



Up-and-Coming or Down-and-Out? Social Media Popularity as an Indicator of Neighborhood Change

Journal of Planning Education and Research
1–12

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0739456X21998445

journals.sagepub.com/home/jpe



Constantine E. Kontokosta¹, Lance Freeman²,
and Yuan Lai³ 

Abstract

By quantifying Twitter activity and sentiment for each of 274 neighborhood areas in New York City, this study introduces the Neighborhood Popularity Index and correlates changes in the index with real estate prices, a common measure of neighborhood change. Results show that social media provide both a near-real-time indicator of shifting attitudes toward neighborhoods and an early warning measure of future changes in neighborhood composition and demand. Although social media data provide an important complement to traditional data sources, the use of social media for neighborhood studies raises concerns regarding data accessibility and equity issues in data representativeness and bias.

Keywords

social media, neighborhood change, urban informatics

Abstract

摘要：通过量化在纽约市274个邻里区域中每个区域的推特的活动和情绪，该研究引入了邻里流行指数，并将该指数的变化与房地产价格相关联，这是衡量邻里变化的一种常用量度。结果表明，社交媒体既提供了近乎实时的对邻里态度转变的指标，又提供了对未来邻里构成和需求变化的预警措施。尽管社交媒体数据是对传统数据源的重要补充，但在邻里研究中使用社交媒体引起了人们在数据代表性和偏见性上的数据可访问性和公平问题的担忧。

Keywords

关键词，社交媒体，邻里变化，城市信息学

Abstract

Mediante la cuantificación de la actividad y el sentimiento de Twitter para cada una de las 274 áreas del vecindario en la ciudad de Nueva York, este estudio presenta el índice de popularidad vecinal y correlaciona los cambios en el con los precios de los bienes raíces, una medida común del cambio de vecindad. Los resultados muestran que las redes sociales proporcionan tanto un indicador casi en tiempo real de las actitudes cambiantes hacia los barrios y una medida de alerta temprana de los cambios futuros en la composición y la demanda de los barrios. Aunque los datos de las redes sociales proporcionan un complemento importante a las fuentes de datos tradicionales, el uso de las redes sociales para estudios de barrio plantea preocupaciones con respecto a la accesibilidad de los datos y la equidad cuestiones de representatividad y parcialidad de los datos.

Keywords

Redes sociales, cambio vecinal, informáticas urbanas

Introduction

The development and preservation of healthy and inclusive neighborhoods has long been a goal of planners, and a key component of fostering and preserving neighborhood vitality is a thorough and timely understanding of the conditions that communities face. This need, however, has long bedeviled planning practitioners and policymakers who have struggled to collect the localized data that would facilitate analyses of neighborhood conditions over time

Initial submission, February 2020; revised submissions, June and November 2020; final acceptance, January 2021

¹New York University, New York City, NY, USA

²Columbia University in the City of New York, New York City, NY, USA

³Massachusetts Institute of Technology, Cambridge, MA, USA

Corresponding Author:

Constantine E. Kontokosta, New York University, New York City, NY 11201, USA.

Email: ckontokosta@nyu.edu

(Ellen and O'Regan 2010; Sawicki and Flynn 1996). Consequently, planners relied on ad hoc social surveys in the nineteenth century and standardized census tract data through much of the twentieth century to identify and assess neighborhood socioeconomic and demographic shifts. The increasing digitization of much of public life and the advent of social media have produced a plethora of data, some of which has the potential to complement traditional surveys to develop higher resolution models of neighborhood change that can be used to predict shifting demand, monitor potential warning signs of gentrification processes, and assist planners with a more comprehensive understanding of neighborhood dynamics (Chapple and Zuk 2016; Glaeser, Kim, and Luca 2017; O'Brien et al. 2015). These data, however, do not come without their own limitations in terms of accessibility and representativeness (Blank and Lutz 2017; Hargittai 2020).

In this paper, we consider the emergence of social media data as a new resource to study neighborhood change at high spatial and temporal resolutions. We develop a near-real-time measure of neighborhood dynamics based on social media activity that can (1) serve as an "early warning" or momentum indicator of neighborhood change and (2) complement conventional sources of planning data to create a more comprehensive and timely understanding of ongoing and future shifts in neighborhood composition and demand. We demonstrate that social media data can be an important addition to the planner's analytical toolkit to identify previously unobservable (or difficult to observe) perceptions and trends of neighborhood dynamics. However, the use of social media data for social science research and policymaking poses a number of challenges related to data access, computing resources, and representativeness bias that must be acknowledged and addressed.

We begin by discussing the role of social media data in urban research and studies of neighborhood change, focusing on use of space and sentiment analysis. Then, using data extracted from more than 13 million tweets between 2010 and 2017, we quantify Twitter activity and sentiment for each of 274 neighborhood areas in New York City (NYC) as defined by Zillow real estate submarket boundaries. We integrate these measures of tweet volume (attention) and reputation (sentiment) to create the Neighborhood Popularity Index (NPI), and correlate changes in this index with real estate prices, a widely used indicator of neighborhood change. A case study of the Gowanus neighborhood in Brooklyn—a community experiencing rapid change over the study period—demonstrates the potential application of the index as a hyperlocal indicator of neighborhood dynamics. We conclude with a discussion of the limitations and potential equity concerns in using this type of data for neighborhood change studies, and opportunities to further advance the science and practice of community planning in the context of new data resources.

Social Media in the Planning Context

Kaplan and Haenlein (2010) define social media as "Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content." Social media has generally been driven by private concerns seeking to capitalize on the human desire to connect with one another or communicate with specific groups that share common interests or opinions. As of 2018, the top five most widely used social media sites in the United States were Facebook, YouTube, Instagram, Twitter, and Reddit (Kallas 2018), with many sites, such as Pinterest and Snapchat, relying on mobile devices as their primary engagement platform (Statista 2018). Beyond these apps, there are also forums on the Internet that focus on issues pertinent to community development, although their usefulness for planners may be limited (Afzalan and Evans-Cowley 2015). These include websites such as City-Data, which houses forums where people post about various neighborhoods and cities, and Yelp, an online review platform where users rate and review business establishments. Taken together, social media provide an emerging data source for understanding use of space (by whom and for what) and users' sentiments toward particular places.

