Who Visits the University Village?
Gentrification and Social Capital in a Decade of Tweets

By Aaron Chen *

Keywords: Gentrification, Social Capital, Mobility

* acc7919@stern.nyu.edu. This project is partially funded by the NYU Dean's Undergraduate Research Fund and the Holten Family Research Scholarship. I would like to thank Professor Petra Moser for endorsing the Undergraduate Research Fund and Professor Arpit Gupta for advising this project. Submitted in fulfillment of the NYU Stern Rosenwald Scholarship Program.

> *"Once this process of 'gentrification' starts in a district it goes on rapidly until ... the whole social character of the district is changed."*
> — Ruth Glass, *Aspects of Change* (1964)

The University of South California Development Plan of 2010 lays the groundwork for what would become the largest private infrastructural investment in Los Angeles over the past decade. This $700 million project aims to transform the University Village at the intersection of Hoover Street and Jefferson Boulevard into a Hogwarts-style hub, featuring a shopping center, cinema, hotel, and dormitories for 40 percent of the student population. Opening in August 2017, the redevelopment replaces fixture food stalls like Sandwich Island with Cava, the Dollar Store with Target, and Superior Grocers with Trader Joe's. The "character" of the University Village has changed.

The USC Village presents a unique case study on gentrification. In a private university, investment decisions are primarily driven by the board as well as faculty and students. Interests of these stakeholders are arguably orthogonal to market dynamics of residential demand and amenity supply. The decision to redevelop the University Village is a rare case of an exogenous amenity shock.

How do amenity shocks through lump-sum private investments such as the University Village contribute to gentrification? Specifically, how does the opening of the University Village change who visits the neighborhood? What are the mechanisms? Are changes in mobility better explained by the spillover of physical improvements or shifts in "social capital", including the quality of local ties and interactions? Despite the neon sign at Starbucks reading "coffee is togetherness", does the University Village bring the community together?

I study these questions by constructing a comprehensive panel on residential visits and amenity supply within the City of Los Angeles between 2010-19. I exhaust over 22 million publicly available Tweets dated between 2010-19 with a GPS coordinate within the City of Los Angeles and link them with the closest address. I further harmonize timestamped, geo-tagged annotations from the building permit catalog, the business registry, and OpenStreetMap GIS layers with the North American Industry Classification System (NAICS) to capture the distribution of public and private amenities at the address-level over time. I complement address-level data with tract-level estimates from the American Community Survey on demographic and housing characteristics.

Exploiting the exogenous nature of the University Village investment, I use a difference-in-difference framework to investigate the effects of the re-opening on residential and amenity outcomes in neighboring tracts. To ensure the plausibility of parallel trends, I propensity-score match tracts in the vicinity of the University Village against other non-neighboring tracts with comparable residential demographics and property valuations before the start of construction.

Difference-in-difference estimates indicate that visits to residential-type amenities (including public parks, civic centers, other recreational spaces, and health amenities) decrease by 89 percent as a result of the University Village opening, relative to pre-construction differences with non-neighboring tracts of comparable demographic and property characteristics before construction. Meanwhile, visits to consumer and touristic amenities (including fast food, full-service restaurants, bars, and gyms) increase 7-fold. These reduced-form patterns in residential and touristic sorting are consistent with predictions by recent structural models (Almagro and Domínguez-Iino, 2024). Event studies show that the increase in visits to consumer amenities and the decrease in visits to residential amenities begin during construction and become more pronounced after the re-opening of the University Village. Effects stabilize by the end of the first year and persist for at least two years. Results are robust to dis-aggregating estimates by specific amenities.

To investigate whether changes in amenity visits are driven by changes in user composition, I identify tourists and residents based on patterns in their geospatial activity (including the duration of observation, number of visits to airports and hotels, as well as the spatial area of travel during daily working hours). I compute the number of visits and average travel distance by residents and tourists lodging in the USC area. These results show that tourists and short-term lodgers in the area visit amenities more frequently and travel shorter distances per visit, particularly in the categories of education, fast food, and gyms, which are directly offered within the University Village. Conversely, residents make 80 percent fewer visits to all places across the city and travel farther per visit to the same amenities. The mobility gap suggests that tourists and short-term lodgers absorb the benefits of convenience offered by the University Village while displacing residents, possibly from the city altogether.

Ding, Hwang and Divringi (2016); Diamond (2016); Davis et al. (2019) suggest that changes in mobility occur along racial and socioeconomic lines and are driven by the share of "ultra-rich" and college-educated. To investigate this mechanism, I identify user characteristics (including race, ideology, and occupation) by parsing keywords in user profiles using tools from Natural Language Processing (NLP). These results indicate that the composition of users has not changed significantly along racial or ideological lines due to the University Village opening. Interestingly, the share of users reporting a university-type role decreases by 25 percent, while the regional share of self-identified influencers (including vloggers and individuals in marketing or entertainment) increases by 48 percent.

I check whether changes in amenity visits and residential mobility are driven by patterns in physical improvements in the neighborhood using 1) Zillow single-family home prices, 2) the valuation of active building permits, 3) the number of permits and business openings, 4) the number off building additions and demolitions, and 5) the number of home renovations. These results are mostly statistically insignificant. Unlike what is suggested by Guerrieri, Hartley and Hurst (2013), changes in mobility due to the University Village redevelopment do not correspond with changes in property value in neighboring tracts.

Following White (1986); Gentzkow and Shapiro (2011); Davis et al. (2019), I construct segregation indices to investigate whether changes in residential sorting indicate increasing levels of fragmentation of local social networks. Difference-in-difference estimates indicate that, due to the University Village redevelopment, ideological segregation between self-identified liberals and conservatives increase by 100 percent. These changes are primarily driven by users visiting places disproportionately visited by members sharing the same ideology. Segregation along racial and religious lines is inconclusive.

Using granular textual data, this paper provides the first panel-level analysis of the effects of gentrification on social attitudes. Motivated by the literature on ideological segregation and misinformation (Gentzkow and Shapiro, 2011; Athey et al., 2023), I proxy for the absence of trust by identifying hateful and offensive speech. I finetune a large language model (Nguyen, Vu and Nguyen, 2020) to simultaneously detect the prevalence of hate, the type of offense (derogation, animosity, dehumanization, and physical threats), and the target identity group (race, religion, nationality, gender, sexuality). The model achieves benchmark performance on state-of-the-art datasets. I use this model to identify hateful or offensive speech within the Los Angeles dataset. Difference-in-difference estimates show that the share of Tweets that express derogation, animosity, and dehumanization increases by 11-21 percent due to the University Village redevelopment. The share of targeted speech increases and is most significant towards gender and sexual identities. These results provide empirical support for the effects of gentrification on informal elements of social organization, particularly the absence of trust (Putnam, Leonardi and Nanetti, 1993; Durlauf and Fafchamps, 2005). These factors may contribute to observations of socioeconomic connected-ness (Chetty et al., 2022).

This paper makes multiple contributions to the empirical literature. By leveraging the exogenous nature of the University Village investment and directly collecting data on social attitudes, I avoid the need to use structural estimation. Despite recent econometric advancements (Durlauf, 2004; Bramoullé, Fortin and Djebbari, 2020), structural models of gentrification and social interaction continue to suffer from identification issues, either due to the endogenous relationship between residential demand and amenity supply or between individual and group-level decisions (Manski, 1993). On the other hand, reduced-form estimates from this paper provide support for structural mechanisms that may drive residential mobility patterns, including segregation (Davis et al., 2019) and homophily (Golub and Jackson, 2011; Goldsmith-Pinkham and Imbens, 2013). In particular, formalizing mechanisms of social organization warrants further research and may lead to new avenues for identification.

