

COMPARING TWO METHODS FOR URBAN COMPLEXITY CALCULATION USING THE SHANNON-WIENER INDEX

JESÚS LÓPEZ BAEZA^{*1,3}, DAMIANO CERRONE^{†2}, KRISTJAN MÄNNIGO^{‡3}

¹University of Alicante, Spain

²Tampere University of Technology, Finland

³SPIN Unit, Estonia

ABSTRACT

This study will compare the results of measuring Urban Complexity using the Shannon-Wiener index in two different methods. Using a joint dataset retrieved from Foursquare API, we will measure the degree of urban complexity of every street: 1) relating every amenity to the closest street segment in a computational way and then applying the calculation to the segments; and 2) applying the calculation to every cell of a grid that will be combined with the street network afterwards. The selected case study is the city of London, and the dataset employed will be retrieved from Foursquare. Over 79,000 venues were collected and classified in over 660 categories. In order to proceed to the analysis, these 660 categories will be reduced to 10 based on the classification of activities observed in the public space from the traditional urban discipline. Then the urban complexity index of each street segment of London will be measured as a simultaneous calculation of the density and diversity of collected and classified economic activities.

Keywords: urban complexity, Shannon-Wiener, social media, London.

1 INTRODUCTION

The use of social media data when analysing the city is nowadays becoming popular in the urban planning discipline [1]. GIS technologies allow us not only to visualize but also operate with data retrieved from social networks and online servers. Furthermore, due to the popularization of social media among smartphone users, data samples are greater in surface and in number of points than any other data sources so far [2], [3] and they have a higher spatial and temporal resolution [2].

In this paper, we aim to study the intangible relations between city form and the distribution of activities hosted by amenities. We speculate that compact urban forms foster a higher variety of amenities capable to host a mix of different activities, which are beneficial to life in the city. Perceivably dense urban areas with small narrow streets -such as European old towns- generally cluster a high amount of venues with a high level of complexity.

Scholars have attempted to measure the urban complexity index through the application of standardized indexes. Specifically, this article uses the Shannon-Wiener index taking points relative to venues registered in Foursquare for the calculation. This data source has been utilized through the so-called *Grid* calculation method, capable of determining an index of complexity for a given area divided in equal cells [4]. Notwithstanding, we aim to measure the complexity within the street. Thus, in this paper we will obtain a value of urban complexity attached to every street segment of the street network of London city refining the calculation for a given surface (two dimensional) into a unit of length (one dimension). To do so, we will test two different methods: Closest Point (CP) and Grid method.

* ORCID: <http://orcid.org/0000-0002-4092-1782>

† ORCID: <http://orcid.org/0000-0002-4853-9887>

‡ ORCID: <http://orcid.org/0000-0002-6592-9588>



We expect to obtain similar results. However, when measuring with Closest Point (CP) calculation these venues would be divided in smaller groups –relating every one of them to the closest street- and the complexity value would be lower. On the contrary, sprawl fabric with a low density of amenities might obtain a low value of complexity when measured through Grid calculation, but a higher value when measured through CP calculation, since points would be linked to *the closest street* without taking a maximum distance into account.

1.1 Place-based social media

We can define place-based social media as those social networks that record and store the contents shared on their platforms and the geographic location they were shared from. This data can often be accessed by third parties from API services. Since our research focuses on the metadata regarding the content's location rather than the personal information of users, our methodology is about which place people are posting from rather than who is posting.

The use of place-based applications in spatial planning research is generally linked to the study of landscape or in particular, to observe a particular area in order to quantify the impact of certain factors [5]–[7] to describe intangible urban phenomena.

Place-based social media such as Google Places has been used by urban studies scholars as a tool to measure urban complexity [8], [4], [9] due to its ability to represent the totality of businesses registered in Google with geographic coordinates.

In this way, place-based social media offer a virtual representation of the heterogeneity of the economic activities of a certain area, constantly updated and revised by the users themselves. In addition to location, they offer other information such as name, address, and business category [9], allowing the classification of activities according to type. Although this data can be easily accessed and it is storing the exact geographic coordinates of businesses, the reliability on the content itself it may be hard to verify.

