The Contribution of Urban Parks to Social Ties between Neighborhoods.

A Study of Online Friendship Networks in New York City.*

Timur Abbiasov†

Columbia University‡

September 28, 2021

Abstract

Although the role of open spaces in promoting social ties between neighborhoods has been widely discussed by social scientists and urban design scholars, data sets from online social networks present unexplored opportunities to quantify this link on a larger scale. In this paper I use data on friendship links between Facebook users across New York city zip codes to show that two neighborhoods with higher density of green space between them are more likely to have stronger social ties. In particular, when controlling for demographic differences and zip-code level fixed effects, I find that a 1 p.p. increase in the percentage of land allocated to parks between two given zip codes is associated with a 1.2% higher chance of online social connection between their residents. Comparing the effects of park density for different types of parks, I further document that the presence of community parks, flagship parks, and playgrounds are all significant predictors of higher social connectedness between zip codes. Notably, the largest estimated effect is for playgrounds, indicating a 33% higher probability of connection per 1 p.p. increase in density.

*I am very grateful to Donald Davis, Reka Juhasz and David Weinstein for their invaluable guidance and continued support. I also thank Dmitry Sedov and Dmitry Sorokin for helpful discussions and comments. Special thanks to Michael Bailey for facilitating access to the Facebook data. All errors are mine.

†Email: t.abbiasov@columbia.com

‡Department of Economics, 420 W. 118th St., New York, NY 10027

1 Introduction

Urban settings with stronger neighborhood social ties have previously been associated with improved psychological well-being (Gu, 2020), higher crisis resilience (Klinenberg, 2018), lower levels of social disorder (Hirschfield and Bowers, 1997) and better economic development (Burchardi and Hassan, 2013). Public spaces and, in particular, urban parks have been recognized as providing opportunities for encounters that enable different forms of social interactions and offer the promise of improved social connectedness between neighborhoods. While social scientists and urban design scholars have accumulated a considerable body of knowledge about different types of shared spaces and how their physical and subjective properties affect social contacts (see e.g. Mehta, 2009; Mehta and Bosson, 2010; Aelbrecht, 2016; Peters et al., 2010), existing empirical studies typically rely on qualitative data collection methods[1]. In this paper, I use the more ubiquitous data on online social connections between New York zip codes obtained from Facebook to investigate the role of local parks in facilitating social contacts between neighborhoods on a larger scale.

In his seminal work "Life Between Buildings", Gehl (1987), emphasizes the role of parks and other public and semi-public spaces as the breeding ground for various types of social activities. These can include brief conversations, family gatherings and children at play, flocks of cyclists, or people who brought their dogs to run around with each other. Most common activity, Gehl adds, is simply seeing and hearing other people. Being among others, in turn, can create opportunities for new forms of contact such as spontaneous friendships or brief exchanges between people crossing each other's paths every once in a while. Furthermore, for neighbors these cursory connections and repeated visual contacts in parks, gardens and plazas can contribute to the formation of local social ties that are instrumental to maintaining a well-functioning community (Ross and Jang, 2000). While the role of urban green spaces in supporting neighborhood social contact has been theorized and studied by scholars across disciplines (for an overview, see Kazmierczak, 2013; Moulay et al., 2017; Wan et al., 2021), large-scale analysis has typically been unfeasible due to lack of comprehensive city-wide data on social contacts. By exploiting the vast universe of online links

[1] Such as interviews, surveys, photo-elicitation and direct observation.

between Facebook users that reside across different zip codes in New York city, this paper provides a first attempt in quantifying the contribution of local parks to social connections between neighborhoods using big data from social media.

The main dataset used in this paper is identical to the one first described in Bailey, Farrell, Kuchler, and Stroebel (2020). More specifically, based on a de-identified snapshot of all active Facebook users from March 2018 with active geolocation tracking, Bailey et al. (2020) derive counts of the number of connections – i.e. friendship links – between each zip code in the New York CSA and other regions outside of New York area. From these, authors construct a measure of *social connectedness* that represents the relative probability of a Facebook friendship link between a given user in zip code i and a given user in zip code j. After obtaining the information on the online social connections between the city's zip codes, I use the digital maps of New York City's parks to measure how much green space is available at each location.

I establish two main results. First, I show that a 1 p.p. increase in park density in the area between each pair of zip codes (measured as the share of land area covered by parks) is associated with 1.2% higher chances of online social contact between the two zip codes, when controlling for zip-code level fixed effects and differences in relevant demographics in each pair. Second, I use similar specifications to consider the role of different types of urban parks and find that the presence of community parks, playgrounds and flagship parks between neighborhoods are all significant predictors of higher social connectedness. In particular, a 1 p.p. increase in density of playgrounds is associated with 33% higher chance of a friendship link between two zip codes, while for community parks and flagship parks the same increase leads to 2.6% and 0.9% higher chances of social connections, respectively.

The results presented in this paper come with important limitations. While controlling for zip-code level fixed effects allows to account for unobserved heterogeneity on the zip-code level, it cannot completely rule out endogeneity due to factors at the zip-pair level that influence both the availability of parks between neighborhoods and their tendency to socialize with each other. To alleviate this concern, I include various measures of demographic differences between neighborhoods, including income, education and racial composition.

