

RESEARCH

Open Access



Street context of various demographic groups in their daily mobility

Ariel Salgado^{1*} , Weixin Li², Fahad Alhasoun³, Inés Caridi¹ and Marta Gonzalez^{3,4}

*Correspondence:

arielofalsalgado@gmail.com

¹ Instituto de Cálculo,

University of Buenos Aires,

Buenos Aires, Argentina

Full list of author information is available at the end of the article

Abstract

We present an urban science framework to characterize phone users' exposure to different street context types based on network science, geographical information systems (GIS), daily individual trajectories, and street imagery. We consider street context as the inferred usage of the street, based on its buildings and construction, categorized in nine possible labels. The labels define whether the street is residential, commercial or downtown, throughway or not, and other special categories. We apply the analysis to the City of Boston, considering daily trajectories synthetically generated with a model based on call detail records (CDR) and images from Google Street View. Images are categorized both manually and using artificial intelligence (AI). We focus on the city's four main racial/ethnic demographic groups (White, Black, Hispanic and Asian), aiming to characterize the differences in what these groups of people see during their daily activities. Based on daily trajectories, we reconstruct most common paths over the street network. We use street demand (number of times a street is included in a trajectory) to detect each group's most relevant streets and regions. Based on their street demand, we measure the street context distribution for each group. The inclusion of images allows us to quantitatively measure the prevalence of each context and points to qualitative differences on where that context takes place. Other AI methodologies can further exploit these differences. This approach presents the building blocks to further studies that relate mobile devices' dynamic records with the differences in urban exposure by demographic groups. The addition of AI-based image analysis to street demand can power up the capabilities of urban planning methodologies, compare multiple cities under a unified framework, and reduce the crudeness of GIS-only mobility analysis. Shortening the gap between big data-driven analysis and traditional human classification analysis can help build smarter and more equal cities while reducing the efforts necessary to study a city's characteristics.

Keywords: Mobility, Network, Street, Images, Context, Segregation, GIS, CDR, AI, CNN

Introduction

Network science has already established itself as an interdisciplinary field, with tools that have been applied in a myriad of fields (Costa et al. 2011). Ecological systems (Provata et al. 2008), social groups (Vega-Redondo 2007), and software (Decan et al. 2019), are only a few examples of applications. Its use for spatial characterization (Barthélemy 2011) comprises a sub-field that provides methodologies for analyzing

urban areas (Jiang and Claramunt 2004; Cheng et al. 2013). One of the most used representations is the street network, where nodes play the role of street intersections and edges the role of streets. This representation has been fruitful [see section B1 in Barthélemy (2011) for a detailed review], allowing to characterize a city's possibilities in terms of potential mobility. Incorporating human mobility information (Jiang et al. 2013) (for example in the form of Call Detail Records, CDR) can help contrast the potential view given by the street network with actual traveling. Estimating travel demand (Toole et al. 2015), measuring congestion during peak hour traffic (Çolak et al. 2016), and road demand in connection with departing car sources (Wang et al. 2012) are some examples. Combining the best of both views can help to understand the relevance of streets (Toole et al. 2015; Wang et al. 2012), both in terms of the possibilities they offer, and the demand they receive.

Understanding the street user's view of the street network improves with the addition of mobility information, as it makes possible to weight each street based on the use each street receives. Still, multiple dimensions regarding the perception of the places and the opportunities they provide fall behind this type of data. New technologies provide solutions to this type of characterizations using street imagery-based analysis (Gebru et al. 2017; Naik et al. 2014). Methodologies based on street imagery allow researchers to detect land use (Cao et al. 2018), street context (Alhasoun and González 2019), measure income of a neighborhood (Gebru et al. 2017), perceived safety (Naik et al. 2014), or even detect car accident risk (Kita and Kidziński 2019). While manual image classification is costly, artificial intelligence (AI) can help to massively classify them, making the procedures scalable to multiple orders of magnitude without extra effort.

The topic of racial segregation in U.S. cities has a long history (Iceland and Sharp 2013; Iceland and Weinberg 2002), and has been deeply studied by multiple authors. Segregated regions of the town consist of areas where mainly only one racial/ethnic group inhabits. The process in which these segregated regions arise is not trivial, but the result of multiple factors regarding work opportunities, social connections, and the city's building (Sherman 2014; Sanchez et al. 2003; Fox 2017; Galster 1988). The possibilities offered by big data and machine learning methodologies can help to detect on greater scale aspects of the urban life in which racial inequalities in the opportunities and segregation manifest (Benthall and Haynes 2019; Gebru et al. 2017; Naik et al. 2014; Candipan et al. 2021; Hofstra and de Schipper 2018; Zhou et al. 2020).

Here we propose a data analysis framework based on the ease of managing the information layers provided by the network representation and the power of AI to manage big amounts of images (Alhasoun and González 2019). We study each racial/ethnic group's street demand of Boston city using a CDR-based model [TimeGeo (Jiang et al. 2016)]. TimeGeo allows us to represent human mobility through synthetically generated daily trajectories based on real CDR data. We use the street network to construct paths between different regions of the city. We add an extra information layer to the street network, representing each street's context, using both human and AI-based labeled street imagery. We characterize each group's street context during its daily activities using the algorithm presented in Alhasoun and González (2019). The addition of mobility with massively gathered data characterized using AI points the direction of further studies on

combining contextual information, like the recorded by mobile devices, with the regions through which persons move daily.

