

GIS and Data: Three applications to enhance Mobility

Andy S Alic¹, Jussara Almeida², Wagner Meira Jr.², Dorgival Guedes²,
Walter dos Santos², Ignacio Blanquer¹, Sandro Fiore³, Nádia P. Kozevitch⁹,
Nazareno Andrade⁴, Tarciso Braz⁴, Andrey Brito⁴, Carlos Eduardo Pires⁴,
Nuno Antunes⁵, Marco Vieira⁵, Paulo Silva⁵, Danilo Ardagna⁶, Keiko Fonseca⁹,
Daniele Lezzi⁷, Donatello Elia³, Regina Moraes⁸, Tania Basso⁸, Wilian H. Cavassin⁹

¹Universitat Politècnica de València - CSIC

²Universidade Federal de Minas Gerais (UFMG), Brazil

³Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Italy

⁴Universidade Federal de Campina Grande (UFCG), Brazil

⁵CISUC, University of Coimbra, Portugal

⁶Politecnico di Milano, Milan, Italy

⁷Barcelona Supercomputing Center (BSC), Spain

⁸University of Campinas (UNICAMP), Brazil

⁹Universidade Tecnológica Federal do Paraná (UTFPR), Brazil

asalic@upv.es, {jussara,meira,dorgival,walter}@dcc.ufmg.br,
iblanque@dsic.upv.es, sandro.fiori@unisalento.it, nadiap@utfpr.edu.br,
{nazareno,tarciso,andrey}@computacao.ufcg.edu.br, cesp@dsc.ufcg.edu.br,
{nmsa,mvieira,pmsgilva}@dei.uc.pt, danilo.ardagna@polimi.it,
keiko@utfpr.edu.br, danieli.lezzi@bsc.es, donatello.elia@cmcc.it,
{regina,taniabasso}@ft.unicamp.br, wiliancavassin@alunos.utfpr.edu.br

Abstract. *The increasing urban population sets new demands for mobility solutions. The impacts of traffic congestions or inefficient transit connectivity directly affect public health (e.g. emissions and stress) and the city economy (e.g. deaths in road accidents, productivity, and commuting). In parallel, the advance of technology has made it easier to obtain data about the systems which make up the city information systems. This paper takes advantage of GIS and real-time data to present: 1) a web application integrating multiple services; 2) an android application for bus visualization and prediction and 3) a dashboard focused on applying exploratory data analysis techniques on ticketing data.*

1. Introduction

The steady growth of urban centers poses several challenges to human well-being, many of which are associated with urban mobility (traffic jams, longer travel times, health

issues due to emissions, stress, etc.) requiring new approaches to overcome them. Thus, it is necessary to provide new tools for managing a city, reconciling the functioning of several systems so that their performance fit to better serve its inhabitants. In this scenario, the concept of *smart city* has emerged along of its several approaches of smartness [Husár et al. 2017], usually linked to efficiency in the use of natural resources [Souza et al. 2015, Azambuja 2016].

In particular, public transportation is one of the most critical areas of smart cities. In Brazil, the vehicle fleet in major cities grew more than the road structure¹. Mobility challenges have already gained attention of the Computer Science society in Brazil². The efficiency of its performance helps to reduce its operation costs and also provides social integration (such as the use of government applications and crowd-sourcing). Some people have the public transport system as their only mean of displacement among their daily trips [Weigang et al. 2001].

The increasing availability of city open data provides opportunities to explore new applications or innovative data exploitation, along with GIS techniques to enhance the cities mobility. However, processing big volume of raw data in limited time to provide timely information for services with an acceptable quality poses several challenges.

This paper describes three applications (Routes4People, Melhor Busão and City Administration Dashboard) based on GIS, cloud and parallel computing technologies to enhance mobility, not only from the citizen perspective, but also from the perspective of the city administration. The applications were developed under the *EUBra-BIGSEA* project (Europe-Brazil Collaboration of Big Data Scientific Research Through Cloud-Centric Applications), where all developments are available under Open Source licenses³. The rest of the paper is organized as follows: Section 2 presents related work. Section 3 offers an overview of the applications platform. Section 4 details the applications and their features. Finally, Section 5 presents the conclusion.

