

Gender Homophily and Segregation Within Neighborhoods

Gregorio Caetano and Vikram Maheshri*

June 27, 2017

Abstract

Homophily generates segregation, reducing diversity in peer groups and leading to narrower social interactions. Using novel data from Foursquare, a popular mobile app that documents the activity of millions of people, we document robust, highly localized gender homophily: over half of the gender segregation of individuals' recreational and commercial activities in thousands of venues (e.g., shops, restaurants, parks, museums) in eight major US cities occurs within census blocks. Gender segregation is mostly driven by venue offerings, not discriminatory preferences. A higher variety in the supply of venues on a block attracts more gender-balanced visitors, but, perversely, more intense sorting across those venues ultimately reduces the actual exposure of individuals to gender diversity in venues. Using employment data from the US Census, we find suggestive evidence that these homophilic forces may contribute to the gender gap in labor force participation. Our analysis also suggests that localized homophily along other demographic dimensions may be similarly prevalent. JEL: R1, R2, R3, J1, J3. Keywords: Gender Segregation, Homophily, Peer Groups, Urban Sorting, Diversity.

1 Introduction

Homophily, or the tendency of similar people to associate with each other (McPherson et al. (2001)), is a pervasive, gravitational social force that leads to segregated peer groups. Segregation as a social phenomenon has been widely studied in a number of important contexts, such as residential neighborhoods, schools and workplaces (Card et al. (2008a); Boustan (2012); Echenique et al. (2006); Fernandez et al. (2000)). While segregation at these levels partially determines peer groups, many further daily choices may expose people to very different social interactions. For

*University of Rochester and University of Houston. We are extremely grateful to Foursquare, Inc. for generous access to their restricted, anonymized data, and to Michael Li and Blake Shaw for helpful discussions about the data. We thank Alexei Alexandrov, Dionissi Aliprantis, Dani Arribas-Bel, Donald Davis, Joan Monras, Richard Murphy, Romans Pancs, Stephen Ross, and various seminar and conference participants for helpful discussions. We also thank Riley Hadden and Hao Teng for excellent research assistance. All errors are our own.

instance, neighbors may shop at different supermarkets, students may select different extracurricular activities, and coworkers may exercise at different gyms.

Although these mundane decisions may influence peer group formation, they are difficult to account for due to data availability. This difficulty is compounded by what we term the *paradox of diversity*: as individuals are supplied with a more diverse set of choices, they will tend to be exposed to a less diverse set of peers. The ideal environment for this paradox to prevail, one that is densely populated with diverse individuals and options, is precisely that in which neighborhood effects have been most widely studied: large, metropolitan areas.

In this paper, we exploit a unique data set from a prominent location-based social network, Foursquare, that documents how individuals in eight major US cities¹ sort by gender across tens of thousands of commercial and recreational venues such as shops, restaurants, parks, churches and museums that offer the activities that constitute much of people’s social lives. We find evidence of substantial gender homophily in individuals’ venue choices, which results in more highly segregated peer groups than would otherwise be measured with residential data alone. We also find strong evidence for the paradox of diversity. Gender segregation at the finer, venue level is facilitated by the urban landscape, as neighborhoods rich in a variety of offerings encourage homophilic forces. Although each of the decisions that we observe are trivial in isolation, cumulatively, they may have measurable effects on important economic and social outcomes.² For instance, we provide some suggestive evidence that gender segregation in venues may contribute to the gender gap in labor force participation.

Our first finding is that gender segregation is highly localized: 80-90 percent of such segregation in venues is observed within census tracts, and over half of it is observed within census blocks. Thus, the level of actual gender segregation to which people are exposed is substantially higher than can

¹Our analysis covers New York City, Los Angeles, Chicago, Dallas, Washington DC, San Francisco, Atlanta and Philadelphia.

²There is a small but growing experimental literature that finds the gender of peers with whom individuals make casual contact in venues has measurable effects. For instance, Kniffin et al. (2016) show that when dining in restaurants, men eat significantly more (nearly twice as much) in the presence of women than in the presence of men. And marketing studies of retail venues (e.g., Tifferet and Herstein (2012)) have found that women exhibit higher levels of impulse buying than men, and being observed by other women has been identified as a relevant sensory cue for shoppers (Meyers-Levy and Sternthal (1991)). From a theoretical perspective, Akerlof and Kranton (2000) build a theory of gender identity, developed in part through casual encounters, by which, “...society-wide changes are necessary to change gender norms... The model predicts many implications of such changes. Womens’ participation in the labor force will increase. Occupational segregation will decrease...” (p. 90) and “...gender norms significantly influence the division of labor and leisure” (p. 92).

be measured from residential neighborhood data. Put another way, the level of gender diversity to which people are exposed in day-to-day activities is at least 25 percent (one standard deviation) lower than the reported aggregate levels of residential diversity for roughly half of all neighborhoods in our sample.

Given this robust finding, we ask how gender segregation arises and what policies, if any, can affect local diversity. Segregated peer groups might potentially arise from two homophilic forces: *active* segregation might occur because individuals prefer the company of similar peers, and *passive* segregation might occur because similar individuals prefer similar activities, which leads them to visit the same venues. Although active segregation has been more widely discussed in the broad literatures on segregation and discrimination (Schelling (1969); Bruch and Mare (2006); Bobo et al. (2012); Boustan (2012)), passive segregation can also be an important driver of segregation (Banzhaf and Walsh (2013); Caetano and Maheshri (2017)). Empirically, we find that the segregation patterns in our data are most consistent with passive homophilic forces: men and women simply tend to prefer different types of activities. We present a simple model of venue choice in the spirit of Hotelling (1929) to illustrate how variety in the supply of venues in a neighborhood affects diversity at both the neighborhood and the venue levels, and we show that the directions of both of these effects are theoretically ambiguous. We then estimate these effects with three alternative and complementary identification strategies that all support the paradox of diversity: greater venue variety attracts more gender-diverse visitors to a neighborhood, but once there, individuals tend to self-segregate more intensely across venues, thereby *reducing* the amount of gender diversity to which they are exposed. As a result, denser urban areas may actually foster narrower social interactions by providing more opportunities for people to sort into specific venues.

While casual interactions in shops and recreational venues may seem trivial, cumulatively they may be influential. We attempt to illustrate one potential consequence of gender segregation in venues by connecting our analysis to the literature on gender gaps in the labor market. Loury (2006) finds that female informal contacts have a lower impact on employment outcomes than male informal contacts, implying that gender segregated referral networks may contribute to the gender gap. Indeed, Bayer et al. (2008) show that interactions among neighborhood residents are gender segregated, and they seem to contribute to the gender gap in labor force participation. We hypothesize that similar networks may develop between neighborhood residents and venue visitors

(who might or might not reside in the same neighborhood) through their day-to-day interactions and adapt the identification strategy directly from Bayer et al. (2008) to estimate the effects of such interactions. Among low wage employees, and only among low wage employees, we find evidence that venue gender segregation contributes to gender gap in labor force participation even after accounting for many confounding factors. These results are consistent with theoretical and empirical evidence of the “strength of weak ties” (Granovetter (1973); Montgomery (1992)) and the notion that informal social networks are particularly valuable to individuals who are less attached to the labor market (Montgomery (1991); Fernandez et al. (2000); Ioannides and Loury (2004)).

Methodologically, our paper attempts to contribute to a growing literature that leverages user-generated data to study behaviors that previously proved impossible to observe (Couture (2014); Davis et al. (2014)). Although these datasets offer much promise, they are often plagued by selection issues that make it difficult to extract a meaningful, externally valid signal. Many broad questions in social science are difficult to approach comprehensively because in practice, one cannot observe all of the choices that jointly determine individuals’ social interactions. Moreover, segregation is an end result of homophily along many potential dimensions, many of which are difficult to observe. Thus, any study like the one conducted in this paper is bound to use data that is both incomplete and unrepresentative. With these obstacles in mind, we develop an empirical approach throughout the paper that attempts to reach only conservative qualitative conclusions, i.e., all of our conclusions would plausibly strengthen with access to more detailed and complete data. Such an approach could be useful for other studies facing similar issues; to that end, we provide a detailed sensitivity analysis in the appendix including a detailed Monte Carlo study of the implications of potential selection issues in our user-generated data and a variety of other checks for measurement error in the spirit of Carrington and Troske (1997) and Allen et al. (2015).

The remainder of the paper is organized as follows. In Section 2, we describe our data set, and in Section 3, we show widespread evidence of gender segregation in location choices. In Section 4, we explore the causes of this phenomenon with a simple model of sorting across venues, and we show empirically that the variety of venues that are available in neighborhoods impacts both the levels of diversity in venues and in neighborhoods but in opposite directions. In Section 5, we show suggestive evidence that gender segregation in venues may contribute to the gender gap in labor force participation. In Section 6, we discuss the external validity of our findings to other

environments and along dimensions other than gender. We conclude in Section 7.

2 Data

For a local analysis of individual interactions, we require comprehensive, disaggregated data of their whereabouts across a large number of locations within small neighborhoods; this is difficult to observe directly. We circumvent this issue with novel, proprietary data from Foursquare, Inc., creators of the eponymous mobile app and social network that allows users to document their precise whereabouts electronically. Upon arriving at a venue, Foursquare identifies the venue by GPS on the user’s mobile phone, and the user can electronically “check in”. We use information on the demographic composition of Foursquare users in each venue to construct a proxy for the actual demographic composition of all individuals (i.e., Foursquare and non-Foursquare users) in the venue. Although this raises important concerns of sample selection, we develop an empirical approach with these concerns specifically in mind. We show that our empirical approach allows us to extract a meaningful signal about the sorting of all individuals across venues from this novel dataset. A comprehensive sensitivity analysis concerning the potential sources of measurement error that might exist in our data is provided in the appendix.³

Ours is the first paper to use this large and highly detailed database of venue visitors to study diversity within neighborhoods.⁴ Foursquare is particularly suitable for our analysis because it is a prominent location-based social network that boasts a large number of active users (over 50 million worldwide checking in over 6 billion times as of March 2015), which makes for a highly detailed catalog of activity.

Our data set contains information on all Foursquare activity in venues in eight major US cities: Atlanta, Chicago, Dallas, Los Angeles, New York City, Philadelphia, San Francisco and Washington, DC. Our specific sample regions are defined as the counties in which these cities are primarily located.⁵ For each of the 76,377 venues that are tracked in these cities, Foursquare has directly

³In the appendix, we show that our main results will not change even if the propensity to “check-in”, conditional on visiting the venue, differs by gender. We also conduct a Monte Carlo study that shows that our results are, if anything, conservative. That is, in worst case scenarios of measurement error (i.e., when propensity to “check-in” varies by gender as a function of the female share of actual visitors) our main results change only slightly and, if anything, become stronger in the direction of our conclusions.

⁴A small number of studies (e.g., Arribas-Bel and Bakens (2014)) have begun to use Foursquare data obtained indirectly via the Foursquare API (application programming interface). Foursquare data obtained via the API unfortunately does not disaggregate check-ins along any demographic dimension.