Materials and methods

The proposed data analysis framework uses three sources of information: city information in GIS, street view imagery, and daily trajectories over the city. City information regarding population demographic and housing partition is obtained through the 2010 Census (2016), and the street network is obtained through OpenStreetMaps (2017). The street imagery is obtained through Google Street View (Anguelov et al. 2010; Alhasoun and González 2019). The daily trajectories are generated with the CDR-based model TimeGeo (Jiang et al. 2016).

Street view imagery

To provide a context for every street in Boston city, we worked with two sets of images, downloaded from Google Streets View (Anguelov et al. 2010). The first set H consists of 23, 927 images, and the second set A consists of 5998 images. We use the street network geographic representation to obtain images over the streets based on their latitude–longitude locations. Downloaded images represent a horizontal view with an ocular opening of 45° each. To have a uniform sample over the streets, we sampled a subgroup of the GIS points representing the streets and downloaded two images at each of these points. This guarantees an equal representation of street images over all the geography represented by the streets.

To classify the set H , we followed the San Francisco methodology (Newsom 2010), which assigns to each image one out of ten contexts: Alley, Downtown Commercial (Downtown Comm.), Commercial Throughway (Commercial Trwy.), Highway, Highway Ramp (Highway Rmp.), Industrial, Neighborhood Commercial (Neighborhood Comm.), Neighborhood Residential (Neighborhood Res.), Park and Residential Throughway (Residential Trwy.).

The classification process can be summarized as follows (Alhasoun and González 2019):

- Side use context is determined based on land use information. This information separates images on commercial and residential use.
- Transportation context is determined based on the characteristics of the street. This includes highway, throughway, downtown, and neighborhood.
- Special conditions on the street lead to categories like Park, Alley, and Industrial.

The number of images of each type can be found in the Table 1:

The AI-based labeling methodology follows Alhasoun and González (2019) using the set A of images. Images in set A were downloaded in pairs corresponding to the same latitude–longitude location, each one comprising the half of a 90° view. We combine the paired images to cover the complete angle of 90° covered by them following the procedure indicated in Alhasoun and González (2019). Images in set A have been selected in a way that both images in each pair have the same context, as required by Alhasoun and González (2019). We split set A into two sets, A_T and A_P , with sizes 4804

Table 1 Number of images in each street context, in the three sets: H (human-labeled), A_T (human-labeled set used to train AI), and A_P (AI-labeled set)

Set	Alley	Commercial Trwy.	Downtown Comm.	Highway	Highway Rmp.
H	1141	2964	2208	2531	2031
A_T	179	535	530	469	421
A_P	39	104	136	119	108
Set	Industrial	Neighborhood Comm.	Neighborhood Res.	Park	Residential Trwy.
H	1993	2408	4063	1646	2942
A_T	355	352	1115	367	481
A_P	79	102	277	109	121

Table 2 Fraction of correctly assigned (F.C.A.) images for each category in the training set A_T

Category	Alley	Commercial Trwy.	Downtown Comm.	Highway	Highway Rmp.
F.C.A.	0.983	0.972	0.989	0.985	0.983
Category	Industrial	Neighborhood Comm.	Neighborhood Res.	Park	Residential Trwy.
F.C.A.	0.975	0.972	0.982	0.997	0.952

and 1194. Images in A_T are manually labeled in the same manner explained for set H . We use images in set A_T to train a convolutional neural network (CNN) (Szegedy et al. 2016). This CNN identifies patterns in groups of pixels, seeking identifiable objects associated with each street context [see Fig. 6 in Alhasoun and González (2019)]. Several layers allow the CNN to associate parks with trees, borders of highways with highways, big streets, and certain shops with neighborhood commercial zones and other particular patterns for each category. Finally, we use the resulting CNN to predict the labels of images in A_P . After manually checking images in A_P , we found 97% of correctly assigned categories in the label prediction considering all categories together (the fraction of correctly assigned images per category in training set A_T can be found in Table 2). The resulting number of images in each context for sets A_T and A_P can be found in Table 1. The overall result is above 95% and the mean above 97%.

GIS from census and street network

We used the US 2010 Decennial Census (2016) tracts and their associated information regarding the number of self-reported persons in the racial/ethnic groups of White, Black, Hispanic, and Asian. We categorized each census tract based on the population fraction in each group (fraction of self-reports in each racial/ethnic group). A tract is labeled after group g if the biggest population fraction corresponds to g and also that fraction is over 0.5. Tracts which do not meet these conditions are left unlabeled. The resulting group labels are shown in Fig 1A, while Table 3 presents the amount of tracts and city area of each group. The city presents typical segregation patterns (Iceland and Weinberg 2002). Whites are majoritarian with more than half of the labeled tracts and half the area of the city in their group. Blacks follow, with around 15% of the Boston