2. Related Work

Several applications are already available for mobility, such as Crowdbus [Sousa Junior et al. 2014], Bus Brasil⁴, Cadê o Ônibus⁵, Itibus⁶ and Moovit⁷.

Crowdbus uses resources of crowdsourcing technology to provide data about the public transportation in Recife and Maceió where the crowdsourced data are collected by users support and user's smartphones functionalities (e.g., compass, GPS, and accelerometer). Crowdbus uses data from speed, routes to generate a quality standing for each route of public transportation, processing the data to provide, in the future application,

¹<http://www1.folha.uol.com.br/cotidiano/2014/08/1503030-frota-de-veiculos-cresce-mais-rapido-que-a-estrutura-viaria-no-pais.shtml> - Last accessed on Nov 24, 2017.

²<http://www.sbc.org.br/documentos-da-sbc/send/141-grandes-desafios/802-grandesdesafiosdacomputaonobrasil> - Last accessed on Nov 24, 2017.

³<http://github.org/eubr-bigsea> - Last accessed on Nov 24, 2017.

⁴<http://www.busbrazil.com.br/> - Last accessed on Ago 22, 2018.

⁵ <http://www.cadeoonibus.com.br/CoO/SiteV2> Last accessed on Nov 4th, 2017.

⁶<https://www.urbs.curitiba.pr.gov.br/mobile/itibus> - Last accessed on Ago 22, 2018.

⁷ <https://moovit.com/> - Last accessed on Ago 28, 2018.

measurements of time to travel between bus stops [Sousa Junior et al. 2014].

Bus Brasil uses an application for Android smartphone that stores bus time tables from various cities in Brazil. The application provides schedules for the buses, with the closest bus coming from a determined place, such as a Bus Terminal. The application can store data in the device, providing some functionalities while the device is not connected to the Internet.

Cadê o Ônibus was developed in cross-platform modal (Android, iOS and Windows Phone). This application detects the position of buses in real-time, allowing a variable number of search requests by the user. This allows user collaboration (feature that is only showed in the Moovit and the Crowdbus apps), thanks to the use of the real-time tracking can also predict the arrival of the bus on the bus stop.

Itibus is a web-plataform application which provides the schedule of lines, lines by its code or label, itinerary of the lines in a map, real-time location of the bus, near location of bus stops and lines by stops. Besides that, the application can provide news and the balance of the client's transport card.

Moovit is another example of application which uses GIS and processes data from external sources to generate knowledge. The application operates in more than 2,500 cities, with more than 200 million users. Under the concept of Urban Mobility Analysis and Mobility as a Service (MaaS), the system provides a list with buses lines, various types of search, lines by stops, route creation, and buses that accept transport card. The system can predict the arrival and departure times of the lines in stops and terminals.

Compared to these online applications (listed in Figure 1), our three approaches present the following advantages: open source licenses, integration with several data sources (e.g. Waze, twitter, mobility open data) along with Big Data and Cloud services (among others).

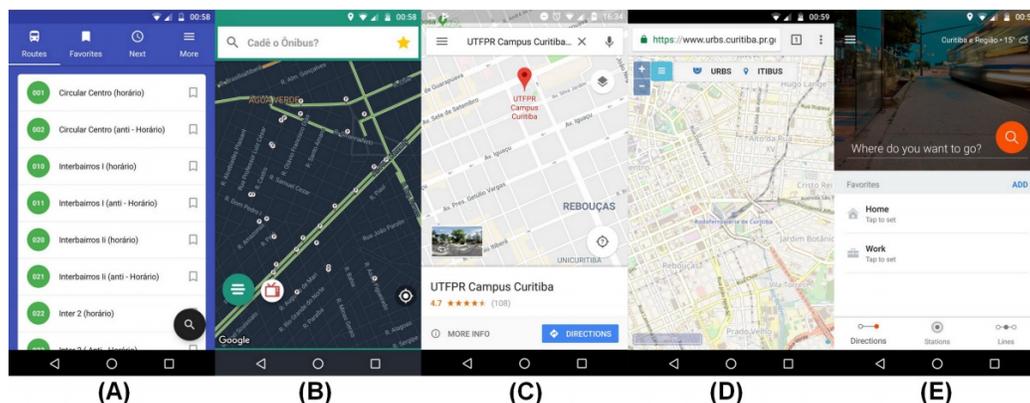


Figure 1. Home Screen of the applications: (A) Bus Brasil. (B) Cadê o Ônibus?. (C) ItiBus. (D) Moovit. Source: [Calandre et al. 2018]

