more deaths. The very density that makes cities more diverse, more creative, and better economically is what creates more exposure to others, more contagion, and, thus, more infections. And before the pandemic, cities were worse in terms of diseases related to health risks like less physical activity, more exposure to pollution, and co-morbidities like obesity or diabetes.

Although there is a significant genetic component in some health outcomes, our behavior is equally and sometimes more critical. This became evident during the pandemic when our mobility was directly associated with the possibility of being infected. Most social-distancing interventions were deployed to reduce people's mobility and hence exposure to other people. Mobility data was then an invaluable source to understand who, when, and where that mobility was reduced and, if not, where transmission risk was higher. Using our large datasets of mobility data and epidemiological models during the pandemic, we could characterize, for example, the transmission of COVID-19 across different settings. During the pandemic's first wave, we estimate that only 18% of individuals produce most infections (80%), with about 10% of events that can be considered superspreading events (SSEs). Although mass gatherings present an important risk for SSEs, we estimate that the bulk of transmission occurred in smaller events in settings like workplaces, grocery stores, or food venues. Once again, the returning "fast" repetitive behavior was responsible for most of the infections during the first wave, primarily due to the impossibility of replacing it[39]. Since the mobility behavior of people changed during the first and subsequent waves, we also found that places most important for transmission change during the pandemic and that they were different across cities. For example, the emphasis that some policies put on restaurants or other venues was not justified in specific moments or in some cities. Most of the infections after the lockdowns were happening in grocery stores or at home. Our results and methodology putting together high-resolution mobility data with epidemiological models allow for a real-time data-driven analysis that connects non-pharmaceutical interventions, human behavior, and the transmission of COVID-19 to provide quantitative information that can aid in defining more targeted and less disruptive interventions not only a local level but also to assess whether local restrictions could trigger undesired effects at nearby locations not subject to the same limitations.

Another prevalent behavior associated with health outcomes in our cities is food consumption. Poor diets are a leading cause of morbidity and mortality in the world. Exposure to low-quality food environments, including those with a large fraction of fast-food outlets (FFO), is hypothesized to impact diet and health negatively. Low-quality built food environments are generally categorized into two types. "Food deserts" are defined as areas with low access to healthy foods, while "Food swamps" are areas saturated with less healthy food outlets, often defined as neighborhoods that have a higher number of fast-food outlets (FFO) and convenience stores or a high ratio of these outlets relative to healthier food outlets. While exposure to both types of food environment has been associated with increases in unhealthy eating and diet-related disease, overall, findings are mixed and predominantly null. Billion-dollar interventions in the US to change the food environments around people's homes have demonstrated no meaningful impact on diet quality or diet-related disease outcomes. The limited focus on residential and static food environments may be one explanation for these mixed results, given that a growing proportion of food acquisition and consumption occurs miles from our homes. In our research using large-scale mobility data, we found, for example, that people travel on average 6.5km to get fast food, while they only go 0.75km for groceries[40]. Thus, the mobile food environments people are exposed to and food outlets they visit as they move through the day are quite different from their neighborhood environments. Since food consumption is repeated every day, mobility data extracted from digital traces can tell us not only where food is acquired but also about the food environment where those decisions are made. Moreover, for most users, we have different events and contexts where those decisions are made (around the workplace, around their home, in the mall, etc.), so we can use that data to even investigate the effect of food environments even on the individual level. Using that granularity and semi-causal machine learning techniques, we were able to mimic the experimental situation of having the same user exposed to different food environments to detect their impact on going to fast-food places. We found that 10% more fast-food places in an area increase the odds of people visiting a fast-food place by approximately 20%. And since most users get food outside their neighborhoods, those mobile food environments have a more significant impact on their diet than what they have around their homes.

This finding allows us to investigate better places in food environments to promote more healthy food consumption. Using our mobility data, we found optimal locations for intervention are a combination of where i) the prevalence of fast-food places is the highest, ii) most decisions about food are made, and most importantly, iii) where visitors' food decisions are more susceptible to the environment. As a result, most efficient interventions may be further from people's homes and not even closer to "food deserts" or "food swamps." Our data revealed that more successful interventions could be deployed in areas like airports and other transportation hubs, workplaces, or shopping areas, up to two times more efficient in reducing visits to fast food restaurants.

Behavioral spaces

Large-scale mobility data has permitted us to understand the behavior of urban areas and their citizens at an unprecedented scale. Although the focus of most urban interventions, urban design,

39 Aleta, Alberto, David Martín-Corral, Michiel A. Bakker, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi et al. "Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas." *Proceedings of the National Academy of Sciences* 119, no. 26 (2022): e2112182119.

