Where to catch 'em all? – a geographic analysis of Pokémon Go locations

Levente Juhász and Hartwig H. Hochmair

Geomatics Program, University of Florida, Fort Lauderdale, FL, USA

ABSTRACT

In 2016, Niantic Labs released Pokémon Go, an augmented reality smartphone game that attracted millions of users worldwide. This game allows users to “catch” Pokémons through their mobile cameras in different geographic locations that often correspond to prominent places. This paper analyzes the distribution of PokéStops, Pokémon gyms, and spawnpoints in selected urban areas of South Florida and Boston. It identifies which socioeconomic variables and land-use categories affect the density of PokéStops, and how PokéStops and gyms cluster relative to each other. Using nearest neighbor analysis, this paper assesses also how actual PokéStop locations are reflected in Yelp’s “PokéStop nearby” attribute. Results show that black and Hispanic neighborhoods are disadvantaged when it comes to crowd-sourced data coverage, that PokéStops occur more frequently in commercial, recreational and touristic sites and around universities, and that PokéStops tend to cluster around gyms. The latter suggests that these point sets were generated by a similar location selection process. To mitigate geographically linked biases, future versions of augmented reality and geo-games should aim to make them equally accessible in all areas, for example by placing extra resources, such as points of interest, in neighborhoods that are currently underrepresented in data coverage.

1. Introduction

Recent years experienced an increased popularity of augmented reality (AR) applications and location-based games embedded in different devices, such as smartphones or tablets (Johnson et al. 2011; Neustaedter, Tang, and Judge 2013). AR is often used for educational purposes as it offers new learning opportunities through combining computer-assisted contextual layers with relevant real-world information (de Lucia et al. 2012; Wu et al. 2013). Pokémon Go, released in July 2016, is a location-based AR game that quickly became the most downloaded smartphone app on both Android and iOS in history (BBC 2016; Reisinger 2016). The game allows anyone with a smartphone to collect virtual Pokémon characters on the screen which appear to be positioned at the same location as the player. To do so, the player needs to navigate to certain points on a map. These points are often placed at prominent places, such as landmark buildings or statues. Locations are based on the crowd-sourced data-set of an earlier AR game called Ingress which was mostly collected by male, tech-savvy players, leading to a concentration of virtual landmarks in commercial and downtown areas and fewer in non-white or residential areas (Akhtar 2016). Pokémon Go users can directly interact with three different point data-sets in the game, which are PokéStops, Pokémon gyms, and spawnpoints. PokéStops are points in the geographic space, associated with a landmark, such as a building or monument. Players need to visit these locations and perform some actions, such as flipping a coin. In return, players are rewarded with items, such as a Pokéball, which they need to capture Pokémons later. Gyms are virtual locations where Pokémons can be trained and that are also associated with landmarks. Pokémons can “pop-up” at distinct locations (spawnpoints) during the game for 30 min. During this time-window, users have the chance to visit these locations and catch the specific Pokémon, which, upon capturing, will appear in the inventory of the player. Since all these locations provide a player with opportunities to perform beneficial activities for their game, an area with a higher density of PokéStops, gyms, or spawnpoints means also more advantages to the player (Colley et al. 2017).

Pokémon Go was in the center of media attention for an extended period of time with several media articles reporting issues and findings about the game. Residents of certain neighborhoods reported that they felt that there was a smaller number of PokéStops in their areas than in other areas of the same city. The lack of PokéStops in these disadvantaged neighborhoods eventually prevented these residents from effectively participating in the game. Figure 1 visually supports these inequalities of Pokémon-related point densities.
between metropolitan Miami downtown area (Figure 1(a)) and the nearby municipality of Hialeah, which has a significant Hispanic population (Figure 1(b)). The difference becomes most obvious for PokéStops and gyms.

Using a set of 600 gyms, over 5000 PokéStops, and over 18,000 spawnpoint locations within parts of South Florida, including Miami, and Boston, one of the goals of this study is to analyze how these Pokémon-related point locations cluster in geographic space and, more specifically, what kind of demographic or land-use bias persists. Quantifying the latter would help players to adjust their strategy of participating in a game by becoming more effective in finding the related point of interest (POI).