More broadly, this paper contributes to the vast policy literature on the value capture of amenity investments (Baum-Snow, 2007; Ahlfeldt et al., 2015; Gupta, Van Nieuwerburgh and Kontokosta, 2020; Su, 2022). Baum-Snow (2007) shows that a highway passing through a central city reduces congestion by increasing the appeal of suburbs. Su (2022) shows that shocks on working hours in the labor market shift workers' valuations of commute time, which in turn impacts the appeal of central city neighborhoods. (Ahlfeldt et al., 2015) shows that economies of agglomeration and dispersion also shape worker decisions and amenity stock. (Gupta, Van Nieuwerburgh and Kontokosta, 2020) shows that local real estate prices absorb the benefits of transit spillovers from subway extensions. This paper offers a counterpoint insofar as increases in the convenience of commute due to amenities offered in the new University Village do not always align with changes in residential visits and travel distance. Instead, reductions in visits and increases in travel distance by incumbent residents suggest that valuations of amenities are likely influenced by non-market mechanisms, including the "social character" of a neighborhood.

Section I outlines the empirical strategy. Section II describes the data collection. Section III reports the effect of the University Village redevelopment on amenity visits and user demographics. Section IV explores social mechanisms that may explain the changes in residential mobility. Section **??** concludes.

I. Empirical Strategy

A. *Event studies*

To investigate the effects of the University Village investment, I estimate the following event studies by the method of generalized linear models:

$$F^{-1}(y_{it}) = \alpha_i + \gamma_t + \sum_{\tau \neq -1} \beta^{\tau}(Nexus_i \times D_t^{\tau}) + \varepsilon_{it}, \tag{1}$$

where y_{it} is the outcome of interest in tract i at time t and $F^{-1}(\cdot)$ is the corresponding link function. When y_{it} represents the number of visits, I assume that y_{it} is Poisson-distributed so that $F^{-1}(\cdot)$ is the natural logarithm. $Nexus_i$ is an indicator for tracts in the USC Nexus Study Area. These tracts are identified by the Los Angeles Housing Department and Department of City Planning as part of the 2012 Nexus study to design appropriate land use regulations in communities plausibly affected by the construction of the University Village.[1] I focus on the effects of the University Village on these Nexus tracts. D_t^{τ} is an indicator for observations in calendar quarter t measured τ quarters before the start of construction (if negative) or after the end of construction (if positive). $\tau = 0$ for

[1]See City Council File 08-2620.

all quarters during construction, from 2015 Q1 to 2017 Q2. The quarter before construction $\tau = -1$ is the excluded period. α_i and γ_t are tract- and calendar quarter- fixed effects, and ε_{it} are unobserved residuals. I cluster standard errors on the tract level to allow for serial correlation in outcomes within a tract over time. When y_{it} represents amenity visits, auto-correlation of visits may occur due to the accumulation of goodwill.

$F(\beta^r) - F(0)$ are the quantities of interest. When regressing amenity visits, the quantity $100 \times (\exp(\beta^\tau) - 1)$ estimates the percentage difference in visits during event quarter τ, between Nexus tracts and other tracts plausibly unaffected by the construction, relative to the difference in the quarter before construction. For table estimates, I estimate a difference-in-difference version of equation (1):

$$F^{-1}(y_{it}) = \alpha_i + \gamma_t + \beta(Nexus_i \times Open_t) + \delta(Nexus_i \times Construct_t) + \varepsilon_{it}, \tag{2}$$

where $Open_t$ is an indicator for calendar quarters after the end of construction, and $Construct_t$ is an indicator for calendar quarters during construction.

B. *Identification*

Identification in event studies requires (a) no anticipation, (b) no spillovers of treatment, and (c) parallel trends. First, patterns in visitation and amenity stock in the Nexus area should not change in anticipation of construction. This is a plausible assumption because the University Village only replaced existing property owned by the school, meaning that no takings in other Nexus tracts commenced in preparation for the construction of the University Village. The only instance of land acquisition occurred in the quarter before construction in February 2015 between USC and the City, when the University exchanged land for the obsolete Fire Station 15 located on the same block.

The assumption of no treatment spillover requires that visits and amenity stocks in control tracts should not be affected by the construction of the University Village. To this end, I preclude first-order neighbors of the Nexus tracts from the control pool. In Appendix **??**, I test different thresholds for geographic proximity in constructing the control sample.

The parallel trends assumption further requires that Nexus and control tracts exhibit comparable slopes in outcomes before the start of construction. I address possible violations of parallel trends using propensity-score matching (Rosenbaum and Rubin, 1983). First, I estimate a propensity-score model by regressing an indicator for being in the Nexus area on pre-construction tract characteristics that are plausibly correlated with Twitter visits or amenity stock. These covariates include American Community Survey property valuations and demographic characteristics such as age, education attainment, occupation, race and ethnicity, mode of commute, and internet access (see Appendix **??**). I match Nexus tracts with control tracts that exhibit comparable predicted probabilities ("propensity scores") from the model. I test a variety of parametric and non-parametric models (including logistic regression, random forests, and multi-layered perceptions), matching algorithms (including k-nearest neighbors and caliper-matching) as well as distance metrics (including Euclidean and Mahalanobis distances). I choose models and hyper-parameters that minimize the average absolute standardized differences ("effect sizes") across covariates (see Appendix Figure A1).

I match twenty Nexus tracts with 135 of 988 control tracts (Figure 1). Appendix Figure A1 shows that Nexus and control tracts exhibit similar demographic and property characteristics in the pre-construction period after propensity-score matching.[2] Conditional on covariate balance, I assume that user valuations of amenities and tract-specific marginal costs are independent across locations. Therefore, outcomes of interest are plausibly parallel across Nexus tracts and propensity-matched controls.

II. Data

A. *Tweets*

I capture residential mobility using geo-tagged social media posts ("Tweets") on Twitter. Survey estimates indicate that one in five American adults uses Twitter (Smith and Anderson, 2018). Nearly half of users use Twitter daily, with nearly one-third using the platform multiple times per day. According to Twitter, roughly 1-2 percent of Tweets are tagged with precise GPS coordinates from the transmitting device. At the time of data retrieval, geo-tagged Tweets provide a relatively inexpensive source of mobility data for a representative segment of the American population.

Between December 2021 and October 2022, I collect an exhaustive list of 22,050,394 publicly available Tweets with a GPS coordinate within the City of Los Angeles. Specifically, I retrieve Tweets within 1.5 times the radius of 956,446 active parcels in the City's address registry. These Tweets date between 2010-01-01 and 2019-07-01, just after Twitter discontinued the collection of geo-coordinates. My approach abides with Twitter's Academic Research API guidelines at the time of retrieval. The data includes the original user text, GPS-coordinates, timestamp, username, as well as metadata such as the language of writing and number of likes and re-posts. I link each Tweet to the closest address. Figure 2 plots the volume of Tweets by census tract between 2010-19.

To ensure the integrity of the geo-location data, I use the state-of-the-art Botometer API (Sayyadiharikandeh et al., 2020) to verify the identities of 45,535

[2]The main alternative to propensity-score matching is the use of synthetic controls. I use propensity-score matching over synthetic control methods in the main regressions to ensure consistency of the control sample across different regression outcomes. In my case, synthetic control methods lead to sub-optimal matching because the control pool of tracts is relatively large (988 tracts). The computational burden also increases depending on distributional assumptions imposed on the outcome in equation (1).

