Mobility Networks for Predicting Gentrification

Oliver Gardiner(✉) and Xiaowen Dong

Department of Engineering Science, University of Oxford, Oxford, UK
owgardiner@gmail.com, xdong@robots.ox.ac.uk

Abstract. Gentrification is a contentious issue which local governments struggle to deal with because warning signs are not always visible. Unlike current literature that utilises solely socio-economic data, we introduce the use of large-scale spatio-temporal mobility data to predict which neighbourhoods of a city will gentrify. More specifically, from mobility data, which is associated with the exchange of ideas and capital between neighbourhoods, we construct mobility networks. Features are extracted from these mobility networks and used in gentrification prediction, which is framed as a binary classification. As a case study, we use the Taxi & Limousine Commission Trip Record Data to predict which census tracts would gentrify in New York City from 2010 to 2018, and show that considering network features alongside socio-economic features leads to a significant improvement in prediction performance.

Keywords: Gentrification · Mobility networks · Urban computing · New York City

1 Introduction

The precise definition of gentrification remains a topic of open debate, but at its core the term refers to a period of rapid change in a previously disadvantaged neighbourhood. Regardless of exact terminology, the impact of gentrification on neighbourhoods is undeniable. Moreover, public opinion, sometimes manifested as protests, demonstrates that it is a problem city governments are struggling with [26]. Fundamentally, local governments struggle to deal with gentrification because by the time obvious signs of gentrification appear, the process is already in full flow: 'The tide of living expenses in a given neighbourhood may already be rising so fast... If you're poor or working class, it's just time to leave' [8].

This has motivated a number of recent quantitative studies to understand how different factors contribute to gentrification, and to predict which neighbourhoods will gentrify. These studies have used either regression [22,25] or binary classification [1,5,14]. Studies which use regression face the natural challenge that there is no obvious continuous variable that can be used to represent gentrification. Binary classification, using a clear definition of gentrification, is therefore a more logical approach and is adopted in this paper. All of these studies, however, utilise only socio-economic data and face the significant limitations

R. M. Benito et al. (Eds.): COMPLEX NETWORKS 2020, SCI 944, pp. 181–192, 2021.
https://doi.org/10.1007/978-3-030-65351-4_15

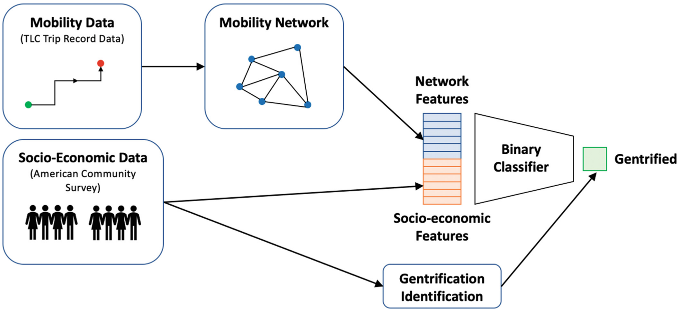


Fig. 1. Proposed analysis pipeline.

that such data often has poor spatial and temporal resolution, and fails to account for other factors, such as human behaviour and tastes, which undoubtedly also contribute to gentrification.

We investigate the use of mobility data, specifically taxi trajectory records, to overcome these limitations. The usage of taxis is documented to vary according to household income [24], and so we hypothesise it should be useful in predicting gentrification. Moreover, mobility data is also interesting with respect to gentrification because the movement of people leads to interactions between neighbourhoods which, in turn, lead to the exchange of ideas, opportunities, and capital, all of which are of key importance to any urban process. Such spatio-temporal data streams, however, do not offer obvious features that can be related to gentrification. To address this, we take inspiration from previous examples of spatial networks in urban computing. Liu et al. showed that a spatial network inferred from taxi trip data in Shanghai encoded useful information about the city structure [16]. Hristova et al. used Twitter and Foursquare data to infer spatial and social networks in London [12]. These networks were interconnected and node statistics were used to measure the social diversity of each neighbourhood. They then demonstrated a correlation between these statistics and the change in socio-economic well-being. We explore the hypothesis that node statistics from a *mobility network* (i.e. a spatial network inferred from mobility data) are useful in predicting gentrification. Unlike the analysis in [12], we formally define the gentrification process using census data, and quantitatively evaluate the predictive performance of the proposed method.

