Could Data from Location-Based Social Networks Be Used to Support Urban Planning?

Rodrigo Smarzaro
Universidade Federal de Viçosa
Rio Paranaíba, MG, Brazil
smarzaro@ufv.br

Tiago França Melo Lima
Universidade Federal de Ouro Preto
Ouro Preto, MG, Brazil
tiagolima@decsi.ufop.br

Clodoveu A. Davis Jr
Universidade Federal de Minas Gerais
Belo Horizonte, MG, Brazil
clodoveu@dcc.ufmg.br

ABSTRACT

A great quantity of information is required to support urban planning. Usually there are many (not integrated) data sources, originating from different government bodies, in distinct formats and variable properties (e.g. reliability, completeness). The effort to handle these data, integrate and analyze them is high, taking to much time for the information to be available to help decision making. We argue that data from location-based social networks (LBSN) could be used to provide useful information in reasonable time, despite several limitations they have. To asses this, as a case study, we used data from different LBSN to calculate the Local Availability Index (IOL) for a Brazilian city. This index is part of a methodology to estimate quality of urban life inside cities and is used to support urban planning. The results suggest that data from LBSN are useful and could be used to provide insights for local governments.

Keywords

LBSN; Urban Planning; Urban Informatics

1. INTRODUCTION

Most of world population lives in cities and many challenges arise from this growing urbanization [23]. Local governments require information about the city in time to support urban planning and decision making processes. The use of metrics and indicators summarizing information helps to measure and monitor performance of aspects such as the availability of services and the quality of urban life.

From the creation and establishment of the Human Development Index (HDI) [22], many efforts have being made to produce indicators that can be used to evaluate the performance of countries, regions and cities. The quality of urban life is a frequent concern in research initiatives [1, 5, 6, 10, 14, 17, 19, 25].

Nahas [12] proposed the Urban Quality of Life Index (IQVU, acronym for *Indice de Qualidade de Vida Urbana* in Portuguese) to spatially quantify inequalities in the supply and access to services by the population. It is composed by ten dimensions (such as education, housing, and urban security) and calculated for city subdivisions, therefore it can be used to establish priorities for public investment. IQVU was initially created and used to support urban planning in Belo Horizonte, a 2.5-million-people city in Brazil.

A major problem for IQVU use is the lack of regularity and timeliness in updating index values. It was calculated for the years 1994, 2000, 2006, 2010 and 2012. This deficiency limits its potential use as an effective tool to support urban planning, and reflects the difficulty for recalculating the index as originally proposed. Some possible causes include differences in temporal granularity among the various sources of information, difficulties in obtaining data from some governmental agencies, methodological changes in data generation, and other political reasons.

We argue that data from location-based social networks (LBSN) can be used as a source for the calculation of metrics and indexes supporting urban planning and decision making in cities. To this end, a case study was developed using data from LBSNs to estimate the Local Availability Index, a component of IQVU that measures the availability of services inside a geographic region. Results suggest that the use of LBSN data to infer quality of urban life indicators is promising, and can lead to the formulation of new metrics, indexes and methods, for instance, allowing the use of *nowcasting* models [2, 3] to support urban planning.

2. BACKGROUND

This section presents some basic concepts related to Quality of Urban Life Index, Location-Based Social Networks and approaches using LBSN data to estimate urban features such as quality of life.

2.1 IQVU and the Local Availability Index

The Quality of Urban Life Index of Belo Horizonte (IQVU-BH) is a multidimensional indicator created by a multidisciplinary group from government and academia. It aims to spatially quantify the inequality of services available for the population, and thereby, to be a tool to support the distribution of public resources [12].

The IQVU calculation method uses a subdivision of Belo Horizonte in regions called Units of Planning (UPs). IQVU is based on the availability and accessibility of facilities and services, classified in ten dimensions (called variables): food supply, culture, education, sports, habitation, urban infrastructure, environment, health, urban services and public security. Georeferenced data from several government agencies

WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3051700

are used to calculate the indicators, which are aggregated into components and variables. For instance, *health centers* (the number of public health facilities in a UP per thousand inhabitants) is an indicator of the *health care* component, which is part of the *health* variable. The steps to calculate the IQVU are shown in Figure 1.