1.2 Foursquare

In this study, we are going to focus on Foursquare, a social media platform that allows users to add “Venues” worth visiting and expressing preferences. The app also allows “check-ins” for users who intend to state their presence in a specific venue. Originally Foursquare was designed to be used as a game, where users with most check-ins in one venue could earn badges and popularity [10]. While the Foursquare mobile app still allows users to check-in and review venues, the ludic component has been removed from the service. In the past years, venues were mostly trendy places such as bars, restaurants and clubs, but only recently Foursquare's coverage became broader and more detailed. Today the list of venues is no longer limited to few urban amenities but to the overall set of locations, places and spaces present in one city - including open areas, infrastructures and buildings hosting real estate or small businesses.

Foursquare is understood by Komninos et al. [11] as a collective knowledge tool through which users use information from third parties when making decisions about areas of interest in the urban context [11] such as how to spend their free time (32.5%), what venues it's worth visiting (27.7%), and when to go to visit the selected venues (26%) [12]. There are several studies that discuss Foursquare users' location choice and their relation to the game-like interface capable of influencing individual's decision making in the city [13]–[15]. In addition to the progress that users can make by checking-in venues and leaving reviews, users' behaviour is an aspect of the individual spatial self, which can be interpreted as a representation of lifestyle in order to create or maintain social relationships [16].



The base of this platform is a series of public or private places listed by users themselves, in which anyone can register a visit by checking-in [17]. These places are classified through categories that are predetermined by the application. Thus, data obtained from the searches through this app are likely to be filtered [18].

The main implementation of Foursquare in the field of urbanism is framed in the ability to (1) list places and register them with a digital identification and coordinates, and (2) adding metadata to every registered place. Specifically, when measuring user preferences for both Foursquare application and the physical space of the city, we will place particular emphasis on data about the number of visitors and check-ins. In other words, Foursquare, with its particular game-like interface allows the user to reveal where he/she has been and how many times. In this way, the information obtained from this app allows “*to understand human mobility in its spatial, temporal, social, and content aspects*” [2, p. 781].

There are several articles that relate the use of the Foursquare to the observation of the diversity of economic activities in a given area [18] listing registered places and observing distribution patterns.

The dataset retrieved from the Foursquare API corresponds to every point obtained in the selected query area would correspond to a physical place (facility, venue, amenity) in the urban environment registered by the app. In this way, for each place we obtain the following data:

1. Spatial coordinates Latitude and Longitude
2. Business Name Identification
3. Category and Sub-categories
4. Address, street and number
5. Average rating
6. Number of visits or different users who have visited the site (one visit per user)
7. Number of check-ins or sum of all visits that all users have made on-site.

2 CATEGORIES OF PLACES

The geographic distribution of amenities is topical to the definition of spatio-temporal patterns of urban and rural settlements. Numerous studies have revealed the spatial relations that govern the location choice of these establishments although it is primal to map and understand these relations in their socio-cultural context [19], [20]. In particular we are going to focus on clustering and dispersal phenomena to learn how the spatial relations between socio-economic establishments are typical to the urban and regional contexts of one city.

Gehl et al. [21] makes a classification of observed activities performed by people within the urban space. These activities were classified by nature in Optional, Necessary and Social being Optional those performed when one has free time (gym, museum, spa, etc.), Necessary those integrated in the daily routine (school, office, market, etc.) and Social would include those activities that involve interaction with others (bar, café, club, etc.) [22].

In this context, we may understand Social activities as a part of those performed when one has free time, so we included them into the category Optional. Thus, we firstly divide all activities in these two major groups, Necessary and Optional.

In order to be able to operate with categories in a more detailed way and define uses, it is needed to define activity types more accurately, but at the same time to let these new categories be flexible enough to include all kind of Foursquare venues in them. To do so we use the SPIN Unit Urban activity Wheel [22].

The novelty of this approach is the classification of amenities not by type or function but according to the spectrum of human activities they can host. In this way, we are going to

perform an indirect observation of activity patterns that can be scaled from local to regional scale – something that is traditionally performed with long terms and micro-scale only surveys a field observation.

Following this discourse, we created categories according to the SPIN Unit Urban activity Wheel (Fig. 1).

This classification method was created from empirical observations of people performing activities in public and private spaces to study the difference between what the city offers (Supply) through its amenities and what people do in the city (Demand). This method is designed to categorize Foursquare venues to map and measure the Supply and other datasets such as Instagram and Twitter to measure human activity patterns in the city (Demand). In this paper, we reclassified Foursquare data only to measure the complexity of activity patterns offered by the city and study if their distribution has a clear relation to the perceived form of the city.