However, even after controlling for these differences it is still possible that unobserved factors at the zip-pair level confound the causal interpretation of the estimated coefficients. As such, this paper serves as first step in attempting to study the relationship between open spaces and local social ties using the geospatial data from online social networks, and warrants further research along these lines.

In addition to the research already mentioned, this paper contributes to a recent literature that has used data from online services or mobile phones to understand how people interact in cities (Davis, Dingel, Monras, and Morales, 2019; Athey, Ferguson, Gentzkow, and Schmidt, 2020; Xu, Belyi, Santi, and Ratti, 2019), and also builds on my earlier paper, where I investigate the causal link between access to green space and racial diversity experienced in daily encounters between the residents of New York using the data from Twitter (Abbiasov, 2021).

The remainder of the paper is organized as follows. Section 2 describes the data. In Section 3, I explore the link between the overall access to parks and neighborhood social connectedness, and then consider differences in park types and how these relate to the number of social contacts between neighbors. Section 4 concludes.

2 Data

To quantify social ties between neighborhoods in New York I use the data set initially introduced in (Bailey et al., 2020). Derived from the snapshot of friendship links between Facebook users that were active[2] and had geolocation tracking enabled as of March 2018, this data set measures the level of social connectedness between nearly all zip codes in New York city.

Since its launch in 2004, Facebook has became the most popular social media platform both in the US and globally. According to an independent survey cited in Bailey et al. (2020), more than 68% of the U.S. adult population and 79% of online adults in the U.S. used Facebook in 2015. The platform's usage rates are also consistently high across income groups, education, and race, and among urban, rural, and suburban residents, while declin-

[2]i.e. had interacted with Facebook over the 30 days prior to the date of the snapshot.

ing slightly in age (Greenwood et al., 2016). Moreover, as discussed in Bailey et al. (2020), Facebook mainly serves as a platform for real-world friends and acquaintances to interact online, and that the networks formed on Facebook more closely resemble real-world social networks than those on other online platforms. This data set, hence, presents a unique opportunity for researchers to study the role of shared spaces in facilitating social contacts.

Due to concerns about user anonymity, the data on friendship links is available only for zip codes with a total population of at least 500 people and that are above the 5th percentile in the number of eligible Facebook users within the New York CSA. The final data set includes 182 zip codes in New York city and 1,181 zip codes across the entire metropolitan area.

To measure the availability of parks across different neighborhoods I use the data provided by the New York City Department of Parks & Recreation describing the locations and features of all parks operated by the department, including geographic boundary, total area and park category. Finally, I combine these with the data on demographics of zip codes from the 2018 Census Bureau 5-year American Community Survey (ACS). In particular, I obtained zip-code level measures of total population, median income, educational attainment, and racial composition.

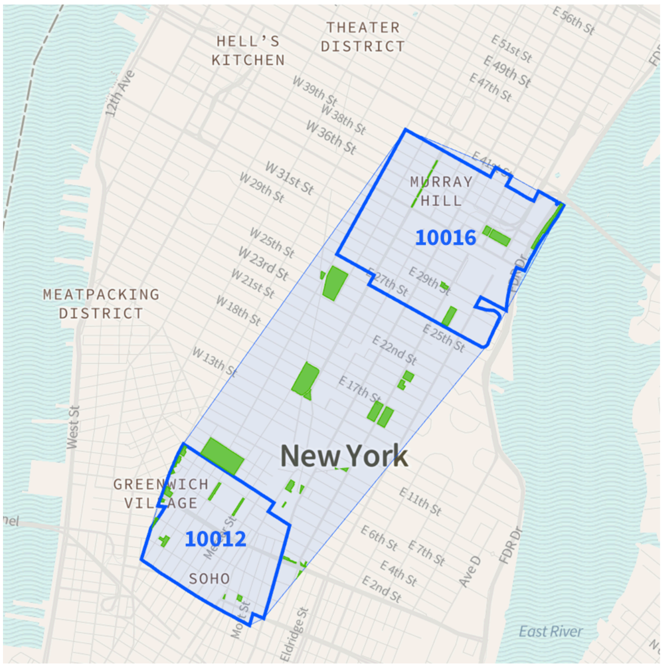
2.1 Social connectedness index

The primary variable describing the relative probability of a friendship link between a given user in zip code i and a given user in zip code j is defined as follows:

$$SocialConnectedness_{i,j} = \frac{FB\_Connections_{i,j}}{FB\_Users_i \times FB\_Users_j} \tag{1}$$

In the above, $FB\_Connections_{i,j}$ refers to the total number of connections between Facebook users living in zip code i and users living in zip code j, and FB_Users_i is the number of eligible Facebook users in zip code i. For more details about the construction of the social connectedness index, please, see the original paper by Bailey et al. (2020).

FIGURE 1: MEASUREMENT OF PARK DENSITY BETWEEN NEIGHBORHOODS



Note: Thick blue lines indicate the boundaries of two given zip codes. Light blue shaded area is the convex hull of the two zip code geographies. Light green shaded areas show all park properties within the convex hull. The resulting density measure is obtained by dividing the total park acreage within the hull by its total land area.