As a case study, we explore the Taxi & Limousine Commission (TLC) Trip Record Data, a data set of taxi journeys in New York City (NYC) [18]. We then consider a number of different network definitions, each defined in Sect. 3.1. From the mobility network we extract a set of network features, which are used alongside socio-economic features, to train a binary classifier to identify which census tracts in NYC would gentrify from 2010 to 2018. There is no commonly used definition of gentrification; for this paper we use the same definition for gentrification as the Urban Displacement Project [6], which provides our labels for classification (the definition is given in Sect. 4.1). We show a significant increase in the performance of binary classification, measured by the area under the receiver operating characteristic (AUROC), compared to using only socio-economic features. An overview of this approach is shown in Fig. 1.

In summary, the main contributions of this paper are:

– We propose a novel framework to use large-scale spatio-temporal mobility data for understanding and predicting gentrification.
– As a case study, we present a technical methodology and results for using both taxi trajectory records and socio-economic data in NYC to predict gentrification.
– We present a qualitative discussion of the network features particularly important to gentrification prediction.

2 Data Sources

We consider two main data sources: the TLC Trip Record Data [18] and the American Community Survey (ACS) [27]. The TLC Trip Record Data details over two billion taxi trips in NYC from 2009 to present. The ACS provides socio-economic and demographic information on each census tract in NYC. We also use Google Maps to provide the travel time via subway of each tract to downtown (Union Square Park) [10].

We consider both yellow taxis and green taxis, for the years where available, from the TLC Trip Record Data in the time period of 2011–2014 (so as not to overlap with the ACS data). The data set was cleaned by removing data points that were obviously erroneous (e.g. have a negative travel time). Following this, the pick-up/drop-off location of each trip was assigned to the census tract within which it was located, using a shapefile of 2018 NYC census tracts [28]. Some summary statistics from the Trip Record Data are shown in Table 1.

The NYC TLC Trip Record Data data set has already been the subject of interest from a networks perspective [7,19,29]. We build on these studies via a discussion of different definitions of a mobility network (Sect. 3.1) and by presenting a formal methodology for the use of these networks in predicting gentrification (Sect. 4).

Table 1. TLC trip record data: summary statistics.

Statistic	2011	2012	2013	2014
Number of trips	170,941,180	173,097,854	169,884,723	176,295,717
Mean travel time (±std)	12:25 (±11:34)	12:24 (±10:01)	12:39 (±15:37)	13:32 (±19:33)

We use ACS 5-year estimates for the periods 2006–2010 and 2014–2018 (the most recent available at the time of analysis). The socio-economic data from the ACS is used to both provide features for gentrification prediction and identify which census tracts have gentrified. Pre-processing was needed before the ACS could be used; some data categories are only recorded to a certain value, we replaced these data points with their limit (i.e. 250,000+ was replaced by 250,000). We also removed tracts that had no population in either 2010 or 2018.

3 Mobility Networks

3.1 Network Definition

In NYC we define a spatial network in which each census tract (excluding those with no population) is represented by a node. Thus, the network has 2114 nodes. Below we describe the different network definitions we explored. For consistency between definitions, all of the networks defined here are undirected.

Origin-Destination Network. In an *origin-destination* network an edge is defined between two nodes if there is a taxi trip between the two tracts, similar to other studies on mobility networks [7,16,19,29]. The weight of an edge is the total number of trips between the two tracts in the time period considered (Eq. 1, where $t_{i\to j}$ is the number of trips from tract i to tract j in the time period considered, and $\mathbf{A}$ is the adjacency matrix).

$$\mathbf{A}_{i,j} = \mathbf{A}_{j,i} = t_{i\to j} + t_{j\to i} \tag{1}$$

This definition follows from the idea that a trip from tract i to tract j likely leads to interactions between individuals in those tracts. We consider two time periods: 2011–2014 and 2014.