In this paper Over 79,000 venues were collected and classified in over 660 categories. In order to proceed to the analysis, these 660 categories were reduced to 10 based on the *SPIN Unit Urban Activity Wheel* (Fig. 2).

3 URBAN COMPLEXITY

The concept of Complexity in urban studies starts to be related to the economic activities in urban spaces registered in online applications and web services in the mid-2010s [9]. However, a standardized methodology to measure it has not been developed yet. To do so, first of all, it will be necessary to define this concept and identify methods through which it can be observed and measured.

Other previous authors such as Ruiz Sánchez [23, p. 4] have defined the city as “*a complex self-regulating system*” in which all the elements that shape it in their multiple scales are related to each other in a complex network, in which the minimum modification in one of its elements will be a readjustment on a global scale. Blanco and Subitrats [24] associate this heterogeneous complexity with “*a real and potential conflict*” an effect of increasing size and consequent complexity of cities, making them unable to be managed [24], or becoming a challenge per se for policy makers [25].

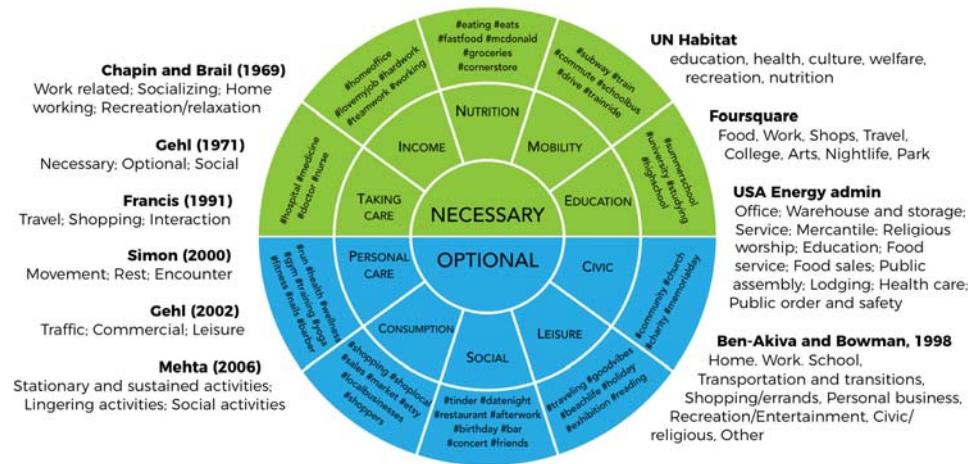


Figure 1: SPIN Unit Urban activity Wheel describing activities in urban space [22].



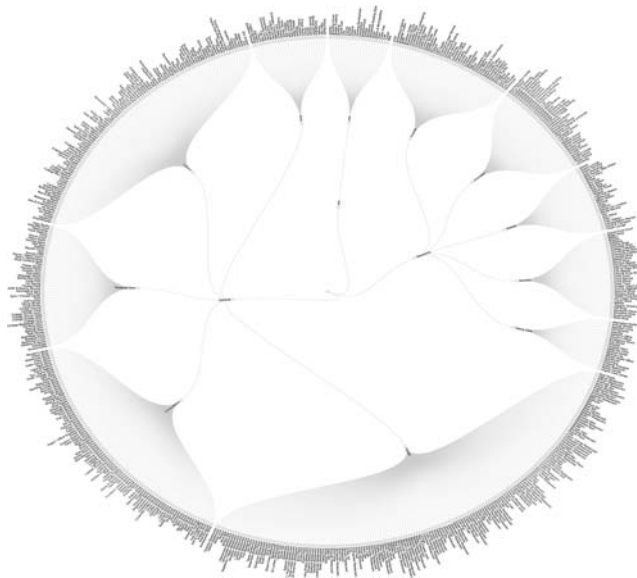


Figure 2: Categories in Foursquare re-classified into *Urban Activity Wheel* of activities in urban space. (Self-elaborated.)

From a pragmatic point of view, Rueda [26] refers to complexity as “*the measure of the degree of organization of the urban system*” in terms of “*the diversity of the mix of uses and services*” [9, p. 5]. Complexity defined as the set of relations between all elements that constitute the city, the “complex self-regulating system” shows the ecosystemic nature of the same. For this reason, Rueda [26] states that this degree of complexity can be obtained through the application of the equation corresponding to the Shannon-Wiener complexity index developed to measure the degree of entropy of an ecosystem [27].