Use of Space

Social media users are often motivated by a desire to share their experiences with others (Oh and Syn 2015). As such, social media can offer a sense of how individuals use and interact with physical space, and provide an indicator of the significance of a particular place to a social media user based on the frequency of activity (Poorthuis 2018). Whether reporting their experiences at the local coffee shop, walking their dog, or at the post office, part of the social media experience is to report one's mundane yet idiosyncratic experiences while carrying out day-to-day activities. Instagram and Twitter posts are often geotagged (either through geographic coordinates extracted from a user's mobile device or by inferring from the contents of a post), giving a lens into how particular spaces are used (Hawelka et al. 2014). For instance, Shelton et al. (2015) used Twitter and Foursquare to study how and why Louisville residents cross neighborhood boundaries in their daily routines. They associate the preponderance of tweet activity with a Twitter user's home neighborhood, using a simple method of defining "home" as where 50 percent or more of their tweets originated. Then, by following Twitter users across space, they could discern how residents of highly segregated neighborhoods (in terms of residence) crossed neighborhood boundaries. Their findings show that users originating in the predominantly black and presumably isolated West side of Louisville crossed over to the "White" side of Louisville much more frequently than vice versa. The use of tweets enables the analysis of the

timing, as well as the location, of activity, which has been used to demonstrate that some spaces, while heterogeneous spatially, are segregated temporally. For example, the area around Churchill Downs horse racing track shows tweet activity from the predominantly black West End neighborhood most of the time, except during the racing season when tweets from the east end abound. Exploring another way to detect neighborhood change through the use of space, Steentoft et al. (2018) examined changes in the socioeconomic profile of Twitter users in a neighborhood to predict house prices. Users' socioeconomic profiles were inferred by the median household income of the neighborhoods where the respective user spent the most time. Although limited by the relatively low proportion of geotagged tweets and the coarse method used for defining home neighborhoods, their approach proved fairly accurate in predicting subsequent price changes.

Neighborhoods play a variety of roles as spaces where people live, work, shop, go to school, and recreate, among many other activities. Conventional neighborhood indicators tell us much about the people who live in neighborhoods, but little about how people move and use space within and across neighborhoods. Transportation scholars have used social media to estimate "home," work, and entertainment locations based on the frequency and timing of geotagged posts (e.g., posts originating every morning at 6 a.m. are assumed to be the home location; Luo et al. 2016; Mahmud et al. 2014). Thus, geotagged social media allows one to infer what activity the user was engaged in at the time of posting. The ability to detect these variations highlights the importance of high spatial and temporal resolution data that can be used to more fully understand the dynamics of place and transitions between neighborhoods beyond what can be learned from static, survey-based methods alone. With over 1.3 billion global users, Instagram and Twitter posts are ubiquitous and generate trillions of individual data points; therefore, these platforms are likely to generate greater spatial coverage of activity, particularly in urban environments. Usage of both platforms, however, tend to skew toward the young, and an unrepresentative subset of the population are the most frequent users of social media (Efthymiou and Antoniou 2012). Nevertheless, geotagged posts emanating from mobile technologies can reveal how space is being used and how these uses vary over time.

The perceptive reader can anticipate some of the pitfalls of using social media to study the use of space. To the extent that only posts from the most active social media users inform an analysis, the results will be biased to reflect their behavior in a particular neighborhood. More importantly, the socioeconomic, demographic, and cultural characteristics that influence the use of social media by different groups, if not recognized and accounted for, can lead to an incomplete understanding of place that obscures the needs of vulnerable and under-served populations (Blank and Lutz 2017). These representativeness issues become more pronounced when

only using a small sample of social media data, such as what is available for free through Twitter's API (Morstatter, Pfeffer and Liu 2014). Despite this, when compared with other approaches for studying neighborhood dynamics, which tend to entail costly and obtrusive surveys, time diaries, or tracking devices, geotagged social media is an accessible and timely resource that warrants further investigation.

Sentiment Analysis and Perceptions of Place

Because social media content is often created as a way for individuals to present themselves to the larger world, it provides real-time information on users' self-described activities and sentiments. The ability to discern people's perceptions of place has thus attracted the attention of scholars who study neighborhoods (Hollander et al. 2016). For example, Schweitzer (2014) investigated social media sentiment toward public transit using Twitter data and suggested that planners have opportunities to leverage social media for more engaging and reactive civic services. In addition to online forums and social media that allow posters to record their experiences and opinions in real time, Yelp, the online reviewing site, has also proved to be a rich source of data on neighborhood attributes. For instance, Zukin et al. (2017) used narratives that accompany posted reviews to parse perceptions of gentrification by analyzing how reviewers referred to two gentrifying neighborhoods, one predominantly black and the other predominantly white. Their analysis found the black neighborhood to be viewed in much less positive terms and the reviewers more sympathetic to the threats posed by gentrification in the predominantly white neighborhood. Although the representativeness of Yelp reviewers is limited, Yelp coverage of businesses is extensive, with one study finding that it captured more active businesses than the Census Bureau's County Business Patterns (Glaeser et al. 2017). The same study used Yelp reviews to predict changes in house prices and economic activity at the ZIP code level. Their analyses suggest that the number of Yelp reviews are positively associated with economic activity and, in turn, demographic and house price changes, providing further evidence that social media data can be a viable resource to supplement conventional data used to measure neighborhood change.

While spatial distributions and temporal patterns of social media usage reflect the "digital traces" of collective human mobility in cities, such data do not indicate public attention and social media content sentiment. Social listening is a novel approach to collect and analyze content relevant to different topics by analyzing social media mentions (Crawford 2009). This is based on the fact that social media as a tool for expressing popular opinions and public attention may provide timely socioeconomic and political insights (Qualman 2012). In recent years, social media analytics has been widely integrated as a critical part of business intelligence for digital marketing and customer engagement (Fan and Gordon

2014). Similar approaches could provide new opportunities for urban planning and management. For example, analyzing the shifting frequency of posts and sentiment about specific social media topics can support quantitative insights to monitor the impacts of specific events, such as a sport or cultural venue (Balduini et al. 2013). If social media, at some level, reflects public opinions and reactions—what people are talking about and how they feel about it—can we use such novel information sources to quantify and monitor perceptions of neighborhoods?

Data and Methods

We develop an indicator of neighborhood change—the Neighborhood Popularity Index (NPI)—using Twitter data to capture both sentiment and volume of social media activity about a particular place. The objective is to create a transparent, reproducible index that can be used by planners as a complementary measure of neighborhood change, while overcoming the spatial and temporal constraints of using conventional data sources alone. This tool could be used as an early warning indicator of shifting attitudes or demand for specific neighborhoods or signal incipient economic distress. To begin, we analyze 13 million tweets originating within the boundaries of NYC using a natural language processing (NLP) algorithm and time-series clustering to create a neighborhood index based on tweet volume and sentiment derived from tweet text. We use the 274 Zillow real estate submarkets to define neighborhood areas, and examine the relationship between our index values and a widely used indicator of neighborhood change: real estate prices.

Data Sources

We first extract Twitter data from Crimson Hexagon, a social media analytics company (Crimson Hexagon 2018), through a data sharing agreement. Using its pre-defined NLP algorithm, the text of each tweet is classified by its overall sentiment as “positive,” “negative,” or “neutral” (for a full description of the methodology, please see Hopkins and King 2010). Here, sentiment is analyzed for each tweet referencing a specific place that we define based on common neighborhood names and derivations. We adopt neighborhood naming conventions and spatial boundaries defined by Zillow, an online property listing platform for sales and rentals (Zillow 2018a). Zillow collects information on approximately 17,000 neighborhood areas in major U.S. cities based on multiple datasets and real estate agents’ perception of neighborhoods (Zillow 2018b). In NYC, Zillow defines a total of 274 unique neighborhoods. We acknowledge that this convention may differ from how local residents perceive the boundaries and names of their own communities; however, the Zillow definitions enable standardized keyword searches for sentiment analysis and linked geographies for real estate price data. We

use these neighborhoods to conduct a comprehensive web search to identify common colloquialisms for each neighborhood name (e.g., “Bedford-Stuyvesant” as “Bed-Stuy”) to expand our keyword query of the Twitter dataset. Crimson Hexagon has a built-in function to specify data queries within certain spatial extents (such as a state or a city). Considering the possibility that neighborhoods outside of NYC may have similar names as those within it, we set NYC as the spatial boundary so tweets posted outside the City limits are excluded. Using approximately 13.6 million individual tweets from 2010 to 2017, we then calculate two daily social media measures—attention and reputation—by neighborhood.