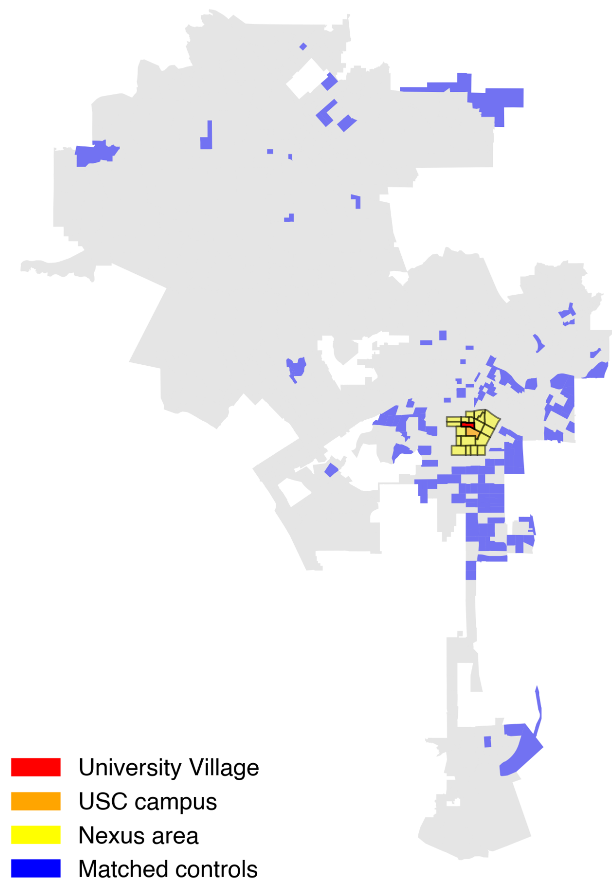


Figure 1. Map of USC Nexus Area and Propensity-matched Controls

Notes: This map plots the locations of the University Village, the University of South California campus, the Nexus study area (18 tracts excluding USC), and 135 propensity-matched control tracts with comparable demographics and property valuations before the construction of the University Village.

out of 806,392 total users (5.6 percent) who are plausibly bots.[3] Specifically, I

[3]Documentation is available at https://github.com/osome-iu/botometer-python. As of May 2023,

check users who either 1) rank in the top ten percentile of total or monthly Tweet count, 2) post consecutively for more than six months, or 3) belong to a one percent random sample of users with more than ten geo-tagged Tweets. By parsing a user's Tweets, profile, and metadata, the model outputs predicted probabilities of a user being a fake follower, spammer, self-promoter, or self-declared bot, among other statistics. I manually check the profiles for two percent of English-speaking users whom the model predicts is a bot with more than 0.5 probability. I tune the tolerance threshold of the model to maximize the validation F1. Appendix **??** describes the pipeline and model performance in full detail. Using the final model, I identify 7,801 bots (0.9 percent of all users) and remove their Tweets from the dataset.

B. Amenity Annotations

I annotate every address in the City of Los Angeles using building permit[4] and business registry data[5] maintained by the Bureau of Engineering. Specifically, I pool together annotations from the datasets using the North American Industry Classification System (NAICS), which uses a six-digit hierarchical coding system to classify all economic activity into twenty industry sectors. Using the geo-coordinates and timestamps of permit issues, business openings, and closings, I identify the types of amenities available at each address at each point in time between January 2010 and July 2019. To detect public amenities and residences not identified by building permits or the business registry (including post offices and community centers), I use the Osmium API[6] to retrieve monthly Point of Interest (POI) extracts from OpenStreetMap, the leading open-source alternative to Google Maps. I manually harmonize over 700 POI labels with the NAICS to retrieve 70,153 additional amenity annotations. Pooling together all sources, I construct 1,017,690 timestamped amenity annotations linked with 729,205 addresses in the City of Los Angeles.

III. Effects of the University Village Opening

A. Amenity Visits

Figure 3 plots the effect of the University Village opening on Twitter visits to different amenity types based on equation (1), estimated on a propensity-matched

Twitter changed API endpoints so that Botometer is no longer functional.

[4]Building permit data retrieved on February 23, 2023 from: https://data.lacity.org/A-Prosperous-City/Building-Permits/nbyu-2ha9.

[5]Business registry data retrieved on March 18, 2023 from:

- https://data.lacity.org/Administration-Finance/Listing-of-Active-Businesses/6rrh-rzua.
- https://data.lacity.org/Administration-Finance/All-Closed-Businesses/tkh9-tssh.

[6]https://osmcode.org/libosmium/

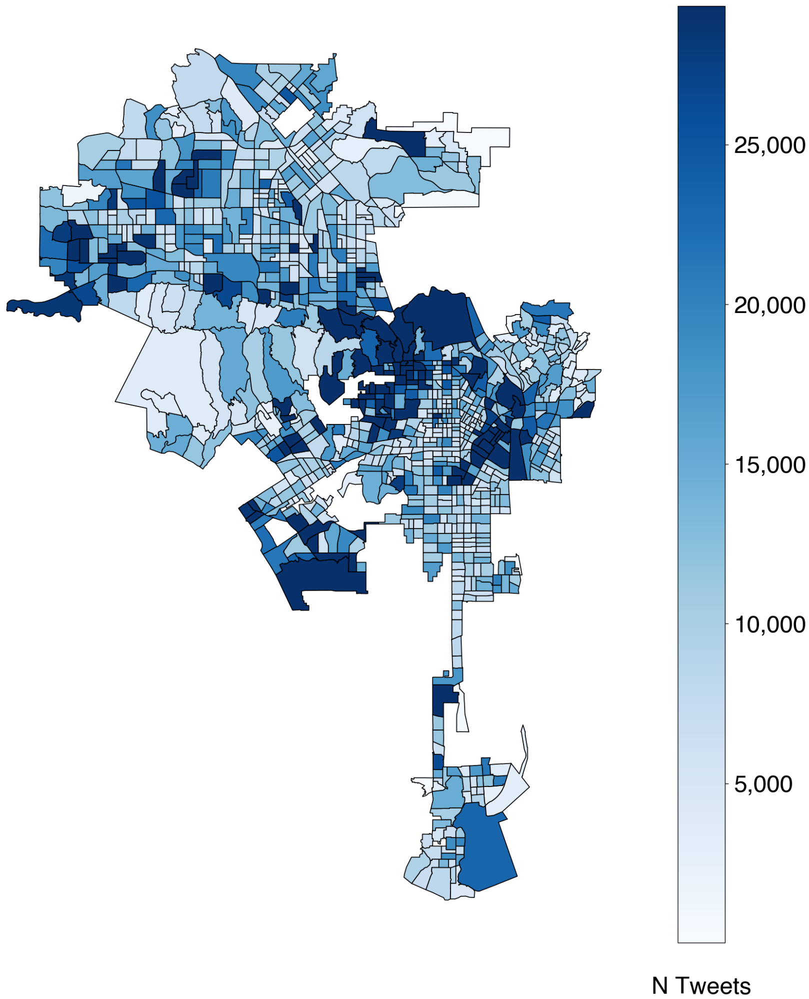


Figure 2. Volume of Tweets by census tract, 2010-19

sample of tracts with comparable demographics and property valuations pre-construction. Coefficients before construction are close to zero and statistically insignificant, supporting the assumption of parallel trends in amenity visits.