Florida [28] will later relate this urban complexity of cities to economic growth through what he calls the “creative class”, which he places in specific spaces characterized by tolerance, openness, and creativity of business [28]. To this end, he elaborates a series of indicators capable of measuring particular aspects that directly influence what he calls “social complexity”.

Within the field of urban studies and in relation to social networks and web services, Nolasco-Cirugeda and García-Mayor [4] use data collected from Google Places to evaluate the functional urban complexity of economic activities that directly affect urban space [4, p. 4]. They concluded that the functional appeal linked to the tourism experience and indicators of sustainability in its case study – Benidorm, Spain – is proportional to the degree of complexity through the application of the Shannon-Wiener index. Therefore, they also corroborate the possibility of applying this index as a truthful indicator of urban complexity.

Bustos Hernández [8] also deals with complexity within the field of urbanism, drawing relationships between it – and concepts such as density, diversity or specialization – with specific cases of urban regeneration. Like Nolasco-Cirugeda and García-Mayor [8], he relates the improvement of the space experience with the increase of activities and relations resulting from the increase of urban complexity [8].

Therefore, we could define of urban complexity as a simultaneous measure of both the density and diversity of economic activities in a given urban space, which is directly related

to spatial experience and can be measured quantitatively through the application of the Shannon-Wiener diversity index.

4 METHODOLOGY

The following section will explain the two methods used for calculating urban complexity by applying the Shannon-Wiener index using location data retrieved from Foursquare. The first method (Grid method) divides the land into a grid of cells and calculates for each cell an index of complexity that will later be transferred into every street segment contained in the cell. The second method (Closest-Point method) relates every geolocation to the nearest street segment and calculates the complexity index for the street segment per unitary length value. The results obtained from both methods will be interpreted and visualized through a GIS platform from which maps will be plotted.

The formula employed for calculation is given below, where: H' = Urban complexity index; S = number of categories; p_i = number of amenities with category 'i' divided by number of amenities.

$$H' = -\sum_{i=1}^S p_i \log_2 p_i. \quad (1)$$

4.1 Grid method

After every point is re-categorized using the *Urban Activity Wheel* of categories as a reference, points are laid out on a map using the location data stored as metadata in form of geographic coordinates. Subsequently, the area (the city of London in our case) is divided into a grid where each cell has dimensions of 100x100 meters. The formula used does not refer directly to any unit of area or longitude. This grid is the basis for comparing cells with others with one same dimension. In this paper, it has been taken the value of 100x100 meters as unitary taking as a precedent the work of Nolasco-Cirugeda and Garcia Mayor [4]. Hence, Foursquare points are bounded inside every grid cell boundary with their category values. If no points were found in a cell, it would be removed from further calculation and assigned a null value. As noted previously, the formula takes into account the number of categories and the number of items for each category inside every cell. After running the Shannon-Wiener index calculation, the values obtained are transferred from every cell to every street segment comprised within that cell. The results obtained will mirror the value of urban complexity in every street segment per unitary measure of area.

4.2 Closest Point (CP) method

The closest point has accessibility information involved, although not precise all the time.

The closest street to any venue is not always the one from which one can access. However, it is true in most of the cases. Therefore, urban complexity index calculated through CP method in an ideal scenario, would measure the level of complexity of one street segment regarding all the venues to which one can access from that segment.

Similarly, as described above, we firstly need to re-classify the categories and plot de dataset in the map through the geographical coordinates attached to every Foursquare point as metadata. The street network with which we will operate should be larger than the area covered by Foursquare points to avoid *edge* effect in which points located on the edge or near the edges are finding their closest street only from inside of the clipping boundary whereas the actual closest street could be located outside the operation area. Firstly, we calculate the closest street segment for every Foursquare point in the dataset.

The results of this calculation are displayed for every street segment containing the information about the amount of Foursquare points related to it. In similarity with the previous method described, all streets unrelated to any point are removed from the calculation and assigned a null complexity value. Due to the variety of length between segments, the complexity value should be divided by the length of the segment in order to obtain a unitary value. Thus, the Shannon-Wiener formula is applied for every street segment and the results obtained will represent the value of complexity per unitary measure of length.