⁵The counties are Fulton (Atlanta), Cook (Chicago), Dallas, Los Angeles, New York, Philadelphia, and San

provided to us in fully anonymized form the number of daily check-ins by male and female users from August 1, 2012 to July 31, 2013. This data is aggregated to the venue level, hence we cannot observe any characteristics of individual Foursquare users, nor can we track a particular individual’s activity. We restrict our sample to venues that experienced at least 10 check-ins during the sample period to improve our measurements of the gender compositions of venues.⁶ In total, these venues experienced 49.6 million check-ins during the sample period with the average venue in our sample experiencing 649 check-ins. Each venue in our data set is also geo-coded by latitude and longitude, which allows us to link venues to unique census tracts, block groups and blocks using neighborhood definitions from the 2010 Decennial Census.

In Table 1, we summarize our sample by city and by venue classification. Not surprisingly, larger cities such as New York and Los Angeles have more venues and check-ins. Males tend to check in slightly more than females on average, but there is substantial and robust variation in the gender composition of venues in all cities. It is immediate that there is more variation in the average gender composition of venues across categories than across cities and more variation in gender composition within categories than within cities.⁷ The 9 categories of venues are further classified into 225 narrow subcategories; detailed summary statistics disaggregated by subcategory can be found in the appendix.

Because we observe daily check-ins at each venue, we can assess whether there are any dynamic trends in our data over the sample period. As shown in the first panel of Figure 1, there is substantial day-of-week variation in check-ins since venues are more highly frequented on weekends, but the gender composition of check-ins is nearly constant. This suggests that we can aggregate the data at least to the weekly level to analyze gender diversity. We do so and check for aggregate weekly trends in our data in the second panel of Figure 1. There is no systematic weekly variation in check-in frequency and no discernible seasonality or aggregate trend. More importantly, the gender

Francisco respectively. We treat the entire District of Columbia as the “county” for Washington. Most of the cities in our sample are entirely contained in their corresponding county with the notable exception that New York County only contains the borough of Manhattan.

⁶We also only consider check-ins from users who have specified their gender. These restrictions do not seem to bias our results (see appendix).

⁷For each city in our sample, check-ins across venues are approximately distributed log-normally.

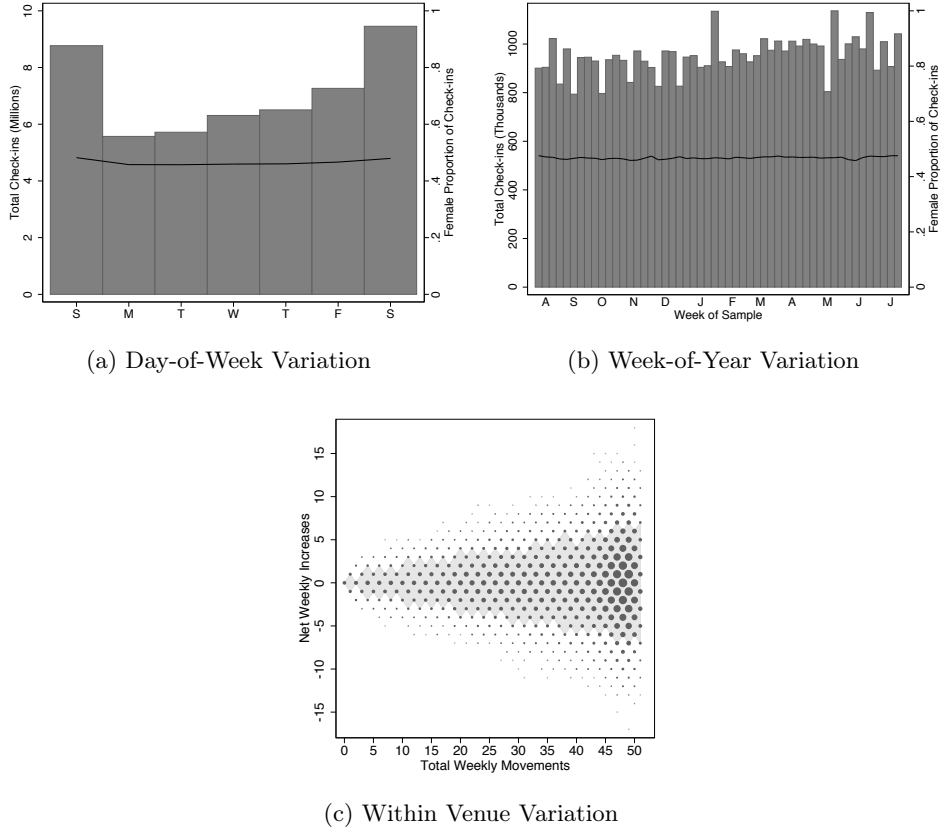
Table 1: Summary Statistics

City	Venues	Check-ins	μ	σ	$p_{75} - p_{25}$	Tracts	B. Groups	Blocks
Atlanta	4,115	2.84	0.46	0.17	0.19	180	361	1,307
Chicago	13,665	8.11	0.49	0.16	0.19	1,100	2,235	6,237
Dallas	5,065	2.40	0.45	0.16	0.19	421	774	1,986
Los Angeles	23,108	10.2	0.46	0.15	0.18	1,902	3,584	9,182
New York City	16,203	16.2	0.49	0.17	0.19	282	945	2,501
Philadelphia	3,933	2.10	0.47	0.16	0.19	301	568	1,757
San Francisco	6,601	4.78	0.42	0.15	0.16	182	440	1,898
Washington, DC	3,687	2.98	0.43	0.16	0.17	152	272	1,069
Category	Unique Subcategories							
Food	31,398	16.6	0.45	0.13	0.17	65		
Shops/Services	20,903	9.97	0.52	0.21	0.28	66		
Bars	6,441	6.52	0.44	0.12	0.13	20		
Outdoors	4,795	4.62	0.44	0.16	0.21	22		
Cafes	4,483	3.88	0.47	0.14	0.18	3		
Entertainment	4,189	4.08	0.46	0.13	0.15	29		
Hotels	1,798	2.24	0.40	0.11	0.13	5		
Gyms	1,625	1.41	0.49	0.23	0.34	12		
Spiritual	745	0.29	0.48	0.17	0.23	3		

Notes: Check-ins reported in millions. μ and σ refers to the mean and standard deviation of the proportion of females in venues, and p_{25} and p_{75} refer to the 25th and 75th percentiles of the proportion of females in venues.

composition of check-ins is roughly constant throughout the sample period. This suggests that we can aggregate the data set to the annual level to analyze gender diversity without loss of generality.

Figure 1: Check-ins and Gender Composition Over Time



Notes: (a), (b): Bars represent total check-ins, lines represent gender composition of aggregate check-ins. The 53rd week of the sample is omitted because it only contains a single day. (c): In this scatter plot of venues in our data, larger dots correspond to a greater numbers of venues. A venue experiences a weekly increase (decrease) in gender composition if the proportion of female check-ins rises (falls) by at least one percentage point.

To further support this choice of aggregation, we check whether the gender compositions of individual venues follow a trend over time. For each venue, we compute the net number of week-on-week increases (increases minus decreases) in the proportion of female check-ins over the sample period, and we plot them against the total number of changes in the proportion of female check-ins in Figure 1.⁸ Larger dots represent more venues in the sample, and the shaded region is defined to include 95% of all venues. It is immediate that most venues experience roughly as many relative increases in female popularity as relative decreases in female popularity. Because the gender composition of

⁸A venue is defined to experience a week-on-week increase (decrease) in the female share if its female share increases (decreases) by a threshold of at least one percentage point over consecutive weeks. The total number of changes in the proportion of female check-ins is equal to the sum of increases and decreases. We replicated panel (c) of Figure 1 with alternative thresholds of 5, 10 and 15 percentage points and obtained qualitatively similar results.

a venue tends to vary around a fixed value, it is appropriate to interpret longitudinal variation in check-ins as measurement error, which we minimize by aggregating our data to the annual level in order to focus on the more relevant cross-sectional variation in our data.⁹

3 Measuring Gender Homophily and Segregation In Neighborhoods

Homophily will lead the gender compositions of the venues to diverge from one another as individuals sort across them. In the extreme case, if some venues are only visited by females and others are only visited by males, then the venues are fully segregated and exhibit no gender diversity. One important and widely used measure of segregation is the Theil (1967) index.¹⁰ Formally, if s_{jk} is the share of females at venue j located in neighborhood k , then the Theil index of neighborhood k is given by

$$T_k = \frac{1}{n_k} \sum_{j \in k} \left(\frac{s_{jk}}{\bar{s}_k} \cdot \log \frac{s_{jk}}{\bar{s}_k} \right) \quad (1)$$

where n_k is the number of venues in the neighborhood and \bar{s}_k is the simple average of s_{jk} across all venues in the neighborhood.¹¹ If the neighborhood is fully integrated (i.e., no observable homophily and hence maximal diversity), then all of its venues will have the same gender composition as the neighborhood overall, and $T_k = 0$. Neighborhoods with less diverse venues have larger values of T_k .¹² In practice, k can correspond to the entirety of a city (c), a census tract (t), a census block group (g) or a census block (b), so T_k represents the extent to which venues in k are segregated by gender.

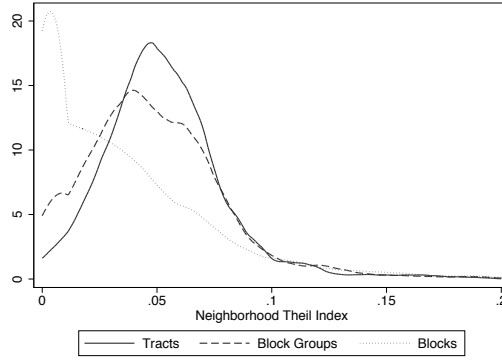
⁹As a robustness check, we replicated all main results of the paper by month-of-year and by day-of-week and found similar results (see appendix).

¹⁰Weitzman (1992) proposes a general, recursively defined measure of diversity that satisfies numerous attractive mathematical, economic and conceptual properties. In certain contexts, he shows it to be equivalent to the widely used Shannon index, which measures the amount of “true diversity” or the effective number of different types of “objects”. In our application, objects correspond to venues by demographic composition, and the Shannon index reduces to the Theil index up to an additive constant.

¹¹Our results are virtually unchanged if we denote s_{jk} as the share of men in venue j in neighborhood k .

¹²The maximum value that the Theil index can take is $\log n_k$, which varies with the density of venues in a neighborhood. Where applicable, our results using the Atkinson (1970) index (the Theil index divided by $\log n_k$, thus normalized to values between 0 and 1) are all qualitatively equivalent. As we explain below, we use the Theil index instead of the Atkinson index in our analysis because of its decomposability properties.

Figure 2: Densities of Theil Indices for Various Neighborhood Definitions



Notes: All densities are estimated using a bandwidth of 0.005 and an Epanechnikov kernel. For clarity, we present the density only for values of the domain less than 0.2; fewer than 1% of neighborhoods of any type have a Theil Index in excess of 0.2. Theil Indices are pooled across neighborhoods in all cities.