The opening of the University Village attracts touristic visits and reduces res-

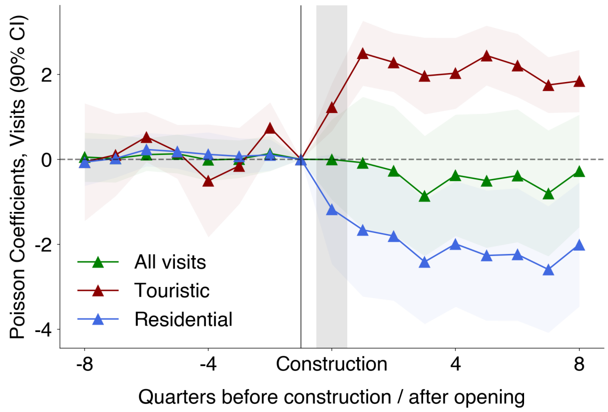


FIGURE 3. EFFECTS OF THE UNIVERSITY VILLAGE OPENING ON AMENITY VISITS

Notes: The figure plots estimates and 90% confidence intervals for β^{τ} from equation (1), which is a Poisson regression of the number of Twitter visits (by amenity type) on tract fixed-effects, calendar-quarter fixed effects, and event quarter indicators interacted with a treatment indicator for tracts in the USC Nexus area. The quarter before the start of construction is the excluded period. The control sample contains propensity-matched tracts that are not directly bordering the USC Nexus area with comparable demographic characteristics and property valuations pre-construction. Standard errors are clustered on the tract level.

TABLE 1—EFFECTS OF THE UNIVERSITY VILLAGE OPENING ON AMENITY VISITS

	Point estimate	Standard error	Control count	Pct Δ
All	-0.48	(0.90)	1036.8	-38.3
Touristic amenities	2.08***	(0.27)	53.9	700.0
Fast food	1.57***	(0.60)	17.9	382.7
Restaurants	2.79***	(1.08)	3.5	1527.8
Bars	1.29***	(0.33)	11.4	262.9
Gyms	2.79***	(0.46)	21.1	1525.3
Residential amenities	-2.20***	(0.97)	1171.1	-88.9
Parks & civic	-2.52***	(0.97)	279.9	-92.0
Other recreation	-1.91**	(0.90)	302.2	-85.2
Health & family care	-3.37***	(1.07)	279.8	-96.6
Residences	-2.53**	(1.15)	309.1	-92.0

Note: dependent variable is the number of visits to each type of amenity. The point estimate is the coefficient from a Poisson regression model. The standard error is robust. The control count is the number of control tracts in the model. The percent change is calculated as $100 * (exp(\beta) - 1)$.

idential visits. Difference-in-difference estimates from equation (2) show that visits to residential amenities in the USC Nexus area (including public parks and civic centers, other recreational spaces, health amenities, and residences) decrease by 88.9 percent (point estimate of -2.20, significant at the 1% level) as a result of the University Village opening, relative to other non-neighboring tracts with comparable demographics and property valuations pre-construction

(Table 1). Specifically, visits to parks decrease by 92.0 percent (point estimate of -2.52, significant at 1%), other recreational amenities decrease by 85.2 percent (point estimate of -1.91, significant at 5%), and health and family care decrease by 96.6 percent (point estimate of -3.37, significant at 1%). This suggests that some amenities have become obsolete. Conversely, visits to tourist and consumer amenities (including fast food, restaurants, bars, and gyms) increase by 700 percent (point estimate of 2.08, significant at 1%). Particularly, visits to restaurants and gyms increase fifteen-fold (both point estimates of 2.79, significant at 1%). Event studies show that the increase in touristic visits and decline in residential visits start during construction and become more pronounced after the opening of the University Village (Figure 3). Effects stabilize by the end of the first year after opening and persist for at least two years. Event study results are robust to dis-aggregating estimates by specific amenities (Appendix Figure A2).

B. User Demographics

Are changes in amenity visits driven by changes in the behavior of residents? To investigate the effects of the University Village on resident mobility patterns, I identify tourists and residents based on patterns in their geo-location. I classify a user as a tourist if they have at least five geo-tagged Tweets, is a short-term visitor to the City[7], and meets at least three of the following conditions: 1) visits the airport or train station at least once, 2) visits an accomodation-type amenity (e.g., hotel) at least once, 3) visits a touristic amenity (including fast food, restaurants, bars, beaches, gyms, museums, and shopping centers) at least once, and 4) never visits a place of residence (excluding registered accommodations such as hotels and AirBnBs). By this definition, I identify 23,384 tourists on Twitter between 2010-19 in the city of Los Angeles.

I classify the most frequently visited accommodation as a tourist's place of stay. I identify a non-tourist's place of residence based on the spatial and temporal distribution of their visits. I classify an address as a user's place of residence if a) it is one of the most frequently visited locations overall,[8] and b) it is the most frequently visited location outside of daily working hours.[9] Consistent with Almagro and Domínguez-Iino (2024), I assume that residential decisions are made on a per-year basis.

Table 2 reports estimates from equation (2) of the effect of the University Village

[7]I classify a user as a short-term visitor if they meet any two of the following criteria: 1) observed in at most three different months, 2) the date range between the first geo-tagged Tweet and last geo-tagged Tweet is under six months, 3) never observable for more than two months consecutively, and 4) has less than thirty geo-tagged Tweets.

[8]Such an address must satisfy any of the three requirements: 1) in the top two most frequently visited locations during the weekends, 2) in the top two most frequently visited locations of any type, or 3) in the top two most frequently visited residential amenities.

[9]I identify users as either working during the day or the night. Assuming that users must commute to work, a user works during the day if the spatial area spanned by the convex hull of their geo-tagged visits is greatest between 6AM and 6PM; a user works during the night otherwise.

TABLE 2—MOBILITY OF RESIDENTS AND TOURISTS LODGING IN THE USC NEXUS AREA

	Residents				Tourists			
	Visits	Pct Δ	Dist. (log km)	Pct Δ	Visits	Pct Δ	Dist. (log km)	Pct Δ
All	-2.03**	-86.9	0.06	5.6	0.41	50.3	-0.29	-29.3
	(1.02)		(0.07)		(0.39)		(0.49)	
Bars	-0.22	-20.1	0.17	17.3	-0.09	-8.4	-0.05	-5.1
	(0.97)		(0.24)		(0.35)		(0.47)	
Civic	0.01	0.5	0.96**	95.8	0.64***	90.2	-0.26	-26.3
	(0.91)		(0.45)		(0.17)		(0.34)	
Education	-0.29	-25.1	0.09	9.3	0.79***	120.6	0.24	23.6
	(0.85)		(0.31)		(0.23)		(0.55)	
Fast food	-0.55	-42.2	-0.11	-11.1	0.52*	68.4	0.19	18.9
	(1.03)		(0.31)		(0.28)		(0.38)	
Grocery	-1.40	-75.4	-0.19	-19.1	1.13	208.6	0.26*	26.1
	(0.97)		(0.56)		(1.12)		(0.15)	
Gym	0.95	159.1	-0.09	-9.1	1.09***	196.0	-0.53	-52.5
	(1.02)		(0.47)		(0.31)		(0.35)	
Health	-2.71**	-93.3	0.27	26.9	0.24	27.5	-0.33	-33.4
	(1.07)		(0.23)		(0.32)		(0.31)	
Parks	-2.94***	-94.7	0.32*	32.0	-0.10	-9.3	-0.55	-54.8
	(1.06)		(0.19)		(0.70)		(0.43)	
Recreation	-0.49	-39.0	0.59*	59.2	0.43*	53.7	-0.06	-6.2
	(1.02)		(0.34)		(0.23)		(0.25)	
Restaurants	-0.79	-54.7	0.26	26.1	0.47	60.2	-0.03	-3.1
	(1.16)		(0.17)		(0.96)		(0.42)	
Urban transit	0.02	2.3	0.66	66.5	1.58**	383.7	-1.27**	-126.9
	(0.93)		(0.43)		(0.67)		(0.59)	