2.2 Park availability measurement

One of the common approaches used in the literature to measure the availability of parks across locations is the density measure (see discussion in Panduro et al., 2018). In this paper I adopt a similar approach to measure the availability of parks between neighborhoods. In particular, I compute the share of parkland in the geographic area between each pair of zip codes. As illustrated in Figure 1 the density measure is obtained by constructing a convex hull for each pair of zip-code boundary geometries, and then dividing the total park acreage within that hull by its total land area. Hence, it most closely corresponds to the density of parks located directly between each pair of zip codes. In the Appendix A.3 I discuss an alternative measure using a concave hull[3] of the two geometries and show that the results presented in the following section are robust to using this approach.

To provide some reference on the magnitudes of park densities between pairs of zip codes, Table 1 summarizes the distribution of park densities across zip code pairs. For an

[3]See details on the concave hull technique here: `http://www.bostongis.com/postgis_concavehull.snippet`

Table 1: Density of parks between zip code pairs

Park Type	Density, % (by zip pair) Mean	Median	St. dev.
All Parks	9.37	8.59	5.02
Community Park	1.80	1.38	1.76
Neighborhood Park	0.70	0.59	0.48
Flagship Park	3.95	2.93	4.26
Playground	0.37	0.34	0.20
Garden	0.04	0.03	0.05
Recreation Field/Courts	0.37	0.15	0.63
Triangle/Plaza	0.08	0.05	0.09
Other	2.07	1.58	1.97

Note: The density measure is obtained by constructing a *convex* hull for each pair of zip-code boundary geometries, and then dividing the total park acreage within that hull by its total land area. Playground category is comprised of two distinct groups of parks defined by the Parks Department: ”Playgrounds” and ”Jointly Operated Playgrounds”. Other category includes: ”Buildings/Institutions”, ”Cemetery”, ”Historic House Park”, ”Lot”, ”Mall”, ”Managed Sites”, ”Nature Area”, ”Parkway”, ”Strip”, ”Undeveloped” and ”Waterfront Facility”

average pair, I find that roughly 9.4% of land between them is allocated to parks with a standard deviation of 5 p.p. Furthermore, Table 1 includes a breakdown of park densities for the major park categories used in my subsequent analysis. Additionally, Table A.1 provides a summary similar to Table 1, but for park densities across individual zip codes. For reference, Appendix A.1 includes the list of categories with the corresponding definitions used by the Parks Department.

3 Parks and social ties between neigborhoods

In this section, I explore how access to local parks relates to the levels of social contact between neighborhoods in New York City. In Section 3.1, I estimate the elasticity of the level of social connectedness between pairs of zip codes with respect to the density of park space in the geographic area between them, while controlling for differences in demographic variables and also accounting for zip-code level fixed effects. Further, in Section 3.2 I compare these elasticities for different park types, such as gardens, community parks, playgrounds and others.

3.1 Access to green space and social ties between neighborhoods

To establish whether the availability of local parks can facilitate social contacts between residents of zip codes i and j, I estimate the following linear model:

$$\log(SocialConnectedness_{i,j}) = \beta_0 + \beta_1 ParkDensity_{ij} + \beta_2 \log(d_{ij}) + \beta_3 X_{i,j} + \epsilon_{i,j} \quad (2)$$

The dependent variable is the log of social connectedness (see Equation 1 for definition), and $ParkDensity_{ij}$ denotes the land percentage of parks located between i and j. Control variables $X_{i,j}$ (following Bailey et al. (2020)) include measures of dissimilarity of the two zip codes across income groups (the absolute difference in median income across the zip code-pair), education levels (the absolute difference in the shares of population with a BA degree or higher across the zip code-pair), and racial compositions (the absolute difference in the non-Hispanic white shares of the residents across the zip code-pair), and d_{ij} denotes the distance in miles between each zip-code pair.

The first two columns of Table 2 present the results of estimating the model described in Equation 3. Column 1 is the baseline specification with no control variables, and column 2 includes socioeconomic differences between each pair of zip codes. The obtained results imply that 1 p.p. increase in percentage of park space in the area between given two zip codes is associated with 0.44% higher social connectedness between them. Moreover, as shown in column 2, controlling for sociodemographic disparities between the zip codes in each pair does produces a similar estimate of 0.55%.

The estimates presented in columns 1 and 2, however, can not be interpreted as causal. One major concern is due to residential selection: if proximity to parks is correlated with the unobserved characteristics of zip code residents that affect their social behaviors, the estimates in (1) and (2) will be biased. For example, this could occur if people who prefer living closer to parks are on average more sociable. To alleviate this concern, in columns 3 and 4 I present estimates for the following fixed effects model that can account for the unobserved zip-code level heterogeneity:

Table 2: Park density and social connectedness between neighborhoods

	Log social connectedness (pairs of zips)			
	OLS		*FE*	
	(1)	(2)	(3)	(4)
Park Density (% of land area)	0.0044***	0.0055***	0.0040	0.0120***
	(0.0010)	(0.0009)	(0.0043)	(0.0040)
Log(distance in miles)	−0.8517***	−0.7957***	−1.0068***	−0.9060***
	(0.0071)	(0.0067)	(0.0287)	(0.0290)
Δ Med Income (zip pair diff, k$)		0.0007***		−0.0035***
		(0.0001)		(0.0011)
Δ Share White (zip pair diff, %)		−0.0075***		−0.0093***
		(0.0002)		(0.0012)
Δ Share Pop BA and higher (zip pair diff, %)		−0.0073***		−0.0032**
		(0.0003)		(0.0014)
Zip code fixed effects	No	No	Yes	Yes
Dyadic cluster-robust SE	No	No	Zip-Zip	Zip-Zip
Observations	15,225	15,225	15,225	15,225
R^2	0.4983	0.5782	0.6505	0.7301
Adjusted R^2	0.4982	0.5781	0.6423	0.7237