We compute the Theil index for each tract, block group and block in the cities in our sample and present the densities of these indices in Figure 2. The bulk of the density of T_t lies away from zero, which reveals that individuals sort within tracts. Similarly, the bulk of the density of T_g lies away from zero, which reveals that individuals also sort within block groups. The density of T_b is close to zero for approximately 10% of the sample, so roughly 90% of blocks in these cities are further sorted by gender in venues. Mathematically, $T_b \leq T_g \leq T_t$ for all $b \in g \in t$. Because these three densities roughly coincide for higher values of the Theil index, all of the sorting in highly homophilous tracts and block groups occurs within their constituent blocks as opposed to across them.

The Theil index possesses the attractive property of being additively decomposable, which allows for segregation in an entire city to be split into one term that captures segregation within neighborhoods and another term that captures segregation across neighborhoods.¹³ Formally, we

¹³Although the Theil index is not the only such measure that is additively decomposable, it is the only one that is homogeneous of degree zero (Bourguignon (1979)), which makes it invariant to rescaling. This is important in our application because males may be more or less likely to check in on Foursquare than females; hence in order to maintain the external validity of our estimates we should make only relative comparisons of homophily. In addition, as Shorrocks (1980) points out, other commonly used measures of segregation, diversity, exposure or inequality with other attractive properties which are based on the Herfindahl index (e.g., the index of segregation introduced in Ellison and Glaeser (1997)) or the Gini coefficient are not additively decomposable, so they are less useful and appropriate in our context.

can decompose the total Theil index of city c into within- and across- tract components as

$$T_c = \underbrace{\sum_{t \in c} \alpha_t \cdot T_t}_{\text{within-tracts}} + \underbrace{\sum_{t \in c} \alpha_t \cdot \log \frac{\bar{s}_t}{\bar{s}_c}}_{\text{across-tracts}} \quad (2)$$

where the weights $\alpha_t = \frac{n_t s_t}{n_c s_c}$ correspond to the contribution of each tract to overall venue diversity in c (s_k represents the share of females across all venues in neighborhood k). T_c can be similarly decomposed to the block group or block levels. The key benefit of this simple decomposition is that we can analyze neighborhood segregation (and hence homophily in venues) independently of how individuals sort across neighborhoods. In Table 2, we present the proportion of city-wide gender segregation in venues that is attributable to homophily within neighborhoods, i.e., the contribution of the first term of equation (2).¹⁴ Intuitively, this captures how much of the variation in the gender composition of city venues is “local.” It is immediate that the majority of homophily and resulting gender segregation in city venues is highly localized.

Table 2: Venue Sorting Within Neighborhoods

	Proportion of city-wide segregation attributable to homophily within:		
	Tracts	Block Groups	Blocks
Atlanta	0.89	0.83	0.59
Chicago	0.82	0.74	0.47
Dallas	0.79	0.71	0.48
Los Angeles	0.83	0.74	0.50
New York City	0.92	0.88	0.78
Philadelphia	0.85	0.78	0.50
San Francisco	0.83	0.78	0.57
Washington, DC	0.88	0.84	0.61

Note: Bootstrapped standard errors for all entries in all cities are less than 0.005 and are omitted for clarity.

To better interpret the measures in Table 2, we can benchmark the observed gender compositions of venues against the gender compositions of venues that would be hypothetically observed if there

¹⁴We calculated bootstrapped standard errors with 500 repetitions for the means of T_t , T_g and T_b for each city separately. All means are statistically significantly different from zero at the 99% level.

was no homophily within neighborhoods.¹⁵ This exercise reveals how much additional segregation we can measure because we can observe sorting across venues within neighborhoods as opposed to only sorting across neighborhoods (as in the vast majority of studies). By observing sorting at the more disaggregated venue level, we are able to detect 2-4 times more homophily than in data aggregated to the block level, and 4-12 times more homophily than in data aggregated to the tract level.¹⁶ For Manhattan, these numbers are on the higher end: we are able to detect 4 (12) times more homophily than we would have with data aggregated to the block (tract) level.¹⁷

Finally, because we find gender homophily in all of the choices that we are able to observe – men and women systematically visit different tracts within a city, different block groups within a tract, different blocks within a block group, and different venues within a block – it is likely that we are underestimating the extent to which homophily actually mitigates exposure to diversity in peer groups. For example, individuals may sort to the same restaurant at different times of the day, to different tables in the same restaurant, or even to different conversations at the same table.

Neighborhood Residents vs. Visitors

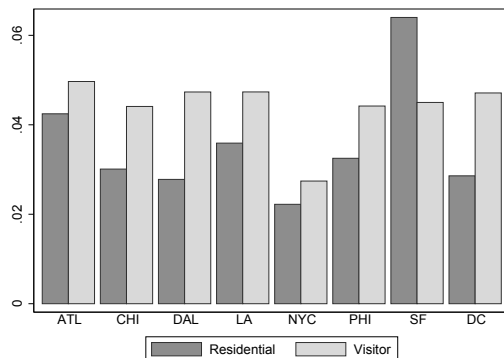
Typically researchers can observe only the broad location choices that individuals make such as the neighborhoods where they reside. Because we also observe the choices of which neighborhoods people visit, we can compare the relative amounts of homophily among residents and visitors.

¹⁵We also conduct a falsification exercise in which individuals are not allowed to sort within blocks to validate this benchmarking exercise and ensure that our results are not simply artifacts of sampling error. The details and results of this exercise are provided in the appendix.

¹⁶To obtain these figures, we take the reciprocal of the proportion of observed venue sorting due to homophily within neighborhoods (e.g., $(1 - 0.89)^{-1} = 9.09$ for Tracts in Atlanta).

¹⁷The amount of gender segregation in venues that we find is comparable to the amount of residential segregation along other demographic dimensions found in Fischer et al. (2004). After appropriately rescaling all measures for comparison, we find that in a representative tract with a Theil of 0.05, the extent to which women are segregated in venues is the following percentage of the extent to which these various demographic groups are segregated residentially on average: 65% for Blacks, 93% for Whites, 125% for Hispanics, 224% for foreign born individuals, 154% for top quintile earners, 181% for bottom quintile earners, 107% for homeowners, 330% for married households, 801% for households with children under 15, 362% of households headed by somebody aged 18-29, and 374% for households headed by somebody older than 64.

Figure 3: Residential Homophily vs. Visitor Homophily



Note: Residential homophily is calculated as the Theil index of the gender composition of block residents from the 2010 Census. For comparability, visitor homophily is calculated as the Theil index of the gender composition of check-ins in blocks. Bootstrapped standard errors for all estimates are below 0.005 and are omitted for clarity.

In Figure 3, we compare how residents sort across blocks with how visitors sort across blocks for each city in our sample. Residential homophily is calculated as the Theil index of the gender composition of block residents for each city from the 2010 Census. Visitor homophily is calculated as the Theil index of the gender composition of block visitors for each city from our data, which is equivalent to the second term in a block level decomposition of T_c according to equation (2). We find that for all cities except one, there is less residential homophily than visitor homophily.¹⁸

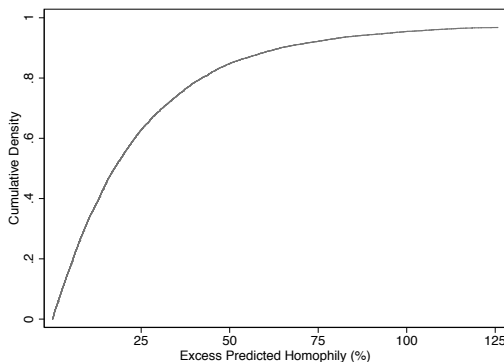
Our findings suggest that by understating homophily, studies that rely on residential data alone will substantially overstate individuals' exposure to diversity. For example, if we were to observe residential data only, then we would estimate that the average woman in neighborhood k would randomly encounter another woman with probability equal to the female share of residents in that neighborhood. However, if we were to observe data disaggregated to the venue level, then we could better estimate that the average woman in k would randomly encounter another woman with probability $\sum_{j \in k} \frac{f_{jk}}{f_k} \cdot s_{jk}$ where f_{jk} is the number of women visiting venue j , and f_k is the total number of women visiting neighborhood k .

To provide some empirical context for this thought experiment, we calculate how much more likely we would estimate that a woman would encounter a woman in venue level data than in residential data, and present the empirical cumulative distribution of this difference (which corresponds

¹⁸The exception is San Francisco. We speculate that this is due to San Francisco's sizable gay population, which concentrates residentially in certain neighborhoods whose venues are visited by a very gender diverse population. Indeed, San Francisco, like all other cities in the sample, exhibits less residential homophily than visitor homophily by age (see appendix).

to the extent to which residential data fails to capture homophily that is observable in venue data) in Figure 4. To be conservative, we define neighborhoods as census blocks, which are the smallest residential geographic units that are used by researchers. In roughly 40 percent of the blocks, we would underestimate exposure to peers of the same gender by at least 25 percent (or about one standard deviation of the probability that a woman would encounter another woman in a venue) if we used residential data instead of venue data. In roughly 20 percent of the blocks, we would underestimate exposure to peers of the same gender by at least 50 percent.

Figure 4: Excess Predicted Homophily in Venue Data



Note: In this figure, we present the empirical cumulative distribution of how much more likely we would predict that a woman would encounter another woman in a census block using venue level data than using residential data.

4 Local Determinants of Gender Diversity

4.1 A Simple Model of Venue Choice

The substantial gender homophily in venue choices creates a wedge between the levels of gender diversity that are observed in venues and in neighborhoods. In order to explore how variety in the supply of venues might be a determinant of this wedge, we consider a simple, stylized model of how individuals choose between venues within and across different neighborhoods in the spirit of Hotelling (1929).

On the supply side, we model a neighborhood k as a collection of venues indexed by j , each of which possess a single particular characteristic x_j that lies on the unit interval and differentiates the venues. This characteristic can be thought of as the venue's particular type of offering. The spatial

distribution of venues on the unit interval corresponds to the variety of venues in the neighborhood. For example, a Mexican restaurant and a Chinese restaurant would lie closer to each other on the interval than a Mexican restaurant and a shoe store. More generally, when venues are more spread out, they collectively represent a greater variety of venue offerings, which we denote as V_k . To simplify our analysis as much as possible, we consider the simplest setting that could feature sorting across venues: a single neighborhood with two fixed venues (i.e., $j \in \{1, 2\}$). In such a neighborhood, $V_k = |x_1 - x_2|$.¹⁹

On the demand side, we assume that there is a mass of individuals, each of whom are indexed by i and possess a utility function over venues $U_i(x; \delta_i)$. U_i is assumed to be a single peaked function around the point δ_i , which represents individual i 's ideal point (e.g., $U_i(x) = u - (\delta_i - x)^2$). Once again, we consider the simplest specification of demand that could feature sorting across venues: individuals belong to one of two equally sized groups of potential venue visitors: males and females. The δ_i are drawn from different distributions depending on i 's gender. Each individual is assumed to choose at most one venue that maximizes their utility provided $U_i > 0$.²⁰ If more than one venue offers an individual maximal positive utility, then ties are broken randomly.