Note: dependent variable is the logged share of users. The point estimate is the coefficient from an OLS regression model. The standard error is robust. The control count is the number of control tracts in the model. The percent change is calculated as $100 * \beta$.

opening on visits and travel distance by residents and tourists lodging in the USC Nexus area. Visits to amenities across the City of Los Angeles by residents in the USC Nexus area decrease by 86.9 percent (significant at the 5% level). The decline in visits is most drastic for health amenities (-93.3 percent, significant at the 5% level) and parks (-94.7 percent, significant at the 1% level). Except for gyms, resident visits to all other amenities decrease or remain close to the same level as before, though these estimates are statistically insignificant. Meanwhile, overall tourist visits are unaffected. Tourist visits increase across various categories, including civic centers (increase by 90.2 percent, significant at the 1% level), education institutions (120.6 percent, significant at the 1% level), fast food (68.4 percent, significant at the 10% level), gyms (196.0 percent, significant at the 1% level), recreational amenities excluding parks (53.7 percent, significant at the 10% level), and urban transit (383.7 percent, significant at the 5% level).

While changes in travel distance are mostly statistically insignificant, the signs of the coefficients for residents and tourists are nearly always opposite of one an-other (Table 2). Whereas residents travel to farther parks, health and recreational amenities, restaurants, and urban transit options, tourists who lodge in the USC Nexus area travel to closer amenities as a result of the University Village opening. Conditional on the same locations of residence and the same target amenity type, the opposite travel behaviors indicate the presence of heterogeneous sorting among residents and tourists. Notably, the travel distance for residents only decreases for fast food, grocery, and gyms, which are amenities offered directly within the University Village. This supports the hypothesis that residents in the Nexus area venture out less into the City due to the University Village.

As an alternative measure of demographics, I cluster users using account bios available for 609,621 of 798,591 users (76 percent). Specifically, I encode account bios using a Large Language Model (LLM) and retrieve vectorized representations of each account bio. I k-Means cluster the user representations from that model into 100 clusters. I choose $k-$Means clustering because it yields more coherent centroids in the data relative to other clustering techniques (see Appendix **??**). Moreover, I show that the quality of the clusters is robust to the choice of k. Following Grootendorst (2022), I retrieve representative keywords for each cluster using a Bag of Words (BoW) with class-based Total Frequency-Inverse Document Frequency (c-TF-IDF) approach. From these keywords, I manually infer demographic information such as ideology, race, religion, and occupation. Appendix **??** describes the pipeline in full detail.

Table 3 reports difference-in-difference estimates from equation (2) on the effects of the University Village opening on the log of the percentage share of users in different occupations and interest groups.[10] As a result of the University Village opening, the share of influencers and users working for media platforms in the USC Nexus area increase by 24.2 percent (significant at 5%) and 18.9 percent (significant at 10%) respectively, relative to other non-neighboring tracts with comparable census demographics and property valuations pre-construction. The share of occupations in art, design, and marketing also increase, though estimates are not statistically significant. The share of users interested in cars, photography, and wellness increase by 28.2 percent (significant at 1%), 5.7 percent (significant at 10%), and 17.7 percent (significant at 10%) respectively. These occupations and interests are arguably touristic.

The share of students and university workers ("University") in the Nexus area decrease by 14.8 percent (significant at 10%). The share of researchers, entrepreneurs, brokers, and users in the tech industry also decreases, though estimates are not statistically significant. These occupations mostly require higher education and are likely driven by residents of the University of South California area. The change in user composition supports the hypothesis that residents venture less into neighboring tracts.

[10] I transform user shares (in pp.) by the map $x \mapsto \log(x+1)$ to account for possible zeros.

TABLE 3—EFFECTS OF THE UNIVERSITY VILLAGE OPENING ON USER COMPOSITION

	Point estimate	Standard error	Control pct share	Pct Δ
Race				
Black	0.17	(0.20)	10.8	17.2
Hispanic	-0.03	(0.23)	9.8	-3.3
Asian	0.03	(0.10)	3.7	3.1
White	0.02	(0.14)	4.3	2.4
Other	-0.07	(0.06)	0.8	-6.7
Ideology				
Liberal	-0.13	(0.21)	11.8	-13.2
Conservative	-0.06	(0.08)	1.7	-6.1
Religion				
Christianity	-0.11	(0.21)	14.3	-11.4
Islam	0.02	(0.03)	0.3	1.8
Judaism	-0.17**	(0.09)	2.1	-17.5
Buddhism	-0.04	(0.09)	0.7	-3.8
Hinduism	-0.23***	(0.08)	4.0	-22.7
Other	0.24	(0.24)	13.4	23.7
Occupation				
Influencer	0.48*	(0.28)	31.4	48.2
Business	-0.07	(0.19)	10.8	-7.3
University	-0.25**	(0.11)	3.0	-25.1
Travel	0.14	(0.12)	3.7	14.1
Art	0.16	(0.20)	7.9	16.4

Note: dependent variable is the logged share of users. The point estimate is the coefficient from an OLS regression model. The standard error is robust. The control count is the number of control tracts in the model. The percent change is calculated as $100 * \beta$.

IV. Mechanisms

Are changes in user mobility driven by physical improvements in surrounding neighborhoods, or changes in the level of "social capital", including the quality of local networks and trust?

A. *Property Value and Physical Improvements*

I measure physical improvements through 1) Zillow single-family home prices, 2) the dollar valuation of active building permits, 3) the number of permits and business openings by residential or touristic use, 4) the number of building additions, alterations, or demolitions, and 5) the number of home renovations (including window, pool, bedroom, kitchen, patio and roof covering, and tenant improvements).

The effect of the University Village on Zillow prices and permit valuations in the USC Nexus area is inconclusive. Appendix Figure A3 plots estimates from equation (1) of the effect on the logged first difference of single-family home prices and building permits as a result of the University Village construction. The logged first differences in single-family home prices (Panel 3(a)) and permit valuations

Table 4—Effects of the University Village Opening on Physical Improvements

	Point estimate	Standard error	Control count	Pct Δ
N Permits	-0.04	(0.20)	5.6	-3.5
Touristic	-0.43*	(0.26)	0.2	-35.0
Residential	0.13	(0.12)	3.9	13.5
N Openings	-0.21***	(0.07)	5.0	-19.0
Touristic	0.08	(0.19)	0.8	8.4
Residential	0.02	(0.16)	0.4	2.1
Building construction				
Addition	0.36	(0.31)	0.6	44.0
Alteration	-0.13	(0.17)	3.5	-12.5
Demolition	0.30	(0.42)	0.4	34.9
Home renovations				
Tenant	0.25	(0.37)	0.2	28.5
Kitchen	0.52**	(0.26)	0.4	67.8
Bedroom	-0.17	(0.21)	0.3	-15.9
Roof cover	0.03	(0.20)	1.2	3.1
Windows	-0.11	(0.28)	0.5	-10.6
Patio cover	-0.40	(0.51)	0.2	-32.6
Pool	-0.42	(0.94)	0.0	-34.5

Note: dependent variable is the logged share of users. The point estimate is the coefficient from an OLS regression model. The standard error is robust. The control count is the number of control tracts in the model. The percent change is calculated as $100 * \beta$.