Andrade et al. [Andrade et al. 2014] propose how the combination of open data, geographic information systems and cluster algorithms can bring benefits to urban mobil-

ity. This paper has the time spent in a displacement within the city of Belo Horizonte as a case study. By the cluster algorithm analysis it is possible to determine clusters of any radius and decide the best vehicle to be chosen on a given trip within the city, as well as the best place to be carried out.

In the same way, Monteiro et al. [Monteiro et al. 2017] suggest how the bus stops can be distributed in a way that bus stops are not too far apart from each other. In other words, their algorithm can be used to determine the best places to have a pick off point. To do this, they have used open data from public transportation and a Simulated Annealing algorithm.

Several other works also uses transportation scenarios, but in different contexts, such as location modeling [Li and Tong 2017], optimization [Yang et al. 2000], complex network metrics [da Silva et al. 2016, De Bona et al. 2016], and exploratory analysis [Kozievitch et al. 2016, Vila et al. 2016].

3. Infrastructure Overview

The infrastructure used to support the applications included 19 components (Figure 2). In summary, there are five different layers:

1. Modules for resource configuration, prediction of resource usage, scheduling of jobs and proactive policies for vertical and horizontal elasticity. This layer has the following items:
 - Infrastructure Manager configures the underlying infrastructure with the software required to execute the jobs from the Programming Models layer;
 - Elastic Compute Clusters in the Cloud provides the interface to deploy self-configurable scalable clusters. This is the main tool for deploying the *EUBra-BIGSEA* infrastructure, and interacts directly with IM;
 - DagSim simulator and Lundstrom predictor are two components that use information from the logs of COMPSs and Spark applications to create predictor models for estimating running time under different resource conditions;
 - Proactive Policies have two implemented components: Marathon and Chronos Framework for dealing with QoS, which adjusts the amount of resources allocated to match the expected QoS and the component to adjust the CPU CAP on hypervisors (working in both OpenNebula and OpenStack) to meet the expected deadlines.
2. Programming Models, which provide the means to write parallel data analytics programs on top of the *EUBra-BIGSEA* platform. This layer has the following items:
 - COMPSs is a programming framework that infers the inner parallelism of sequential applications dynamically, executing the different steps in parallel and taking care of data dependencies. In the frame of *EUBra-BIGSEA*, it has been extended to work as a Mesos Framework and to use HDFS as a back-end, facilitating the execution on distributed environments;
 - LEMONADE is a platform for the visual creation and execution of data analysis workflows, which produces Spark and COMPSs code.

3. Security and privacy mechanisms provide a homogeneous Authentication, Authorization and Accounting (AAA) mechanism and privacy policies for data access and processing. This layer has the following items:
 - AAAaaS is a module of Authentication, Authorization and Accounting as a Service for the *EUBra-BIGSEA* Project;
 - PRIVAAaaS is a set of libraries and tools that allows controlling and reducing data leakage in the context of Big Data processing and, consequently, protecting sensible information processed by data analytics algorithms, with multiple types of anonymization techniques.
4. Big Data services is composed by the following items:
 - Ophidia - which exploits advanced parallel computing techniques and a hierarchical, distributed storage organization to execute intensive OLAP-based analysis over multi-terabyte datasets;
 - Data Quality as a Service (DQaaS) is a tool able to provide information about the quality of the analyzed Big Data sources;
 - Entity Matching as a Service (EMaaS) is a service that supports the detection and measurement of matching problems related to the linkage of large data sources
5. High-level services are composed by the following items:
 - Traffic Congestion Prediction: aims to identify traffic jams using data provided by Waze. To this end, a probabilistic graphical model equipped with Gaussian latent nodes is formulated;
 - Trip Duration Prediction: is a tool that aims to predict bus trips duration based on historical bus GPS data. We train the model using Machine Learning techniques (Support Vector Regression and Lasso Regression) on historical bus trips data, and use it to predict future trips;
 - Sentiment Analysis: transforms social media data (textual) into a quantitative estimation of the citizens expressed sentiment. Such analysis targets a specific subject, for example, traffic status or city services;
 - Trip Crowdedness Prediction is a tool that aims to predict the number of passengers (crowdedness) of a bus trip in the future, based on historical bus location and ticketing data;
 - People Paths: is an application which performs a descriptive analysis on bus GPS and passenger ticketing data, finding paths taken by urban Public Transportation users in a time period, and matching the paths origin/destination locations with city area social data.