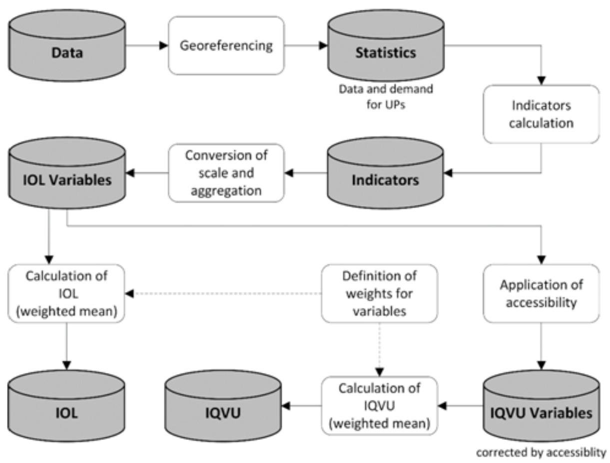


Figure 1: **IQVU Methodology.** Steps to calculate the IQVU index (adapted and translated from [13]).

In short, after data gathering and calculation of indicators, three main results are obtained:

- The **Local Availability Index** (IOL, acronym for *Índice de Oferta Local* in Portuguese) is calculated for each dimension on each UP. First, each indicator's availability is calculated with a specific measure, generally, a count per thousand inhabitants (e.g. number of hospitals / population x 1000). Then, the indicator values are normalized between 0 and 1 by the equation

  $$I_c = 1 - e^{-(f.v)}$$

  where I_c is the normalized indicator value, v is the original indicator value, and f is given by

  $$f = -ln(0.05)/L_{ref}$$

  where L_{ref} is the reference value for the indicator, which is assumed to be the 95% percentile. This adjustment compensates for higher values, so they cause less impact on the index [9]. The IOL of a dimension is given by the simple mean of the normalized value of its indicators, and the IOL for the UP is given by the weighted average of all dimensions.

- The **IQVU Variables** (sector indexes) are calculated by applying an accessibility index on the IOL. The accessibility index is based on an estimate of the time a citizen needs to move between each pair of UPs using public transportation. The accessibility index can increase or decrease the variable's IOL value, based on how easy it is for the population of an UP to access the services available in other UPs. For details on the procedure to incorporate accessibility measures, see [9].

- The **IQVU index** for each region is given by a weighted average of IQVU variables. The weight of each dimension was established by the specialists group responsible for the creation of IQVU.

With IQVU results, local governments can identify which regions need more attention. It has already been used to support decision processes for public investment, but we do not have information on its current use. The index underwent methodological changes and was not published regularly. Thereby, improvements are required in order to use similar methods to effectively help local governments.

In this work we will focus on calculating IOL for two main reasons. First, IOL should describe the availability of services from governmental data, but we perceive differences between the official record of activities and actual functioning businesses. Second, although it is possible to estimate the accessibility using services such as Google Maps, this step was postponed and will be executed in a future work, addressing a new index entirely based on data from LBSNs and other open sources of urban data.

2.2 Location Based Social Networks (LBSN)

Online Social Networks (OSN), like Facebook and Twitter, enable users to create a network of friends and share any content to all or a group of their contacts. Smartphones and other portable devices equipped with Global Positioning System (GPS) and connected to the Internet motivated the creation of Location-Based Social Networks (LBSN) [15]. In LBSNs, geographical location can be shared along with the regular content of interest.