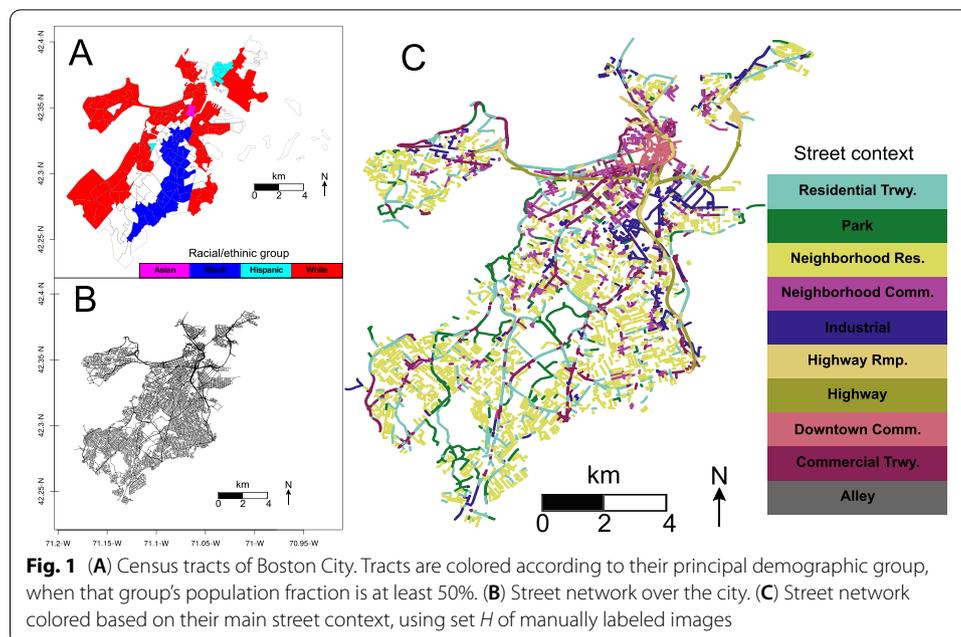


Table 3 Number of tracts and area for Boston and for each group

	Boston	White	Black	Hispanic	Asian
Tracts	180	88	30	8	2
Area (km ²)	128.4	65.5	18.7	2.3	0.5

area. Most of the Hispanic group is concentrated in the northern region of the city, while Asians are associated with only two tracts in the downtown.

Figure 1B shows the street network of the city. Notice how Hispanic and White tracts at the north are heavily dependent on the highways connecting them to the rest of the city. To associate street context to the streets, we assign a buffer radius of 10 m to each position where an image was sampled, and associate the corresponding image to every street intersecting the buffer radius. This buffer was selected to acknowledge the width of the various types of streets. This process results on a set of categorized images for each street. We define the street context of street e as a vector u_e , where component u_e^i represents the fraction of images associated with street section e and context i . Figure 1C shows the street network after labeling each street using the images in set H . Each street is labeled after the context i that maximizes u_e^i , when u_e has a maximum value over 0.5. Streets in contexts like Highways and Commercial Trwy. connect the city.

Daily mobility and TimeGeo trajectories

TimeGeo (Jiang et al. 2016) is a primarily CDR-based Jiang et al. (2013) mechanistic modeling framework that generates urban mobility patterns with a resolution of 10 min and approximately 400 m. It generates daily trajectories, representing aspects like activity stay duration, visited Location Points (LPs) per day, and daily mobility networks. TimeGeo divides LPs into three categories: *home*, *work*, and *other*, depending

on the activity realized in that LP. Both *home* and *work* locations are uniquely defined, meaning that each agent has only one associated *home* LP, and one associated work LP. However, they may transit to both of them many times during their daily trajectory. Other-type activities have different behavior. Agents may arrive at multiple different *other* LPs during the day. The spatial choices of visited locations are modeled by a rank-based exploration and preferential return mechanism (Jiang et al. 2016). These locations are originally selected from a grid over the city, with 400-m side cells representing the region where the activity occurs.

For this work's purpose, we do not consider the different types of activities performed by each user. Furthermore, we convert the original trajectories consisting of latitude–longitude pairs to sequences of tracts, associating each point to the census tract containing it. So a trajectory consists of a sequence representing the tracts where the activities are performed. We discard from the trajectories sequential activities realized on the same tract. Our focus is to consider the effects over the four different groups. Thus we only consider trajectories departing from tracts with a defined group label, and label each trajectory associated with that group. Table 4 presents the number of considered trajectories per group. Note that TimeGeo generates trajectories proportionally to the tract population. As tracts are approximately equal in population, then the number of trajectories per group is proportional to the number of tracts in that group.

To relate the sequences of census tracts with the street network geographic representation, we use the OSMnx (Boeing 2017) python package which downloads information from OpenStreetMaps (2017). We construct the street path between two census tracts using the graph representation of the city streets. The network edges represent the street segments and the nodes represent the street intersections, so we refer to the edge corresponding to street segment e with the same letter. When going from census tract A to census tract B , the starting and ending points are selected as the centroids of tract A and B , respectively. Then, we identify the geographically nearest street network nodes to the centroids using OSMnx, and calculate the shortest path between them in the street network's graph representation. This process converts a travel between tract A and tract B into a sequence of edges of the network (with each edge being associated with a street segment) used to reach the centroid of A to the centroid of B . This process ultimately converts every TimeGeo trajectory in a set of street network edges representing the shortest path among the tracts' centroids in which the trajectory's activities occur.

While considering that all the trajectories from a tract start from its centroid conflates all of them in a unique trajectory, this reduction can still be realistic. The majority of the trajectories are expected to follow the shortest and fastest routes between locations. Given that a trajectory departs from a location in tract A to reach another site in tract B , we can expect that a significant portion of the route connecting both locations will use of the fastest route connecting A and B , which can be seen as the

Table 4 Number of trajectories considered for each group

White	Black	Hispanic	Asian
218,038	68,945	17,053	4785

one connecting their centroids. This approximation becomes better as the distance between locations becomes greater.