Note: The density measure is obtained by constructing a *convex* hull for each pair of zip-code boundary geometries, and then dividing the total park acreage within that hull by its total land area. *p<0.1; **p<0.05; ***p<0.01

$$\log(SocialConnectedness_{i,j}) = \beta_0 + \beta_1 ParkDensity_{ij} + \beta_2 \log(d_{ij}) + \beta_3 X_{i,j} + \psi_i + \xi_j + \epsilon_{i,j} \tag{3}$$

The results of estimating the model with zip code fixed effects are presented in columns 3 and 4 of Table 2. Compared to column 1, the inclusion of fixed effects in the specification without controlling for demographic differences renders the effect of park density on social connectedness statistically insignificant. However, as shown in column 4, with the addition of the relevant control variables, the estimated effect of park density on social connections is significant and its magnitude is roughly twice as large compared to the model without the fixed effects. Namely, the results in column 4 imply that 1 p.p. increase in the share of land covered by parks between two given zip codes leads to 1.2% higher chances of a friendship link between their residents. The difference between columns (3) and (4) suggests that the former specification is susceptible to omitted variables bias and, hence, the results in column

(4) are the preferred estimates.

The results in Table 2, therefore, suggest that a higher density of parkland in the geographic area between two given zip codes in New York City is associated with higher levels of social connectedness between their residents. This finding hence falls in line with the previous literature arguing that shared spaces and parks, in particular, can serve as the breeding ground for social ties between neighborhoods.

3.2 Differences between park types

To further explore what types of parks are more or less conducive to social contacts, I next estimate a model, where the independent variable is replaced with a vector corresponding to densities of major parks categories in the area between each pair of zip codes. In particular, I consider neighborhood and community parks, gardens, flagship parks and playgrounds, and others (see Appendix A.1 for the full list of definitions). Similar to Section 3.1, I first estimate a simple OLS model, and then re-estimate it with the addition of zip code fixed effects to account for the unobserved heterogeneity at the zip-code level. Hence, I first estimate the following equation:

$$\log(SocialConnectedness_{i,j}) = \gamma_0 + \gamma_1 ParkDensityByType_{ij} + \gamma_2 \log(d_{ij}) + \gamma_3 X_{i,j} + \epsilon_{i,j} \tag{4}$$

In the above, $ParkDensityByType_{ij}$ is a vector of densities of each park type located in the geographic area are directly between i and j, and similarly to Eq. 3, X_{ij} controls for differences in demographic factors in each pair of zip codes, and d_{ij} denotes the distance in miles between i and j.

The results presented in columns 1 and 2 of Table 3 describe the results of estimating Equation 4. First, I find that for one of the most popular types of parks, community parks[4], a 1 p.p increase in park density between two zip codes is associated with 4% higher chances of a friendship link between them. When controlling for differences in median income, educational attainment and racial composition, the corresponding estimate is slightly reduced

[4]Community parks account for the second largest share in total acreage among all park types, and have the largest total number of visitors according to a study of Twitter users's geolocations in New York (Abbiasov, 2021).

to 3.3 %.

Another type of parks that has been considered in the literature as particularly conducive to social cohesion (Teig et al., 2009), gardens, display a markedly stronger association with the social connectedness between neighborhoods. In the baseline model, the estimate corresponds to 69% higher relative chances of a friendship link between two zip codes for a 1 p.p. increase in the percentage of land between them covered by gardens. Controlling for socioeconomic discrepancies between the two zip codes further increases the corresponding estimate to 115%.

On the contrary, for neighborhood parks[5], also one of the most widespread types of parks, estimating Equation 4 produces a negative coefficient: a 1 p.p. increase in density of neighborhood parks between two zip codes is associated with 4.4% lower chances of social contact between them. However, after accounting for differences in demographic variables the corresponding estimate is no longer statistically significant.

Next, the estimates in columns 1 and 2 suggest that the presence of playgrounds is negatively associated with social connectedness – a 1 p.p. increase in playground density corresponds to 23% lower chances of an online contact between two users in a given pair of zip codes (and 22% lower when controlling for differences in income, education and racial composition in each pair). As previously noted, these results should be interpreted with caution since the model in Equation 4 is susceptible to endogeneity bias. For example, if people living in neighborhoods with higher density of playgrounds systematically differ along some of the dimensions affecting social life (e.g. presence of young kids, amount of free time) not accounted in the model, this could bias the estimates downwards. Later in this section, I discuss the results of estimating a model similar to 4 with the addition of zip code fixed effects that allow me to account for this type of unobserved selection at the neighborhood level.

The estimates of the model in 4 suggest that the presence of triangles and plazas between two zip codes is positively associated with the chances of an online friendship link between them. In particular, 1 p.p. increase in density of these types of parks between neighborhoods

[5] Parks department defines neighborhood parks as "Parks that are intended to serve the direct neighborhood in which they are located". Neighborhood Parks are typically up to 50 acres and may include passive/active recreational areas.