Zheng (2011) [27] uses the role location plays on LBSNs to classify them into three groups: geo-tagged-media-based, point-location-driven and trajectory-centric. Geo-tagged-media-based add the location to the media being shared (e.g. photos, videos and tweets). Flickr, Youtube and Twitter can all be considered LBSNs in this category. Point-location-driven services encourage users to share their current location in real time. For example, a user can arrive at a shopping mall and share his location and post his opinion about the place. Foursquare and Yelp are examples of point-location-driven LBSNs. Trajectory-centric services enable users to share a route, which is a sequential connection of point locations. Information like distance, speed, duration, altimetry and others the user can provide (such as tags, photos, opinions) are also shared. Sports logging services like Garmin Connect, Nike+ and Strava are representative of this category. Our interest in this work is on point-location-driven LBSNs as sources of POIs in a city.

2.3 Metrics, Indicators and LBSN Data

Many studies aim to use alternative (non-governmental) data sources to investigate urban problems such as deprivation, diversity and availability of services. Venerandi et al.(2015) [24] present a methodology to measure urban deprivation from user-generated content. Using data from Foursquare and OpenStreetMap (OSM), they quantitatively describe neighborhoods using a metric called Offering Advantage, which is then used to infer urban deprivation. Also, they use the UK Index of Multiple Deprivation (IMD), a composite score that is calculated as the weighted means of seven distinct domains (e.g. income deprivation and crime).

Yuan et al.(2012) [26] propose a framework to identify regions performing different functions in urban areas, which is important for planning and predicting city development. Initially, partitions of the selected area, based on its major roads are created. Then, they use human mobility data (taxi trajectories) to infer each partition's functionality (e.g., business, commercial, entertainment). However, obtaining representative data on mobility is a challenge.

Quercia and Saez (2014) focus on ascertaining if social media offers an alternative data source for studying the relationship between resources and neighborhood deprivation. They use Foursquare data from London users and employ classification algorithms to infer land-use information, in order to verify the relationship between socioeconomic deprivation and the presence of specific economic activities.

Shelton et al. (2015) [20] present an extensive analysis of geotagged tweets from Louisville, USA, aiming to provide useful insight for urban planning and geographic research. Their approach combines relational socio-spatial theory and GIScience to show issues like segregation between neighborhoods, mobility and inequality within the city. Although very interesting, an issue is the bias of the study, enhanced by the exclusive use of geolocated Twitter data and the application in a single city, with very specific features.

De Nadai et al. (2016) [4] studied the relationship between urban vitality and diversity for six Italian cities, following the ideas proposed by Jane Jacobs [7]. They used call data records to extract proxies for urban vitality and web data from public and commercial entities to assess urban diversity. Data from census, urban mapping and LBSNs (OSM and Foursquare) were also used. They proposed a set of metrics and used a regression model to evaluate the relationship between structural diversity and activity density. As stated by the authors, one major limitation is the difficulty to fully replicate the study without call records, which are hard to obtain.

3. METHODS

This section describes the process for calculating IOL from LBSN data, including data sources, collection and analysis.

3.1 Official Data Sources

In this work, we used data from three official data sources: (i) the most recent IQVU-BH dataset, (ii) Brazilian Census data, (iii) urban geographic data of Belo Horizonte.

The IQVU-BH dataset was obtained from Belo Horizonte's Web portal [1]. It is available as a spreadsheet that contains the values of indicators, components and variables of IQVU for each UP in 1994, 2000, 2006, 2010 and 2012. The raw data and calculation formulas are not included in the file.

Almost all indicators are normalized by the population of each UP. For that, official demographic data from 2010 were used, as in previous IQVU calculations. Census data provides the population count for each census sector, while IQVU uses Units of Planning (UP), which are spatial aggregations of census sectors. Therefore obtaining demographic data for each UP is straightforward.

3.2 LBSN Data Sources

The LBSNs Facebook Places, Foursquare, Google Places and Yelp were used as data sources, as they maintain publicly available datasets on services and businesses, which can be used as a source for IQVU indicators. They were chosen because they (i) offer a public API that allows data gathering, (ii) contain a large number of users and places, and (iii) allow volunteered data contributions.