5 RESULTS

After the application of both calculation methods, the results obtained have been represented in two comparative maps. The case study represented is the city of London, UK.

The center of the city confined between Regent's park, Hyde Park and Victoria park, together with the south bank of the Thames presents a higher urban complexity index than the rest of the study area in both cases. In addition, a series of commercial axes arranged radially present the high presence of establishments of commercial activity. This axis is common to both results and they conform a network in which the nodes (i.e.: Elephant & Castle station, Kennington park, Cramberwell Green park) have an exceptionally high rate of urban complexity.

Quantitatively, we observe how the results obtained from the Grid method represent more uniform values with a variation between 0.31 and 3.05, whereas values obtained from CP method vary between 0.3 and 67.6. However, the unconventionally high values happen in specific locations, which will be later discussed in the "Discussion" section.

Qualitatively, the GRID method provides uniform values by proximity to the location, while the CP method presents the value of each segment independently of the nearest ones. In this way, the GRID method calculation colored spots with the same complexity value are observed due to the superficial nature of the calculation at origin, while in the CP method larger discontinuities and differences between neighboring streets are observed since each segment is analyzed independently.

6 DISCUSSION

Both calculation methods display similar overall results. However, due to the different nature of them (one-dimension vs two-dimensions) each of them contains certain bias, which are described below:

6.1 Grid Bias: data loss

CP method avoids all data loss from calculation. There are no venue points lost from the case study area, since every point is related to the closest street without taking this distance into account. On the contrary, Grid calculation method misses all points located in areas without any street segment capable of representing the value of the grid. An example would be a park with an area over 100x100 meters surrounded by streets. Even if there were any venues inside the park, it would not be measured by any grid that at the same time includes a street segment, so the park would appear as empty and the streets would represent a lower complexity value.

6.2 CP Bias: edges and cul-de-sacs

CP method produces a bias in which some small segments collect a large amount of venues. Those segments have almost no length and they are generally plotted as edges or *cul-de-sacs*. In these cases, these small segments would collect an exceptionally high amount of venues





Figure 3: Calculation of Shannon-Wiener index per street segment using Grid Method. *(Self-elaborated.)*



Figure 4: Calculation of Shannon-Wiener index per street segment using CP Method. *(Self-elaborated.)*

because of the way in which the geometry of the network is plotted. Due to their petite length, these segments would maintain an exceptionally high level of complexity after unifying the value.

There is no such problem with Grid method, since it does not take into account the length of the street but it is essentially independent of any geometry.

7 CONCLUSIONS

This paper proved that complexity of urban space can be measured through Shannon-Wiener index and represented along every linear street segment within the street network of a city. Both methods tested – CP and GRID method – seem relevant and offer a unique representation of the city. Being these two methodologies unique, each of them present a different bias that the other could solve. However, the overall story told by both results is consistent and coinciding.

REFERENCES

- [1] Lazzarini, L. & López Baeza, J., “The Mushrooms’ Lesson: Instagram as a tool to evaluate users’ perception of urban transformations.” *Un nuovo ciclo della pianificazione urbanistica tra tattica e strategia*, Roma-Milano: Planum Publisher, pp. 178–184, 2016.
- [2] Kheiri, A., Karimipour, F. & Forghani, M., “Intra-urban movement flow estimation using location based social networking data.” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, **40**(1W5), pp. 781–785, 2015.
- [3] Silva, T.H., Vaz De Melo, P.O., Almeida, J.M. & Loureiro, A.A., “Large-scale study of city dynamics and urban social behavior using participatory sensing.” *Wirel. Commun. IEEE*, **21**, pp. 42–51, February 2014.
- [4] Nolasco-Cirugeda, A. & García Mayor, C., “Aplicación de los indicadores de complejidad urbana a través de las redes sociales y TIG: El caso de los paseos marítimos de Levante y Poniente en Benidorm.” *XVI Congreso Nacional de Tecnologías de la Información Geográfica. Tecnologías de la información para nuevas formas de ver el territorio*, 2014, pp. 995–1004.
- [5] Boardman, J., “The value of Google Earth™ for erosion mapping.” *Catena*, **143**, pp. 123–127, 2016.
- [6] De Vries, S., Buijs, A.E., Langers, F., Farjon, H., Van Hinsberg, A. & Sijtsma, F.J., “Measuring the attractiveness of Dutch landscapes: Identifying national hotspots of highly valued places using Google Maps,” *Appl. Geogr.*, vol. 45, pp. 220–229, 2013.
- [7] Pulighe, G. & Lupia, F., “Mapping spatial patterns of urban agriculture in Rome (Italy) using Google Earth and web-mapping services.” *Land use policy*, **59**, pp. 49–58, 2016.
- [8] Bustos Hernández, M., “Análisis de la complejidad urbana en la ciudad turística: El caso de La Pineda (Vila-seca, Tarragona).” *Int. Conf. Reg. sicence Innov. Geogr. spillovers New approaches Evid.*, 2015.
- [9] Tomás Castelló, M.P., *La clase creativa y la complejidad urbana de actividades en Barcelona. Un análisis a través de datos de redes sociales* [dissertation] Universidad de Alicante, 2016.
- [10] Saker, M. & Evans, L., “Everyday life and locative play: an exploration of Foursquare and playful engagements with space and place.” *Media, Cult. Soc.*, 2016.
- [11] Komninos, A., Stefanis, V., Plessas, A. & Besharat, J., “Capturing urban dynamics with scarce check-in data.” *IEEE Pervasive Comput.*, **12**(4), pp. 20–28, 2013.