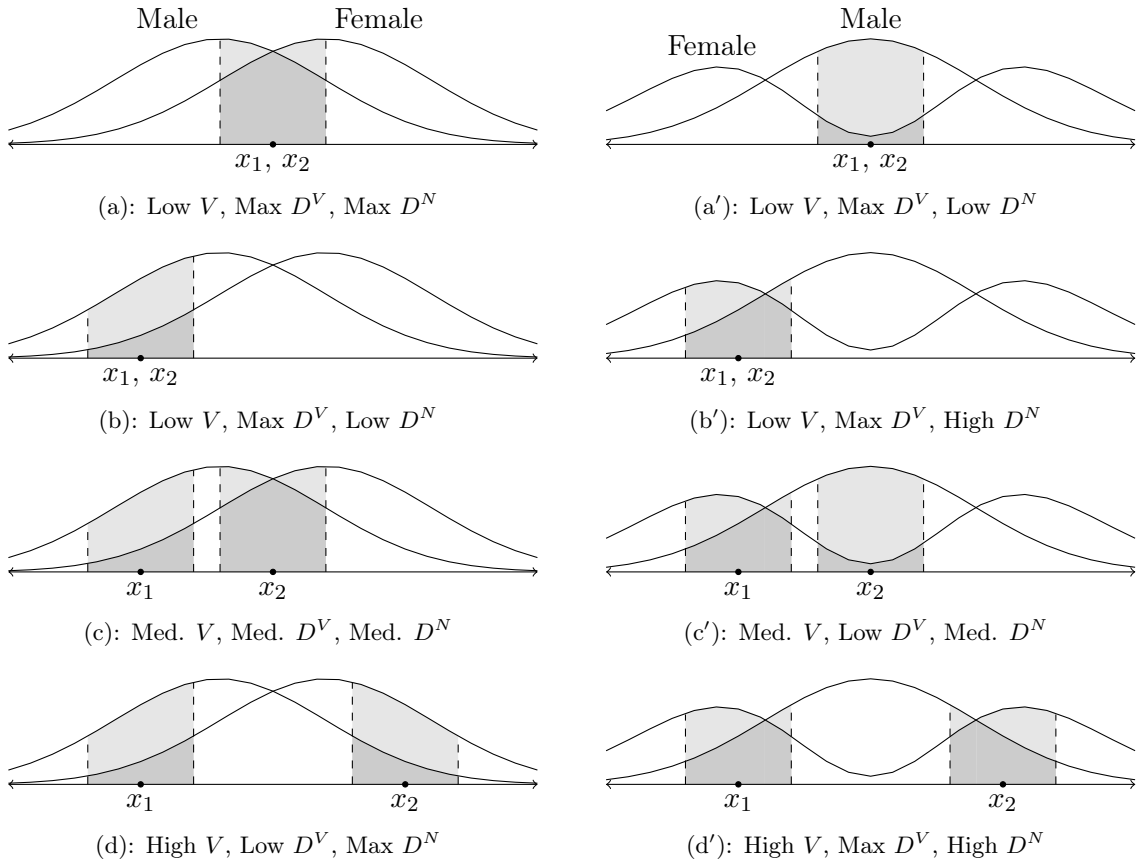
We combine the supply and demand sides to define equilibrium venue diversity and neighborhood diversity. Venue diversity in a neighborhood is measured by the negative Theil index of the gender composition of venues, i.e., $D_k^V = -T_k$, since higher levels of T_k correspond to less diversity. The overall amount of diversity in neighborhood k can be understood as how representative the gender composition of *actual* neighborhood visitors is relative to *potential* neighborhood visitors. Because the groups are of equal size, the latter is equal to $\frac{1}{2}$, so we can define neighborhood diversity as $D_k^N = -\left|\frac{f_1 + f_2}{f_1 + f_2 + m_1 + m_2} - \frac{1}{2}\right|$ where f_j and m_j are the numbers of female and male visitors to venue j respectively.

We use this simplified model to illustrate the relationship between neighborhood venue offerings and diversity in a series of diagrams. In Figure 5, we consider four different neighborhoods in order of increasing venue variety (a) - (d), which have counterparts (a') - (d') that are identical except for

¹⁹In general, x_j could also refer in part to the physical locations of venues. In such a formulation, connected subsets of the unit interval could correspond to physical neighborhoods, and we could study sorting across neighborhoods as well. For simplicity, we abstract away from this formulation because our empirical analysis exploits only very local variation in venue variety.

²⁰This condition accommodates an outside option within the model. Individuals for whom $U_i \leq 0$ for all available x_j should be understood to choose the outside option, which reflects visiting another neighborhood or staying at home.

Figure 5: Venue Variety and Diversity



the demands that they face. In each of the neighborhoods in the first column, men’s ideal points tend to be lower than women’s ideal points. In neighborhood (a), there is no venue variety, as $x_1 = x_2$. As a result, there is no sorting, so the neighborhood exhibits maximal venue diversity (D^V). Also, since the venues attract equal numbers of men and women, there is maximal neighborhood diversity (D^N). In neighborhood (b), $x_1 = x_2$ as before, so there is still no venue variety or sorting, and hence maximal D^V . However, D^N is low since venue visitors are unrepresentative of the population at large. In neighborhood (c), $x_1 \neq x_2$, so this neighborhood has a moderate level of venue variety, which is accompanied by a moderate amount of sorting (and hence moderate levels of D^V). As a result, this neighborhood has a moderate overall level of D^N relative to neighborhoods (a) and (b). Finally, neighborhood (d) features a high level of venue variety, which is accompanied by a high level of sorting and hence low D^V . However, because the two venues cater to symmetric groups of consumers, an equal number of men and women go to one of the venues, and hence the neighborhood has maximal D^N .

The four analogous neighborhoods in the second column, (a') - (d') face different demands. In these hypothetical neighborhoods, women can be classified into two groups with fairly disparate taste for activities (say, hanging out in cafes and shopping) whereas men tend to be more homogenous in their tastes for activities (say, dining at restaurants). Mathematically, while men’s and women’s average ideal points are now both located at $\frac{1}{2}$, women prefer venues with low and high x_j ’s whereas men tend to prefer venues with moderate x_j ’s. The resulting levels of neighborhood and venue diversity as venue variety increases in neighborhoods (a') - (d') are quite different from the levels of diversity in their counterparts (a) - (d). For instance, an increase in venue variety from (b') to (c') *reduces* both D^V and D^N , whereas a further increase in venue variety from (c') to (d') *increases* both D^V and D^N , yielding the paradox of diversity.²¹

We draw three conclusions from this stylized analysis. First, sorting is made possible only by venue variety; it is trivial to note that there will be no sorting across venues with identical x_j ’s (and minimal sorting across venues with very similar x_j ’s). Accordingly, venue variety is an

²¹The paradox of diversity sets up an interesting tradeoff between serving the narrower needs of consumers and enriching society more broadly by increasing their exposure to diversity. Waldfogel (2009) introduces the concept of the “tyranny of the market” in which small-scale markets can fail to serve individuals with niche preferences. While sufficiently “thick” markets do not suffer from the tyranny of the market, our analysis suggests that they will instead suffer from a lack of exposure to diversity at venues. On the other hand, “thin” markets that fall prey to the tyranny of the market are less vulnerable to a lack of exposure to diversity at venues. We thank an anonymous referee for suggesting this connection.

attractive candidate for a determinant of the wedge between venue diversity and neighborhood diversity that we have established in Section 3. Second, the relationship between venue variety and venue diversity is theoretically ambiguous. In the neighborhoods $a - d$, venue variety and venue diversity are inversely related to each other, but in neighborhoods $a' - d'$, this relationship no longer holds. Third, the relationship between venue variety and overall neighborhood diversity is also theoretically ambiguous. In neighborhoods $a - d$, venue variety and neighborhood diversity are directly related to each other, but in neighborhoods $a' - d'$, this relationship no longer holds. The latter two implications suggest that we must empirically determine the relationships between venue variety and venue and neighborhood diversity in order to determine the extent to which venue variety creates this wedge.

4.2 A Proxy for Venue Variety

In order to generalize the model and take it to the data, we need a measure of venue variety. Intuitively, venue variety should be lower in neighborhoods with more substitutable venues whose characteristics are more similar. One important characteristic of a venue is its location. All else constant, venues located farther from each other should be less substitutable. In addition, the offerings of a venue can be proxied for by its categorization in our data.

Because the subcategories of venues are so narrowly defined, we can interpret them as proxies for x_j provided that we compare venues only in narrow geographic areas (i.e., the same location). Thus, we can recast the first conclusion drawn above in terms of something that is measurable with our data: The proportion of sorting within a neighborhood that is due to sorting across subcategories should be high if neighborhoods are narrowly defined. In Table 3 we present the proportion of sorting within neighborhood that occurs across venue types for each neighborhood definition. Our findings are consistent with the model. The bulk of sorting within neighborhoods occurs across subcategories; for instance, about 90% of sorting within census blocks occurs across subcategories.²² However, much less sorting within entire cities occurs across subcategories. This suggests that location is a better proxy for x_j when comparing venues that far from each other.

²²Similarly, between 50% and 60% of the sorting within blocks occurs across categories depending on the city.

Table 3: Proportion of Within-Neighborhood Sorting by Gender Due to Sorting Across Subcategories:

	City	Tracts	Block Groups	Blocks
Atlanta	0.26	0.78	0.83	0.91
Chicago	0.26	0.84	0.89	0.94
Dallas	0.27	0.82	0.86	0.92
Los Angeles	0.20	0.83	0.87	0.92
New York City	0.31	0.70	0.81	0.90
Philadelphia	0.22	0.81	0.85	0.94
San Francisco	0.28	0.76	0.82	0.92
Washington, DC	0.26	0.75	0.82	0.91

Note: Subcategories (225) are defined in the appendix. Bootstrapped standard errors for all entries are less than 0.005 and are omitted for clarity.

Remark 1. There are two potential explanations for our findings of gender segregation in Section 3: (a) men (women) prefer to be in the company of other men (women) in venues (active segregation); and (b) men and women systematically prefer different types of venues (passive segregation). For simplicity, the model we consider here allows for only the second explanation, as the first explanation does not seem to be empirically important in our context. Indeed, the first explanation should generate a “social contagion” effect which would result in dynamic trends (and possibly tipping behavior) in the gender compositions of otherwise similar venues (Schelling (1971)); certain venues of a particular type would tend to become increasingly male while others of that same type would tend to become increasingly female. The results above are inconsistent with this explanation: nearby venues of the same type have very similar gender compositions, whereas nearby venues of different types have very different gender compositions. (Moreover, we do not find systematic trends in the gender composition of individual venues as shown in panel (c) of Figure 1.) This suggests that the reason we observe most individuals going to venues filled with others of the same gender is not because they actively seek similar people; instead, men and women just tend to prefer different activities.