There is evidence that other services and local businesses try to utilize the popularity of Pokémon Go by organizing Pokémon-related special offers or developing some functionalities in their sites that would be of interest to players. For example, vape shops began to monetize Pokémon Go by announcing that their store was a PokéStop, or for giving away a prize for the best Pokémon caught in a shop (Kirkpatrick et al. 2017). Another example is Yelp, which is a location-based service that gives users the ability to rate, review, and browse businesses, such as restaurants and stores. In July 2016, Yelp added a functionality to its smartphone apps and website where users could mark businesses with a “PokéStop nearby” attribute. This attribute was incorporated into the Yelp search functionality, so that users could find businesses in the proximity of PokéStops and therefore combine two leisure activities, which are Pokémon hunting and eating out in a restaurant. This process of accessing crowd-sourced information from one platform (i.e. PokéStops in Pokémon Go) and transferring it to another (Yelp) has been referred to as “cross-viewing” (Juhász and Hochmair 2017). In this study, we assess the correctness of Yelp’s “PokéStop nearby” attribute by comparing the spatial proximity of PokéStop-labeled Yelp businesses to their nearest PokéStops with the spatial proximity of all Yelp businesses to their nearest PokéStops. This approach will demonstrate if and how AR applications affect crowd-sourced data platforms.

Based on the previous considerations, the following four research objectives can be formulated:

- **R1**: Identify the effect of socioeconomic factors and land-use categories on the number of PokéStops in a census block group;
- **R2**: For PokéStops, gym locations, and spawnpoints, quantify the point distributions across different land-use categories;
- **R3**: For PokéStop and gym locations, compare their relative spatial clustering patterns;
- **R4**: Determine the quality of the “PokéStop nearby” attribute on Yelp.

The remainder of the paper is structured as follows. The next section reviews previous work on Pokémon Go user activities as well as the distribution of Pokémon

![Figure 1. Pokémon-related point data-sets (a) around Downtown Miami and (b) in Hialeah.](Image)
Go locations and crowd-sourced POIs. This is followed by a description of the study setup in Section 3, and a presentation and discussion of analysis results in Section 4. Section 5 summarizes the findings and provides directions for future work.

2. Previous work

Several media sources reported a new phenomenon around the release date of Pokémon Go, showcasing masses of people in public places watching their smartphones while walking around. Indeed, Pokémon Go was found to motivate people to go outside and become more physically active (McCarty 2016; Nigg, Mateo, and An 2017; Xu et al. 2017). According to Clark and Clark (2016), the quick update of the Pokémon Go app demonstrates that health promotion should include a social dimension. The authors note that academic research oftentimes lacks the fast pace of technological developments in mobile industry, and that collaborations are needed between academia and industry to develop future apps for key populations, such as patients with chronic diseases or poor mental health. Besides having positive effects on people’s health, Pokémon Go encourages players to explore new areas, which can even lead to the identification of new species (Nature 2016). A related example is the recent detection of a new species of the pygmy devil grasshopper through a photo on Facebook (Skejo and Caballero 2016). The fact that millions of users are participating in this game is also appealing for business owners and other services since they can generate revenue from this. For example, Yelp allowed its users to tag businesses near PokéStops which, in turn, allows Yelp users to search for businesses to visit in the proximity of PokéStops. Kondamudi, Proto, and Alhoori (2017) conducted a multi-year comparison of the number of Pokémon points, we selected 17 hexagon-shaped study areas with different neighborhood characteristics, such as downtown areas, suburban, touristic, and rural/agricultural areas. Figure 2 shows the study areas in South Florida (Figure 2(a)) and Boston (Figure 2(b)).

3. Study setup

3.1. Study areas

The focus of this research was on selected urban areas in South Florida and Boston, MA. Since visual inspection suggested that the underlying geography affects the number of Pokémon points, we selected 17 hexagon-shaped study areas with different neighborhood characteristics, such as downtown areas, suburban, touristic, and rural/agricultural areas. Figure 2 shows the study areas in South Florida (Figure 2(a)) and Boston (Figure 2(b)).