Results

Network weighting by edge demand

We use the sequence of edges constructed from each trajectory to weigh each street network edge based on the number of times that edge appears on the trajectories of each group. The weight of an edge e , d_e^g equals the number of times edge e is included in the trajectories departing from tracts in group g . d_e^g represents the demand over edge e of group g . The total demand over edge e is calculated as $d_e = \sum_g d_e^g$. To consider how the demand over an edge is shared among the groups, we compute $f_e^g = d_e^g / d_e$, the fraction of the total demand over edge e corresponding to group g . To detect if an edge has a particularly high fraction of demand, we compute the total demand of each group $d^g = \sum_e d_e^g$, the total demand $d = \sum d^g$, and the corresponding fraction of demand $f^g = d^g / d$. We set a threshold to separate the demand levels into three categories. This threshold measures the deviation of the fraction of demand over an edge (f_e^g) from what could be expected by chance (f^g). We consider that the demand for an edge e is high if $\log_2(f_e^g / f^g) > 0.2$, low if $\log_2(f_e^g / f^g) < -0.2$ and average otherwise.

Figure 2 presents the street network for each racial/ethnic group, with edge width representing the square root of group demand d_e^g , and colors representing demand (high,

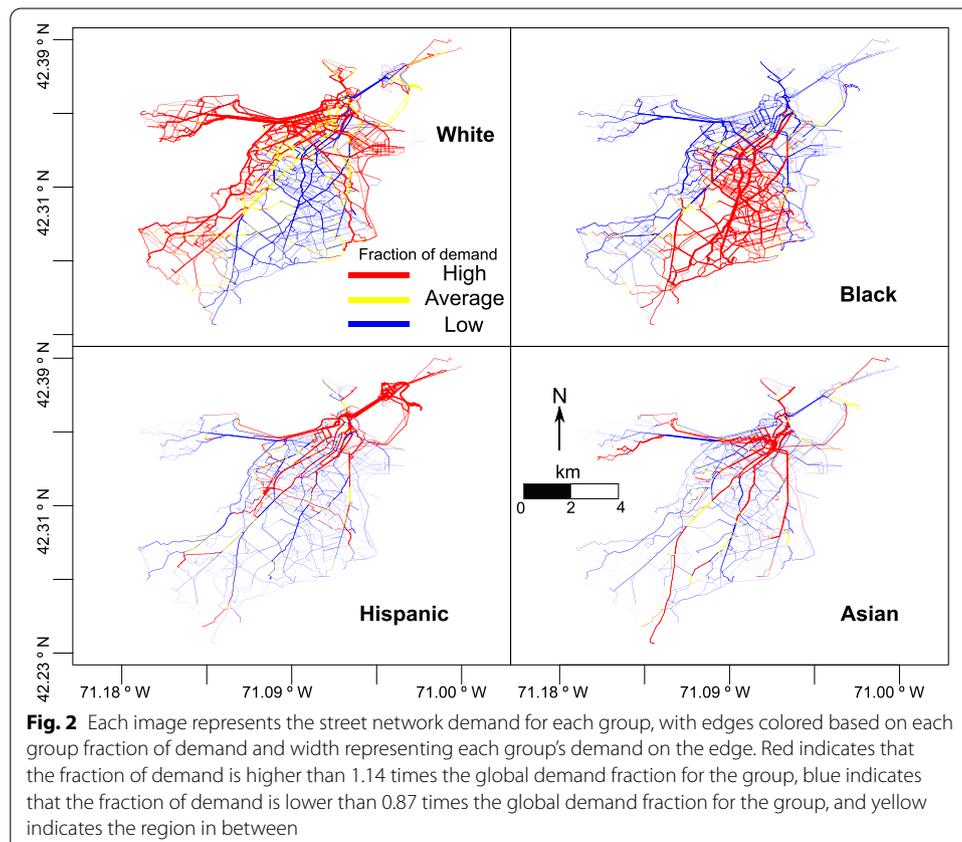


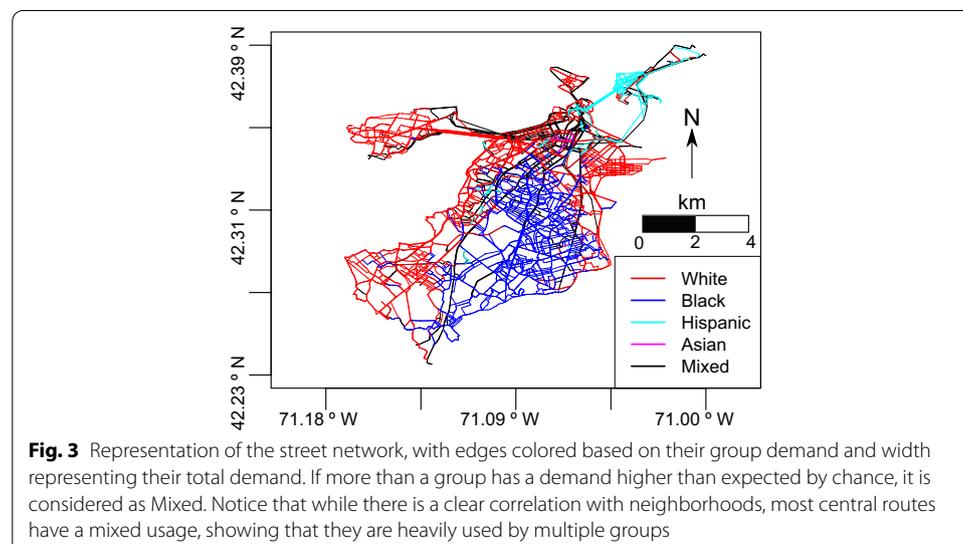
Fig. 2 Each image represents the street network demand for each group, with edges colored based on each group fraction of demand and width representing each group's demand on the edge. Red indicates that the fraction of demand is higher than 1.14 times the global demand fraction for the group, blue indicates that the fraction of demand is lower than 0.87 times the global demand fraction for the group, and yellow indicates the region in between