Co-work Location Network. This definition is inspired by the use of gathering events by Psorakis et al. [21]. It follows the assumption that two trips originating from tracts i and j and terminating at tract k do not necessarily imply a link between tracts i or j and k, but instead may imply a link between tracts i and j. An obvious example of this is two colleagues commuting to the same office. A *gathering event* is defined when there are many trips to a single tract during a short period of time, with the assumption that during this event there is a higher probability of interactions between individuals.

We use the morning commute as an obvious candidate for a gathering event. Not only is it easy to define, but the growth of service sector employment is also commonly noted in gentrified neighbourhoods [13]. Specifically, we define a gathering event at each tract in Manhattan (which has a high density of offices) on weekday mornings between the hours of 07:00–10:00 am for all of 2014.

The strength of the relationship between tracts that send trips to the gathering event is given by the association factor, $r_{ij,e}$. Taking inspiration from ecological networks [9], we define the association factor between tracts i and j for gathering event e:

$$r_{ij,e} = \frac{x_i \; x_j}{(\sum_{k=1}^{n} x_k)^2} \tag{2}$$

where x_i is the number of trips to the gathering event from tract i, and n is the total number of tracts in NYC. The association factor is, therefore, the number of possible interactions between individuals from tracts i and j divided by the maximal number of interactions possible between individuals at the event. To infer a network over some time period, we take the mean association factor (Eq. (3), where E is the total number of gathering events in a time period).

$$\mathbf{A}_{i,j} = \mathbf{A}_{j,i} = \frac{\sum_{e=1}^{E} r_{ij,e}}{E} \tag{3}$$

Weighted and Binary Networks. The edge weights of the origin-destination network and co-work location network face the inherent problem that the Trip Record Data undoubtedly contains noise. To combat this we opt for a relatively simple solution of creating a *binary* network: the weight of a link is set to 1 if it is greater than the median edge weight and 0 otherwise. We explore both weighted networks and their binary counterparts in the remainder of this paper.

3.2 Network Visualisation and Summary Statistics

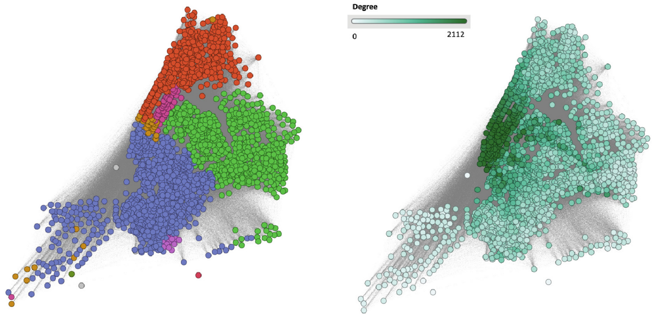
Using these definitions, we inferred six mobility networks from the Trip Record Data. Some summary statistics of these are shown in Table 2 and a visualisation of the 2011–2014 Weighted Origin-Destination Network is shown in Fig. 2.

Table 2. Network summary statistics. The degree is the number of edges each node has (i.e. for weighted graphs ignoring the weight of an edge) and the edge density is the number of edges in the graph divided by the total number of potential edges.

Network	Mean degree	Edge density
2011–2014 Weighted Origin-Destination	881	0.42
2011–2014 Binary Origin-Destination	334	0.16
2014 Weighted Origin-Destination	720	0.34
2014 Binary Origin-Destination	265	0.13
2014 Weighted Co-Work Location	474	0.22
2014 Binary Co-Work Location	200	0.09

The communities detected in Fig. 2a confirm that the networks contain useful information as they closely match what would be expected in NYC[1]. To determine which definition is the most useful for gentrification prediction all of the inferred networks are investigated in Sect. 4.

[1] The authors have discussed the networks and communities with Dr. Gerard Torrats-Espinosa, Assistant Professor in the Department of Sociology at Columbia University, New York City (in conversation 29 April 2020), and Charlie Dulik, Tenant Organizer at the Urban Homesteading Assistance Board, New York City (in conversation 23 April 2020).



(a) Communities (b) Node degree visualisation

Fig. 2. 2011–2014 Weighted Origin-Destination Network (visualisation created using Gephi (version 0.9.2) [2]). Each node is positioned at the centroid of the corresponding tract. In Fig. 2a the colours represent communities detected via modularity optimisation [3].