Table 3: Types of parks and social connectedness between neighbors

	Log social connectedness (pairs of zips)			
	OLS		*FE*	
	(1)	(2)	(3)	(4)
Community Parks (% of land area)	0.0402***	0.0336***	0.0250**	0.0259***
	(0.0026)	(0.0024)	(0.0102)	(0.0098)
Gardens (% of land area)	0.6945***	1.1502***	−0.5795	−0.5602
	(0.1454)	(0.1411)	(0.5430)	(0.5217)
Neighborhood Parks (% of land area)	−0.0436***	−0.0077	0.0067	0.0069
	(0.0134)	(0.0127)	(0.0619)	(0.0568)
Playgrounds (% of land area)	−0.2319***	−0.2234***	0.1490	0.3341**
	(0.0368)	(0.0341)	(0.1959)	(0.1502)
Triangles/Plazas (% of land area)	0.2316***	0.3504***	0.1499	−0.2494
	(0.0502)	(0.0515)	(0.2070)	(0.1698)
Flagship Parks (% of land area)	−0.0103***	−0.0075***	−0.0007	0.0090*
	(0.0011)	(0.0011)	(0.0048)	(0.0047)
Recreation Fields/Courts (% of land area)	−0.0589***	−0.0227***	−0.0499**	0.0076
	(0.0071)	(0.0062)	(0.0234)	(0.0260)
Log(distance in miles)	−0.8541***	−0.7887***	−0.9961***	−0.9060***
	(0.0075)	(0.0071)	(0.0308)	(0.0314)
Δ Med Income (zip pair diff, k$)		0.0004***		−0.0033***
		(0.0002)		(0.0011)
Δ Share White (zip pair diff, %)		−0.0072***		−0.0092***
		(0.0002)		(0.0012)
Δ Share Pop BA and higher (zip pair diff, %)		−0.0073***		−0.0038***
		(0.0003)		(0.0014)
Zip code fixed effects	No	No	Yes	Yes
Dyadic cluster-robust SE	No	No	Zip-Zip	Zip-Zip
Observations	15,225	15,225	15,225	15,225
R^2	0.5111	0.5870	0.6505	0.7292
Adjusted R^2	0.5109	0.5867	0.6463	0.7259

Note: The density measure is obtained by constructing a *convex* hull for each pair of zip-code boundary geometries, and then dividing the total park acreage within that hull by its total land area. *p<0.1; **p<0.05; ***p<0.01

corresponds to 23% higher chances of social connection between them (and 35% higher chances when controlling for sociodemographic differences).

Finally, according to columns 1 and 2 of Table 3, the presence of flagship parks and sports fields/courts between zip codes is negatively associated with the social connectedness between them. For flagship parks, I find that 1 p.p. increase in density corresponds to 1% lower social connectedness between two given zip codes. Controlling for differences in income, education and racial composition further reduces the magnitude of the estimate to 0.75%. For sports fields/courts the corresponding estimates are slightly more pronounced: -5.9% without the additional controls, and -2.3% when I control for demographic differences in each zip code pair.

While the estimates of the model in Equation 4 provide some insights into how the presence different types of parks relates to the levels of social contact between neighborhoods, they are susceptible to different types of endogeneity biases. In particular, as discussed in the previous section, one major concern is the presence of unaccounted residential selection across neighborhoods: if people living closer to particular types of parks differ in their unobserved characteristics affecting their social behavior, this selection will confound the causal interpretation of the coefficients. In order to alleviate this concern, I next estimate a model similar to 4 augmented with fixed effects at the zip-code level. Including fixed effects allows me to account for unobserved differences between zip codes that affect their tendency to establish social contacts with other neighborhoods, and to separate these effects from the effects of park density. Hence, I next estimate the following model:

$$\begin{aligned} \log(SocialConnectedness_{i,j}) = \gamma_0 + \gamma_1 ParkDensityByType_{ij} + \gamma_2 \log(d_{ij}) \\ + \gamma_3 X_{i,j} + \psi_i + \xi_j + \epsilon_{i,j} \end{aligned} \tag{5}$$

The results of estimating the model with zip code fixed effects are presented in columns 3 and 4 of Table 3. Notably, compared to columns 1 and 2, the addition of zip-code level fixed effects renders estimates for many of the considered park types insignificant. As shown in column 3, estimating Equation 5 without controlling for demographic differences in each zip code pair produces only two significant estimates (at 5% confidence level). First, column 3 indicates that 1 p.p. increase in the density of community parks between two zip codes leads to 2.5% higher probability of an online social connection between them. Second, for 1 p.p. increase in the density of recreation fields/courts I find a negative effect of -5%. However, unaccounted demographic differences between neighborhoods are likely to cause omitted variables bias in the results shown in column 3. Therefore, in column 4 I present the results of estimating the fixed effects model with the addition of controls for differences in income, educational attainment and racial composition.