4.3 Identifying The Causal Effects of Venue Variety on Venue and Neighborhood Diversity

The stylized model described above reaches ambiguous conclusions about the effects of venue variety on venue and neighborhood diversity, so we identify these causal effects empirically. Consider two small, nearby neighborhoods that are otherwise similar except for their levels of venue variety. For instance, one neighborhood may feature only restaurants, whereas another neighborhood may feature both restaurants and shops (compare to neighborhoods (a) and (c) in Figure 5). Given their small sizes and proximity, it is reasonable to consider their locations and the demands that they face to be approximately the same, except for their venue offerings. Thus, differences in venue and neighborhood diversity across these neighborhoods can be reasonably attributed to the difference in their venue variety. We implement an identification strategy that makes this comparison.

Following the model, the amount of local diversity in venues in block b can be measured by the negative Theil Index, $D_b^V = -T_b$, and the overall amount of neighborhood diversity in b can be measured by how representative the distribution of visitors in b are of the distribution of visitors in the whole city. As a generalization of the model, if f_{jb} and m_{jb} represent the total number of females and males in venue j in block b , and $s_b = \frac{\sum_{j \in b} f_{jb}}{\sum_{j \in b} (f_{jb} + m_{jb})}$, then we can define

$$D_b^N = -|s_b - s_c| \quad (3)$$

to be the overall amount of diversity in block b in city c . Finally, because we compare only small neighborhoods that are close to each other, we can take advantage of the classification of venues in our data to generalize the model above and define venue variety, V_b , as either the number of unique categories or subcategories of venues that are on offer in that block.

We estimate the regression equations:

$$D_b^V = \beta^V V_b + \alpha_g^V + X_b \delta^V + R_b \lambda^V + \epsilon_b^V \quad (4)$$

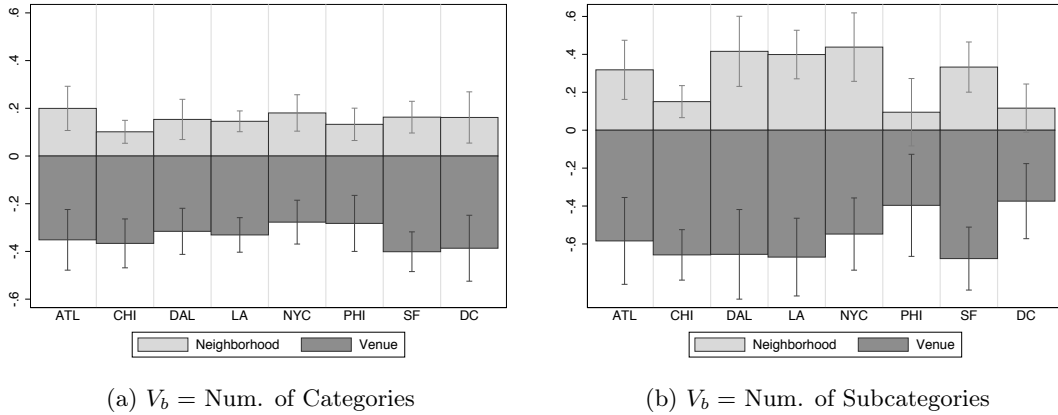
$$D_b^N = \beta^N V_b + \alpha_g^N + X_b \delta^N + R_b \lambda^N + \epsilon_b^N \quad (5)$$

where α_g are fixed effects at the block group level for $b \in g$, and X_b represents a set of block control

variables that includes the total number of venues and the amount of checkin activity in b , R_b represents a set of residential control variables that includes the total number and the female share of residents in b , and ϵ_b^V represents an error term.²³ β^V and β^N are the coefficients of interest. For interpretation, we normalize all variables by their standard deviations, so β^V and β^N correspond to the effects of a one standard deviation increase in venue variety on venue and neighborhood diversity respectively (in units of their standard deviations).

In Figure 6, we present estimates of β^V (darker bars) and β^N (lighter bars) along with their corresponding 95% confidence interval for each city separately, and for V_b defined as either the number of unique categories or subcategories. We systematically find that $\hat{\beta}^V < 0$ and $\hat{\beta}^N > 0$. This implies that any increase in neighborhood diversity due to a an increase in venue variety will generate more intense sorting between venues within the neighborhood, thereby reducing the exposure to diversity at the venue level. Indeed, a one standard deviation increase in venue variety will lead to roughly a 0.2 standard deviation increase in neighborhood diversity and roughly a 0.4 standard deviation decrease in venue diversity.²⁴

Figure 6: $\hat{\beta}^V$ and $\hat{\beta}^N$ By City



Notes: The dark bars represent estimates of $\hat{\beta}^V$ from equation (4), and the light bars represent estimates of $\hat{\beta}^N$ from equation (5). 95% confidence intervals are also shown from robust standard errors clustered at the block group level. The number of observations for each of the 16 regressions is equal to the number of census blocks in each city (see Table 1), and the R^2 of each regression varies from 0.33 to 0.50.

²³The residential control variables are obtained from the 2010 Census Summary File 1 (SF1).

²⁴Our estimates of β^N corroborate the idea advanced by Glaeser et al. (2001) and Couture (2014) that the variety of venues and activities on offer is a primary amenity to urban consumers.

Can We Interpret $\hat{\beta}^V$ and $\hat{\beta}^N$ as Causal?

The causal parameters β^V and β^N are identified under the assumptions $\text{Cov}(\epsilon_b^V, V_b | \alpha_g^V, X_b, R_b) = 0$ and $\text{Cov}(\epsilon_b^N, V_b | \alpha_g^N, X_b, R_b) = 0$ respectively. Because we conduct our analysis at the block level, we explicitly consider small neighborhoods, and the inclusion of block group fixed effects α_g ensures that we only compare neighborhoods that are located near each other, which holds constant all determinants of the demand and supply that vary at the block group level. Still, certain neighborhood amenities that are correlated to venue variety might attract different groups of people to different nearby blocks, so we control for X_b to ensure that we compare blocks that have similar numbers of venues and levels of foot traffic, and we control for R_b to ensure the number of residents of each gender is similar across these blocks.

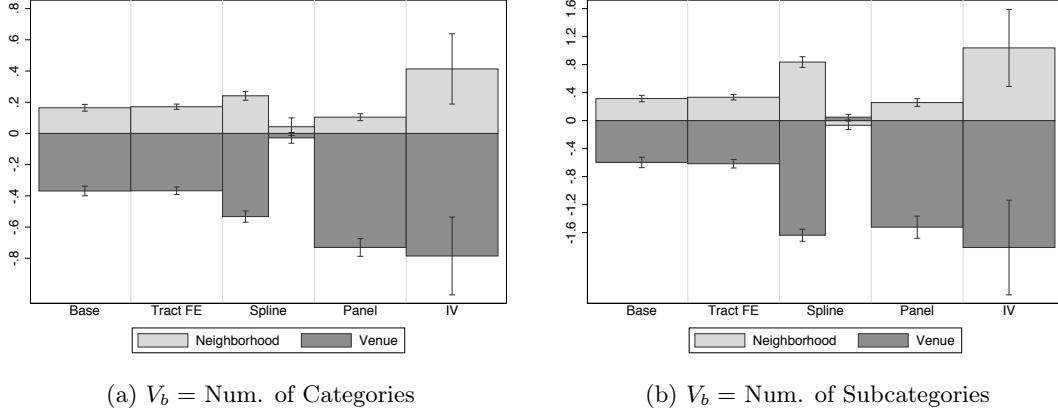
The remaining concern is that some unobserved neighborhood amenities that cannot be controlled for by these covariates may be correlated to venue variety. For instance, one might worry about simultaneity bias: different venues may decide to locate in neighborhoods that attract more diverse visitors, i.e. demand for venues causes supply of venues, rather than the other way around. The fact that neighborhoods are both small and close to each other in our context helps allay such concerns, as this could only be an issue if venues had control over and preferences for locating in specific blocks of a given block group. This seems implausible since locating in a particular block requires a commercial vacancy and the blocks are similar in venue density, foot traffic, and location.²⁵

Nevertheless, we provide four robustness checks that address these and other concerns. The results of these four robustness checks are shown in Figure 7, where we compare the baseline estimates of β^V and β^N from equations (4) and (5) pooled over all eight cities with estimates from four alternative specifications.²⁶ In the first set of bars, we define venue variety as the number of distinct categories in a neighborhood, and in the second set of bars, we define venue variety as the number of distinct subcategories in a neighborhood.

²⁵The motivation for this identifying assumption is analogous to the one made by Bayer et al. (2008) for residents. If the housing market is not too dense at all points in time (as appears to be case even in large metropolitan areas), then it is difficult for a venue owner to choose an exact census block in which to locate.

²⁶We also conducted these robustness checks for each city separately and obtained similar results, which are reported in the appendix.

Figure 7: $\hat{\beta}^V$ and $\hat{\beta}^N$: Alternative Identification Strategies



Notes: The dark shaded bars represent $\hat{\beta}^V$, and the light shaded bars represent $\hat{\beta}^N$. 95% confidence intervals are also shown from robust standard errors clustered at the block group level. The first bars correspond to baseline estimates from equations (4) and (5). The second bars replace the block group fixed effects in the baseline estimates with tract fixed effects. The third set of bars correspond to estimates of the parameters specified as a linear b-spline with a knot at 3 subcategories. The fourth bars correspond to estimates from equations (6) and (7) where the dataset is disaggregated to a monthly panel, and the block group fixed effects are replaced with block fixed effects. The fifth bars correspond to 2SLS estimates of the baseline regressions with zoning instruments.

In our first robustness check, we re-estimate equations (4) and (5) with tract fixed effects instead of block group fixed effects. Tracts typically encompass two or more block groups, so these fixed effects no longer control for unobserved amenities varying across block groups within tracts, which may confound our estimates. The results (denoted as “Tract FE”) are virtually unchanged, which suggests that after controlling for X_b and R_b , amenities and local demand varying across block groups within tract are uncorrelated to V_b . It is difficult to conceive of unobservables that are correlated to V_b , that vary across blocks within block groups but do not vary across block groups within tracts.²⁷

Second, we re-estimate equations (4) and (5) using linear B-splines in V_b , which allows us to estimate separate marginal effects of venue variety on diversity for neighborhoods with three or fewer subcategories and for neighborhoods with four or more subcategories. If $\hat{\beta}^V$ and $\hat{\beta}^N$ are causal estimates, then they will likely decline in magnitude as we compare nearby blocks with higher levels of venue variety.²⁸ In contrast, if these estimates reflect confounding factors that are

²⁷For instance, simultaneity could only be a concern if venues had more control or preference over their choice of which block within a block group to locate relative to their choice of which block group within a tract to locate.