We use the Zillow Home Value Index (ZHVI) as a proxy for neighborhood real estate values. The ZHVI is based on a real-time database of more than 110 million property transactions representing 95 percent of the U.S. housing stock by market value (Zillow Research 2019). The ZHVI reports several monthly indicators at the national, state, ZIP code, and neighborhood levels, including median home value and median listing price, as well as rental rates.

Methodology

We define two parameters—*attention* as a measure of tweet volume and *reputation* as a measure of sentiment positivity—to develop a unified index of neighborhood popularity. Attention measures the number (volume) of tweets referencing a particular neighborhood, which reflects social media activity (positive or negative) about a given place that can fluctuate with local media coverage, emergencies, or other special or anomalous events (Kavanaugh et al. 2012). Reputation reflects users’ mood, emotions, and opinions in response to social, political, and cultural events associated with a neighborhood, which collectively captures user sentiment (Bollen et al. 2011; Stieglitz and Dang-Xuan 2013). For a particular neighborhood i , its social media attention is given by the total social media mentions at time t divided by the citywide total tweet volume in the same period. Reputation is a normalized measure based on all non-neutral mentions of each neighborhood. The product of the normalized social media attention and reputation measures generates a single value, which we define as the NPI.

Using the NPI, we then categorize long-term neighborhood popularity change by time-series classification. This classification process proceeds in the following steps. First, we calculate each neighborhood’s annual moving average NPI (window = 12, minimum periods = 3) to smooth monthly fluctuations. Each neighborhood NPI is converted to a z score to standardize time-series values, such that each has a mean of zero and standard deviation of one to account for variations in geographic scale and population of individual neighborhoods. We identify a threshold that defines an “unchanged” neighborhood based on observed fluctuations in the standardized time-series. To do so, we sort all neighborhoods by their sum absolute deviation (total absolute

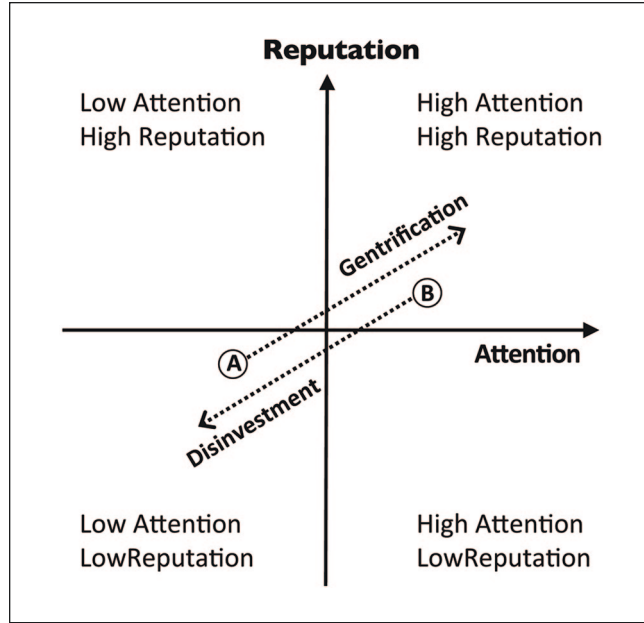


Figure 1. Neighborhood popularity measured by social media attention and reputation.

deviation from its mean) and identify a threshold using the knee-point locator method. Neighborhoods with low sum absolute deviations are those with relatively stable popularity levels. For neighborhoods whose popularity has been found to change over time, k -means clustering is used to classify similar groups of neighborhoods based on underlying temporal patterns in social media attention and reputation. A Silhouette analysis is used to identify the optimal number of clusters. Specifically, for $k \in [2, 10]$, a Silhouette score is computed for each iteration to measure inter-cluster variance and within-cluster similarity. The highest Silhouette score indicates the optimal number of clusters, which is found to be $k = 2$, yielding two distinct groups of neighborhoods based on respective trends, one positive and one negative.

To aid in the visualization of popularity over time, and its relationship to neighborhood change, Figure 1 shows potential temporal shifts in social media attention and reputation mapped on a coordinate plane with attention rank on the x -axis and reputation rank on the y -axis. We group ordinal rankings of NPI sub-components to label each neighborhood across one of four quadrants representing *low attention–low reputation*, *low attention–high reputation*, *high attention–low reputation*, and *high attention–high reputation*. Although time-series clustering identifies trends in popularity over time, it does not account for a neighborhood’s initial state of popularity. A popularity shift can be mathematically expressed as a vector p originating from initial popularity $p_{x,y}$ at time t_0 to popularity $p'_{x,y}$ at time t_1 . We then measure each neighborhood’s popularity shift by calculating the net difference between 2010 and 2017. As Figure 1 shows, we can consider the temporal shift represented by neighborhood A to be

consistent with increasing popularity and neighborhood B to be a signal of potential neighborhood disinvestment or economic distress. Other quadrant shifts, such as from *low attention–low reputation* to *high attention–low reputation*, also represent potential indicators of neighborhood change processes that may be of interest to planners.

To explore correlations between the NPI and conventional neighborhood change indicators, we examine the empirical relationship between historical real estate market performance and neighborhood popularity between 2010 and 2017. Using Zillow neighborhood monthly home value estimations, we adjust nominal dollar values for inflation using the monthly Consumer Price Index (CPI) and calculate annual home value percentage change.

Results

Social Media Popularity

We begin by analyzing neighborhood net sentiment positivity aggregated at week, month, quarter, and annual timescales. The results indicate significant variation in sentiment at high frequencies (e.g., daily) and a dominance of positive posts compared with negative, with the majority classified as “neutral”. These findings are consistent with previous research analyzing social media data in the Netherlands and Greater London (Daas and Puts 2014; Kovacs-Gyori et al. 2018). Therefore, a low neighborhood social media reputation rank does not necessarily indicate a negative public impression, but rather a “less positive” sentiment when compared with other neighborhoods. Since 2011, there is an overall growth in neutral social media sentiment, possibly reflecting the increasing proportion of neighborhood-related social media posts used for information sharing, such as local news reporting or advertising, rather than personal opinions.