3.2. Data extraction

3.2.1. Pokémon Go locations

Pokémon Go is a commercial product that does not provide an open application programming interface (API) for data access on its backend. However, a discussion started on reddit (https://www.reddit.com/r/pokemon-godev) resulted in open source software libraries that accessed the communication flow between the smartphone apps and Niantic servers, enabling information extraction. We used the pgoapi (https://github.com/tejado/pgoapi) and PokemonGo-Map (https://github.com/scottstamp/PokemonGo-Map) libraries along with multiple accounts (player profiles) in different geographic areas (Figure 2), to log on to the game programmatically and to obtain Pokémon, PokéStops, and gym locations. The method used from these libraries
3.2.3. Supplemental data sources

The analysis for R1 was conducted at the US Census Block Group level. This aggregation level was chosen as it avoids too many zero-count areas with regard to Pokémon locations (as would have been the case with smaller census blocks) while still providing a sufficient sample size and a detailed enough spatial granularity to capture local variability in socioeconomic and land-use-related variables between analyzed areas (which would not have been the case with, for example, larger census tracts). For each block group, 2016 projections of percentage (%) of African-American and percentage (%) of Hispanic population were obtained from the Business Analyst of Environmental Systems Research Institute, Inc. (Esri) (Esri 2016). The presence of parks and higher education institutions in block groups was extracted from the OSM OverpassAPI and coded as dummy variables. More specifically, parks were extracted through

simulates the behavior of a smartphone user who moves around the city. This process is illustrated in Figure 3 where cyan circles represent zones that were scanned in each hexagon-shaped area. Our agents (one for each hexagon) started in the middle of the hexagon and systematically spiralled outwards through all zones. The entire scan process for a hexagon was optimized so that it was faster than the duration that a Pokémon is visible (30 min) at a spawnpoint. This increases the chance to record a Pokémon that popped up. Scanning an area once provides a snapshot of currently visible Pokémon, which is only a subset of all available spawnpoints. To overcome this limitation, our agents were continuously collecting data between late July and early August, 2016, and unique spawnpoint locations were extracted from a large set (137,917) of Pokémon encounters. PokéStops and gym locations are static, therefore a complete list can be obtained with a single scan. We stored all locations of PokéStops and gyms as well as Pokémon encounters in SQLite databases. The final data-set contains 600 gyms, 5017 PokéStops, and 18,257 spawnpoint locations spread across South Florida and Boston.

3.2.2. Yelp businesses

Yelp provides an API to access its services and data, which was used to extract Yelp business information within bounding box queries (Juhász, Rousell, and Arsanjani 2016), where the API returns up to 20 businesses for each query request. To build a complete data-set of Yelp businesses within a chosen geographic area, our algorithm inserted locally refined bounding boxes whenever this download threshold was reached, resulting in a geographically nested sequence of queries (Figure 4). This means that areas with a higher density of businesses, such as strip malls, required a refined grid pattern of bounding boxes to obtain all business locations within the original bounding box. In an additional step, we extracted all businesses tagged with the “PokéStop nearby” attribute using a query filter.

Figure 2. Study areas in (a) South Florida and (b) Boston, MA.

Figure 3. Example of scanned area for extracting Pokémon Go point data-sets in Miami Beach.

3.2.3. Supplemental data sources

The analysis for R1 was conducted at the US Census Block Group level. This aggregation level was chosen as it avoids too many zero-count areas with regard to Pokémon locations (as would have been the case with smaller census blocks) while still providing a sufficient sample size and a detailed enough spatial granularity to capture local variability in socioeconomic and land-use-related variables between analyzed areas (which would not have been the case with, for example, larger census tracts). For each block group, 2016 projections of percentage (%) of African-American and percentage (%) of Hispanic population were obtained from the Business Analyst of Environmental Systems Research Institute, Inc. (Esri) (Esri 2016). The presence of parks and higher education institutions in block groups was extracted from the OSM OverpassAPI and coded as dummy variables. More specifically, parks were extracted through
transportation category in Boston combines roads and infrastructure, so that all Pokémon Go features falling onto streets or transportation infrastructure will fall into the transportation and none in the road category. We decided to keep road and infrastructure land-use categories separated for Miami (and not to aggregate them into one class) to obtain more refined information about the prevalence of Pokémon Go features in different land-use classes.