(Panel 3(b)) remain insignificant and close to zero leading up to construction. In the first quarter after opening, the price of single-family homes increases, while the valuation of active building permits decreases. These effects are transient and statistically insignificant after the second quarter.

Table 4 reports difference-in-difference estimates from equation (2) on the effect of the University Village opening on the number of physical improvements in Nexus tracts. The overall level of issued permits decreases, though estimates are statistically insignificant. The decline is statistically significant for the subcategory of touristic permits (including fast food, restaurants, bars, gyms). These amenities are offered within the University Village, so permit declines in these categories may simply reflect the expiration of building permits upon the completion of the University Village. This is plausible, since there are only on average 0.2 touristic permits per quarter before the construction of the University Village. The number of business openings decreases by 19.0 percent (point estimate of -0.21, significant at 1%), though the decline is not explained by any specific amenity category. Except for kitchen improvements, the effect of the University Village on the number of home renovations is insignificant. Effects on building additions, alterations, and demolitions in neighboring tracts are insignificant. Changes in mobility patterns do not correspond with changes in the level of physical improvements.

B. Segregation

Do changes in user mobility reflect changes in the underlying social network? Specifically, I investigate whether changes in user composition indicate increasing levels of segregation. Does the opening of the University Village encourage users to visit places disproportionately visited by members of the same social strata? I quantify the presence of this mechanism using the "isolation index" (e.g., White (1986), Gentzkow and Shapiro (2011), Davis et al. (2019)). The isolation index for user group g is defined as

$$I_g = \sum_j \frac{v_{gj}}{v_g} \cdot \left(\frac{v_{gj}}{v_j}\right) - \sum_j \frac{v_{g^c j}}{v_{g^c}} \cdot \left(\frac{v_{gj}}{v_j}\right), \tag{3}$$

where v_{gj} is the number of visits to amenity j by group g and $v_{g^c j}$ is the number of visits to amenity j by non-members. Other variables are defined analogously. The first summand is an interaction between the share of group g that visits amenity j and the share of visits to amenity j comprising of that same group. The second summand is an interaction between the share of other groups g^c who visit amenity j and the share of visits to amenity j comprising group g. I_g measures the difference between the average exposure to group g by members of g and the exposure to group g by non-members.[11] The index varies from 0 (all user groups visit the same amenity) to 1 (each group only visits an amenity composed completely of members of the same group).

Table 5 reports estimates from equation (2) on the effect of the University Village opening on ideological, racial, and religious segregation. The first set of columns report changes in the isolation index defined by (3). The second set of columns report changes to the first summation in (3), which measures within-group exposure. The last set of columns report changes to the second summation in (3), which measures the average exposure to a group by other groups. The level of ideological segregation increases by 103.5 percent due to the University Village opening (point estimate of 1.04, significant at 5%). Subgroup estimates are positive for both liberals and conservatives, suggesting that self-isolation occurs on both ends of the ideological spectrum. Estimates are most significant for conservatives, who experience an increase in isolation by 68.8 percent (point estimate of 0.69, significant at 10%). The isolation of conservatives is mainly driven by increases in within-group exposure (point estimate of 0.58, significant at 5%) rather than changes to exposure by liberals.

The isolation index rises for all racial groups except Asians, though point estimates are not statistically significant. Within-race exposure rises for all racial

[11] To correct for finite-sample bias, I use $I_g = \sum_j \frac{v_{gj}}{v_g} \cdot \left(\frac{v_{gj}-1}{v_j-1}\right) - \sum_j \frac{v_{g^c j}}{v_{g^c}} \cdot \left(\frac{v_{gj}}{v_j-1}\right)$, which computes the exposure to others minus oneself.

Table 5—Effects of the University Village Opening on Segregation

	Isolation index	Pct Δ	Within-group exposure	Pct Δ	Cross-group exposure	Pct Δ
Ideology	1.04**	103.5	0.81*	81.2	0.26	25.9
	(0.41)		(0.41)		(0.41)	
Liberal	0.75	75.1	0.70	70.2	0.25	24.9
	(0.43)		(0.41)		(0.42)	
Conservative	0.69*	68.8	0.58**	58.5	0.10	9.5
	(0.31)		(0.26)		(0.44)	
Race	0.10	9.7	0.18	17.7	0.49*	48.8
	(0.22)		(0.23)		(0.25)	
Black	0.03	2.7	0.45	44.6	0.42	42.4
	(0.26)		(0.31)		(0.31)	
Hispanic	0.13	13.4	0.37	37.5	0.44*	44.2
	(0.32)		(0.31)		(0.23)	
Asian	-0.04	-3.9	0.15	14.9	0.65	64.7
	(0.25)		(0.39)		(0.41)	
White	0.13	13.5	0.07	7.1	0.32*	32.2
	(0.23)		(0.22)		(0.18)	
Other	0.43	42.7	0.71**	71.3	0.39*	38.7
	(0.36)		(0.33)		(0.21)	
Religion	0.28	28.1	0.38	38.2	0.33*	33.3
	(0.27)		(0.26)		(0.19)	
Christianity	-0.30	-29.8	-0.17	-16.5	0.15	15.1
	(0.24)		(0.28)		(0.23)	
Islam	0.01	1.2	0.19	18.9	0.40	40.2
	(0.17)		(0.32)		(0.40)	
Judaism	-0.36	-36.1	-0.49**	-49.1	-0.20	-19.6
	(0.28)		(0.22)		(0.21)	
Hinduism	-0.42	-41.7	-0.55*	-54.7	-0.11	-10.8
	(0.29)		(0.29)		(0.27)	
Buddhism	0.07	7.2	-0.08	-8.3	0.07	7.5
	(0.21)		(0.30)		(0.32)	
Other	0.43	42.7	0.71**	71.3	0.39*	38.7
	(0.36)		(0.33)		(0.21)	

Note: dependent variable is the logged share of users. The point estimate is the coefficient from an OLS regression model. The standard error is robust. The control count is the number of control tracts in the model. The percent change is calculated as $100 * \beta$.

groups. However, the exposure to each race by other races also rises. The effects on cross-race exposure are significant at the 10% level (point estimate of 0.49), in particular for Hispanics (0.44), Whites (0.32), and other races excluding Blacks and Asians (0.39). Since both within-race and cross-race exposure rise to varying extents, changes to racial segregation are inconclusive.

Changes to religious segregation are statistically insignificant and the direction of effect is inconclusive across religious groups.

C. Trust and Hate Speech

Elements of social organization, including levels of segregation, determine the level of trust and types of norms formed in a community (Putnam, 2000; Durlauf and Fafchamps, 2005), which in turn generate externalities for its members (Coleman, 1990; Putnam, Leonardi and Nanetti, 1993). Does the opening of the University Village affect the formation of trust?