3.2.1 Data Collection

In order to collect the full set of services for Belo Horizonte, an iterative procedure was implemented. First, a regular grid of points separated by 25 meters was generated, covering the bounding box of Belo Horizonte's city limits, provided by the city's GIS. Then, a geometric intersection between the grid and the city limits polygon was used to select the points inside the municipal territory. This operation resulted in a set of 530,044 distinct reference points. For each point, a call to the LBSN's API was made using the point's coordinates and a radius of 25 meters as parameters. The radius size of 25 meters was considered small enough to get all the available places for each social network, since a larger radius might cause the APIs to return only a subset of the available places instead of the full list.

Crawlers for each LBSN were implemented using Python and data were collected from October 2 to October 25, 2015. Crawlers had to pace themselves to comply to the API's enforced limit on the number of daily requests. Yelp, for instance, allows 25,000 search requests per day, while Foursquare allows 5,000 requests per hour. The API returns JSON data that were initially stored as a single text file. Then the file was processed to extract, for each location, its ID, latitude, longitude and category.

3.2.2 Data Analysis and Characterization

Data gathering produced a dataset containing a large number of entries for each LBSN. Duplicate entries are collected, since there are intersections between areas generated for each point of the reference grid, and must be eliminated. Furthermore, many places collected from the LBSNs are not relevant to our study, since they are not used to calculate IQVU indicators. For example, the dataset has information on laundries and clothing stores, but these services are not considered in the IQVU methodology. Therefore, it was necessary to match the categories of locations in the datasets with those used by the IQVU. The approach used relied on relating each category offered by each LBSN with one of the IQVU indicators, using manual classification (Table 1).

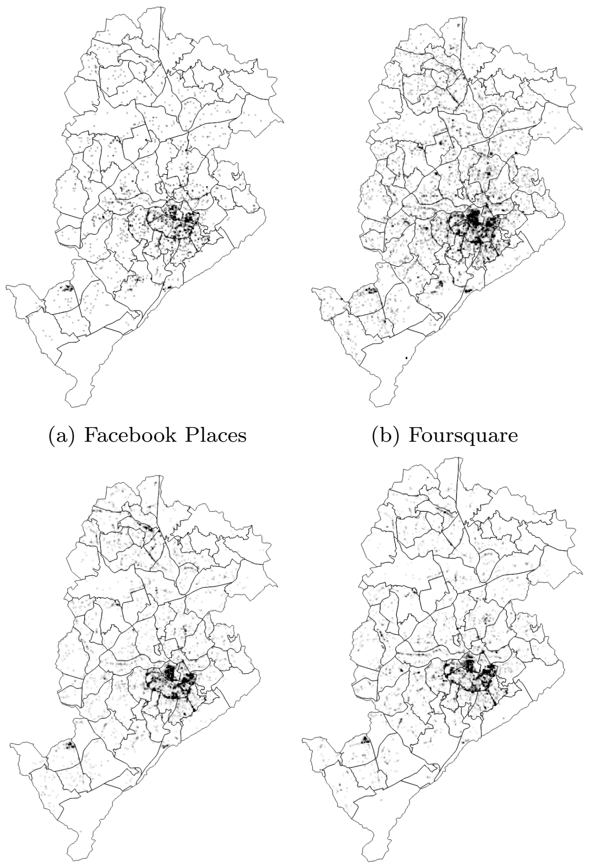
Table 1: Quantity of places (points) of each step on data source cleaning process

Operation on data	Facebook	Foursquare	Google	Yelp
Data Gathering (raw data)	115,137	286,227	1,389,061	354,278
Remove duplicate entries	42,551	91,816	169,814	57,840
Remove entries without compatible category	2,214	6,827	6,456	7,669

Besides the differences in the quantity of unique entries in each LBSN, their spatial distribution is also uneven. Figure 2 shows the spatial distribution of the resulting points from Facebook Places, Foursquare, Google Places and Yelp across Belo Horizonte. There is a concentration in central regions, although some other places also have good coverage. Discrepancies such as these were expected, since downtown concentrates business places, other regions are specialized (e.g., concentrate hospitals), and many are clearly more deprived. Furthermore, the distribution of places among categories is

[1]http://portalpbh.pbh.gov.br/pbh/

also not homogeneous (e.g. the number of markets is much higher than the number of hospitals).