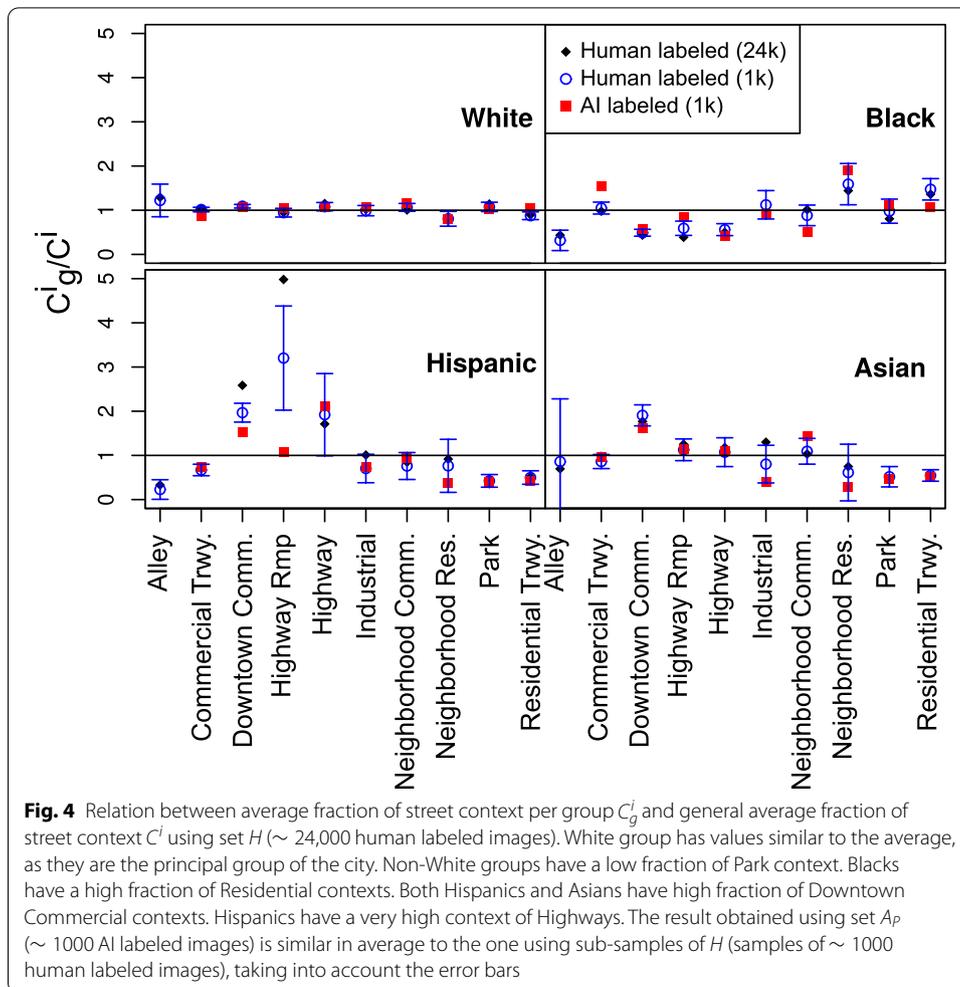
average, or low). Notice that while there is a clear correlation with neighborhoods, Hispanic and Asian groups have great demand on the long routes over the city center. Also, while groups may have high demand over an edge (compared to the rest of the edges), they may not be their principal users (as other groups may have a higher demand over them).

Considering edge e , we assign it to group g if $f_e^g > f^g$ only for group g . If $f_e^g > f^g$ for more than one group, we consider it as Mixed. The result of this partitioning can be found in Fig. 3. What we see for Whites and Blacks can be expected from the neighborhood distribution of Fig. 1A, as their most demanded streets correspond geographically with their neighborhoods. Hispanics and Asians show an interesting pattern of dispersion from the tracts associated with them to the rest of the city, through the principal routes of the city. Hispanics heavily use the highways connecting the north of the city.

Street context

Once we defined each street network edge's demand, we can include the information regarding the different types of street contexts over each street. As mentioned, we associate to each street the images from set H within 10 meters distance. We represent this process's result with a vector u_e^i , indicating the proportion of images representing context i and intersecting edge e . We calculate the average fraction of context i for group g as $C_g^i = \frac{1}{d^g} \sum_e d_e^g u_e^i$. C_g^i indicates the average fraction of streets in context i that a group member g meets during their daily activities. To compare with the general population, we also compute $C^i = \frac{1}{d} \sum_e d_e u_e^i$, the average fraction of context i in the total population considered. The relation C_g^i/C^i is presented in Fig. 4 for each group, calculated using the human and AI labeled images. The White group has a distribution of C_g^i similar to the average, as they are the city's main group. Non-White groups have a low fraction of Park context compared to the average, while White group has a slightly greater fraction than the average. The Black group has a high fraction of Residential contexts compared to the average. Both Hispanic and Asian groups have a high fraction of Downtown Commercial contexts compared to the average.