The applications on the framework work at two levels. The final-user developed applications offer a Graphic User Interface that exposes the outcome of other components in the *EUBra-BIGSEA* platform. On the other hand, applications for descriptive and predictive data models run on top of the infrastructure. Further details of each component are presented at the *EUBra-BIGSEA* site (<http://www.eubra-bigsea.eu/>).

4. The Applications

All the applications have as use case the data from Curitiba Municipality⁸. In summary, data covering DPGS trajectories, ticketing, Weather, Social media, routes, timetables,

⁸<http://www.curitiba.pr.gov.br/dadosabertos/>



Figure 2. The 19 components from the infrastructure and its layers.

sociodemographic and environment (among others) in several formats (CSV, DOC, SHP, data from Twitter, etc.) were manipulated in several databases (PostgreSQL, MongoDB, among others) for the applications. Further details of the data sources, acquisition and integration can be found here ⁹. The application ecosystem is presented in Figure 3: note that different types of data, such as file sources and databases are used.

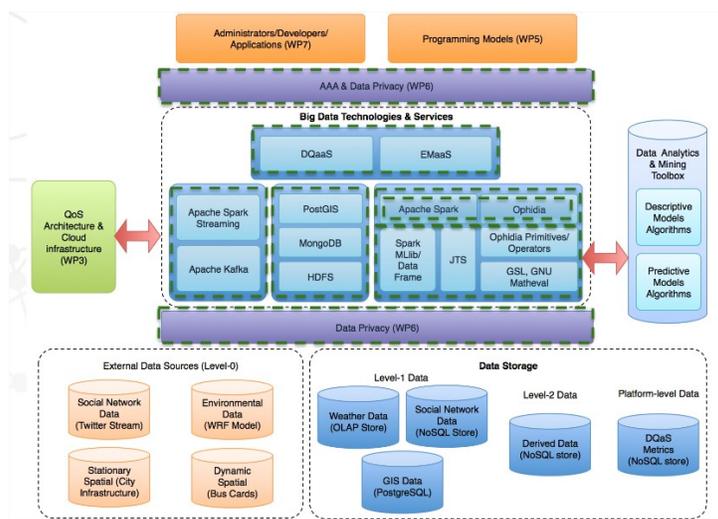


Figure 3. Integrated ecosystem.

On the other hand, the applications use several software components, as shown in Figure 4. Links (lines) between the relations represent the module relation to other

⁹ http://www.eubra-bigsea.eu/sites/default/files/EUBRra-BIGSEA_D7.2_GES3DataIntegration_v1.pdf – Last accessed on Nov 24, 2017.