(a) Facebook Places (b) Foursquare

(c) Google Places (d) Yelp

Figure 2: **Unique POIs.** Distribution of collected points over Belo Horizonte for each LBSN.

Some of the data used to calculate IOL and IQVU are not available in the LBSNs, such as the number of registered students in schools and criminal occurrences. Finally, the data availability provided by LBSNs varies, as the distribution of collected places over the indicators is not uniform.

3.3 IOL Calculation

The Local Availability Index (IOL) aggregates information from indicators that measure the availability of services (e.g. supermarkets), into IQVU variables (e.g. food supply). Then, results are weighted according to the importance of each IQVU variable.

To calculate IOL variables, the number of places collected from the LBSNs inside each UP was counted for each indicator. This number is divided by the population of the UP. Since it was not possible to get LBSN data to calculate all indicators of IQVU, data from IQVU 2012 were used as a basis, in order to allow a comparative analysis. For IOL variables calculation, indicators for which there is no LBSN data were replaced by official data. We were able to calculate 15 of the 36 indicators using only data from LBSNs. Using data exclusively from each LBSN, we were able to calculate the IOL for four of the ten IQVU variables. If combined with official data, six of the ten variables can be calculated. Indicator values were normalized between 0 and 1, and then aggregated on IOL variables for each UP by applying a simple average.

4. RESULTS AND DISCUSSION

First, we evaluate IOL indicators calculated using data from LBSNs. For this, values obtained were discretized into the same intervals used to present IQVU official results: [0,0.5), [0.5, 0.6), [0.6, 0.7), [0.7, 0.8), [0.8, 1], labeled "1", "2", "3", "4" and "5", respectively. After classifying the results on the intervals, accuracy, precision and recall were calculated for IOL indicators using a multi-class approach [21], testing all datasets against official 2012 IOL results.

Mixed results for accuracy were obtained, considering indicators and data sources as shown on Table 2. *Green Area* indicator performed quite well in all datasets, with an accuracy varying from 0.772 (GPlaces) to 0.911 (Yelp). *Post Office* indicator also performed well on four datasets, but Gplaces had the worst accuracy (0.658) while Facebook had the best (0.759). *Cultural Equipment* indicator also performed well on Yelp (0.898) and Foursquare (0.746) but not so on Facebook (0.518), and no data were available on that from GPlaces. The *Health Care Services* indicator also performed well on Facebook (0.708), GPlaces (0.746) and Yelp (0.848) but no data were available from Foursquare. *Dental Services* indicator had similar results on Facebook (0.721) and Yelp (0.886), but slightly worse from GPlaces (0.645), and also lacked data from Foursquare. Yelp had the best accuracy results for eight of the indicators considered (see bold values on Table 2. Facebook and Foursquare were best on three indicators each, while GPlaces did not have the best result in any indicator. This results indicate that the datasets can be used to complement each other.

In general, good accuracy was achieved, but results for precision and recall were poor. One possible cause is that intervals "1" and "5" have a wider range of values, taking into account 70% of the possible values, while intervals "2","3" and "4" only answer for 30%. This unbalanced distribution and the five intervals considered could decrease the mean value of recall and precision. This suggests that if a binary classification is used, for example, to identify deprived and well-off regions, better results could be achieved.

In order to allow a visual inspection of the spatial patterns of the results, maps were generated to compare the IOL values for the variables that were calculated using LBSN. Figure 3 shows IOL results of variable *Food Supply.*

4.1 Limitations

This work has three limitations that need to be observed. First, LBSN data can be biased by being crowdsourced. LBSNs and other online social networks data are tied to the profile of their users, generally young and technology-friendly [16]. This aspect can cause the coverage of those locations to be better around neighborhoods where young population live, work or have fun [18]. Not having many locations on a LBSN does not necessarily indicate that the ground truth is likewise deprived. However, as smartphone usage increases, it should be increasingly important for businesses to participate in LBSN catalogs such as the ones used in this work.