4. Result analysis

4.1. Analysis methods

Different analysis methods were applied for the different research objectives. The relationship between PokéStop counts and socioeconomic and land-use-based factors (R1) was explored using a negative binomial regression model. Prevalence of Pokémon Go features in different land-use categories (R2) was analyzed using a relative count index and subsequent chi-square tests of independence. Cross K-functions were used to analyze the relative clustering between PokéStops and gyms (R3), and nearest neighbor distances were used to analyze the clustering of PokéStops around Yelp businesses (R4).

4.2. R1 – Relationship between PokéStop counts and neighborhood variables

A negative binomial regression model (NBRM) that relates PokéStop counts in census block groups with ethnicity, race, and land-use-based factors was developed through a manual stepwise approach where variables were added and removed in an exploratory manner to improve the model fit (measured by the Akaike information criterion) while avoiding multicollinearity between predictor variables. The number of Panoramio photos within a block group was used as a proxy for tourist activities, and the number of Yelp businesses per census block group was included as a proxy for economic activity. Block group area was used as a control variable. Other predictor variables, such as population count, median age, median household income, number of jobs, or overlap with central business district were also considered in the manual stepwise approach. However, these were non-significant and therefore not shown in the final results. It is possible that with a larger sample size of observations and a larger variation of attribute values of predictor variables some of the non-significant variables turn significant, e.g. by extending the analysis area. This could, however, not be tested due to a limited Pokémon Go data-set available for our analysis. Results of the final model estimation are listed in Table 1, where the p-values indicate significance of the corresponding coefficient at $p < 0.001$.

Table 1 shows that parks and universities are associated with an increase in PokéStops. This is in-line with

Figure 4. Scanning for Yelp businesses in part of Miami-Dade County, FL.
newspaper articles reporting that Pokémon Go is played by college crowds (university) (Parry 2016) or for recreational purposes (parks) (Grande 2016; Khalid 2016). Areas with business opportunities and tourist activities are also found to be positively associated with PokéStop numbers, supporting earlier notions about a higher density of crowd-sourced points for Ingress in commercial, downtown, and higher income areas (Akhtar 2016). Areas with a higher percentage of African-American and Hispanic population had a fewer PokéStops, supporting the notion of “redlining”, which describes a community being cut off from essential services based on its racial or ethnic group (Kooragayala and Srini 2016).

### 4.3. R2 – Counts of Pokémon Go points on land-use categories

This research question examines in detail how different land-use categories affect the abundance of PokéStops, gyms, and spawnpoints. Using study areas that are combined for the Miami-Dade County and Boston regions, point counts on different land-use categories were compared to count numbers that can be expected under complete spatial randomness (CSR). The expected count number for a land-use category is computed as the total number of points in a region multiplied by the proportion of area covered by the land-use category in question. Table 2 juxtaposes observed and expected point counts for the three point types of Pokémon Go points. Chi-squared tests of independence (lower portion of Table 2) were performed to determine if there was a significant difference between the observed and expected counts.