Turning to the volume of tweets about particular neighborhoods, the average neighborhood attention share is 0.38 percent across NYC. Some popular neighborhoods draw significantly larger attention shares, such as Williamsburg (7.8%), Midtown (6.7%), Astoria (5.7%), Bushwick (3.2%), and Tribeca (3.2%). Such a long tail distribution indicates that neighborhood-related tweet volume follows a rank-size scaling. After integrating the attention and reputation measures, Figure 2A visualizes NPI scores (eight-year average: 2010–2017) by quantile. We use *max-min* normalization to transform the popularity index values to a one to hundred scale for ease of interpretation. Neighborhoods with the highest popularity index values, in addition to those with high attention shares, include Prospect Park, the East Village, DUMBO, Coney Island, and Greenpoint. Popularity correlates with distance from the Manhattan CBD, reflecting the association between population density and social media mentions, as well as the spatial patterns of frequent Twitter users. Using the eight-year average popularity, we

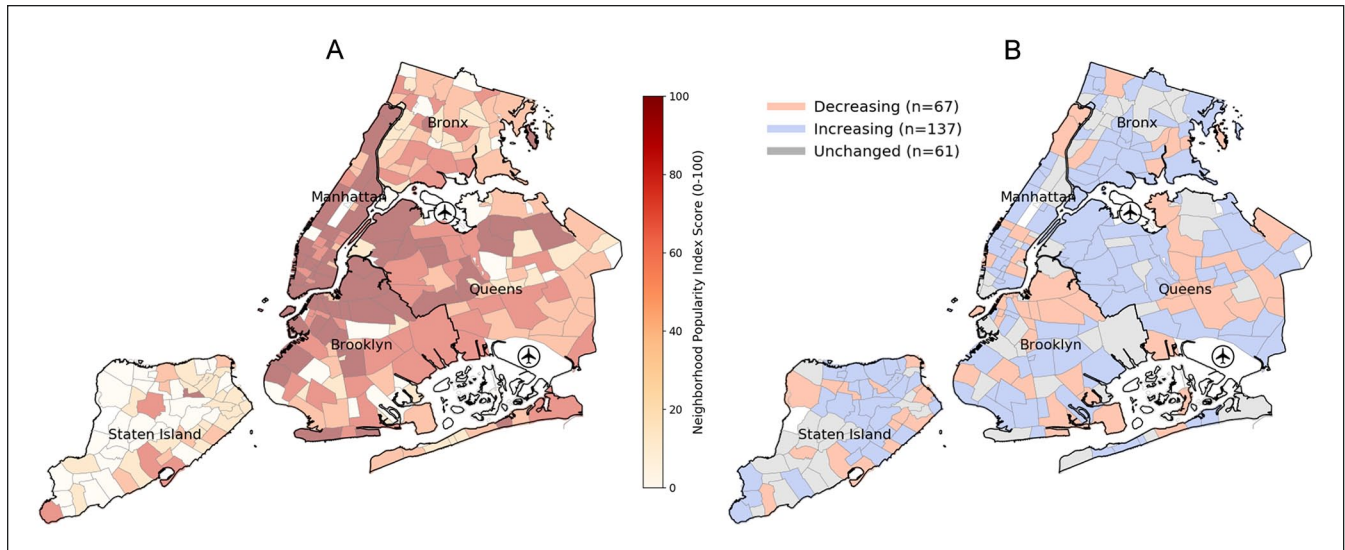


Figure 2. New York City neighborhood social media popularity: (A) eight-year average popularity index score (2010–2017) and (B) popularity change.

compute a Moran's I statistic to test for underlying spatial auto-correlation in neighborhood popularity. The p value is statistically significant ($p = .006$) and the z score is positive ($z = 0.19$). Thus, we can reject the null hypothesis that neighborhood popularity is spatially random. We use American Community Survey (ACS) data to estimate correlations between neighborhood popularity and population, but find that population density is not strongly correlated (Pearson's coefficient = .14, p value = .027). We believe this is, in part, because ACS only captures residents by home location, thus missing a large portion of the worker and visitor population. Alternatively, we estimate local population size based on the Longitudinal Employer-Household Dynamics (LEHD) survey by aggregating census block level worker and resident population data into neighborhood population estimates. We find significant correlations between neighborhood popularity with total local population (Pearson's coefficient = .50, p value < .001), as well as land area (.23, <.001) and total population density (.28, <.001). Figure 2B illustrates neighborhoods by their respective change in popularity based on time-series clustering. More than half of NYC's neighborhoods ($n = 136$) have gained popularity from 2010 to 2017, while the remaining have experienced a decrease ($n = 66$) or remained unchanged ($n = 52$). However, this trend requires a careful interpretation within the context of each neighborhood's initial popularity in 2010. Specifically, some of the most popular neighborhoods in Manhattan and Brooklyn become less dominant in social media mentions, which is labeled a "decreasing" trend.

House Prices and Neighborhood Popularity

Figure 3 provides a scatterplot of the relationship between neighborhood social media attention, reputation, and real

estate values, based on eight-year average measures. There is a clear positive relationship between home values and neighborhood popularity measures, as neighborhoods with higher average attention and reputation scores (those in the upper right quadrant) have higher median home values.

A time-series analysis reveals greater insight into the association between changes in neighborhood popularity and home values during the study period. Figure 4 presents the relationship between neighborhood popularity, popularity change, and local real estate prices based on initial popularity in 2010 and the trend defined by time-series clustering. Initial high popularity is significantly correlated to above-average home value growth over the time period (Pearson's coefficient = .43, p value < .001). Therefore, neighborhoods with high popularity at the beginning of the study period exhibit greater price growth over time than low popularity neighborhoods, even when accounting for increasing or decreasing popularity trends over time. This can be explained, in part, by well-established neighborhoods, such as the Upper East Side in Manhattan, that have historically been wealthy enclaves. Despite decreasing popularity, real estate values remain high and continue to increase. However, in less popular neighborhoods, increasing popularity over time is associated with higher real estate growth rates than those low popularity neighborhoods with unchanged or decreasing popularity over the study period. We further analyze the relationship between neighborhood popularity and home values using a panel regression model with fixed effects (see Supplementary Appendix for details). Results indicate a statistically significant positive correlation between the NPI, initial popularity, and the change in home values.

We use the framework of the neighborhood popularity quadrant illustrated in Figure 1 to more closely examine popularity change for individual neighborhoods. Figure 5 shows each neighborhood's social media popularity shift and