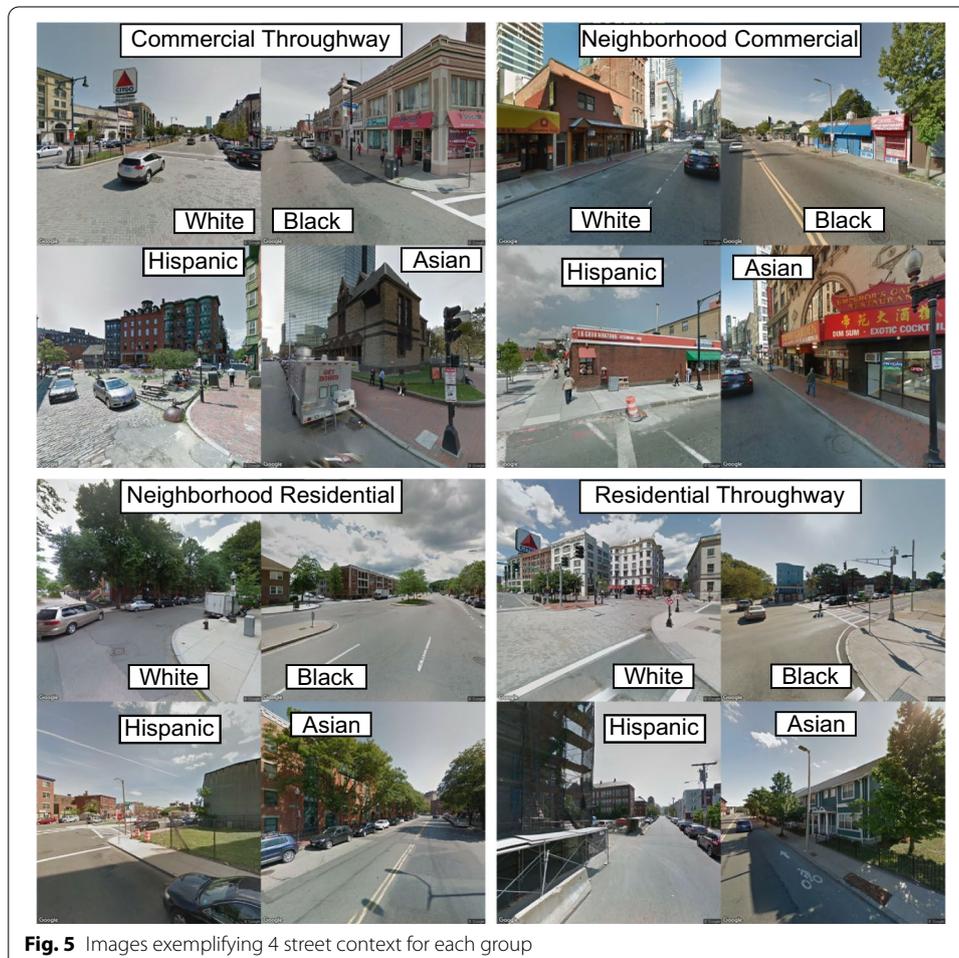
The Hispanic group has a very high context of highways (Highway and Highway Ramp) compared to all other groups. Considering the tracts partition into groups, the Black group mobility occurs at commercial and residential zones of their group. Asian group is concentrated within city’s downtown, with high usage of highways to reach other parts of the city. Compared to the rest, the Hispanic group has the highest context of highways as they have to reach the rest of the city from their principal region at the north of the city.

An important step is to test AI labeling methods, as this is the first step towards extending these types of analysis where street context data is not manually labeled. We applied the same analysis used with set H to set A_p . To compare the results, we randomly sampled 50 subsets of images from H with equal size to A_p , which we call h_k , with k representing the sampled subset. We calculated mean value and standard deviation of C_g^i/C^i over the 50 h_k samples, as presented in Fig. 4. The results are consistent, considering the error bars of the sampled results. White and Black group have the highest demand d and the biggest number of tracts, reducing the variation produced by the smaller image sample size in their C_g^i/C^i value. As Hispanic and Asian groups have lower representation, extreme effects like the Highway Ramp over average context in the

Hispanic group are difficult to capture. Due to the low sample size and the low representation of Alley context in the city, the Alley fraction of context is not well captured by set A_p . Still, the general behavior is maintained at both sample sizes, and it is consistent considering the standard deviation. Increasing the sample size can reduce the variation by increasing the representation of each context.

The quantitative street context analysis points general differences among the four groups, as their context distribution has evident differences. Still, the power of images does not stop here. To point further possibilities which can be explored using other AI algorithms (Cao et al. 2018; Gebu et al. 2017; Naik et al. 2014; Kita and Kidziński 2019), we include images from set H from 4 contexts (Commercial Twry., Neighborhood Comm., Neighborhood Res. and Residential Twry.), one image per group, in Fig. 5. These images were selected from each group's 18 most demanded streets in the particular context.

Comparing the images from Commercial Twry., for Whites we see a wide and newly made cobblestone street, with a boulevard and publicity in the background. For the Blacks we see old fashioned shops, asphalt streets, and a mailbox. For Hispanics we see a deteriorated cobblestone street, with big residential buildings and a plaza in the middle. For Asians we see a street with a large building and a mailbox.



For Asians, downtown architecture with modern buildings and a traffic light. Comparing the four, Whites and Asians show many more cars and more modern architecture.

Comparing the Neighborhood Comm., we see local restaurants for each group. Whites show a Coffee place, in a region with big buildings. Blacks shows multiple little shops, one with Caribbean food. Hispanics have a pedestrian street with a Hispanic restaurant. Asians show multiple Asian shops, particularly a big restaurant in a central street. All the building styles are different, with Whites and Asians in a central zone, Blacks in a more residential zone, and Hispanics in a pedestrian street.