4 Gentrification Prediction

4.1 Methods

Gentrification Identification. First, we identify which tracts gentrified from 2010 to 2018; this provides class labels for training a binary classifier. We adopt the same definition for gentrification of a tract as Chapple et al. (the Urban Displacement Project) and Rigolon and Németh [6,25], and from this definition present the number of eligible and gentrified tracts in Table 3. The definition is as follows.
First we identify whether a tract was eligible to gentrify in 2010:

- Owner-occupied home value or gross rent < 80% of NYC median *And (any 3 of 4):*
- % low income households (annual income below $50,000) > NYC median
- % of residents college educated < NYC median
- % of residents who rent > NYC median
- % of residents who are non-white > NYC median

And then whether a tract gentrified from 2010 to 2018:

- Eligible to gentrify in 2010
- Increase in % of college educated residents > NYC median
- Percentage increase in real median household income > NYC median *And (either of):*
- Increase in median real rent > NYC median
- Increase in median value of owner-occupied units > NYC median

Table 3. Number of gentrified tracts by borough in NYC.

Borough	Total number of tracts	Eligible to gentrify in 2010	Gentrified 2010 to 2018	Percentage of eligible tracts gentrified (%)
Manhattan	281	76	42	55.3
Queens	643	80	14	17.5
The Bronx	332	185	51	27.6
Brooklyn	750	118	56	47.5
Staten Island	108	9	1	11.1

Feature Extraction. We investigate a number of socio-economic features and network features. For brevity, here we only present the features used after a feature selection step to remove multicollinearity.

Socio-economic feature selection was guided by numerous existing studies in the literature [5,14,22,25] and the features chosen are shown in Table 4. The 'Distance to Downtown' is highly non-linear and so was log transformed.

The network features chosen are also presented in Table 4 (where V is the set of nodes, $N(i)$ is the set of neighbours of i, N_T is the cardinality of the set of neighbours, R is the set of recently gentrified tracts[2], R_T is the cardinality of the set of recently gentrified tracts, $\hat{w}_{ij}$ is the edge weights between i and j normalised by the maximum weight in the network, $T(i)$ is the number of triangles through i, and $log()$ is the natural logarithm).

Finally, we also include the borough as a one-hot encoded variable to account for unobserved heterogeneity and confounding variables in the data set associated with the boroughs, such as local trends or policies. This also increases our confidence in the analysis of feature importance presented below.

Binary Classification. We investigated two different binary classifiers, logistic regression (LR) and random forest (RF), both implemented using scikit-learn (version 0.22.2) [20]. We only consider tracts eligible to gentrify in 2010, and a positive label is assigned to tracts which gentrified from 2010–2018.

For both classifiers the data is split into training and test sets in a 80%:20% stratified split. The hyperparameter(s) is chosen using a random grid search and 4-fold cross validation on the training data, optimising for AUROC. Finally, the model is then fitted to the training data and the performance evaluated on the test data.

4.2 Results

Prediction Performance. To investigate the utility of each network definition for gentrification prediction, we compare the performance of the classifiers using features extracted from each network. The performance, as measured by AUROC, is shown in Fig. 3. We see that most, although not all, of

[2] 'Recently Gentrified' tracts are those identified to have gentrified from 2000–2010, using the definition of gentrification in Sect. 4.1 (for this different time period), data from the 2000 U.S. decennial census [27] and the 2006–2010 ACS data.

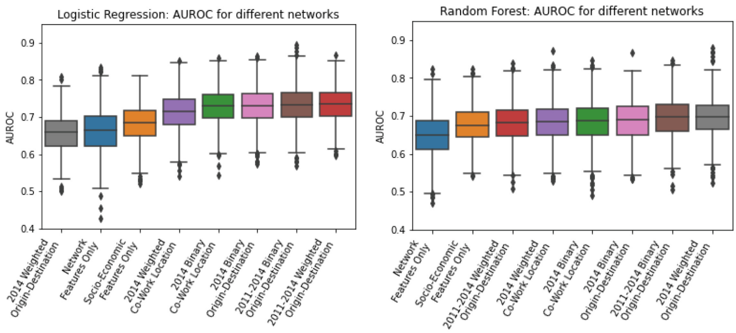
Table 4. Socio-economic and network features used. The network features are mostly calculated using the Python package NetworkX (version 2.4) [11].