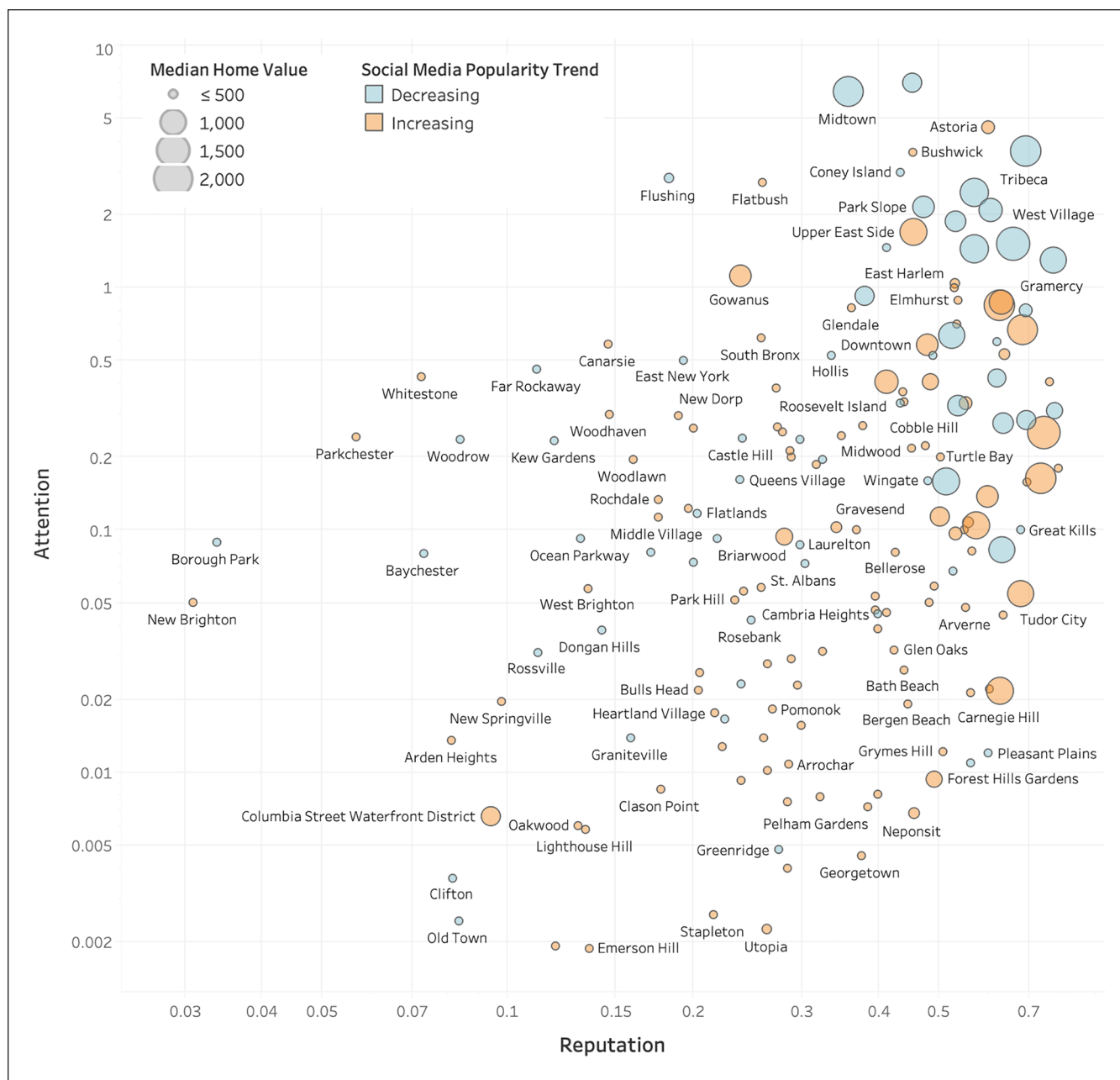


Figure 3. Average neighborhood attention and reputation scores and median home values (\$/sq.ft.).

respective home value percentage change. Neighborhoods with both increasing attention and reputation represent those gaining popularity over the study period. As all neighborhoods have experienced increasing home values since 2010, we categorize the value change into *low-growth*, *medium-growth*, and *high-growth* groups. Visually, most high-growth neighborhoods have high social media attention, reinforcing our finding that changes in home values are a function of the initial state of neighborhood popularity. Some neighborhoods experience relatively large popularity changes over time, so they may shift from one quadrant to another. The

Gowanus neighborhood in Brooklyn provides a good illustration, as it shifted from the *high attention–low reputation* quadrant in 2010 to the *high attention–high reputation* quadrant in 2017. Given this relatively dramatic increase in popularity, we explore this case study in more detail below.

Case Study: Gowanus, Brooklyn

Gowanus is a unique neighborhood in North Brooklyn, surrounded by the established (and gentrifying or gentrified) communities of Park Slope, Carroll Gardens, and Boerum

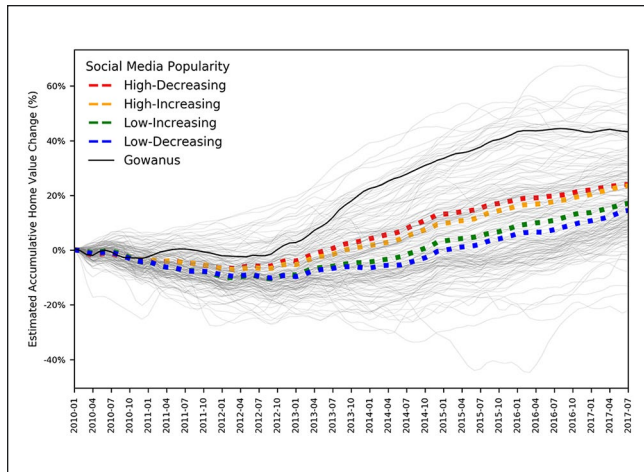


Figure 4. Neighborhood popularity and Zillow-estimated home value change (2010–2017).

Hill. Known for the canal that stretches through the community, Gowanus is home to one of the largest U.S. Environmental Protection Agency (EPA) Brownfield Superfund sites in the country. In 2009, Mayor Michael Bloomberg announced a city-led remediation plan for the Gowanus Canal after the EPA's Superfund designation (Anderson et al. 2012; The City of New York 2009). With a mix of industrial, commercial, and residential land uses, the neighborhood has experienced rapid change in recent years reflecting both the decreasing presence of manufacturing in NYC and the growing attractiveness of Brooklyn as a residential alternative to Manhattan. A previous study investigated gentrification and immigration in Brooklyn from 2000 to 2008 and found this area to have a significantly wealthier and better educated population than other neighborhoods in Brooklyn (Mason 2011). In 2013, a community-led visioning process known as “Bridging Gowanus” set out to define a sustainable and inclusive future for the Gowanus area. Following more than two years of community engagement, the NYC Department of City Planning began the formal process of reshaping the neighborhood's land use and development plans with the launch of the Gowanus Neighborhood Planning Study. These initiatives, spurred by the remediation of the Gowanus Canal and growing demand for Brooklyn real estate, reinforced the attractiveness of the area to new development and residents. According to ACS data, residents in Gowanus are increasingly well-educated, non-Hispanic White millennials. From our social media analysis, Gowanus is characterized by growing social media popularity with both increasing attention and reputation over time. This change coincides with a demographic shift reflected in ACS data: from 2013 to 2017, there was an eight-percentage point increase in individuals with higher education (42% to 50%) and a nine-percentage point decline in the low-income population (32% to 23%).