²⁸Extending the intuition of the model presented above, in a block with greater number of venues with distinct

present irrespective of the level of V_b , then we should find that these effects do not decline for higher V_b . Indeed, we find that nearly all of these effects (denoted as “Spline”) operate at low levels of venue variety in all cities in our sample.²⁹

Third, we exploit the longitudinal variation in our data to estimate β^V and β^N using an alternative identification strategy. We re-specify equations (4) and (5) as

$$D_{bt}^V = \beta^V V_{bt} + \alpha_b^V + \alpha_{ct}^V + X_{bt}\delta^V + \epsilon_{bt}^V \quad (6)$$

$$D_{bt}^N = \beta^N V_{bt} + \alpha_b^N + \alpha_{ct}^N + X_{bt}\delta^N + \epsilon_{bt}^N, \quad (7)$$

respectively. The key difference is that all of our main explanatory variables and controls now vary by month. By doing so, we can identify β^V and β^N using only within-block variation in venue variety that arises due to the entry and exit of venues over time. We implement this identification strategy by including block fixed effects (α_b^V and α_b^N) that additionally control for all unobserved determinants of diversity that vary across blocks within block groups that were not controlled for in equations (4) and (5). The fixed effects α_{ct}^V and α_{ct}^N control for city level amenities that may vary by month in order to absorb any seasonality that varies across cities. Our results (denoted as “Panel”) suggest that our baseline estimates of β^V are conservative, which is consistent with our sensitivity analysis in the appendix.

Finally, we re-estimate β^V and β^N in equations (4) and (5) with a third, distinct identification strategy that uses variation in zoning laws across blocks within block groups as instrumental variables (IVs) for venue variety. We only use identifying variation in the variety of venues that stems from regulations that restrict the location of certain venues in certain blocks. This IV approach deals with any remaining simultaneity concerns and any remaining confounders that are uncorrelated to zoning laws such as most kinds of measurement error. Specifically, we use the share of lots in the block that are zoned to residential, commercial and mixed uses as instruments; hence, we effectively compare diversity in nearby blocks that are zoned differently and thus have different

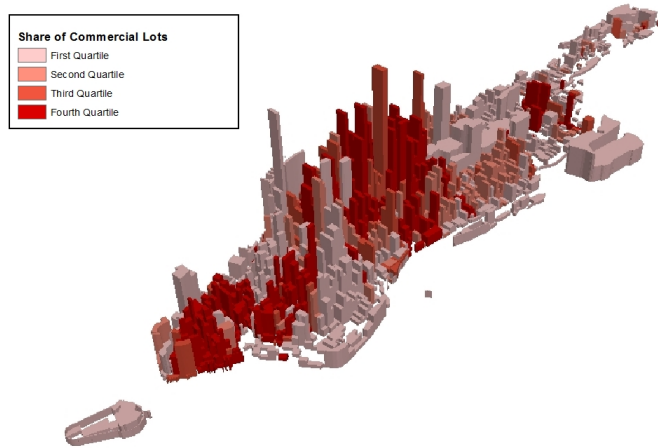
x_j ’s, more of the support will be covered by venue visitors. As a result, a marginal increase in venue variety will have a smaller effect on both D_b^V and D_b^N since the additional venue will draw increasingly from individuals who were otherwise planning to go to another venue on the same block.

²⁹These results are virtually unchanged when we place the knot at 2, ..., 5 subcategories.

levels of venue variety (but a similar number of venues, overall traffic and number of residents of each gender).³⁰

Differences in zoning laws are found to generate differences in the variety of venues in nearby census blocks. In Figure 8, we spatially illustrate the “first-stage” relationship between commercial zoning (here categorized in quartiles for visual clarity) and venue variety (number of unique venue subcategories) in Manhattan census blocks, which is clearly positive. More formally, a joint F-test of the significance of the three instruments for the number of unique venue categories and unique venue subcategories yields $F_{3,5779} = 25.00$ (0.00) and $F_{3,5779} = 12.27$ (0.00), respectively, where the p-values shown in parentheses are much smaller than 0.01. Our estimates (denoted as “IV”) are, if anything, larger in magnitude than all OLS estimates, which suggests that the OLS estimates may be attenuated by measurement error. As a result, our findings that $\beta^V < 0$ and $\beta^N > 0$ should be understood to be conservative. This interpretation is consistent with our Monte Carlo study of measurement error described in the appendix.³¹

Figure 8: Commercial Zoning and Venue Variety (First Stage)



Notes: Each bar represents a census block in Manhattan. The height of each bar corresponds to V_b , the number of unique venue subcategories in b . Darker bars represent blocks with a greater proportion of commercially zoned lots.

³⁰We obtained lot level data on zoning for each city from their respective planning offices. Lots can be zoned for other uses than the three that we use for IVs (e.g., manufacturing or parks), but our results were unchanged when using additional IVs.

³¹In order to ensure that $\hat{\beta}^V$ was not contaminated by the effect β^N and vice versa, we also implemented a robustness check where we added D_b^N as a control variable in the equation of D_b^V (equation (4)), and D_b^V as a control variable in the equation of D_b^N (equation (5)). Our results were unchanged.

5 Gender Homophily and the Labor Force Participation Gap

Social networks and the interactions that they facilitate have the potential to shape labor markets profoundly (Montgomery (1991)). One of the most studied features of the labor market is the persistent gender gap in labor force participation that is observed in many countries over most time periods. A considerable literature has analyzed the impacts of job referral networks on individual employment and wages (e.g., Ioannides and Loury (2004); Bayer et al. (2008); Schmutte (2015)) and found that social interactions play an important role in explaining labor market outcomes. In particular, Bayer et al. (2008) (hereafter, BRT) find that these networks can operate at a highly localized level: residents of the same block form stronger network ties than residents of nearby blocks, particularly when individuals are similar on observable characteristics (ostensibly due to homophily). We complement their analysis with suggestive evidence that their findings may also extend to interactions between block residents and block visitors as mediated by the gender segregation in venues studied in this paper.

Our identification strategy is similar to the one used in BRT. Specifically, we compare the gender gap in labor force participation among residents of otherwise identical, nearby census blocks that differ only in the diversity of their visitors at both the block and venue levels. While block diversity might proxy for the exposure to diversity outside of venues in the block, venue diversity proxies for the exposure to diversity inside venues. Hence, differences in this gender gap can be understood to be mediated through social interactions between block residents and venue visitors either inside or outside of venues. We obtain block-level employment statistics from the Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics, or LODES. For each block in our sample, we observe the numbers of male and female residents who are employed³² in each year from 2012 to 2013.³³ These statistics are further disaggregated into low wage (less than \$1,250 per month), medium wage (between \$1,250 and \$3,333 per month) and high wage (at least \$3,333

³²The LODES data cover approximately 95 percent of wage and salary jobs, excluding a small number of employees in the military, security-related federal agencies, postal workers, employees at non-profit and religious institutions, informal workers and the self-employed (Graham et al. (2014)). Because we do not want to identify referral effects for part-time employment, which may operate along different social networks, we focus our analysis on the subsample of primary jobs; however, all results are qualitatively similar in the sample of all jobs, and in a subsample of only private-sector jobs.

³³We use two years of data from LODES even though our measures of diversity do not vary over time because these measures are constructed from Foursquare check-in data that spans parts of both 2012 and 2013.

per month) groups. For our measure of the gender gap in labor force participation, we construct $GAP_{wbt} = \frac{M_{wbt}}{MR_b} - \frac{F_{wbt}}{FR_b}$ where M_{wbt} and F_{wbt} are the numbers of male and female employees in wage group w in block b in year t from LODES, and MR_b and FR_b are the adult male and female populations of block b from the 2010 Census. We normalize GAP_{wbt} to have mean 0 and variance 1 across all blocks in the sample.³⁴

We estimate the effects of venue and neighborhood diversity on the gender gap in the following regression equation:

$$GAP_{wbt} = \gamma_w^V D_b^V + \gamma_w^N D_b^N + \alpha_{wg} + X_b \delta_w + \epsilon_{wbt} \quad (8)$$

where D_b^V and D_b^N are defined as before, α_{wg} is a block group-wage group fixed effect, X_b is a vector of controls and ϵ_{wbt} is an error term. The parameters of interest, γ_w^V and γ_w^N , represent the effects of venue and neighborhood diversity, respectively, on the female labor force participation gap among workers in wage group w . The OLS estimates of these parameters can be interpreted as causal effects only if $\text{Cov}(\epsilon_{wbt}, D_b^V | D_b^N, \alpha_{wg}, X_b) = 0$ and $\text{Cov}(\epsilon_{wbt}, D_b^N | D_b^V, \alpha_{wg}, X_b) = 0$.

Our identifying assumption is that the residential sorting within block groups does not vary across people by unobservable determinants of employment differently depending on the gender. This assumption is the same one made by BRT, who provide strong empirical evidence of no residential sorting within block groups along any observable demographic dimension (including gender).³⁵ Because housing markets are very thin at small geographic scales, individuals' ability to choose to live in specific census blocks is severely restricted. Moreover, even if they were able to make such specific choices, individuals might find it difficult to observe highly local (block level) amenities at the time of their residential decision. Under this identifying assumption, the parameters γ_w^V and γ_w^N can be interpreted as the causal effects of diversity among visitors to block b on the labor force participation gap among residents of block b .

³⁴Ideally, we would calculate the gender gap as $\widetilde{GAP}_{wbt} = \frac{M_{wbt}}{MR_{wbt}} - \frac{F_{wbt}}{FR_{wbt}}$, where MR_{wbt} and FR_{wbt} are the numbers of male and female residents in block b in year t who are a match for jobs of group w . Unfortunately, we cannot observe MR_{wbt} and FR_{wbt} . We present several robustness checks in the appendix to allay concerns that this measurement error (i.e., $\widetilde{GAP}_{wbt} \neq GAP_{wbt}$) biases our results.

³⁵In practice, our assumption is weaker than the one made in BRT because the outcome variable of our analysis is a *gap* in the labor force participation rather than a *level* of the labor force participation rate. Hence, our estimates are robust to residential sorting within block groups based on the propensity of employment provided that such localized sorting behavior is uncorrelated to gender.

We present results for our preferred specification of equation (8) in Table 4, which includes block group fixed effects, controls from Foursquare and Census data (including the number of venues, total number of check-ins, as well as age and gender demographics of block residents), all flexibly specified using cubic B-splines. In Appendix A.4, we present results for 20 distinct specifications of controls (X_b), neighborhood fixed effects (α) and estimation subsamples to establish the robustness of our identification strategy to many different endogeneity concerns. We obtain two robust findings. First, we find suggestive evidence that greater gender diversity *inside venues* reduces the employment gap for low wage jobs; a one standard deviation increase in venue diversity shrinks this gap by roughly 1.5 percent of a standard deviation (see below for interpretation). However, we do not find any evidence that venue diversity affects the employment gap for medium and high wage jobs. This is consistent with other findings that job referral effects are stronger for individuals who are less attached to the labor force (see Ioannides and Loury (2004) for a survey). Second, we do not find any statistically significant effects of gender diversity outside of venues on employment gaps for any type of job. Moreover, the point estimates for γ_w^N tend to be much smaller than the point estimates for γ_w^V . This suggests that social interactions inside venues are more important for reducing gender gaps in labor force participation than the social interactions that might take place on streets and sidewalks, at least among the lowest paid workers.