The estimates for the preferred fixed-effects specification (column 4 of Table 3) reveal that the presence of community parks, playgrounds and flagship parks are all significant predictors of higher social connectedness between neighborhoods in New York city. In particular, I find that a 1 p.p. increase in the land percentage of community parks in the geographic area between two given zip codes is associated with 2.6% higher probability of a friendship link between them (significant at 1%). The corresponding estimate for flagship parks is 0.9% (significant at 10%). Finally, as show in column 4, the estimate for playgrounds turns out to be the highest among all park types: 1 p.p. increase in density of playgrounds is associated with 33% higher social connectedness, which is significant at the 5% confidence level.

The contrast between the estimates of the simple OLS model and the fixed effects model points to the different ways residential selection into locations with specific types of parks interacts with social activity. For playgrounds, the inclusion of zip-code level fixed effects results in the sign of the coefficient being flipped from negative to positive, suggesting that people living in proximity to playgrounds, on average, establish less social contacts online due to unobserved selection. Similarly, the sign of the estimated effect flips from negative to positive for flagship parks and sports fields/courts, which suggests that those types of parks are more valued by individuals who, on average, have fewer friends online. On the contrary, for gardens the sign flips the other way around: in the OLS model the estimated effect is positive, while the FE model produces a negative estimate (although statistically insignificant). This change indicates that people living in locations with higher density of gardens are, on average, more inclined to establish online friendships with residents from other neighborhoods, but this happens due to selection rather than due to the presence of gardens themselves.

4 Conclusion

In this paper I provide a case for using big data from online social networks to understand the contribution of local parks to social ties between neighborhoods. The presented results lend new support to the hypothesis that public spaces, and, in particular, urban parks provide city residents with viable opportunities to interact and develop social connections. The paper's empirical approach, however, does not imply that these links are necessarily causal. Nevertheless, by using a comprehensive network of connections across New York's zip codes and controlling for a rich set of demographic factors and zip-code level fixed effects, it suggests that the availability of parks contributes to the strength of social ties between neighborhoods and warrants further research along these lines.

References

ABBIASOV, T. (2021): "Do Urban Parks Promote Racial Diversity? Evidence from New York City," Working paper, Columbia University.

AELBRECHT, P. S. (2016): ""Fourth places": the contemporary public settings for informal social interaction among strangers," *Journal of Urban Design*, 21, 124–152.

ATHEY, S., B. FERGUSON, M. GENTZKOW, AND T. SCHMIDT (2020): "Experienced Segregation," Working Paper 27572, National Bureau of Economic Research.

BAILEY, M., P. FARRELL, T. KUCHLER, AND J. STROEBEL (2020): "Social connectedness in urban areas," *Journal of Urban Economics*, 118, 103264.

BURCHARDI, K. B. AND T. A. HASSAN (2013): "The Economic Impact of Social Ties: Evidence from German Reunification*," *The Quarterly Journal of Economics*, 128, 1219–1271.

DAVIS, D. R., J. I. DINGEL, J. MONRAS, AND E. MORALES (2019): "How Segregated Is Urban Consumption?" *Journal of Political Economy*, 127, 1684–1738.

GEHL, J. (1987): *Life between buildings: using public space*, New York, Van Nostrand Reinhold.

GREENWOOD, S., A. PERRIN, AND M. DUGGAN (2016): "Social media update 2016," *Pew Research Center*, 11, 1–18.

GU, N. (2020): "The effects of neighborhood social ties and networks on mental health and well-being: A qualitative case study of women residents in a middle-class Korean urban neighborhood," *Social Science & Medicine*, 265, 113336.

HIRSCHFIELD, A. AND K. J. BOWERS (1997): "The effect of social cohesion on levels of recorded crime in disadvantaged areas," *Urban Studies*, 34, 1275–1295.

KAZMIERCZAK, A. (2013): "The contribution of local parks to neighbourhood social ties," *Landscape and Urban Planning*, 109, 31–44.

Klinenberg, E. (2018): *Palaces for the People: How To Build a More Equal and United Society*, Random House.

Mehta, V. (2009): "Look Closely and You Will See, Listen Carefully and You Will Hear: Urban Design and Social Interaction on Streets," *Journal of Urban Design*, 14, 29–64.

Mehta, V. and J. K. Bosson (2010): "Third places and the social life of streets," *Environment and Behavior*, 42, 779–805.

Moulay, A., N. Ujang, and I. Said (2017): "Legibility of neighborhood parks as a predicator for enhanced social interaction towards social sustainability," *Cities*, 61, 58–64.

Panduro, T. E., C. U. Jensen, T. H. Lundhede, K. von Graevenitz, and B. J. Thorsen (2018): "Eliciting preferences for urban parks," *Regional Science and Urban Economics*, 73, 127–142.

Peters, K., B. Elands, and A. Buijs (2010): "Social interactions in urban parks: Stimulating social cohesion?" *Urban Forestry & Urban Greening*, 9, 93–100.

Ross, C. E. and S. J. Jang (2000): "Neighborhood disorder, fear, and mistrust: The buffering role of social ties with neighbors," *American journal of community psychology*, 28, 401–420.

Teig, E., J. Amulya, L. Bardwell, M. Buchenau, J. A. Marshall, and J. S. Litt (2009): "Collective efficacy in Denver, Colorado: Strengthening neighborhoods and health through community gardens," *Health & place*, 15, 1115–1122.

Wan, C., G. Q. Shen, and S. Choi (2021): "Underlying relationships between public urban green spaces and social cohesion: A systematic literature review," *City, Culture and Society*, 24, 100383.