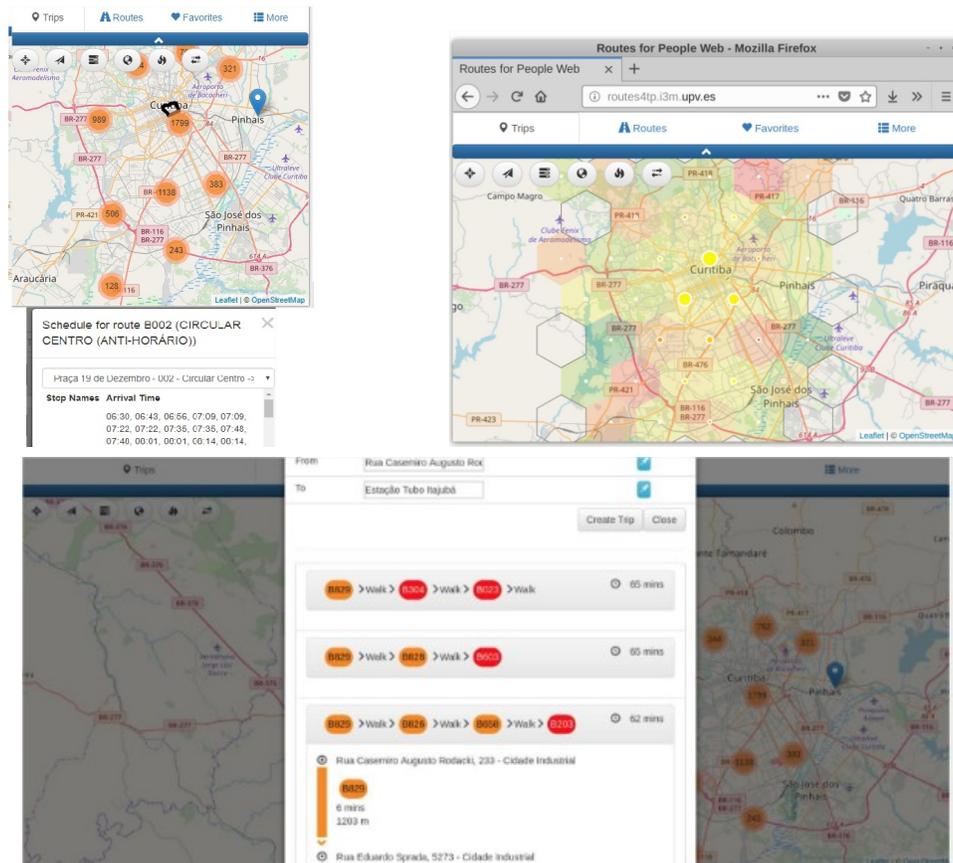


Figure 5. Routes4People application.

Destination matrices, inefficiency analysis, comparison with population, income and literacy rate, implemented in LEMONADE & COMPSs.

The object was not only to present the available bus stops, time table, bus lines, but also present the prediction of the next bus in a specific area (as shown in Figure 6). The code is available online ¹⁴.

4.3. City Administration Dashboard

City Administration Dashboard (video available online ¹⁵) has as main user the decision maker, being an application which has as basis descriptive statistics and visualization techniques in order to assist and facilitate planning and monitoring the system.

As technologies, it uses Ophidia to infer a set of bus usage statistics by means of descriptive analytics algorithms implemented in Python and exploiting the COMPSs programming model. Its modules include also anonymization phases as well as pre-processing steps based on Data Quality and Entity Matching steps.

¹⁴ <https://github.com/eubr-bigsea/bigsea-melhorbusao> – Last accessed on Nov 24, 2017.

¹⁵ <https://youtu.be/AQo5O8YhssA> – Last accessed on July 24, 2018.

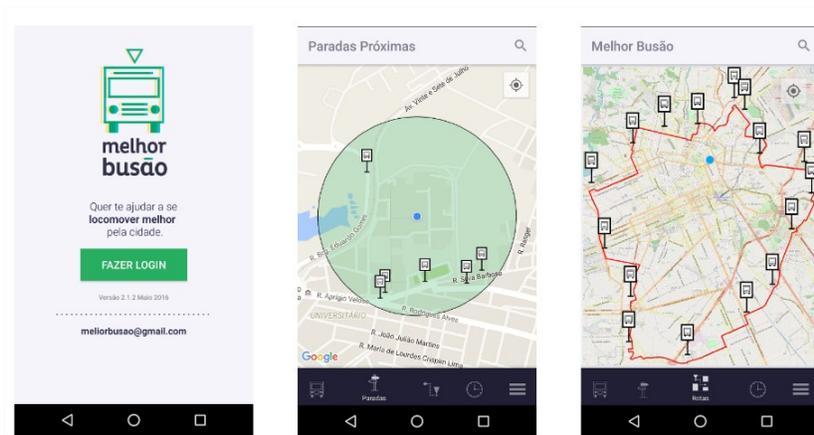


Figure 6. Melhor Busão (Opening Interface, Location of Bus Stops Nearby and Bus Stops available in a Bus Line).

In summary, it presents interactive charts related to a set of 20 statistics in the bus system usage (bus lines, bus stops and passengers) based on bus cards data, along with bus position and shape files (the application focuses on Curitiba city in Brazil) from three perspectives: overall boarding, average passenger boardings and bus stop crowdedness.