<table>
<thead>
<tr>
<th>Land-use</th>
<th>Area (km²)</th>
<th>PokéStop obs. (exp)</th>
<th>Spawnpoint obs. (exp)</th>
<th>Gym obs. (exp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(a) Miami-Dade County, FL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agricultural</td>
<td>13.4</td>
<td>1 (96.1)</td>
<td>179 (881.2)</td>
<td>0 (13.4)</td>
</tr>
<tr>
<td>Commercial</td>
<td>5.5</td>
<td>169 (39.3)</td>
<td>1,216 (360.4)</td>
<td>19 (5.5)</td>
</tr>
<tr>
<td>Industrial</td>
<td>6.1</td>
<td>10 (43.9)</td>
<td>290 (402.6)</td>
<td>2 (6.1)</td>
</tr>
<tr>
<td>Natural</td>
<td>4.8</td>
<td>11 (54.4)</td>
<td>194 (315.4)</td>
<td>2 (4.8)</td>
</tr>
<tr>
<td>Other</td>
<td>14.3</td>
<td>22 (102.0)</td>
<td>178 (935.5)</td>
<td>2 (14.2)</td>
</tr>
<tr>
<td>Public</td>
<td>6.6</td>
<td>237 (47.3)</td>
<td>800 (400.7)</td>
<td>42 (6.6)</td>
</tr>
<tr>
<td>Recreational</td>
<td>5.6</td>
<td>124 (40.1)</td>
<td>801 (367.4)</td>
<td>23 (5.6)</td>
</tr>
<tr>
<td>Residential</td>
<td>39.9</td>
<td>102 (284.9)</td>
<td>4,169 (2,611.9)</td>
<td>10 (39.7)</td>
</tr>
<tr>
<td>Roads</td>
<td>24.8</td>
<td>325 (175.9)</td>
<td>1,161 (1,621.7)</td>
<td>41 (24.6)</td>
</tr>
<tr>
<td>Transportation</td>
<td>2.2</td>
<td>37 (15.5)</td>
<td>131 (142.3)</td>
<td>4 (2.2)</td>
</tr>
<tr>
<td>Water</td>
<td>18.4</td>
<td>2 (131.0)</td>
<td>374 (1,200.9)</td>
<td>0 (18.2)</td>
</tr>
<tr>
<td>Water (open)</td>
<td>4.2</td>
<td>2 (29.7)</td>
<td>52 (272.0)</td>
<td>0 (4.1)</td>
</tr>
<tr>
<td>Total obs.</td>
<td></td>
<td>1041</td>
<td>9545</td>
<td>145</td>
</tr>
<tr>
<td>Pearson's Chi-squared test</td>
<td>719.3</td>
<td>2779.5</td>
<td>114.5</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td></td>
<td>&lt;2.20E−16</td>
<td>&lt;2.20E−16</td>
<td>&lt;2.20E−16</td>
</tr>
</tbody>
</table>

| **(b) Boston, MA** |            |                     |                       |               |
| Agricultural      | 0.0        | 0 (0.9)             | 1 (2.6)               | 0 (0.1)       |
| Commercial        | 6.8        | 589 (290.6)         | 1,249 (809.9)         | 35 (22.9)     |
| Industrial        | 2.1        | 64 (87.4)           | 156 (243.6)           | 7 (6.9)       |
| Natural           | 1.3        | 26 (55.0)           | 93 (153.4)            | 3 (4.3)       |
| Other             | 0.5        | 22 (20.1)           | 50 (58.0)             | 2 (1.64)      |
| Public            | 6.0        | 544 (255.6)         | 1,112 (712.1)         | 41 (20.1)     |
| Recreational      | 2.2        | 178 (91.8)          | 362 (235.9)           | 21 (7.2)      |
| Residential       | 17.2       | 305 (727.3)         | 1,806 (2,026.7)       | 25 (57.2)     |
| Roads             | –          | –                   | –                     | –             |
| Water             | 1.4        | 8 (58.7)            | 62 (163.6)            | 1 (4.6)       |
| Water (open)      | 3.3        | 17 (139.0)          | 162 (387.2)           | 0 (10.9)      |
| Total obs.        | 1,857      | 5,175               | 146                   |               |
| Pearson's Chi-squared test | 557.3   | 504.6               | 42.7                  |               |
| df                | 10         | 10                  | 10                    |               |
| p-value           |            | <2.20E−16           | <2.20E−16              | 5.70E−06      |
Table 2) were performed to examine the association between land-use and point count, which was found to be significant (p < 0.0001) for PokéStops, gyms, and spawnpoints in both analyzed regions. This means that point counts differ significantly from expected counts on different land-use categories.