Socio-economic	Definition
% Black	Percentage of 'Black or African American' residents
% White	Percentage of 'White Alone' residents
Distance to Downtown	Distance from tract centroid to downtown NYC (m)
% College Educated	Percentage of residents with a Bachelor's degree
% Unemployed	Percentage of residents with 'Unemployed' status
% Renters	Percentage of total housing units occupied by rent paying tenant
Subway Travel Time	The expected travel time by subway from tract centroid to downtown (minutes)
Borough	The borough within which the tract is situated included as a one-hot encoded variable
Network	**Definition**
Degree	$k_i = \sum_{j=1}^{n} \mathbf{A}_{ij}$ [17]
Average neighbour degree	$k_{nn,i} = \frac{1}{N_T} \sum_{j \in N(i)} k_j$ [17]
Links to recently gentrified	$l_i = \frac{1}{N_R} \sum_{j \in R} \mathbf{A}_{ij}$
Shannon entropy	weighted: $H(i) = \sum_{j \in N} P(w_{ij}) log\left(\frac{1}{P(w_{ij})}\right)$
Clustering Coefficient	Weighted: $C_c(i) = \frac{1}{k_i(k_i-1)} \sum_{jk} (\hat{w}_{ij}\hat{w}_{ik}\hat{w}_{jk})^{\frac{1}{3}}$ [17] binary: $C_c(i) = \frac{2T(i)}{k_i(k_i-1)}$ [17]

the network definitions lead to an improvement on the baseline of solely considering socio-economic features and that considering only network features performs slightly worse than considering only socio-economic features. Both of these results demonstrate the added value of considering network features.

The performance improvement is much more pronounced for the LR classifier than for the RF. This suggests that a linear relationship between the features and the log odds seems to be an appropriate assumption for the network features.

LR considering features from the 2011–2014 Weighted Origin-Destination network is the best performing model with a median AUROC of 0.73. Having been inferred from a four-year period of data, the 2011–2014 Weighted Origin-Destination network may have captured some temporal trends that are not captured in the networks defined over a shorter period of time. We also note green taxis are only included in the data set from August 2013, which is another difference with the networks inferred from solely 2014. Whilst removing noise, valuable information is also lost in creating the binary network, possibly explaining the slightly better performance of this network over its binary counterpart. We further consider this model in terms of feature importance below.



LR AUROC Results RF AUROC Results

Fig. 3. Boxplots show the AUROC of the classifiers over 1,000 training/testing data set splits. In each, both socio-economic features and features from the named network are considered, apart from 'Socio-Economic Features Only' and 'Network Features Only' (2011–2014 Weighted Origin-Destination Network) which provide baselines against which the performance can be compared. The boxplots show quartiles of the data; outliers are defined as points more than 1.5 times the interquartile range past the upper and lower quartiles. The boxplots are arranged from left to right in order of increasing median.

Feature Importance. Since we Z-transform the features before training the LR classifier, the magnitude and sign of the LR *beta values* (coefficients) give an indication of a feature's importance and relationship with gentrification. The beta values are presented in Table 5; we see that the '% Black', '% White' and '% College Educated' are the most important socio-economic features and the 'Degree' and 'Clustering Coefficient' the most important network features.

Tracts with a lower degree are shown to be more likely to gentrify. This may be explained by the fact that gentrification is often driven by developers buying plots of land and building luxury housing units [15]. Weems et al. showed a positive correlation between the degree and property prices for an origin-destination network in NYC [29]. Thus, less central tracts being more likely to gentrify may be indicative of developers seeking out cheap plots of land for their projects.

We also see that a tract with a lower clustering coefficient is less likely to gentrify. A low clustering coefficient indicates structural holes in the network [17]. In the literature of social network analysis it has been argued that individuals located at structural holes are more likely to have good ideas [4]. It might be expected that tracts which are exposed to good ideas (or perhaps better referred to as trends in an urban context) may be more likely to gentrify. However, this interpretation is seemingly at odds with the relationship discovered here. Considering gentrification as a diffusion-like process, as argued by Redfern [23], may be useful in explaining this discrepancy. Tracts at structural holes, to which there is a limited flow of capital and ideas [17], may be less likely to be involved in this diffusion process.