Under our identifying assumption, our results can be interpreted as causal. However, even if our assumption is valid, it is difficult to identify the exact mechanism underlying this causal effect. Following BRT (p. 1190), we posit that this effect might occur because residents are more likely to interact with visitors in venues of their block than with visitors in venues in nearby blocks (Unger and Wandersman (1985))³⁶, and because female informal contacts may have a lower impact on employment outcomes than male informal contacts (Loury (2006)). Following this logic, men and women who reside in diverse blocks would end up being exposed to more similar referral networks than men and women who reside in adjacent, less diverse blocks. Since personal networks are otherwise assortative along gender lines, this difference may lead to a smaller gender gap in the more diverse block. More concretely, while one-off interactions in venues are unlikely to have much impact on these networks, many venues (e.g., coffee shops, bars, restaurants and speciality shops)

³⁶As discussed in BRT, to the extent that residents of a given block also interact with visitors of nearby blocks, our estimates of γ_w^V and γ_w^N will likely be attenuated.

may develop followings of “regulars” who may interact more deeply. Moreover, interactions between visitors and the employees of venues (who are likely to be of similar gender to the majority of customers in gender segregated venues) may also contribute to these networks. We should also note that it is possible that the effects we find, if indeed causal, do not operate through referrals. Rather, these interactions may develop certain skills in residents that are valued in the labor market (e.g., how to better interact, verbally or non-verbally with people of the opposite gender).

To better interpret our results, we find that a one standard deviation increase in venue diversity decreases the gender gap in labor force participation by 0.21 percentage points. For context, BRT find that a one standard deviation increase in what they define as the average “match quality” of neighborhood residents across several demographic variables reduces the gender gap in labor force participation by 0.9 percentage points (see Table 8 of BRT with calculations described on p. 1190). Hence, the effect of our proxy for venue interactions is $\frac{0.21}{0.9} = 23\%$ of the effect of BRT’s proxy for neighborhood interactions. Our estimates and BRT estimates reflect the impact of similar social interactions that differ for mainly two reasons: (1) we identify potential effects due to interactions between residents and visitors, where these visitors may or may not be residents, while BRT identify potential effects due to interactions between residents, and (2) we focus on interactions that happen only inside venues that we observe, while BRT focuses on interactions that might occur in any venue (including own house, neighbor’s house, etc). While it is plausible that interactions between neighbors occur at higher intensities than interactions between venue visitors, a key insight of the vast “strength of weak ties” literature might explain why our results are of comparable magnitude: weaker ties such as the ones obtained by more casual interactions may have stronger impacts than previously thought on many outcomes including job referrals (Granovetter (1973), Montgomery (1991), Montgomery (1992), Levin and Cross (2004)). Intuitively, while a given “strong” tie (e.g., friend) may have more frequent and intense interactions than a comparable “weak” tie (e.g., friend of a friend), the referral network of the weaker tie may be *more distinct*, so it may offer less redundant referrals than the strong tie. As a result, weaker ties might on average yield marginal impacts on job referrals that are comparable in magnitude to those from stronger ties. It is possible that venue visitors (a plausibly weaker tie than neighbors) may also have more diffuse referral networks from each other and from block residents, which may compensate for their less frequent and less intense interactions, yielding effects of a similar order of magnitude to BRT.

More detailed data on the nature and intensity of interactions in *all* venues (including, for example, residences and workplaces) is needed to better understand the mechanisms underlying these social effects.

Table 4: Effects of Diversity on the Labor Force Participation Gender Gap

	Low Wage Jobs	Medium Wage Jobs	High Wage Jobs
	γ_{low}^V	$\gamma_{\text{med.}}^V$	γ_{high}^V
γ_w^V	-0.015** (0.007)	-0.003 (0.008)	-0.015 (0.012)
γ_w^N	-0.003 (0.007)	0.000 (0.008)	0.013 (0.012)
N		76,236	
Adj. R^2		0.325	

Notes: Low wage jobs pay less than \$1,250 monthly, medium wage pay jobs pay between \$1,250 and \$3,333 monthly, and high wage pay jobs pay more than \$3,333 monthly. We include as controls block group - wage group fixed effects and cubic B-splines (with as many knots as possible) of the numbers of venues, female and male visitors and female and male residents for each block. This amounts to 26 covariates for each group. Robust standard errors clustered at the block level are presented in parentheses. **: 5% significance level

6 Discussion

6.1 Do our findings extend to other environments?

The extensive gender segregation that we measure, which persists down to the venue level, is suggestive of further homophily *within* venues and activities. For instance, men and women may be inclined to sort to different tables within cafes and bars, or engage in different activities within gyms and parks. Hence, we believe that our findings should be understood as conservative estimates of the actual amount of gender segregation in the day-to-day activities of men and women. Moreover, the fact that gender segregation is observed across a wide variety of different recreational and commercial activities suggests that it may also pervade other social settings such as classrooms and workplaces.

Although patterns of venue visitors differ between weekdays and weekends and across cities, our general findings do not. All of our results are similar for each city-day of week combination (see appendix). Similarly, patterns of venue visitors exhibit seasonality, which is city dependent

(e.g., many fewer people visit Chicago parks in the winter months relative to the summer, but this seasonal effect is much weaker in Los Angeles). However, all of our general findings are similar for each city-month of year combination (see appendix). We view the robustness of these results as suggestive that they might also hold in other urban and suburban environments. Of course, more research is needed to understand such additional sources of heterogeneity (e.g., in which venues do peer groups form most effectively, and what are the consequences of those peer groups?).

6.2 Do similar patterns of homophily operate along other demographic dimensions?

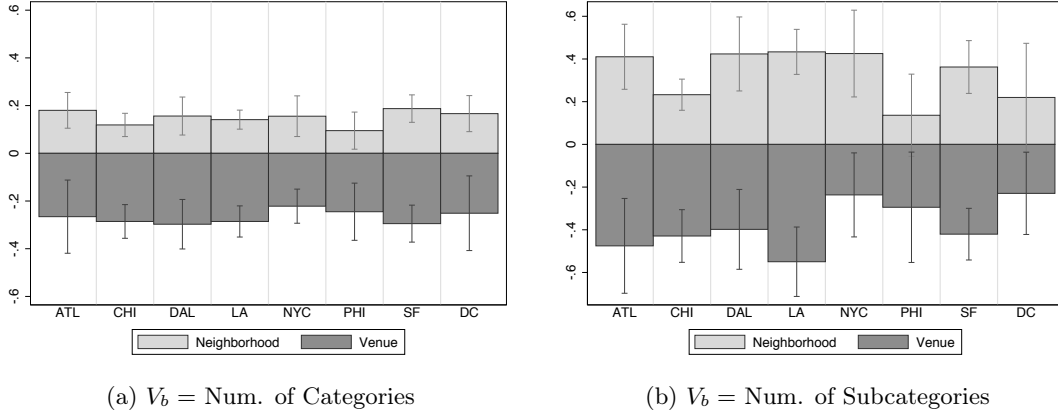
Our Foursquare data allows us to answer this question along only one additional dimension: age. For each venue in our sample, we observe the daily numbers of check-ins from users under 35 years of age and from users 35 years of age or older. With this information, we replicate our entire analysis, substituting for the proportion of females the proportion of youth. Our results are broadly similar to our results on gender, which is not a trivial finding given that gender and age are largely uncorrelated. Although we find roughly half as much age segregation as we do gender segregation, it occurs highly locally as shown in Table 5: from a third to about half of all venue sorting by age in cities occurs within census blocks. As in the case of gender, we find that age homophily is primarily mediated by the fact that people of different ages prefer different activities. Finally, we also find that the causal effects of venue variety on venue and neighborhood age diversity are both qualitatively and quantitatively similar to the respective effects on gender diversity (Figure 9). A full reporting of all results from this replication is provided in the appendix.

Table 5: Venue Sorting Within Neighborhoods By Age

	Proportion of city-wide segregation attributable to homophily within:		
	Tracts	Block Groups	Blocks
Atlanta	0.75	0.68	0.45
Chicago	0.73	0.63	0.36
Dallas	0.76	0.68	0.45
Los Angeles	0.75	0.67	0.43
New York City	0.87	0.81	0.70
Philadelphia	0.68	0.61	0.34
San Francisco	0.83	0.76	0.53
Washington, DC	0.80	0.74	0.47

Note: Bootstrapped standard errors for all Theil indices in all cities are less than 0.005 and are omitted for clarity.

Figure 9: $\hat{\beta}^V$ and $\hat{\beta}^N$ For Age By City



Notes: The dark bars represent estimates of $\hat{\beta}^V$ from equation (4), and the light bars represent estimates of $\hat{\beta}^N$ from equation (5). 95% confidence intervals are also shown from robust standard errors clustered at the block group level. The number of observations for each of the 16 regressions is equal to the number of census blocks in each city (see Table 1), and the R^2 of each regression varies from 0.23 to 0.52.

Although we cannot replicate our analysis along any other demographic dimensions, we conjecture that the robustness of our results across gender and age may be suggestive of similar patterns of homophily in day-to-day activities along other dimensions such as race and income. Indeed, there is reason to conjecture that hidden racial and income segregation may be even greater than the passive gender and age segregation that we observe in our data. For instance, active racial

segregation is likely to contribute to a greater level of racial segregation over and above that which would be implied by passive racially homophilic forces, and venue prices are likely an additional contributor to income segregation.

In a broader sense, individuals sort along multiple demographic dimensions simultaneously. For instance, younger women are plausibly more likely to frequent the same venues as other young women than older women are. This suggests that the true amount of diversity to which individuals are exposed is even more attenuated by endogenous sorting than what we are able to observe, which is an additional reason why our findings on gender and age should be understood to be conservative estimates of the true levels of segregation in peer groups.³⁷

Remark 2. Our findings of highly localized gender and age segregation complement the findings by other researchers of similarly localized segregation along other dimensions. For example, Carrell et al. (2013) document highly localized (within Air Force Academy squadron) segregation by student ability. Among entrepreneurs, Ruef et al. (2003) document segregation along a variety of “status-related dimensions” such as gender, ethnicity and professionalism. Kossinets and Watts (2009) analyze how segregation across a variety of demographic dimensions locally evolves in the university setting along different courses of study and residential choices. And Currarini et al. (2009) document substantial, highly localized (within school) segregation by ethnicity in high school friendships.

6.3 Do gender and age homophily affect other outcomes?

In Section 5 we explored the possibility of a link between gender segregation in venues and local labor force participation gaps. Other links have been suggested in this literature. A large body of research has found that exposure to female peers affects, for example, corporate governance and performance (e.g., Brown et al. (2002); Adams and Ferreira (2009)), student achievement (Hoxby (2000); Lavy and Schlosser (2011); Hill (2015)), substance abuse (Andrews et al. (2002)), the expression of political beliefs (Huckfeldt (1995)), and the level of intimacy in social networks (Verbrugge (1977)). Although peer effects with respect to age have not been widely studied, the systematically different beliefs that people of different ages may hold suggests that age homophily might play a role in the shaping of political preferences and the human capital development. Although all of

³⁷Gender and age are mostly uncorrelated to other characteristics, making them in some ways ideal candidates to provide a conservative conclusion on such inevitably incomplete analysis.

these social interactions are unlikely to occur at all of the different types of venues in our data, we believe it is plausible that repeated exposure to certain peers in venues may accumulate over time, in turn affecting peoples’ beliefs, preferences, social norms, and actions. Identifying these various effects is a difficult proposition that carries heavy data demands and lies beyond the scope of this paper.