Figure 6 plots the monthly change in home values (dashed line) against the change in NPI (solid line) for Gowanus from 2010 to 2017. The neighborhood popularity score for Gowanus first surpasses Brooklyn's median score in 2012, coincident with the large-scale rezoning of the neighborhood, and we observe a significant increase in home values approximately one year later. While the estimated home value in Gowanus has been above the borough-median value since the beginning of the study period, the difference between Gowanus and Brooklyn rapidly rose from \$322/sq.ft. to \$429/sq.ft. during 2013. Beyond the long-term increasing popularity trend, we also observe two anomalous dips in the years 2013 and 2016. These two troughs occurred during the summer months, and can be associated with news, and related social media interest, regarding the canal environmental clean-up. For instance, according to the Google News search engine, the top news articles about Gowanus during the Spring/Summer of 2016 (April to July) were related to environmental hazards, health concerns, stormwater and flooding-related damage, and resident requests tied to the remediation, reflected in the decrease in the popularity index during this period.

The parallel trend between the popularity index and property values provides a validation of the generalized relationship between social media popularity and price change at the neighborhood scale. To explore this trend more fully, we first decompose two time series (monthly social media popularity score and home value) and then conduct a Granger Causality Test (maximum lags = 15) to check if the time series of Gowanus's social media popularity trend is useful in forecasting its home value trend, while taking historical home values into account as regressors. The results show that with a total number of four lags, the p value is less than .05. Thus, we can conclude that Gowanus's social media popularity trend adds statistically significant forecasting power at the 95 percent confidence level.

Discussion and Conclusion

In this study, we explore the use of social media as an emerging data source to measure neighborhood change at high spatial and temporal resolution. By extracting data on Twitter activity and sentiment for each of 274 NYC neighborhoods over seven years, we define the NPI and test the relationship between neighborhood popularity and home values derived from Zillow data and census demographic data from the ACS and LEHD. We demonstrate how social media can be used to provide both a real-time measure of shifting attitudes about neighborhoods and an early warning indicator of future changes in housing demand. By leveraging social media data as a complement to conventional data sources, planners and policymakers can have a near-real-time view of neighborhood dynamics and identify neighborhoods “at-risk” of future shifts in demand or

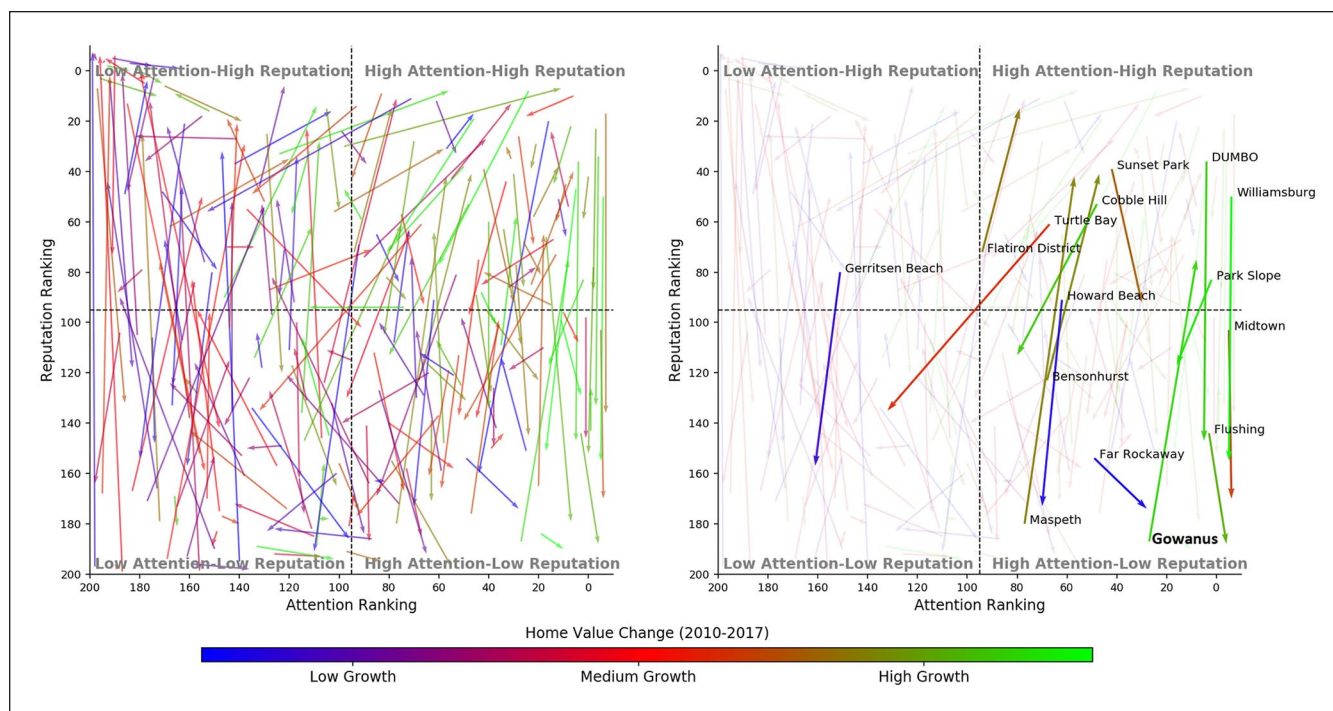


Figure 5. Neighborhood popularity sub-component rank change and home value growth rate.

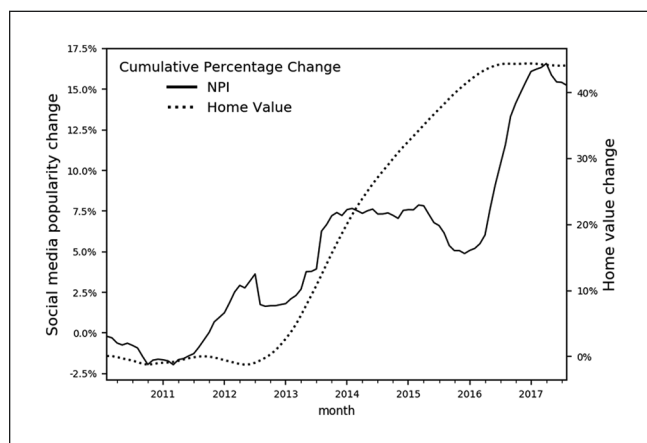


Figure 6. Monthly popularity change and home value in Gowanus, Brooklyn, 2010–2017.

Note: NPI = Neighborhood Popularity Index.

prices. This approach can help to overcome public decision-making constraints that result from an over-reliance on census or real estate price data that suffer from significant time lags in their collection and dissemination.

Critically, social media data exhibit representativeness bias that can limit the ability to generalize analytical results. According to a 2019 survey by the Pew Research Center, social media users tend to skew toward younger and higher income demographic groups, and tweet activity patterns vary significantly from user to user. In the context of understanding neighborhood change, this composition can significantly

obscure the perceptions and priorities of vulnerable and at-risk populations. Notably, however, the racial and ethnic composition of Twitter users aligns closely with the relative size by group among the U.S. adult population (Wojcik and Hughes 2019). Collectively, the demographics of Twitter users, particularly with respect to age, income, and education levels, are consistent with those typically identified as “gentrifiers” in neighborhood change studies. Thus, we are able to capture the shifting perceptions of those that would most likely be the in-movers that would account for changes in real estate prices and demographic shifts at the neighborhood level.