I proxy for the *absence* of trust by identifying polarizing Tweets, in particular hateful and offensive speech. Recent empirical studies of online behavior suggest that there is little evidence of "pure hate"; social media users do not exclusively post hateful comments (Cinelli et al., 2021). Instead, hateful and offensive behavior is often the byproduct of misinformation, either due to the dissemination of strongly emotional speech (e.g., Athey et al. (2023)), or the presence of "echo chambers", when users lack of reliable access to multiple sources of information (e.g., Gentzkow and Shapiro (2011)). These factors are amplified by user segregation and may correlate with distrust.

To test this hypothesis, I leverage a state-of-the-art Large Language Model to identify instances of hateful or offensive speech and characterize the type of offense (including derogation, animosity, dehumanization, and physical threats) and target identity groups (including race, religion, nationality, gender, and sexuality). Specifically, I pool together four datasets of annotated social media posts (Davidson et al., 2017; Kennedy et al., 2020; Mathew et al., 2021; Vidgen et al., 2021), comprising over 200,000 training instances, over 10 different forms of hate or offensive speech, and over 40 different types of target identity subgroups. By comparing the data-generating process across datasets, I create a training dataset ordered by the difficulty of classification. On this dataset, I finetune BERTweet (Nguyen, Vu and Nguyen, 2020), an LLM trained on 850 million Tweets between 01/2012 to 08/2019. The model is tasked to simultaneously identify (a) the prevalence of hate (F1 score of 0.83), (b) the type of hate (macro-F1 score of 0.65), and (c) the target identity group (macro-F1 score of 0.68). I follow recommendations by Yu et al. (2020) to address optimization challenges with multi-task learning. Appendix **??** describes examples of hate speech, the training process, and model performance. I use the fine-tuned model to classify 15,194,795 geo-tagged English Tweets longer than 10 characters. The model identifies 281,886 hateful or offensive Tweets with above 80 percent certainty.

Davidson et al. (2017) observes that some hate lexicons and targets of offensive speech tend to be correlated (e.g., gender and sexuality). To control for correlations between categories of speech, I follow Anderson (2008) and construct a "standardized weighted index" over nine categories (four types and five targets of hate speech). For each category of hate speech, I demean the number of Tweets in each tract i and calendar quarter t and divide by the standard deviation to get a column vector $\tilde{\mathbf{y}}$. This controls for the possibility that some types of speech are more easily identifiable than others. Next, I compute the weighted average outcome in each tract and calendar quarter:

$$\bar{s}_{it} = (\mathbf{1}^T \mathbf{\Sigma}^{-1} \mathbf{1})^{-1} (\mathbf{1}^T \mathbf{\Sigma}^{-1} \tilde{\mathbf{y}}_{it}),$$

where $\mathbf{1}$ is a column vector of ones and $\mathbf{\Sigma}^{-1}$ is the inverse covariance matrix of the speech categories. Each category of speech takes a weight equal to the sum of the row entries of $\mathbf{\Sigma}^{-1}$. This procedure assigns smaller weights to categories of speech that are highly correlated with other categories. The normalized "prevalence index" is computed by demeaning $\bar{s}_{it}$ and normalizing by the standard deviation of each category. Asymptotically, this index has mean-zero and standard deviation one.

TABLE 6—EFFECTS OF THE UNIVERSITY VILLAGE OPENING ON HATE SPEECH

	Point estimate	Standard error	Control pct share	Pct Δ
Prevalence index	0.01	(0.05)		
Speech				
Derogation	0.20*	(0.11)	2.0	19.9
Animosity	0.21**	(0.09)	0.7	20.6
Dehumanization	0.11**	(0.06)	0.2	11.1
Threat			0.0	
Target				
Race	0.12	(0.08)	0.9	11.5
Religion			0.0	
Nationality			0.0	
Gender	0.22**	(0.09)	1.1	22.1
Sexuality	0.06**	(0.02)	0.1	6.2

Note: dependent variable is the logged share of users. The point estimate is the coefficient from an OLS regression model. The standard error is robust. The control count is the number of control tracts in the model. The percent change is calculated as $100 * \beta$.

Table 6 reports the effects of the University Village opening on the dissemination of hateful and offensive speech in the USC Nexus area. Instances of derogation, animosity, and dehumanizing comments increase by 19.9 percent (point estimate of 0.20, significant at 10%), 20.6 percent (point estimate of 0.21, significant at 5%), and 11.1 percent (point estimate of 0.11, significant at 5%) respectively. The rise in targeted speech is most prominent for gender (increase by 22.1 percent, significant at 5%) and sexuality (increase by 6.2 percent, significant at 5%). Levels of race-directed hate speech also increase, though estimates are not statistically significant. There are no instances of physical threats or speech directed at a particular religion or nationality in the data. Estimates of changes in the prevalence index are positive but not statistically significant.

REFERENCES

Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf. 2015. "The economics of density: evidence from the Berlin Wall." *Econometrica*, 83(6): 2127–2189.

Almagro, Milena, and Tomás Domínguez-Iino. 2024. "Location Sorting and Endogenous Amenities: Evidence from Amsterdam." *Working Paper*.

Anderson, Michael L. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103: 1481–1495.

Athey, Susan Carleton, Matias Cersosimo, Kristine Koutout, and Zelin Li. 2023. "Emotion- versus Reasoning-Based Drivers of Misinformation Sharing: A Field Experiment Using Text Message Courses in Kenya." Stanford University Graduate School of Business Research Paper 4489759.

Baum-Snow, Nathaniel. 2007. "Did Highways Cause Suburbanization?" *The Quarterly Journal of Economics*, 122(2): 775–805.

Bramoullé, Yann, Bernard Fortin, and Habiba Djebbari. 2020. "Peer Effects in Networks: A Survey." *Annual Review of Economics*, 12: 603–629.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt. 2022. "Social capital I: measurement and associations with economic mobility." *Nature*, 608(7921): 108–121.

Cinelli, Matteo, Andraž Pelicon, Igor Mozetič, et al. 2021. "Dynamics of online hate and misinformation." *Scientific Reports*, 11(1): 22083.

Coleman, James. 1990. *The Foundations of Social Theory.* Cambridge:Harvard University Press.

Davidson, Thomas, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. "Automated Hate Speech Detection and the Problem of Offensive Language." *ICWSM '17*, 512–515.

Davis, Donald R, Jonathan I Dingel, Joan Monras, and Eduardo Morales. 2019. "How Segregated Is Urban Consumption?" *Journal of Political Economy*, 127(4): 1684–1738.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workersí Diverging Location Choices by Skill: 1980-2000." *American Economic Review*, 106(3): 479–524.

Ding, Lei, Jackelyn Hwang, and Eileen Divringi. 2016. "Gentrification and residential mobility in Philadelphia." *Regional Science and Urban Economics*, 61: 38–51.

Durlauf, Steven N. 2004. "Neighborhood effects." *Handbook of regional and urban economics*, 4: 2173–2242.

Durlauf, Steven N., and Marcel Fafchamps. 2005. "Social Capital." In *Handbook of Economic Growth*. Vol. 1. 1 ed., , ed. Philippe Aghion and Steven Durlauf, Chapter 26, 1639–1699. Elsevier.

Gentzkow, Matthew, and Jesse M. Shapiro. 2011. "Ideological Segregation Online and Offline." *The Quarterly Journal of Economics*, 126(4): 1799–1839.

Goldsmith-Pinkham, Paul, and Guido Imbens. 2013. "Social Networks and the Identification of Peer Effects." *Journal of Business Economic Statistics*, 31(3): 253–264.