- [12] Gretzel, U. & Yoo, K.H., "Use and Impact of Online Travel Reviews." *Inf. Commun. Technol. Tour.*, pp. 33–46, 2008.
- [13] Frith, J., "Communicating through location: the understood meaning of the Foursquare check-in." *J. Comput. Commun.*, **19**(4), pp. 890–905, 2014.
- [14] Jin, L., Zhang, K., Lu, J. & Lin, Y.R., "Towards understanding the gamification upon users' scores in a location-based social network." *Multimed. Tools Appl.*, **75**(15), pp. 8895–8919, 2016.
- [15] Schwartz, R. & Halegoua, G.R., "The spatial self: Location-based identity performance on social media." *New Media Soc.*, Apr. 2014. 1461444814531364.
- [16] Barkhuus, L., Brown, B., Bell, M., Sherwood, S., Hall, M. & Chalmers, M., "From awareness to repartee: sharing location within social groups." In *proceedings of the SIGCHI conference on human factors in computing systems*, pp. 497–506, 2008.
- [17] Roick, O. & Heuser, S., "Location based social networks - definition, current state of the art and research agenda." *Trans. GIS*, **17**(5), pp. 763–784, 2013.
- [18] Agryzkov, T., Alvarez, F., Serrano-Estrada, L., Tortosa, L. & Vicent, J.F., "Different Models to Visualise Geolocated City Data from Social Networks." In *Proceedings - 2015 International Conference on Future Internet of Things and Cloud, FiCloud 2015 and 2015 International Conference on Open and Big Data, OBD 2015*, pp. 687–692, 2015.
- [19] Hillier, B., *Space is the machine: a configurational theory of architecture*. London: Space Syntax, 2007.
- [20] Sevtsuk, A., "Path and place: a study of urban geometry and retail activity in Cambridge and Somerville, MA [doctoral dissertation]," Massachusetts Institute of Technology, 2010.
- [21] Gehl, J., Svarre, B., Press, I. & Steenhard, K.A., *How to study public life*, 2013.
- [22] Cerrone, D., Lehtovouri, P. & López Baeza, J., "The urban activity wheel. Using social media data to map activity patterns," 2016.
- [23] Ruiz Sánchez, J., "Sistemas urbanos complejos: acción y comunicación." *Cuad. Investig. urbanística*, **32**, pp. 1–76, 2001.
- [24] Blanco, I. & Subirats, J., "¿Existen territorios socialmente excluyentes? Contra lo inexorable." In *Respuestas locales a inseguridades globales. Innovación y cambios en Brasil y España*, 2009.
- [25] Wagenaar, H., "Governance, complexity, and democratic participation how citizens and public officials harness the complexities of neighborhood decline." *Am. Rev. Public Adm.*, **37**(1), pp. 17–50, 2007.
- [26] Rueda, S., *Ecología Urbana*. Barcelona: Beta Editorial, 1995.
- [27] Begossi, A., "Use of ecological methods in ethnobotany: diversity indices." *Econ. Bot.*, **50**(3), pp. 280–289, 1996.
- [28] Florida, R., "The rise of the creative class." *Wash. Mon.*, **34**(5), pp. 15–25, 2002.