Comparing Neighborhood Res., for Whites we see a little street with multiple trees and a pedestrian cross street to multiple houses with stairs. For Blacks we see a boulevard with buildings on one side. For Hispanics we see a varying type of architecture, with low and tall buildings, and a wasteland on the street's corner. For Asians we see bigger buildings and the smaller sidewalk of the four groups. While all the images show brick houses, Whites and Asians have many more cars parked, and Hispanics show elements absent in the others (like the wasteland).

Comparing Residential Trwy., for Whites we see again the newly made cobblestone street, from a point of view that shows how the street continues to residential areas. For Blacks, we see a similar building style to the one in Neighborhood Comm. and Commercial Trwy., this time in a boulevard without trees and a park-alike zone at the side. For Hispanics we see a tinier street than for Whites and Blacks, with a big buildings, lots of parked cars and a building in construction. For Asians we see a two ways street with lower houses. Notice that only the Hispanic street does not show any signal for bikes, and also there is a bigger number of cars parked than in the rest.

Overall, we see that Whites and Asians have a similar downtown-alike context, with Asians appearing to be more restricted to the city center. Blacks buildings have an older appearance, with most worn colors. Hispanics are located in more suburban places, including wastelands and old made brick streets, thinner streets and less signals.

This comparison showed the presence (or absence) of unique elements in the images of each group that can be exploited by pattern recognition methodologies specifically designed for that purpose. The comparison also shows how even the same context can look very different based on the particular elements present for each group. State of the cobblestone streets, wastelands, a shop's name language, bike-friendly indications, buildings' age and cars parked can be used as indicators for further analysis beyond the initial categorization.

Conclusions

The presented work points to useful connections between two urban analysis branches: street demand, which is already a combination of street information and daily mobility, and contextual image analysis. This combination shortens the gap between what people actually encounter during their daily activities, and what we are able to capture through the data traces their leave behind. Working with massive information regarding mobility and its connections to the city's possibilities has already proven useful, while its crudeness leaves out aspects regarding the street user's perception, aside from what GIS data sets could offer (in the form of pre-made labels). The inclusion of images and the AI

methodologies to massively characterize them can help to realize comparative studies between cities. This framework can power up urban planning and sociological analysis, allowing to put in a unified framework information from very different regions. We hope these results motivate the use of urban imagery to enrich information and communication data. This would inform applications that relate to improving ambient exposures and urban design considering diverse social groups.

Following the street network demand analysis, we saw long routes connecting different parts of the city. These routes are essential to Hispanic and Asian groups, which are concentrated in a few tracts but still move through the city. Assigning a group to each street results in a street partition similar to the tracts partition, with frontiers represented by streets with mixed demand between groups.

Street context analysis in combination with the mobility patterns provides a novel manner to explore each group's daily experiences. The context of each group reflects mostly the regions where they live. Whites are similar to the average, as they are the city's most populous group. The non-White group's exposure to Park context is below the Whites' exposure to Park context. As the Hispanic group is concentrated in the north of the city, it is heavily exposed to highways, more than any other group. This type of analysis can point out differences in the neighborhood composition and in mobility patterns (for example, the use of highways which are not particularly attached to any neighborhood), helping urban planners conceive more accessible and equal cities.

IA-labeled and human-labeled images provide consistent results in the studied data set. IA-labeling poses a promising methodology, applicable at big scale based on machine learning and network science techniques, making the analysis portable and automatic worldwide. Overall, the proposed framework points to a useful combination of data sets, allowing the quantitative characterization of context-based both in imagery and daily activities.

The comparison among groups for four particular contexts shows the images' usefulness as analysis departing material. Direct inspection shows different styles of restaurants, building construction and street makeup for each group, even while the street context is the same. The presence or absence of some elements (like street construction, bike signals and worn paint) suggests new challenges that other AI methodologies can further exploit by focusing on aspects like comfort, street makeup or the streets' safety.

We believe that this type of analysis can power up urban science methodologies based on street appearance, indicating where to focus (through demand) and reducing working costs (through AI), helping to think more accessible cities and easing the identification of inequalities between different demographic groups.

Abbreviations

CDR: Call-detail record; LP: Location point; GIS: Geographic information systems; Comm: Commercial; Trwy: Throughway; Rmp: Ramp; Res: Residential.

Acknowledgements

We thank Yanyan Xu for kind his help on providing insights for the CDR-based mobility model.

Authors' contributions

AS calculated street demand and context distributions and was a major contributor in writing the manuscript. WL identified the street network shortest paths using OSMnx. FA categorized the street images manually. WL and FA implemented the AI classification. IC and MG coordinated the work. All authors read and approved the final manuscript.

Funding

MCG was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231.

Availability of data and materials

Call-Detail Records are personal information and cannot be shared. Census information can be obtained through the US Census website <https://www.census.gov>. Street information can be obtained through OpenStreetMaps website, <https://www.openstreetmap.org>. Street images can be downloaded contacting with Google <https://developers.google.com/maps/documentation/streetview/overview>. The data sets generated during the current study are available from the corresponding author on reasonable request.

Declaration**Competing interests**

The authors declare that they have no competing interests.

Author details

¹Instituto de Cálculo, University of Buenos Aires, Buenos Aires, Argentina. ²Massachusetts Institute of Technology, Cambridge, MA, USA. ³Department of Civil and Environmental Engineering, UC Berkeley, Berkeley, CA, USA. ⁴Department of City and Regional Planning, UC Berkeley, Berkeley, CA, USA.