40 Garcia-Bulle, Bernardo, Abigail L. Horn, Brooke M. Bell, Mohsen Bahrami, Burcin Bozkaya, Alex Pentland, Kayla de la Haye, and Esteban Moro . "You are where you eat: Effect of mobile food environments on fast food visits." *medRxiv* (2022).

and urban research has been around physical spaces, we have to start incorporating in the description of our cities those "behavioral spaces" where individuals make their decisions. Those spaces are dominated by the balance between routine and exploratory behavior, our lifestyles, or even the temporal and localized influence we get from the physical space when we make decisions. For most people, those behavioral spaces are built around their activity spaces, far away from home. For that reason, traditional census data cannot fully describe them. Complementing our census with behavioral data can yield a much better understanding of our cities.

Interestingly, most of the digital traces used to reconstruct mobility in urban areas come from marketing companies that use that data for products, restaurants, transportation, or other recommendations. Individuals are increasingly using more algorithms to get better recommendations about places to visit, restaurants to eat, goods to be delivered, routes to navigate, or even jobs to apply. Hyper-personalization of those algorithms can unintendedly reinforce the income, racial or political bubble that we live in, creating more segregation in our cities. Or we could increase our visitations to fast-food places if this is what we like. But those algorithms are also the perfect platform in which we can change people's behavior by recommending more integrated places or activities in the city. The interplay between algorithms, people's behavior on physical and online spaces, and policies is going to determine the future of behavior in our cities.

The city is not a tree, is a complex system of interdependencies between urban environment and people's behavior. But those interedepencencies change in time. Cities today look very different as they were before the pandemic. And not only the pandemic. Changes due to new mobility modalities, working from home, or climate change are modifying every day the "social and structural overlap of communities"[41]. Large digital traces about human behavior in the city is the lens that is going to allow us to measure and probably to predict those changes.

Bibliography

Aleta, Alberto, David Martin-Corral, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E. Dean et al. 2020. "Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19." *Nature Human Behaviour* 4, no. 9: 964-971.

Aleta, Alberto, David Martín-Corral, Michiel A. Bakker, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi et al. 2022. "Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas." *Proceedings of the National Academy of Sciences* 119, no. 26: e2112182119.

Alexander, Christopher. 2017. *A city is not a tree*. Sustasis Press/Off The Common Books.

Berke, Alex, Ronan Doorley, Kent Larson, and Esteban Moro. 2022. "Generating synthetic mobility data for a realistic population with RNNs to improve utility and privacy." In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*, pp. 964-967.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014. "Is the United States still a land of opportunity? Recent trends in intergenerational mobility." *American Economic Review* 104, no. 5: 141-47.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment." *American Economic Review* 106, no. 4: 855-902.

Coston, Amanda, Neel Guha, Derek Ouyang, Lisa Lu, Alexandra Chouldechova, and Daniel E. Ho. 2021. "Leveraging administrative data for bias audits: Assessing disparate coverage with mobility data for COVID-19 policy." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*:173-184.

Cook, Cody. 2022. "Heterogeneous Preferences for Neighborhood Amenities: Evidence from GPS data." *Available at SSRN 4212524*.

Deng, Hengfang, Daniel P. Aldrich, Michael M. Danziger, Jianxi Gao, Nolan E. Phillips, Sean P. Cornelius, and Qi Ryan Wang. 2021. "High-resolution human mobility data reveal race and wealth disparities in disaster evacuation patterns." *Humanities and Social Sciences Communications* 8, no. 1: 1-8.

Dong, Xiaowen, Alfredo J. Morales, Eaman Jahani, Esteban Moro, Bruno Lepri, Burcin Bozkaya, Carlos Sarraute, Yaneer Bar-Yam, and Alex Pentland. 2020. "Segregated interactions in urban and online space." *EPJ Data Science* 9, no. 1: 20.

Fritz, Steffen, Linda See, Tyler Carlson, Mordechai Muki Haklay, Jessie L. Oliver, Dilek Fraisl, Rosy Mondardini et al. 2019. "Citizen science and the United Nations sustainable development goals." *Nature Sustainability* 2, no. 10: 922-930.

Garcia, David, Yonas Mitike Kassa, Angel Cuevas, Manuel Cebrian, Esteban Moro, Iyad Rahwan, and Ruben Cuevas. 2018. "Analyzing gender inequality through large-scale Facebook advertising data." *Proceedings of the National Academy of Sciences* 115, no. 27: 6958-6963.