The second limitation regards the lack of some types of information from LBSN. It was not possible to obtain all indicator values from the LBSNs used on this work, but this will not necessarily always be true. Currently, collect-

Table 2: Accuracy (acc.), Precision (prec.) and Recall (rec.) of indicators calculated from Facebook, Foursquare, Google Places and Yelp. The Precision and Recall values are the mean of the five classes considered

Indicators	Facebook			Foursquare			GPlaces			Yelp		
	acc.	prec.	rec.	acc.	prec.	rec.	acc.	prec.	rec.	acc.	prec.	rec.
Hyper and supermarket	0.392		0.330	**0.468**	0.415	0.450	0.417	0.324	0.238			
Grocery store and similar	**0.367**	0.231	0.206	0.126	0.112	0.156	0.341	0.146	0.178	0.303	0.199	0.204
Cultural equipment	0.518			0.746		0.582				**0.898**		
Bookstore and stationery	0.405	0.127	0.152	0.531		0.207	0.379	0.107	0.133	**0.544**	0.196	0.220
Movie rental store	0.518	0.262	0.317				0.481	0.345	0.324	**0.544**	0.384	0.374
Magazine stand				0.569	0.260	0.286				**0.582**	0.287	0.323
Sport court, field	0.379	0.274	0.277	0.481	0.437	0.426	0.392	0.314	0.306			
Green area	0.784			0.835			0.772			**0.911**		
Health centers	**0.367**	0.347	0.291	0.227	0.253	0.201	0.278	0.215	0.212	0.227	0.243	0.201
Other health care services	0.708	0.310	0.368				0.746	0.352	0.428	**0.848**	0.487	0.403
Dental services	0.721	0.383	0.425				0.645	0.307	0.312	**0.886**	0.496	0.468
Bank agency	0.481	0.174	0.195	**0.721**	0.295	0.352	0.658	0.405	0.497	0.506	0.158	0.172
Gas station	0.518	0.261	0.339	**0.594**	0.403	0.398	0.367	0.251	0.270			
Drugstore	0.189	0.079	0.106	0.443	0.231	0.299	0.291		0.142	**0.607**	0.536	0.533
Post office	**0.759**	0.330	0.363	0.734	0.385	0.540	0.658	0.422	0.574	0.734	0.287	0.269

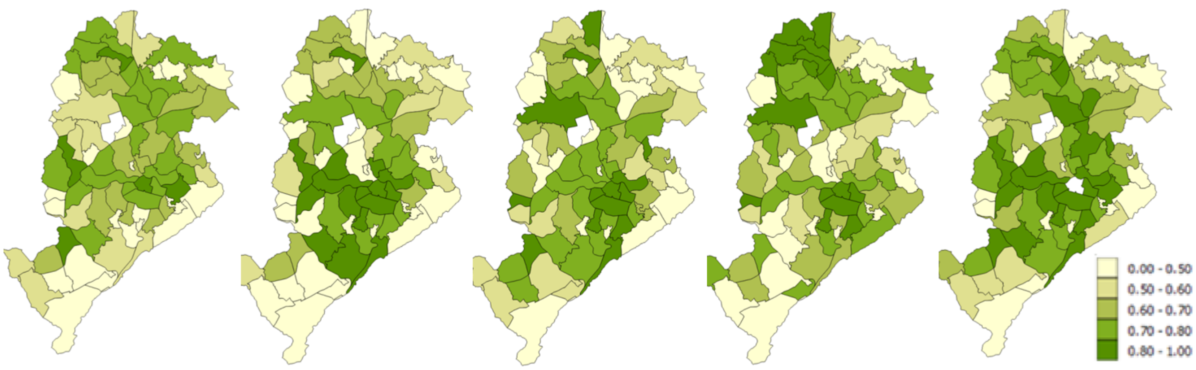


Figure 3: **IQVU Methodology.** Comparison of results for IOL variable Food Supply. From left to right: Official data, Facebook, Foursquare, Google Places and Yelp.

ing the missing data required by this work depends on open data policies in place for government information-producing organizations. The authors believe that this can be fixed with upcoming open government data initiatives and legislation that supports information dissemination in machine-readable formats and services. PDF files containing images of tables and maps do not qualify as such. Active crowdsourcing, in volunteered geographic information initiatives directed at those categories, might fill this gap [11], although new biases may be introduced, as in the first limitation.