7 Conclusion

Peer groups shape our social environment. Homophily leads similar people to associate with one another, and we find that the amount of it that is commonly observed in datasets might only represent the tip of the iceberg when it comes to the actual extent of everyday homophily in people’s lives. Using novel, user-generated data from Foursquare, a popular mobile app, we analyze how individuals sort into neighborhoods and further into venues in eight major US cities. We find that individuals sort by gender and by age across venues that are extremely close to each other and at a similar intensity in a variety of different city types, from the long established, dense, urban cores of New York City and Philadelphia to newer and more diffuse urban areas such as Los Angeles, Dallas and Atlanta. This lends some universality to the widespread, homophilic, endogenous peer group formation that we observe.

Our results echo the central themes of Jacobs (1961): individuals endogenously respond to the urban landscape around them, and it is the diversity of this landscape that gives rise to social interactions. However, they also invite a reassessment of whether mixed-use development in neighborhoods coupled with demographic density, which Jacobs and others have championed, are important ingredients for diversity to emerge. While we find that the resulting variety in the types of venues will lead to more overall diversity in neighborhoods, we also find that it will lead to *less* diversity at the venue level as similar individuals are able to more intensely segregate themselves into venues. Hence, strengthening the social interactions that form the basis for thriving communities may be a more complicated task for policymakers to achieve than previously thought.

Our analysis contributes to the ongoing debate on the ability of cities to offer exposure to a diversity of opinions that might be crucial for the formation of accurate and pro-social beliefs. If similar people tend to hold similar views, then homophily might impact the diversity of opinions to which they are exposed. On the one hand, Sunstein (2009) suggests that physical interactions

in neighborhoods and in venues might be an important source of exposure to diverse views.³⁸ On the other hand, Gentzkow and Shapiro (2011) find that news media (both online and offline) offer more exposure to diverse opinions than neighbors, co-workers and family members do. Our findings help reconcile these two positions: physical interaction may well be a crucial source of exposure to diverse opinions, but most people choose not to be exposed to such diversity, even if inadvertently. They just tend to be drawn to the same activities as other, similar people.

More broadly, the formation of peer groups is a deeply personal choice. Although it is certainly affected by where people live, study and work, people make many smaller decisions on a daily basis that can shape their social environments in profound ways. These might revolve around seemingly insignificant actions such as frequenting a specific venue, making an acquaintance, or joining a conversation, any of which may turn out to be memorable and impactful. While the informal and personal nature of these decisions makes them difficult to observe in standard data sets, the proliferation of user-generated data sets has the potential to offer researchers a window into this rich source of socialization. We view this work as an early step along that path.

References

- Adams, R. B., Ferreira, D., 2009. Women in the boardroom and their impact on governance and performance. *Journal of financial economics* 94 (2), 291–309.
- Akerlof, G. A., Kranton, R. E., 2000. Economics and identity. *The Quarterly Journal of Economics* 115 (3), 715–753.
- Allen, R., Burgess, S., Davidson, R., Windmeijer, F., 2015. More reliable inference for the dissimilarity index of segregation. *The econometrics journal* 18 (1), 40–66.
- Andrews, J. A., Tildesley, E., Hops, H., Li, F., 2002. The influence of peers on young adult substance use. *Health psychology* 21 (4), 349.
- Arribas-Bel, D., Bakens, J., 2014. "the magic's in the recipe": Urban diversity and popular amenities. mimeo.
- Atkinson, A. B., 1970. On the measurement of inequality. *Journal of economic theory* 2 (3), 244–263.
- Banzhaf, H. S., Walsh, R. P., 2013. Segregation and tiebout sorting: The link between place-based investments and neighborhood tipping. *Journal of Urban Economics* 74 (0), 83 – 98.
- Bayer, P., Ross, S. L., Topa, G., 2008. Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy* 116 (6), 1150–1196.

³⁸“The diverse people who walk the streets and use the parks are likely to hear speakers’ arguments; they might also learn about the nature and intensity of views held by their fellow citizens. (...) When you go to work or visit a park (...) it is possible that you will have a range of unexpected encounters” (p. 30).

- Bobo, L., Charles, C., Krysan, M., Simmons, A., Fredrickson, G., 2012. The real record on racial attitudes. *Social Trends in American Life: Findings from the General Social Survey since 1972*, 38.
- Bourguignon, F., 1979. Decomposable income inequality measures. *Econometrica* 47 (4), 901–920.
- Boustan, L., 2012. School desegregation and urban change: Evidence from city boundaries. *American Economic Journal: Applied Economics* 4 (1), 85–108.
- Brown, D. A., Brown, D. L., Anastasopoulos, V., 2002. Women on boards: Not just the right thing... but the "bright" thing. Conference Board of Canada.
- Bruch, E., Mare, R., 2006. Neighborhood Choice and Neighborhood Change. *American Journal of Sociology* 112 (3), 667–708.
- Caetano, G., Maheshri, V., 2017. School segregation and the identification of tipping behavior. *Journal of Public Economics* 148, 115–135.
- Card, D., Mas, A., Rothstein, J., 2008a. Tipping and the Dynamics of Segregation. *Quarterly Journal of Economics* 123 (1), 177–218.
- Carrell, S. E., Sacerdote, B. I., West, J. E., 2013. From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica* 81 (3), 855–882.
- Carrington, W. J., Troske, K. R., 1997. On measuring segregation in samples with small units. *Journal of Business & Economic Statistics* 15 (4), 402–409.
URL <http://www.jstor.org/stable/1392486>
- Couture, V., 2014. Valuing the consumption benefits of urban diversity. mimeo.
- Currarini, S., Jackson, M. O., Pin, P., 2009. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica* 77 (4), 1003–1045.
- Davis, D. R., Dingel, J. I., Monras, J., Morales, E., 2014. Spatial and social frictions in the city: Evidence from yelp. mimeo.
- Echenique, F., Fryer Jr, R., Kaufman, A., 2006. Is school segregation good or bad? *The American economic review*, 265–269.
- Ellison, G., Glaeser, E. L., 1997. Geographic concentration in us manufacturing industries: A dartboard approach. *Journal of Political Economy* 105 (5), 889–927.
- Fernandez, R. M., Castilla, E. J., Moore, P., 2000. Social capital at work: Networks and employment at a phone center. *American journal of sociology*, 1288–1356.
- Fischer, C. S., Stockmayer, G., Stiles, J., Hout, M., 2004. Distinguishing the geographic levels and social dimensions of u.s. metropolitan segregation, 1960–2000. *Demography* 41 (1), 37–59.
URL <http://dx.doi.org/10.1353/dem.2004.0002>
- Gentzkow, M., Shapiro, J. M., 2011. Ideological segregation online and offline. *The Quarterly Journal of Economics* 126 (4), 1799–1839.
- Glaeser, E. L., Kolko, J., Saiz, A., 2001. Consumer city. *Journal of economic geography* 1 (1), 27–50.

- Graham, M. R., Kutzbach, M. J., McKenzie, B., Oct. 2014. Design Comparison of LODES and ACS Commuting Data Products. Working Papers 14-38, Center for Economic Studies, U.S. Census Bureau.
- Granovetter, M. S., 1973. The strength of weak ties. *American journal of sociology* 78 (6), 1360–1380.
- Hill, A. J., 2015. The girl next door: The effect of opposite gender friends on high school achievement. *American Economic Journal: Applied Economics* 7, 147–77.
- Hotelling, H., 1929. Stability in competition. *The Economic Journal* 39 (153), 41–57.
- Hoxby, C., 2000. Peer effects in the classroom: Learning from gender and race variation. Tech. rep., National Bureau of Economic Research.
- Huckfeldt, R. R., 1995. Citizens, politics and social communication: Information and influence in an election campaign. Cambridge University Press.
- Ioannides, Y. M., Loury, L. D., 2004. Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature* 42 (4), pp. 1056–1093.
- Jacobs, J., 1961. The death and life of great American cities. Random House LLC.
- Kniffin, K. M., Sigirci, O., Wansink, B., 2016. Eating heavily: men eat more in the company of women. *Evolutionary Psychological Science* 2 (1), 38–46.
- Kossinets, G., Watts, D. J., 2009. Origins of homophily in an evolving social network¹. *American journal of sociology* 115 (2), 405–450.
- Lavy, V., Schlosser, A., 2011. Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics* 3 (2), 1–33.
- Levin, D. Z., Cross, R., 2004. The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management science* 50 (11), 1477–1490.
- Loury, L. D., 2006. Some contacts are more equal than others: Informal networks, job tenure, and wages. *Journal of Labor Economics* 24 (2), 299–318.
- McPherson, M., Smith-Lovin, L., Cook, J. M., 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415–444.
- Meyers-Levy, J., Sternthal, B., 1991. Gender differences in the use of message cues and judgments. *Journal of marketing research*, 84–96.
- Montgomery, J. D., 1991. Social networks and labor-market outcomes: Toward an economic analysis. *The American economic review*, 1408–1418.
- Montgomery, J. D., 1992. Job search and network composition: Implications of the strength-of-weak-ties hypothesis. *American Sociological Review*, 586–596.
- Ruef, M., Aldrich, H. E., Carter, N. M., 2003. The structure of founding teams: Homophily, strong ties, and isolation among us entrepreneurs. *American sociological review*, 195–222.
- Schelling, T. C., 1969. Models of Segregation. *The American Economic Review* 59 (2), 488–493.

- Schelling, T. C., 1971. Dynamic Models of Segregation. *Journal of Mathematical Sociology* 1, 143–186.
- Schmutte, I. M., 2015. Job referral networks and the determination of earnings in local labor markets. *Journal of Labor Economics* 33 (1).
- Shorrocks, A. F., 1980. The class of additively decomposable inequality measures. *Econometrica: Journal of the Econometric Society*, 613–625.
- Sunstein, C. R., 2009. *Republic. com 2.0*. Princeton University Press.
- Theil, H., 1967. *Economics and information theory*. North-Holland.
- Tifferet, S., Herstein, R., 2012. Gender differences in brand commitment, impulse buying, and hedonic consumption. *Journal of Product & Brand Management* 21 (3), 176–182.
- Unger, D. G., Wandersman, A., 1985. The importance of neighbors: The social, cognitive, and affective components of neighboring. *American Journal of Community Psychology* 13 (2), 139–169.
- Verbrugge, L. M., 1977. The structure of adult friendship choices. *Social forces* 56 (2), 576–597.
- Waldfogel, J., 2009. *The tyranny of the market: Why you can't always get what you want*. Harvard University Press.
- Weitzman, M. L., 1992. On diversity. *The Quarterly Journal of Economics* 107 (2), 363–405.