Garcia-Bulle, Bernardo, Abigail L. Horn, Brooke M. Bell, Mohsen Bahrami, Burcin Bozkaya, Alex Pentland, Kayla de la Haye, and Esteban Moro. 2022. "You are where you eat: Effect of mobile food environments on fast food visits." *medRxiv*. https://doi.org/10.1101/2022.09.20.22280128

41 Alexander, Christopher. *A city is not a tree*. Sustasis Press/Off The Common Books, 2017.

Gauvin, Laetitia, Michele Tizzoni, Simone Piaggesi, Andrew Young, Natalia Adler, Stefaan Verhulst, Leo Ferres, and Ciro Cattuto. 2020. "Gender gaps in urban mobility." *Humanities and Social Sciences Communications* 7, no. 1: 1-13.

Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. "Combining satellite imagery and machine learning to predict poverty." *Science* 353, no. 6301: 790-794.

Kahneman, Daniel. 2011. *Thinking, fast and slow.* Macmillan.

Kandt, Jens, Michael Batty. 2021. "Smart cities, big data and urban policy: Towards urban analytics for the long run. " *Cities* 109: 102992. https://doi.org/10.1016/j.cities.2020.102992

Kryvasheyeu, Yury, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. 2016. "Rapid assessment of disaster damage using social media activity." *Science advances* 2, no. 3: e1500779.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barabasi, Devon Brewer, Nicholas Christakis et al. 2009. "Social science. Computational social science." *Science (New York, NY)* 323, no. 5915: 721-723.

Llorente, Alejandro, Manuel Garcia-Herranz, Manuel Cebrian, and Esteban Moro. 2015. "Social media fingerprints of unemployment." *PloS one* 10, no. 5: e0128692.

Moro, Esteban, Dan Calacci, Xiaowen Dong, and Alex Pentland. 2021. "Mobility patterns are associated with experienced income segregation in large US cities." *Nature communications* 12, no. 1: 1-10.

Pappalardo, Luca, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. 2015. "Returners and explorers dichotomy in human mobility." *Nature communications* 6, no. 1: 1-8.

Salganik, Matthew J. 2019. *Bit by bit: Social research in the digital age*. Princeton University Press.

Schläpfer, Markus, Lei Dong, Kevin O'Keeffe, Paolo Santi, Michael Szell, Hadrien Salat, Samuel Anklesaria, Mohammad Vazifeh, Carlo Ratti, and Geoffrey B. West. 2021. "The universal visitation law of human mobility." *Nature* 593, no. 7860: 522-527.

Silm, Siiri, and Rein Ahas. 2014. "The temporal variation of ethnic segregation in a city: Evidence from a mobile phone use dataset." *Social Science Research* 47: 30-43.

Song, Chaoming, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. "Limits of predictability in human mobility." *Science* 327, no. 5968: 1018-1021.

Wang, Q., Phillips, N. E., Small, M. L., & Sampson, R. J. 2018. Urban mobility and neighborhood isolation in America's 50 largest cities. *Proceedings of the National Academy of Sciences* 115, no. 30: 7735-7740.

Xu, Yang, Alexander Belyi, Paolo Santi, and Carlo Ratti. 2019. "Quantifying segregation in an integrated urban physical-social space." *Journal of the Royal Society Interface* 16, no. 160: 20190536.

Yabe, Takahiro, Bernardo Garcia Bulle Bueno, Xiaowen Dong, AlexSandy Pentland, and Esteban Moro. 2022. "Behavioral changes during the pandemic worsened income diversity of urban encounters." *arXiv preprint arXiv:*2207.06895.

Yanni Yang, Alex Pentland, and Esteban Moro. 2022. "Identifying latent activity behaviors and lifestyles using mobility data to describe urban dynamics." *arXiv preprint arXiv*: 2209.12095. https://doi.org/10.48550/arXiv.2209.12095

Zagheni, Emilio, and Ingmar Weber. 2015. "Demographic research with non-representative internet data." *International Journal of Manpower* 36, no. 1: 13-25. https://doi.org/10.1108/IJM-12-2014-0261

Esteban Moro. Universidad Carlos III de Madrid - MIT Connection Science. Esteban Moro is a researcher, data scientist, and professor at MIT Connection Science and Universidad Carlos III (UC3M) in Spain. He has published extensively throughout his career (more than 100 articles) and has led many projects funded by government agencies and/or private companies. Esteban's work lies in the intersection of big data and computational social science, with particular attention to human dynamics, collective intelligence, social networks, and urban mobility in problems like viral marketing, natural disaster management, or economic segregation in cities. He has received numerous awards for his research, including the "Shared University Award" from IBM in 2007 for his research in modeling viral marketing in social networks and the "Excellence in Research" Awards in 2013 and 2015 from UC3M. Esteban's work appeared in major journals, including Nature, PNAS, and Science Advances, and is regularly covered by media outlets The Atlantic, The Washington Post, The Wall Street Journal, and El País (Spain).