It is important to clarify several methodological decisions embedded in this study. Given our focus on developing the popularity index and its application to planning, we leverage Crimson Hexagon’s pre-existing sentiment analysis algorithm. There is an extensive literature on social media analytics, including platform engineering, data integration, natural language processing, and classification modeling using machine learning (He et al. 2019), and our intent is not to develop yet another sentiment classification algorithm. Furthermore, we do not investigate social media users’ spatial patterns, as the neighborhood in this research context represents a topic of interest, not a spatial territory. We are concerned with individuals’ shifting perceptions of place, and not their physical movements between or within neighborhoods.

The aforementioned caveats notwithstanding, our study shows social media can uncover historically unobservable trends in neighborhood dynamics. Moreover, the persistent, real-time nature of social media can reveal these trends in a

timelier fashion than conventional neighborhood-level data. Given the potential to shape the practice of planning as it relates to neighborhood change, is the type of analysis undertaken here one that could be replicated in the typical planning department? Our exercise was conducted with commonly available university computing infrastructure, using software readily accessible and analytical tools of the type now taught in many planning schools. Graduates of planning, data science, or other social science programs with an analytics focus would likely have the capability to carry out a project like the one described here. Most municipal planning departments in medium-sized cities or larger, and most professional planning consulting firms, would have the capacity to implement our methodology.

This leaves what is perhaps the most important obstacle to undertaking a project such as the one described here—data accessibility. Data from Twitter, like most social media platforms, are ultimately collected by for-profit organizations, and thus the future accessibility of social media data for use by public-serving planners cannot be guaranteed. Urban planning, as a discipline with a professed intent to serve the public interest, certainly has a stake in the accessibility of social media data. If the only entities with access to data such as those extracted for our study are ones who collect data with a profit motive, this could create, and reinforce, power imbalances with respect to the control of data and information to plan and develop our cities. At least two features of social media data should give us pause regarding the wisdom of simply following a proprietary data approach: first, although private concerns are expending resources to collect the data, the data itself are often viewed as semi-private by users, who face numerous obstacles to understanding how their data will be used. Confronted with the daunting task of trying to make sense of terms of service and privacy agreements, many simply agree so as to gain access to the media of choice (Obar and Oeldorf-Hirsch 2018). Therefore, most users are only vaguely aware of the different ways social media data can be used, and by whom.

A second reason for concern beyond the way the data are being used is the fact that control and ownership of such data are coalescing into just a handful of large corporations. Social media are natural monopolies because their reach and utility is greater when everyone is on the same platform. Consequently, social media data are not only privately controlled for the most part, but under the control of relatively few. Given the wide range of potential urban planning uses, leaving such data solely within the realm of private entities operating for profit may therefore exacerbate power differentials between social media users and the firms that control their data and limit public sector innovation.

This work contributes to the growing body of evidence illustrating the myriad ways social media data could be used to understand neighborhood change, among other urban planning-related phenomena. The increasing global

adoption of social media suggests that these data will continue to provide a significant source of granular information on human mobility, behavior, and sentiment. The NPI presented here provides one such indicator to provide planners with additional evidence to support policy decisions and to mitigate the potential negative effects of rapid neighborhood change. The utility of these data sources makes it all the more imperative for urban planners to fully participate in the beginning debates over the way social media data will be collected and curated, ensuring that data can be used widely in the public interest.

Acknowledgments

We would like to thank Crimson Hexagon for providing the social media data used in this study, and the anonymous reviewers for their thoughtful and constructive feedback. All errors remain our own.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This material is based upon work supported by the National Science Foundation under Grant No. 1653772.

ORCID iD

Yuan Lai  <https://orcid.org/0000-0003-0664-5048>

Supplemental Material

Supplemental material for this article is available online.

References

- Afzalan, Nader, and Jennifer Evans-Cowley. 2015. "Planning and Social Media: Facebook for Planning at the Neighbourhood Scale." *Planning Practice & Research* 30 (3): 270–85.
- Anderson, Noel S., Alessandro Busà, Evrick Brown, Jennifer Candipan, Phyllis Conn, Roberta Cordeau, William DiFazio, et al. 2012. *The World in Brooklyn: Gentrification, Immigration, and Ethnic Politics in a Global City*. Lanham: Lexington Books.
- Balduini, Marco, Emanuele Della Valle, Daniele Dell'Aglio, Mikalai Tsytsarau, Themis Palpanas, and Cristian Confalonieri. 2013. "Social Listening of City Scale Events using the Streaming Linked Data Framework." In *International Semantic Web Conference*, edited by H. Alani, L. Kagal, A. Fokoue, P. Groth, C. Biemann, J. X. Parreira, L. Aroyo, N. Noy, C. Welty, and K. Janowicz, 1–16. Berlin: Springer.
- Blank, Grant, and Christoph Lutz. 2017. "Representativeness of Social Media in Great Britain: Investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram." *American Behavioral Scientist* 61 (7): 741–56.
- Bollen, Johan, Huina Mao, and Alberto Pepe. 2011. "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-