To illustrate how Pokémon Go points are over- or underrepresented in different land-use categories, a relative count index (Equation (1)) was calculated as follows:

\[
c(O, E) = \begin{cases} 
0, & \text{if } O + E = 0 \\
 \frac{O - E}{O + E}, & \text{if } O + E > 0
\end{cases}
\]

where \(c\) is the relative count index; \(O\) is the number of observed points falling in a land-use category; \(E\) is the expected number of points falling in that land-use category. The relative count index values range between +1 and −1 (exclusive) where a positive \(c\) means over-representation (i.e. more observed points than expected under CSR) and a negative \(c\) means the opposite. Figure 5 shows the relative count index for PokéStops, gyms, and spawnpoints for the 12 aggregated land-use categories in both study regions. Patterns of over-/under- representation are similar in the three examined point categories and in both analyzed regions.

Three of the 12 land-use categories (commercial, public, and recreational) are overrepresented for all point sets both in Miami-Dade and Boston. For PokéStops, this overrepresentation resembles some of the significant positive coefficients of the NBRM estimation (Section 4.2., # of businesses = commercial, Park = recreational, University = public), suggesting that these land-use categories are indeed the ones where Pokémon Go can be played most effectively. As opposed to this, most Pokémon Go points are underrepresented in agricultural, industrial, natural, residential, and water land-use categories in both cities, reflecting that these are areas with fewer PokéStops, gyms, and spawnpoints.

4.4. R3 – Spatial clustering of Pokémon Go point data-sets

Visual inspection of the study sites and results from the land-use analysis suggest that Pokémon Go point data-sets are spatially clustered. Since PokéStop and gym locations were allegedly generated from the same crowdsourced data-set, it can be hypothesized that these two point groups are similarly clustered throughout both regions, meaning that there is no clustering of PokéStops relative to gym locations and the other way round. This is also suggested by similar relative count index patterns observed for PokéStops and gyms (see Figure 5).

To determine whether PokéStops and gyms cluster similarly the bivariate version of Ripley’s \(K\)-function, known as Cross \(K\)-function (Dixon 2002), can be used. The Cross \(K\)-function can be formulated as:

\[
K_{ij}(r) = \lambda^{-1}E[f(r)]
\]

where \(f(r)\) is the number of type \(j\) events within a distance \(r\) of a randomly chosen type \(i\) event; \(\lambda\) is the density of \(j\) events per areal unit. Under random labeling, \(K_{ij}(r) = K_{ji}(r) = K_{jj}(r)\), where in the context of this paper \(i\) and \(j\) stand for PokéStop and gym, or the other way around. Statistical inference of the difference between the observed Cross \(K\)-function and a Cross \(K\)-function generated by random labeling can be achieved through Monte Carlo simulation. We analyze the cluster behavior of gyms around PokéStops, using gyms as event type \(j\) and PokéStops as event type \(i\) in Equation (2). Within each of the 999 permutations of the Monte Carlo simulation, events were randomly labeled as either PokéStop...
or gym (retaining their observed proportions) and the Cross K-function was calculated. This established and upper and lower simulation envelope for random labeling at a 99.9% confidence level. The Monte Carlo simulation was run for 15 study sites in South Florida and Boston (2 rural sites did not have PokéStops) for distances up to 5000 m. Results show that the observed Cross K-functions fall within the simulation envelope for the whole distance range for most study sites. This implies that PokéStops and gyms are similarly clustered around each other, that no attraction or repulsion between both point types are present, and that these point groups were indeed generated by similar spatial processes.

There are, however a few study sites where the observed Cross K-function falls slightly below the lower simulation envelope (Figure 6). In these areas, gyms and PokéStops are further apart from each other than it would be expected under random labeling at a 0.001 level of significance. For a Pokémon Go player, this means that slightly longer trips may be necessary to cover activities that involve both PokéStops and gyms, compared to other areas that contain the same number of PokéStops and gyms but do not show this effect.