The third limitation is related to the lack of more frequently updated results of IQVU for comparison. All results for this work are based on data collected in the third quarter of 2015, but the latest IQVU results are from 2012, so we are comparing current LBSN data with official data that is three years old. Improving the results of this work, and possibly calculating a quality of urban life index with online and easily available social network and official government data may enable us to produce frequent results in the future.

5. FINAL REMARKS

This work shows the potential use of LBSNs as data sources to calculate a quality of life index for the city of Belo Horizonte. Besides the limitations of LBSN data, results encourage the expansion of research work towards improving the quality of data and methods for the calculation of the index.

One way to improve is to enhance the dataset with data from additional LBSNs and with better integration of the collected data. For this purpose, we plan to develop ways to eliminate duplicates inside each LBSN dataset, and to avoid the redundancies that may be introduced by integrating several data sources. This is not an easy task, as sources do not share a unique identification and classification schemes are diverse. One can explore the location and other attributes to check if two locations represent the same business in two or more datasets. The use of well-established economic activity classifications, ontologies, schema mapping and other spatial data integration techniques are under evaluation in our research group. In addition, it would be useful to provide features for visual exploration and spatial querying (e.g. [8]) on the aggregated data.

Also, much data collected were not used because they did not belong to a category required by some IQVU indicator. We argue that some of these can be related to the quality of urban life, and should be considered along with services that are used daily by many people, such as restaurants. On the other hand, some original IQVU indicators seem outdated (for instance, movie rentals and pay phones) to measure quality of urban life. Future work includes the use of volunteered geographic information as a framework to survey popular perception of quality of urban life, and then dynamically measure and adjust the results according to city dynamics and citizen behavior.

6. ACKNOWLEDGMENTS

The authors acknowledge the support of FAPEMIG and CNPq, Brazilian agencies in charge of fostering R&D.

7. REFERENCES

[1] H. Badland, C. Whitzman, M. Lowe, M. Davern, L. Aye, I. Butterworth, D. Hes, and B. Giles-Corti. Urban liveability: emerging lessons from Australia for exploring the potential for indicators to measure the social determinants of health. *Social Science & Medicine*, 111:64–73, 2014.

[2] M. Bańbura, D. Giannone, M. Modugno, and L. Reichlin. Now-casting and the real-time data flow. In G. Elliott and A. Timmermann, editors, *Handbook of Economic Forecasting, vol. 2*, pages 195–236. Elsevier, 2013.

[3] C. Codeco, O. Cruz, T. I. Riback, C. M. Degener, M. F. Gomes, D. Villela, L. Bastos, S. Camargo, V. Saraceni, M. C. F. Lemos, and F. C. Coelho. Infodengue: a nowcasting system for the surveillance of dengue fever transmission. *bioRxiv*, 2016.

[4] M. De Nadai, J. Staiano, R. Larcher, N. Sebe, D. Quercia, and B. Lepri. The death and life of great italian cities: A mobile phone data perspective. In *Proc. of the 25th International Conference on World Wide Web*, pages 413–423. WWW, 2016.

[5] A. A. Gavrilidis, C. M. Ciocănea, M. R. Niţă, D. A. Onose, and I. I. Năstase. Urban landscape quality index–planning tool for evaluating urban landscapes and improving the quality of life. *Procedia Environmental Sciences*, 32:155–167, 2016.