- Economic Phenomena.” Paper presented at the Proceedings of the International AAAI Conference on Web and Social Media, vol. 5 (1), Barcelona, Spain, July 17–21.
- Chapple, Karen, and Miriam Zuk. 2016. “Forewarned: The Use of Neighborhood Early Warning Systems for Gentrification and Displacement.” *Cityscape* 18 (3): 109–30.
- Crawford, Kate. 2009. “Following You: Disciplines of Listening in Social Media.” *Continuum* 23 (4): 525–35.
- Crimson Hexagon. 2018. “Crimson Hexagon.” Accessed January 1, 2019. <https://www.crimsonhexagon.com>.
- Daas, Piet J. H., and Marco J. H. Puts. 2014. “Social Media Sentiment and Consumer Confidence.” No. 5. ECB Statistics Paper. <https://www.econstor.eu/bitstream/10419/154640/1/ecbsp05.pdf>.
- Efthymiou, Dimitrios, and Constantinos Antoniou. 2012. “Use of Social Media for Transport Data Collection.” *Procedia-Social and Behavioral Sciences* 48:775–85.
- Ellen, Ingrid Gould, and Katherine O’Regan. 2010. “Welcome to the Neighborhood: How can Regional Science Contribute to the Study of Neighborhoods?” *Journal of Regional Science* 50 (1): 363–79.
- Fan, Weiguo, and Michael D. Gordon. 2014. “The Power of Social Media Analytics.” *Communications of the ACM* 57 (6): 74–81.
- Glaeser, Edward L., Hyunjin Kim, and Michael Luca. 2017. “Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity.” No. w24010. National Bureau of Economic Research. <https://www.nber.org/papers/w24010>.
- Hargittai, Eszter. 2020. “Potential Biases in Big Data: Omitted Voices on Social Media.” *Social Science Computer Review* 38 (1): 10–24.
- Hawelka, Bartosz, Izabela Sitko, Euro Beinart, Stanislav Sobolevsky, Pavlos Kazakopoulos, and Carlo Ratti. 2014. “Geo-Located Twitter as Proxy for Global Mobility Patterns.” *Cartography and Geographic Information Science* 41 (3): 260–71.
- He, R., Lee, W. S., Ng, H. T., & Dahlmeier, D. 2019. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. arXiv preprint arXiv:1906.06906.
- Hollander, Justin B., Erin Graves, Henry Renski, Cara Foster-Karim, Andrew Wiley, and Dibyendu Das. 2016. *Urban Social Listening: Potential and Pitfalls for Using Microblogging Data in Studying Cities*. London: Springer.
- Hopkins, Daniel J., and Gary King. 2010. “A method of Automated Nonparametric Content Analysis for Social Science.” *American Journal of Political Science* 54 (1): 229–47.
- Kallas, Priit. 2018. “Top 15 Most Popular Social Networking Sites and Apps.” *Dreamgrow*. Accessed January 1, 2019. <https://www.dreamgrow.com/top-15-most-popular-social-networking-sites/>.
- Kaplan, Andreas M., and Michael Haenlein. 2010. “Users of the World, Unite! The Challenges and Opportunities of Social Media.” *Business Horizons* 53 (1): 59–68.
- Kavanaugh, Andrea L., Edward A. Fox, Steven D. Sheetz, Seungwon Yang, Lin Tzy Li, Donald J. Shoemaker, Apostol Natsev, and Lexing Xie. 2012. “Social Media Use by Government: From the Routine to the Critical.” *Government Information Quarterly* 29 (4): 480–91.
- Kovacs-Gyori, Anna, Alina Ristea, Clemens Havas, Bernd Resch, and Pablo Cabrera-Barona. 2018. “# London2012: Towards Citizen-Contributed Urban Planning through Sentiment Analysis of Twitter Data.” *Urban Planning* 3 (1): 75–99.
- Luo, Feixiong, Guofeng Cao, Kevin Mulligan, and Xiang Li. 2016. “Explore Spatiotemporal and Demographic Characteristics of Human Mobility via Twitter: A Case Study of Chicago.” *Applied Geography* 70:11–25.
- Mahmud, Jalal, Jeffrey Nichols, and Clemens Drews. 2014. “Home Location Identification of Twitter Users.” *ACM Transactions on Intelligent Systems and Technology (TIST)* 5 (3): 1–21.
- Mason, Lorna. 2011. “Mapping a Changing Brooklyn, Mapping a Changing World: Gentrification, Immigration and Outmigration in Brooklyn 2000–2008.” Paper presented at Annual Meeting of the Eastern Sociological Society, Philadelphia, February 24–27.
- Morstatter, Fred, Jürgen Pfeffer, and Huan Liu. 2014. “When Is It Biased? Assessing the Representativeness of Twitter’s Streaming API.” In *Proceedings of the 23rd International Conference on World Wide Web*, edited by Chin-Wan Chung, 555–56. New York: ACM.
- Obar, Jonathan A., and Anne Oeldorf-Hirsch. 2020. “The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services.” *Information, Communication & Society* 23 (1): 128–47.
- O’Brien, Daniel Tumminelli, Robert J. Sampson, and Christopher Winship. 2015. “Econometrics in the Age of Big Data: Measuring and Assessing ‘broken windows’ using Large-Scale Administrative Records.” *Sociological Methodology* 45 (1): 101–47.
- Oh, Sanghee, and Sue Yeon Syn. 2015. “Motivations for Sharing Information and Social Support in Social Media: A Comparative Analysis of Facebook, Twitter, Delicious, YouTube, and Flickr.” *Journal of the Association for Information Science and Technology* 66 (10): 2045–60.
- Poorthuis, Ate. 2018. “How to Draw a Neighborhood? The Potential of Big Data, Regionalization, and Community Detection for Understanding the Heterogeneous Nature of Urban Neighborhoods.” *Geographical Analysis* 50 (2): 182–203.
- Qualman, Erik. 2012. *Socialnomics: How Social Media Transforms the Way We Live and Do Business*. Hoboken: John Wiley & Sons.
- Sawicki, David S., and Patrice Flynn. 1996. “Neighborhood Indicators: A Review of the Literature and an Assessment of Conceptual and Methodological Issues.” *Journal of the American Planning Association* 62 (2): 165–83.
- Schweitzer, Lisa. 2014. “Planning and Social Media: A Case Study of Public Transit and Stigma on Twitter.” *Journal of the American Planning Association* 80 (3): 218–38.
- Shelton, Taylor, Ate Poorthuis, and Matthew Zook. 2015. “Social Media and the City: Rethinking Urban Socio-Spatial Inequality Using User-Generated Geographic Information.” *Landscape and Urban Planning* 142:198–211.
- Statista. 2018. “Most Popular Mobile Social Networking Apps in the United States as of July 2018, by Monthly Users (in Millions).” Accessed January 1, 2019. <https://www.statista.com/statistics/248074/most-popular-us-social-networking-apps-ranked-by-udience/>.
- Stentoft, Aike A., Ate Poorthuis, Bu-Sung Lee, and Markus Schläpfer. 2018. “The Canary in the City: Indicator Groups as Predictors of Local Rent Increases.” *EPJ Data Science* 7:1–15.

- Stieglitz, Stefan, and Linh Dang-Xuan. 2013. "Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior." *Journal of Management Information Systems* 29 (4): 217–48.
- The City of New York. 2009. "PlaNYC Progress Report 2009." http://www.nyc.gov/html/planyc/downloads/pdf/publications/planyc_progress_report_2009.pdf.
- Wojcik, Stefan, and Adam Hughes. 2019. *Sizing up Twitter Users*. Pew Research Center. <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>.
- Zillow. 2018a. "Zillow." Accessed January 1, 2019. <https://www.zillow.com/>.
- Zillow. 2018b. "Zillow Neighborhood Boundaries." Accessed January 1, 2019. <https://www.zillow.com/webtools/labs/neighborhood-boundaries.htm>.
- Zillow Research. 2019. "Zillow Home Value Index." Accessed January 1, 2019. <https://wp.zillowstatic.com/3/ZHVI-InfoSheet-04ed2b.pdf>.
- Zukin, Sharon, Scarlett Lindeman, and Laurie Hurson. 2017. "The Omnivore's Neighborhood? Online Restaurant Reviews, Race, and Gentrification." *Journal of Consumer Culture* 17 (3): 459–79.

Author Biographies

Constantine E. Kontokosta, PhD, is an associate professor of urban science and planning and director of the Civic Analytics program at the New York University (NYU) Marron Institute of Urban Management. He also directs the Urban Intelligence Lab and is associated faculty at the Center for Urban Science and Progress (CUSP) and the Department of Civil and Urban Engineering at the NYU Tandon School of Engineering. His research interests include urban informatics and computation for urban sustainability, resilience, and public health.

Lance Freeman, PhD, is a professor in the Urban Planning program at Columbia Graduate School of Architecture, Planning and Preservation (GSAPP). His research focuses on affordable housing, gentrification, ethnic and racial stratification in housing markets, and the relationship between the built environment and well-being.

Yuan Lai, PhD, is a lecturer in urban science and planning at Massachusetts Institute of Technology. His research interests include urban science, urban informatics, and future connections between computer science and cities to address urban socio-technical complexities.