### Table 3. Descriptive statistics of nearest neighbor distances between Yelp businesses and PokéStops.

<table>
<thead>
<tr>
<th>Distance measurement</th>
<th>Mean (m)</th>
<th>Median (m)</th>
<th>SD (m)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tagged Yelp businesses – PokéStops</td>
<td>52.8</td>
<td>38.1</td>
<td>58.1</td>
<td>896</td>
</tr>
<tr>
<td>All Yelp businesses – PokéStops</td>
<td>102.5</td>
<td>57.8</td>
<td>129.7</td>
<td>10,970</td>
</tr>
</tbody>
</table>

![Figure 6. Computed Cross K-function (black line) with simulation mean (red dashed line) and confidence envelopes (gray area) for random labeling using a Monte Carlo simulation with 999 permutations.](attachment:image.png)

4.5. **R4 – Pokémon-related user tagging of Yelp businesses**

This analysis examines to which extent Yelp users tagged businesses with the “PokéStop nearby” attribute, and how reliable this crowd-sourced information is. Within all analyzed study polygons areas, Yelp users tagged 1392 businesses with the “PokéStop nearby” attribute out of 21,606 total Yelp businesses, revealing that Yelp’s strategy to attract potential customers through targeting Pokémon Go players seemed to work up to a certain extent. To determine whether user tagging of businesses is correct or not, we computed two sets of nearest neighbor distances, namely those measured between businesses tagged with the PokéStop attribute and their nearest PokéStops, and between all Yelp businesses and their nearest PokéStops. For this analysis, only PokéStops within the study site hexagons were considered as potential nearest neighbors. Descriptive statistics for both sets of distance measurements are shown in Table 3, and Figure 7 plots the corresponding histograms for the two distance sets.

Both Table 3 and Figure 7 suggest that businesses tagged with the “PokéStop nearby” attribute are indeed situated closer to a PokéStop than all Yelp businesses. A Mood's median test was performed to determine whether differences between medians were significant between both distance groups, and results confirmed this at a high level of significance ($p < 2.2E-16$). This means that the tagging behavior of Yelp users on this attribute is not random, but that the data contributors tend to annotate this information correctly. A higher tagging intensity of Yelp businesses can be observed in metropolitan areas (Miami Beach: 12.0%, downtown Boston: 12.5%, downtown Miami: 9.7%) than in suburban areas (Hialeah, FL: 1.2%, Homestead, FL: 5.4%, suburban Boston: 2.3%). Two rural areas (Redlands, FL and Immokalee, FL) and two suburban areas (Davie, FL and Brownsville, FL) did not have any businesses tagged with this attribute. Two intertwined explanations for this discrepancy in tagging completeness rates...
Our study also analyzed the interplay of augmented reality gaming and VGI, suggesting that the Pokémon Go user community participates in crowd-sourcing activities, namely adding information to the "PokéStop nearby" attribute on the Yelp business platform. Nearest neighbor analysis suggests that this tagged information tends to be correct, and that it can be used by visitors of the Yelp website to identify businesses that are located near a PokéStop.

The presented research supports earlier findings of a strong geographic and socioeconomic bias in the Pokémon Go data-set. As this bias can affect the user experience in location-based games negatively, future developments and improvements of location-based AR games should address this issue and provide equal access to interactive platforms such as Pokémon Go to all user communities.

Notes on contributors

Levente Juhász is a PhD candidate at the University of Florida where he focuses his research efforts on VGI. He holds a master's degree in Geography from the University of Szeged, Hungary. He is especially interested in contribution patterns of VGI and social media users as well as in how user-generated data are used across different platforms. Previously, he worked as a GIS developer in Hungary, then was a short-term visiting scientist at the Joint Research Centre in Ispra, Italy, before starting his doctoral studies at the University of Florida. He also acts as a data scientist for a geospatial startup, Mapillary, and is an avid contributor of OpenStreetMap and other open data projects.

Hartwig H. Hochmair is an associate professor of Geomatics at the University of Florida where he teaches courses in GIS, Digital Mapping, Adjustment Computations, and Geodesy. He focuses on quality assessment of crowd-sourced data, route planning, wayfinding, and the analysis of transportation networks and travel behavior with a focus on bicycle and public transportation. As part of his interdisciplinary research, he analyzes the spread of invasive species in Southeast Florida, including termites and tegus. His educational background includes Geodesy and Geoinformation, and he obtained his PhD degree from the Technical University of Vienna, Austria.
References