[6] J. F. Helliwell. How's life? Combining individual and national variables to explain subjective well-being. *Economic modelling*, 20(2):331–360, 2003.

[7] J. Jacobs. *The death and life of great American cities.* Vintage, 1961.

[8] C. Kumar, W. Heuten, and S. Boll. Visual overlay on OpenStreetMap data to support spatial exploration of urban environments. *ISPRS International Journal of Geo-Information*, 4(1):87–104, 2015.

[9] M. B. Lemos, O. de Avelar Esteves, R. F. Simões, et al. A methodology for making an urban quality of life index (in Portuguese). *Nova Economia*, 5(2):157–176, 1995.

[10] R. W. Marans. Quality of urban life & environmental sustainability studies: Future linkage opportunities. *Habitat International*, 45:47–52, 2015.

[11] G. V. Mateveli, N. G. Machado, M. M. Moro, and C. A. Davis Jr. Taxonomia e Desafios de Recomendação para Coleta de Dados Geográficos por Cidadãos (in Portuguese). In *Proc. of 30th Brazilian Symposium on Databases*, pages 105–110, 2015.

[12] M. I. P. Nahas. *Theoretical basis, calculation methodology and applicability of intra-urban indicators in the municipal management of the quality of urban life in large cities: the case of Belo Horizonte (in Portuguese).* Phd thesis, Universidade Federal de São Carlos, Jul 2002.

[13] PBH. Relatório geral sobre o cálculo do Índice de qualidade de vida urbana de belo horizonte (iqvu-bh) (in portuguese). Technical report, 2014.

[14] I. A. Pissourios. An interdisciplinary study on indicators: A comparative review of quality-of-life, macroeconomic, environmental, welfare and sustainability indicators. *Ecological Indicators*, 34:420–427, 2013.

[15] T. Pontes, M. Vasconcelos, J. Almeida, P. Kumaraguru, and V. Almeida. We know where you live: privacy characterization of Foursquare behavior. In *Proc. of the 2012 ACM Conference on Ubiquitous Computing*, pages 898–905. ACM, 2012.

[16] D. Quercia and D. Saez. Mining urban deprivation from Foursquare: Implicit crowdsourcing of city land use. *IEEE Pervasive Computing*, 13(2):30–36, 2014.

[17] M. R. Rezvani and H. Mansourian. Developing small cities by promoting village to town and its effects on quality of life for the local residents. *Social indicators research*, 110(1):147–170, 2013.

[18] M. Rost, L. Barkhuus, H. Cramer, and B. Brown. Representation and communication: challenges in interpreting large social media datasets. In *Proc. of the 2013 Conference on Computer Supported Cooperative Work*, pages 357–362. ACM, 2013.

[19] R. Rothenberg, C. Stauber, S. Weaver, D. Dai, A. Prasad, and M. Kano. Urban health indicators and indices—current status. *BMC Public Health*, 15(1):494, 2015.

[20] T. Shelton, A. Poorthuis, and M. Zook. Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, 142:198–211, 2015.

[21] M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, 45(4):427–437, 2009.

[22] United Nations. Human Development Report. http://hdr.undp.org/en/reports/global/hdr1990, 1990.

[23] United Nations. World Urbanization Prospects: The 2014 Revision. https://esa.un.org/unpd/wup/Publications, 2015.

[24] A. Venerandi, G. Quattrone, L. Capra, D. Quercia, and D. Saez-Trumper. Measuring urban deprivation from user generated content. In *Proc. of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 254–264. ACM, 2015.

[25] D. Weziak-Bialowolska. Quality of life in cities–empirical evidence in comparative european perspective. *Cities*, 58:87–96, 2016.

[26] J. Yuan, Y. Zheng, and X. Xie. Discovering regions of different functions in a city using human mobility and pois. In *Proc. 18th ACM International Conference on Knowledge Discovery and Data Mining*, pages 186–194. ACM, 2012.

[27] Y. Zheng. *Location-Based Social Networks: Users*, pages 243–276. Springer, New York, NY, 2011.