Xu, Y., A. Belyi, P. Santi, and C. Ratti (2019): "Quantifying segregation in an integrated urban physical-social space," *Journal of The Royal Society Interface*, 16, 20190536.

A Appendix

A.1 Definitions of park types

For full list of types, see User Guide for Parks Properties[6]:

- **Community Park**: Parks with a combination of active and passive recreational facilities or specialized amenities that service more than one community district. The largest of these parks have a large natural area component. Community Parks are typically between 5 and 250 acres.
- **Garden**: Landscaped plot of land with vegetables/flowers. Planted and maintained by the community and not NYCPARKS
- **Neighborhood Park**: Parks that are intended to serve the direct neighborhood in which they are located. Neighborhood Parks are typically up to 50 acres and may include passive/active recreational areas.
- **Flagship Park**: These parks have a variety of active recreational facilities and include a large natural or landscaped component. Flagship Parks are some of the largest parks in the City and attract users from throughout the metropolitan region. They include, but are not limited to those joint interest areas that traverse multiple community districts. Such sites are designated in the New York City Charter of 1976 as being Central Park, Riverside Park (Manhattan), Prospect Park (Brooklyn), Flushing Meadows-Corona Park, Forest Park (Queens), Bronx Park, Van Cortlandt Park and Pelham Bay Park (Bronx).
- **Playground**: Standalone facilities under DPR jurisdiction and management consisting of playground equipment along with perhaps hard surface or turf sports areas. Typically under 5 acres and more than 50% of the total site.
- **Jointly Operated Playground**: Playgrounds adjacent to public schools jointly operated by NYCPARKS and the Department of Education.

[6] https://nycopendata.socrata.com/Recreation/Parks-Properties/enfh-gkve

- **Recreational Field/Courts**: Sites consisting solely of hard surface/turf sports areas that are operated by DPR.
- **Triangle/Plaza**: A landscaped or paved area usually in conjunction with the arterial or local street system. These sites are primarily developed for passive recreation use or to provide an open space amenity within the neighborhood. These sites are non-linear and may be developed to contain grass, trees, shrubs, cobblestone, fences, monuments, plaques, flagpoles, benches, game or picnic tables and drinking fountains, but they have no active recreational equipment. Smaller sites may be operated as Greenstreets, but are under the jurisdiction of DPR. These sites range in size, but are typically under 1 - acre.

A.2 Descriptive statistics

Table A.1: Density of parks by zip code

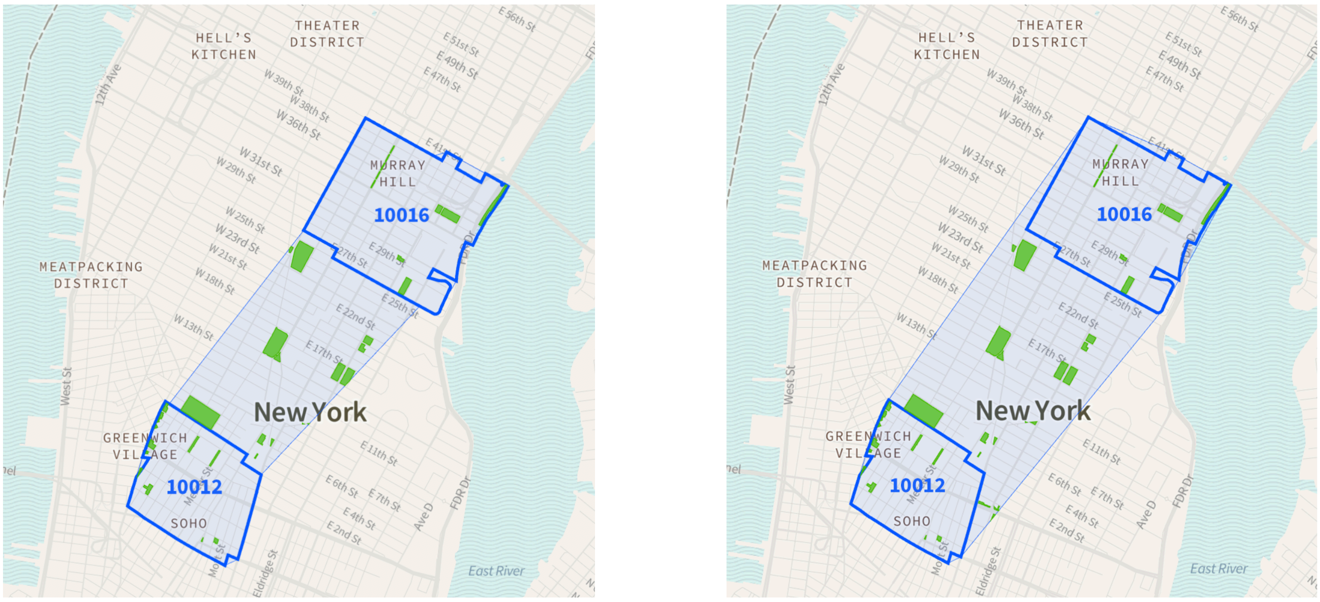
	Land % (by zip code)		
	Mean	Median	St. dev.
Park Type			
All Parks	10.06	5.34	12.40
Community Park	2.81	0.12	6.80
Neighborhood Park	1.27	0.36	5.99
Flagship Park	2.49	0.00	8.50
Playground	0.41	0.29	0.43
Garden	0.04	0.00	0.12
Recreation Field/Courts	0.27	0.00	1.10
Triangle/Plaza	0.18	0.01	1.19
Other	2.59	0.78	4.36