Figure 7 presents the visualization of total number of passengers by hour followed by the heatmap of the passengers bus stops. The objective of this application was to support data visualization, from the municipality point of view, and the code is available online ¹⁶ along with the web interface ¹⁷.

4.4. Discussion

The big and fast data eco-system in this proposal is a platform that can effectively support (i) different types of data processing (i.e. batch and streaming) on (ii) heterogeneous data (i.e. multidimensional, relational, NoSQL) (iii) requiring multiple data analytics and mining features (i.e. descriptive and predictive models), while also (iv) taking security, data privacy and QoS-oriented elastic cloud scenarios into account.

Within the main contributions of the infrastructure, we can mention: 1) specific data services, such as EMaaS [Mestre et al. 2017] (which was fundamental to match entities within the different databases integration) and DQaS [Araújo et al. 2017] (in order to check the data quality); 2) new programming technologies, such as Lemonade [d. Santos et al. 2017] (which simplifies and abstract the infrastructure and the programming task); 3) online available predictive models, such as sentiment analysis ¹⁸ and Trip Crowdedness Prediction ¹⁹, among others.

¹⁶ <https://github.com/eubr-bigsea/ticketing-descriptive-analytics> – Last accessed on Nov 24, 2017.

¹⁷ <https://tarcisob.shinyapps.io/city-admin-dashboard-ctba> – Last accessed on July 24, 2018.

¹⁸ <https://github.com/eubr-bigsea/twitter-sentiment-analysis> – Last accessed on July 24, 2018.

¹⁹ <https://github.com/eubr-bigsea/btr-spark> – Last accessed on July 24, 2018.



Figure 7. City Administration Dashboard.

Within the contributions of the applications, we can mention: 1) their availability over open source licenses (software, documentation and video); 2) their integration with several data sources (Waze, twitter, mobility open data), along with the main challenges, such as data privacy [Basso et al. 2016]; and 3) local panels to discuss mobility applications and mobility solutions, such as MAUI Symposium ²⁰ and Workshop on Secure Cloud and Big Data ²¹.

Further performance evaluation and tests for individual modules or overall applications can be found at the *EUBra-BIGSEA* project ²². Results for the performance evaluation of the Dashboard, for example, show that the platform scalability is adequate enough to process large datasets (related to long time period) and provide useful statistics for the City Administration Dashboard application ²³.

5. Conclusion

In order to accommodate the users and their transportation needs, a city must carefully analyze the several data sources to determine the citizen needs and possible changes in transportation to support those needs. Further analysis should also accommodate different

²⁰ <http://www.prppg.ufpr.br/site/sba-maui/> – Last accessed on July 24, 2018.

²¹ <http://site.sanepar.com.br/noticias/inscricoes-para-evento-sobre-processamento-seguro-de-dados-em-nuvens-inseguras-ate-27> – Last accessed on July 24, 2018.

²² <http://www.eubra-bigsea.eu/menu-deliverables> – Last accessed on Sept. 09, 2018.

²³ http://www.eubra-bigsea.eu/sites/default/files/D2.5_EUBra-BIGSEA_FinalActionPlanReport.pdf – Last accessed on July 09, 2018.

perspectives: city administrators, general visualization and basic statistics, bus timetables, among others. This paper presented three applications resulted from the *EUBra-BIGSEA* project, from the urban mobility perspective. The first one (City Administration Dashboard) presents an historical overview of the data, from the urban traffic management perspective. The second one presents an Advanced Traveler Information System (Melhor Busão), while the third one gathers information from traffic congestion and sentiment analysis (among others) in a web interface. In summary, the three applications took advantage of several infrastructure enhancements (such as elastic computation and proactive policies), programming models (using technologies such as COMPSs and LEMONADE), security and privacy mechanisms, Big Data services (parallel computing and data quality), along with high-level services based on models (such as traffic congestion prediction and trip duration prediction). As future work, we can mention the integration of other historical data, user evaluation of the system, further tests within more data, and integration with additional scenarios (such as accidents and bumps).