Table 5. Mean beta values (over 1,000 training/testing splits) for a LR classifier considering the 2011–2014 Weighted Origin-Destination network. Statistically significant values, with $p < 0.05$, are in bold.

Socio-economic Feature	Beta value	Socio-economic Feature	Beta value
% Black	**0.559**	Manhattan	0.565
% White	**0.577**	Queens	0.460
Distance to downtown	**0.468**	**Network Feature**	**Beta value**
% College Educated	**-0.838**	Degree	**-1.607**
% Unemployed	**-0.343**	Average Neighbour Degree	0.059
% Renters	**-0.509**	Links to Recently Gentrified	-0.130
Subway Travel Time	**-0.487**	Shannon Entropy	0.097
Bronx	0.746	Clustering Coefficient	**2.269**
Brooklyn	1.109		

In fact, the degree and clustering coefficient have the largest-magnitude beta values of the features considered. This emphasises that, in addition to traditional factors, the potential exchange of capital and ideas associated with these network features is important to consider in understanding and predicting gentrification.

Finally, the beta values indicate that, on average, it is more likely for a tract in Bronx and Brooklyn to gentrify than tracts in the other boroughs. However, this tendency is found to not be statistically significant (with p-values of 0.48 and 0.24, respectively). Nevertheless, analysing differences between the boroughs with regards to gentrification would be an interesting future direction.

5 Discussion

We have presented a methodology for using large-scale mobility data alongside socio-economic data for gentrification prediction, using NYC as a case study. As part of this, we presented a discussion of different methods for inferring a mobility network from mobility data (Sect. 3), and showed that features extracted from these networks can be used to improve the performance of gentrification prediction (Sect. 4). Finally, we also provided a qualitative discussion of network features identified as particularly important (Sect. 4.2).

The methodology and results presented in this paper have limitations. The vast majority of taxi trips in NYC take place in Manhattan, so the Trip Record Data may not accurately capture the movement of all the residents of NYC. There is also no single definition of gentrification used by social scientists, which means that results are difficult to compare across studies.

To build upon the research presented in this paper, other forms of mobility data could be considered, for instance bus or subway data. The methodology used here could be adapted to a finer spatial and temporal resolution to further benefit from the advantages of mobility data. The Co-Work Location network definition presented here could be expanded by automatically detecting gathering events. Finally, the methodology presented in this paper should also be applied to different cities and time periods.