Golub, Benjamin, and Matthew O. Jackson. 2011. "How Homophily Affects the Speed of Learning and Best-Response Dynamics." Unpublished manuscript.

Grootendorst, Maarten. 2022. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." *arXiv preprint arXiv:2203.05794*.

Guerrieri, Veronica, Daniel Hartley, and Erik Hurst. 2013. "Endogenous Gentrification and Housing Price Dynamics." *Journal of Public Economics*, 100: 45–60.

Gupta, Arpit, Stijn Van Nieuwerburgh, and Constantine Kontokosta. 2020. "Take the Q-Train: Value capture of public infrastructure projects." *Working Paper*.

Kennedy, Chris J, Geoff Bacon, Alexander Sahn, and Claudia von Vacano. 2020. "Constructing interval variables via faceted Rasch measurement and multitask deep learning: a hate speech application." *arXiv preprint arXiv:2009.10277*.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies*, 60(3): 531–542.

Mathew, Binny, Punyajoy Saha, Seid Muhie Yimam, Chris Biemann, Pawan Goyal, and Animesh Mukherjee. 2021. "HateXplain: A Benchmark Dataset for Explainable Hate Speech Detection."

Nguyen, Dat Quoc, Thanh Vu, and Anh Tuan Nguyen. 2020. "BERTweet: A pre-trained language model for English Tweets." 9–14.

Putnam, Robert D. 2000. *Bowling Alone.* New York:Simon and Schuster.

Putnam, Robert D., Robert Leonardi, and Raffaella Y. Nanetti. 1993. *Making Democracy Work: Civic Traditions in Modern Italy.* Princeton:Princeton University Press.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, 70(1): 41–55.

Sayyadiharikandeh, Mohsen, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2020. "Detection of Novel Social Bots by Ensembles of Specialized Classifiers." *CIKM '20.* ACM.

Smith, Aaron, and Monica Anderson. 2018. "Social Media Use in 2018." Pew Research Center.

Su, Yichen. 2022. "The Rising Value of Time and the Origin of Urban Gentrification." *American Economic Journal: Economic Policy*, 14(1): 402–439.

Vidgen, Bertie, Tristan Thrush, Zeerak Waseem, and Douwe Kiela. 2021. "Learning from the Worst: Dynamically Generated Datasets to Improve Online Hate Detection."

White, Michael J. 1986. "Segregation and Diversity Measures in Population Distribution." *Population Index*, 52(2): 198–221.

Yu, Tianhe, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. 2020. "Gradient Surgery for Multi-Task Learning."

Supplementary Figures and Tables

Figure A1. Balance plot of pre-treatment characteristics

Notes: Standardized differences in pre-treatment characteristics between treated and control groups.

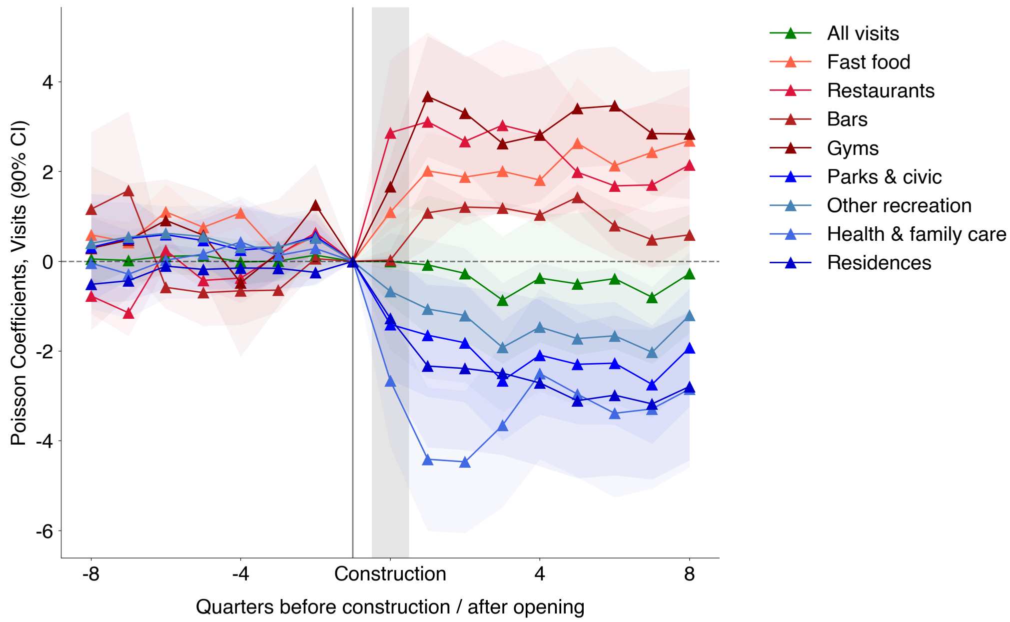
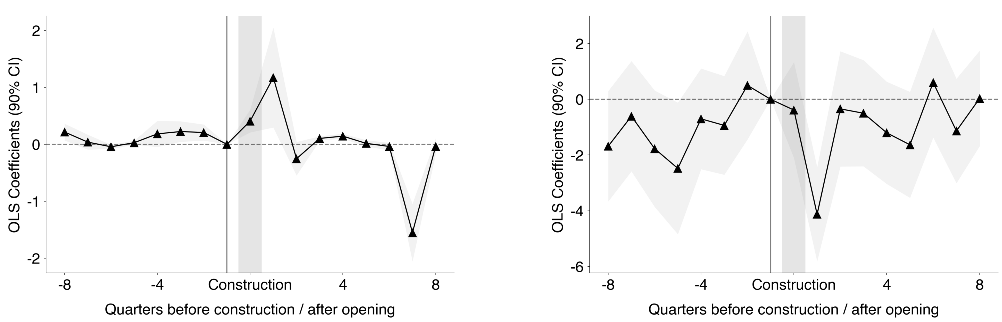


FIGURE A2. EFFECTS OF THE UNIVERSITY VILLAGE ON AMENITY VISITS

Notes: The figure plots estimates and 90% confidence intervals for β^{τ} from equation (1), which is a Poisson regression of the number of Twitter visits (by amenity type) on tract fixed-effects, calendar-quarter fixed effects, and event quarter indicators interacted with a treatment indicator for tracts in the USC Nexus area. The quarter before the start of construction is the excluded period. The control sample contains propensity-matched tracts that are not directly bordering the USC Nexus area with comparable demographic characteristics and property valuations pre-construction. Standard errors are clustered on the tract level.



(a) Zillow Single-Family Home Prices (log of Δ)

(b) Permit Valuations (log of Δ)

FIGURE A3. EFFECTS OF THE UNIVERSITY VILLAGE ON PROPERTY VALUE

Table A1—Effects of the University Village Opening on Prevalence of Political Tweets

	Point estimate	Standard error	Control pct share	Pct Δ
Politicians	0.04	(0.14)	0.9	4.5
Policy	0.08	(0.11)	0.7	11.9
Ideology	-0.03	(0.03)	0.1	-37.6
Partisan	-0.02	(0.05)	0.6	-4.0
Memorial	0.00	(0.08)	0.7	0.1
Community	0.10	(0.18)	1.7	6.3
Poverty	-0.02	(0.09)	0.5	-4.8

Note: dependent variable is the logged share of users. The point estimate is the coefficient from an OLS regression model. The standard error is robust. The control count is the number of control tracts in the model. The percent change is calculated as $100 * \beta$.