Note: Playground category is comprised of two distinct groups of parks defined by the Parks Department: ”Playgrounds” and ”Jointly Operated Playgrounds”. Other category includes: ”Buildings/Institutions”, ”Cemetery”, ”Historic House Park”, ”Lot”, ”Mall”, ”Managed Sites”, ”Nature Area”, ”Parkway”, ”Strip”, ”Undeveloped” and ”Waterfront Facility”

A.3 Locating parks between zip codes

As an alternative to using a convex hull for each pair of zip code geographies to define the area between them, one can use a concave hull instead[7]. Figure A.1 illustrates the difference between the two approaches. As shown in Tables A.2 and A.3, all of the results presented in the paper are robust to using the concave hull measurement.

FIGURE A.1: TWO MEASURES OF PARKS BETWEEN ZIP CODES



(A) CONCAVE HULL　　(B) CONVEX HULL

[7]See details on the concave hull technique here: http://www.bostongis.com/postgis_concavehull.snippet

TABLE A.2: PARK DENSITY AND SOCIAL CONNECTEDNESS BETWEEN NEIGHBORHOODS

	Log social connectedness (pairs of zips)			
	OLS		*FE*	
	(1)	(2)	(3)	(4)
Park Density (% of land area)	0.0049***	0.0053***	0.0025	0.0090***
	(0.0009)	(0.0008)	(0.0033)	(0.0029)
Log(distance in miles)	−0.8529***	−0.7970***	−1.0087***	−0.9111***
	(0.0072)	(0.0067)	(0.0293)	(0.0297)
∆ Med Income (zip pair diff, k$)		0.0007***		−0.0034***
		(0.0001)		(0.0011)
∆ Share White (zip pair diff, %)		−0.0075***		−0.0092***
		(0.0002)		(0.0012)
∆ Share Pop BA and higher (zip pair diff, %)		−0.0073***		−0.0032**
		(0.0003)		(0.0014)
Zip code fixed effects	No	No	Yes	Yes
Dyadic cluster-robust SE	No	No	Zip-Zip	Zip-Zip
Observations	15,225	15,225	15,225	15,225
R^2	0.4986	0.5784	0.6504	0.7293
Adjusted R^2	0.4985	0.5782	0.6422	0.7229

Note: The density measure is obtained by constructing a *concave* hull for each pair of zip-code boundary geometries, and then dividing the total park acreage within that hull by its total land area. *p<0.1; **p<0.05; ***p<0.01

Table A.3: Types of parks and social connectedness between neighbors

	Log social connectedness (pairs of zips)			
	OLS		*FE*	
	(1)	(2)	(3)	(4)
Community Parks (% of land area)	0.0260***	0.0213***	0.0145**	0.0143*
	(0.0021)	(0.0019)	(0.0071)	(0.0079)
Gardens (% of land area)	0.5352***	0.8553***	−0.2745	−0.2415
	(0.1100)	(0.1119)	(0.3586)	(0.3427)
Neighborhood Parks (% of land area)	−0.0513***	−0.0107	−0.0283	−0.0132
	(0.0116)	(0.0108)	(0.0460)	(0.0401)
Playgrounds (% of land area)	−0.1829***	−0.1578***	0.0538	0.2396**
	(0.0310)	(0.0290)	(0.1539)	(0.1169)
Triangles/Plazas (% of land area)	0.1719***	0.2807***	0.1045	−0.1780
	(0.0410)	(0.0427)	(0.1668)	(0.1464)
Flagship Parks (% of land area)	−0.0087***	−0.0062***	−0.0008	0.0084**
	(0.0010)	(0.0010)	(0.0040)	(0.0036)
Recreation Fields/Courts (% of land area)	−0.0304***	−0.0118**	−0.0258	0.0151
	(0.0064)	(0.0057)	(0.0197)	(0.0228)
Log(distance in miles)	−0.8533***	−0.7883***	−1.0048***	−0.9137***
	(0.0076)	(0.0071)	(0.0305)	(0.0308)
Δ Med Income (zip pair diff, k$)		0.0005***		−0.0033***
		(0.0002)		(0.0011)
Δ Share White (zip pair diff, %)		−0.0072***		−0.0092***
		(0.0002)		(0.0012)
Δ Share Pop BA and higher (zip pair diff, %)		−0.0074***		−0.0037***
		(0.0003)		(0.0014)
Zip code fixed effects	No	No	Yes	Yes
Dyadic cluster-robust SE	No	No	Zip-Zip	Zip-Zip
Observations	15,225	15,225	15,225	15,225
R^2	0.5070	0.5841	0.6489	0.7278
Adjusted R^2	0.5067	0.5838	0.6446	0.7245

Note: The density measure is obtained by constructing a *concave* hull for each pair of zip-code boundary geometries, and then dividing the total park acreage within that hull by its total land area. *p<0.1; **p<0.05; ***p<0.01