References

1. Alejandro, Y., Palafox, L.: Gentrification prediction using machine learning. In: 18th Mexican International Conference on Artificial Intelligence. MICAI 2019, Xalapa, Mexico, pp. 187–199. Springer (2019)
2. Bastian, M., Heymann, S., Jacomy, M.: Gephi: an open source software for exploring and manipulating networks. In: BT - International AAAI Conference on Weblogs and Social. International AAAI Conference on Weblogs and Social Media, pp. 361–362 (2009)
3. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech. Theor. Exp. **2008**(10), 1–12 (2008)
4. Burt, R.S.: Structural holes and good ideas. Am. J. Sociol. **110**(2), 349–399 (2004)
5. Chapple, K., et al.: Mapping Susceptibility to Gentrification: The Early Warning Toolkit. Technical report, University of California, Berkeley (2009). https://communityinnovation.berkeley.edu/sites/default/files/mapping_susceptibility_to_gentrification.pdf
6. Chapple, K., Loukaitou-Sideris, A., Chatman, D., Waddell, P., Ong, P.: Developing a New Methodology for Analyzing Potential Displacement. Technical report, University of California, Berkeley (2016)
7. Deri, J.A., Moura, J.M.: Taxi data in New York city: a network perspective. In: 49th Asilomar Conference on Signals, Systems and Computers, pp. 1829–1833 (2015)
8. Frank, A.: What Does It Take To See Gentrification Before It Happens? (2017). https://www.npr.org/sections/13.7/2017/08/29/546980178/what-does-it-take-to-see-gentrification-before-it-happens?t=1587409355525
9. Ginsberg, J.R., Young, T.P.: Measuring association between individuals or groups in behavioural studies. Anim. Behav. **44**(2), 377–379 (1992)
10. Google: Google Maps. https://www.google.com/maps/dir//Union+Square,+New+York,+NY+10003,+USA/@40.7358421,-74.0611234,12z/data=!3m1!4b1!4m9!4m8!1m0!1m5!1m1!1s0x89c259989e14aa8b:0xcd00afc9db20caa4!2m2!1d-73.9910835!2d40.7358633!3e3
11. Hagberg, A.A., Schult, D.A., Swart, P.J.: Exploring network structure, dynamics, and function using NetworkX. In: 7th Python in Science Conference (SciPy 2008), pp. 11–15 (2008)
12. Hristova, D., Williams, M.J., Musolesi, M., Panzarasa, P., Mascolo, C.: Measuring urban social diversity using interconnected geo-social networks. In: 25th International World Wide Web Conference, WWW 2016, pp. 21–30 (2016)
13. Hyra, D.S.: Race, Class, and Politics in the Cappuccino City. The University of Chicago Press, Chicago (2017)
14. Knorr, D.: Using Machine Learning to Identify and Predict Gentrification in Nashville, Tennessee. Technical report, Vanderbilt University (2019)
15. Kohli, S.: Developers have figured out the secret sauce for gentrifying neighborhoods (2015). https://qz.com/408986/developers-have-figured-out-the-secret-sauce-to-gentrification/
16. Liu, X., Gong, L., Gong, Y., Liu, Y.: Revealing travel patterns and city structure with taxi trip data. J. Transp. Geogr. **43**, 78–90 (2015). https://doi.org/10.1016/j.jtrangeo.2015.01.016
17. Newman, M.E.J.: Networks: An Introduction, 1st edn. Oxford University Press 2010)
18. NYC Taxi & Limousine Commission: NYC Taxi and Limousine Commission (TLC) Trip Record Data. https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

19. Patil, N.: Characterizing & analyzing networks : NYC taxi data. https://nilesh-patil.github.io/blog/transportation-graph-nyc-taxi-data/#full-network-analysis
20. Pedregosa, F.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
21. Psorakis, I., Roberts, S.J., Rezek, I., Sheldon, B.C.: Inferring social network structure in ecological systems from spatiotemporal data streams. J. R. Soc. Interface **9**(76), 3055–3066 (2012)
22. Reades, J., De Souza, J., Hubbard, P.: Understanding urban gentrification through machine learning. Urban Stud. **56**(5), 922–942 (2019)
23. Redfern, P.A.: A new look at gentrification: 2. A model of gentrification. Environ. Plann. A: Econ. Space **29**(8), 1335–1354 (1997)
24. Renne, J., Bennett, P.: Socioeconomics of urban travel: evidence from the 2009 national household travel survey with implications for sustainability. World Transp. Policy Pract. **20**(4), 7–27 (2014)
25. Rigolon, A., Németh, J.: Toward a socioecological model of gentrification: how people, place, and policy shape neighborhood change. J. Urban Affairs **41**(7), 887–909 (2019)
26. Santus, R.: How anti-gentrification activists derailed Amazon's New York City plans (2019). https://www.vice.com/en/article/nex34z/how-anti-gentrification-activists-derailed-amazons-new-york-city-plans
27. US Census Bureau: American Community Survey 5-year estimates (2018). https://data.census.gov/cedsci/table?g=0500000US36005,36047,36061,36081,36085&d=ACS%205-Year%20Estimates%20Data%20Profiles&tid=ACSDP5Y2018.DP02&hidePreview=false
28. US Census Bureau Geography Division: TIGER/Line Shapefiles. https://www.census.gov/cgi-bin/geo/shapefiles/index.php
29. Weems, B., Field, E., Ward, T.: New York Taxi Network: Community Structure and Predictive Analysis. Technical report, Stanford (2016). http://snap.stanford.edu/class/cs224w-2016/projects/cs224w-59-final.pdf