Received: 2 March 2021 Accepted: 25 May 2021

Published online: 12 June 2021

References

- Alhasoun F, González M (2019) Streetify: using street view imagery and deep learning for urban streets development. In: 2019 IEEE International conference on big data (big data). IEEE, pp 2001–2006
- Anguelov D, Dulong C, Filip D, Frueh C, Lafon S, Lyon R, Ogale A, Vincent L, Weaver J (2010) Google street view: capturing the world at street level. *Computer* 43(6):32–38
- Barthélemy M (2011) Spatial networks. *Phys Rep* 499(1):1–101
- Benthall S, Haynes BD (2019) Racial categories in machine learning. In: Proceedings of the conference on fairness, accountability, and transparency, pp 289–298
- Boeing G (2017) OSMNx: a Python package to work with graph-theoretic OpenStreetMap street networks. *J Open Source Softw*. <https://doi.org/10.21105/joss.00215>
- Candipan J, Phillips NE, Sampson RJ, Small M (2021) From residence to movement: the nature of racial segregation in everyday urban mobility. *Urban Stud*. <https://doi.org/10.1177/0042098020978965>
- Cao R, Zhu J, Tu W, Li Q, Cao J, Liu B, Zhang Q, Qiu G (2018) Integrating aerial and street view images for urban land use classification. *Remote Sens* 10(10):1553
- Cheng J, Bertolini L, le Clercq F, Kapoen L (2013) Understanding urban networks: comparing a node-, a density-and an accessibility-based view. *Cities* 31:165–176
- Çolak S, Lima A, González MC (2016) Understanding congested travel in urban areas. *Nat Commun* 7:10793
- Costa LDF, Oliveira ON Jr, Travieso G, Rodrigues FA, Villas Boas PR, Antiquera L, Viana MP, Correa Rocha LE (2011) Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Adv Phys* 60(3):329–412
- Decan A, Mens T, Grosjean P (2019) An empirical comparison of dependency network evolution in seven software packaging ecosystems. *Empir Softw Eng* 24(1):381–416
- Fox W (2017) Segregation along highway lines. Thesis, University of Buffalo, History Department & Honor's College
- Galster G (1988) Residential segregation in American cities: a contrary review. *Popul Res Policy Rev* 7(2):93–112
- Gebru T, Krause J, Wang Y, Chen D, Deng J, Aiden EL, Fei-Fei L (2017) Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the United States. *Proc Natl Acad Sci* 114(50):13108–13113
- Hofstra B, de Schipper NC (2018) Predicting ethnicity with first names in online social media networks. *Big Data Soc* 5(1):2053951718761141
- Iceland J, Sharp G (2013) White residential segregation in us metropolitan areas: conceptual issues, patterns, and trends from the us census, 1980 to 2010. *Popul Res Policy Rev* 32(5):663–686
- Iceland J, Weinberg DH (2002) Racial and ethnic residential segregation in the United States 1980–2000. Bureau of Census, Suitland
- Jiang B, Claramunt C (2004) Topological analysis of urban street networks. *Environ Plan B Plan Des* 31(1):151–162
- Jiang S, Fiore GA, Yang Y, Ferreira J Jr, Frazzoli E, González MC (2013) A review of urban computing for mobile phone traces: current methods, challenges and opportunities. In: Proceedings of the 2nd ACM SIGKDD international workshop on urban computing, UrbComp'13
- Jiang S, Yang Y, Gupta S, Veneziano D, Athavale S, González MC (2016) The TimeGeo modeling framework for urban mobility without travel surveys. *Proc Natl Acad Sci* 113(37):5370–5378
- Kita K, Kidziński Ł (2019) Google street view image of a house predicts car accident risk of its resident. arXiv preprint [arXiv:1904.05270](https://arxiv.org/abs/1904.05270)
- Naik N, Philipoom J, Raskar R, Hidalgo C (2014) Streetscore-predicting the perceived safety of one million streetscapes. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 779–785
- Newsom MG (2010) Guide to the San Francisco better streets plan. <https://sfplanning.org/resource/better-streets-plan>
- OpenStreetMap (2017). <https://www.openstreetmap.org>. Accessed 1 Apr 2017
- Provata A, Sokolov IM, Spagnolo B (2008) Ecological complex systems. *Eur Phys J B* 65(3):307–314

- Sanchez TW, Stolz R, Ma JS (2003) Moving to equity: addressing inequitable effects of transportation policies on minorities. The Civil Rights Project (UCLA)/ Proyecto Derechos Civiles. Retrieved from <https://escholarship.org/uc/item/5qc7w8qp>
- Sherman BP (2014) Racial bias and interstate highway planning: a mixed methods approach. *College Undergraduate Research Electronic Journal (CUREJ)*, University of Pennsylvania. <https://repository.upenn.edu/curej/176>
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 2818–2826
- Toole JL, Colak S, Sturt B, Alexander LP, Evsukoff A, González MC (2015) The path most traveled: travel demand estimation using big data resources. *Transport Res Part C Emerg Technol* 58:162–177
- United States Census Bureau (2016) United States Census Bureau: census data. <https://www.census.gov/data.html>. Accessed Oct 15 2016
- Vega-Redondo F (2007) *Complex social networks*, vol 44. Cambridge University Press, Cambridge
- Wang P, Hunter T, Bayen AM, Schechtner K, González MC (2012) Understanding road usage patterns in urban areas. *Sci Rep* 2:1001
- Zhou X, Chen Z, Yeh AG, Yue Y (2020) Workplace segregation of rural migrants in urban China: a case study of Shenzhen using cellphone big data. *Environ Plan B Urban Analyt City Sci*. <https://doi.org/10.1177/2399808319846903>

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