Acknowledgments. We would like to thank the Municipality of Curitiba, IPPUC, CAPES, CNPq, Fapemig and *EUBra-BIGSEA* project (EC/MCTIC 3rd Coordinated Call).

References

- [Andrade et al. 2014] Andrade, T. C., Pereira, M. A., and Wanner, E. F. (2014). Development of an application using a clustering algorithm for definition of collective transportation routes and times. In *XV GEOINFO*.
- [Araújo et al. 2017] Araújo, T. B., Cappiello, C., Kozevitch, N. P., Mestre, D. G., Pires, C. E. S., and Vitali, M. (2017). Towards reliable data analyses for smart cities. In *Proceedings of the 21st International Database Engineering & Applications Symposium, IDEAS 2017*, pages 304–308, New York, NY, USA. ACM.
- [Azambuja 2016] Azambuja, L. S. d. (2016). Dados abertos em cidades inteligentes: portais de dados abertos possibilitando o acesso e uso da informação. B.s. thesis, UFRGS.
- [Basso et al. 2016] Basso, T., Matsunaga, R., Moraes, R., and Antunes, N. (2016). Challenges on anonymity, privacy, and big data. In *2016 Seventh Latin-American Symposium on Dependable Computing (LADC)*, pages 164–171.
- [Braz et al. 2018] Braz, T., Maciel, M., Mestre, D. G., Andrade, N., Pires, C. E., Queiroz, A. R., and Santos, V. B. (2018). Estimating inefficiency in bus trip choices from a user perspective with schedule, positioning, and ticketing data. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–12.
- [Calandre et al. 2018] Calandre, A., Pasquim, B., Santos, E., and Oliveira, J. (2018). *MYURB: Assistente de Mobilidade em Curitiba através do Transporte Público*. monography, UTFPR.
- [d. Santos et al. 2017] d. Santos, W., Carvalho, L. F. M., d. P. Avelar, G., Silva, A., Ponce, L. M., Guedes, D., and Meira, W. (2017). Lemonade: A scalable and efficient spark-based platform for data analytics. In *2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, pages 745–748.
- [da Silva et al. 2016] da Silva, E. L. C., d. O. Rosa, M., Fonseca, K. V. O., Luders, R., and Kozevitch, N. P. (2016). Combining k-means method and complex network analysis

- to evaluate city mobility. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1666–1671.
- [De Bona et al. 2016] De Bona, A., Fonseca, K., Rosa, M., Lüders, R., and Delgado, M. (2016). Analysis of public bus transportation of a brazilian city based on the theory of complex networks using the p-space. *Mathematical Problems in Engineering*, 2016:1–12.
- [Husár et al. 2017] Husár, M., Ondrejčka, V., and Varış, S. C. (2017). Smart cities and the idea of smartness in urban development – a critical review. *IOP Conference Series: Materials Science and Engineering*, 245(8):082008.
- [Kozievitch et al. 2016] Kozievitch, N. P., Gadda, T. M. C., and Marcelo O. Rosa, K. V. O. F., Gomes-Jr, L. C., and Akbar, M. (2016). Exploratory analysis of public transportation data in curitiba. In *XXXVI CSBC*, pages 1656–1666. Sociedade Brasileira de Computação.
- [Li and Tong 2017] Li, R. and Tong, D. (2017). Incorporating activity space and trip chaining into facility siting for accessibility maximization. *Socio-Economic Planning Sciences*, 60:1 – 14.
- [Mestre et al. 2017] Mestre, D. G., Pires, C. E. S., and Nascimento, D. C. (2017). Towards the efficient parallelization of multi-pass adaptive blocking for entity matching. *Journal of Parallel and Distributed Computing*, 101:27 – 40.
- [Monteiro et al. 2017] Monteiro, C. M., Cruzeiro Martins, F. V., and Davis Junior, C. A. (2017). Optimization of new pick-up and drop-off points for public transportation. In *XVIII GEOINFO*, pages 222–233.
- [Sousa Junior et al. 2014] Sousa Junior, S. R. d., Lima, R. d. S., and Cunha, R. A. H. d. (2014). Crowdbus: Aplicativo crowdsourcing para informação, localização, avaliação e fiscalização de frotas de ônibus. *SEGeT : Simpósio de Excelência em Gestão e Tecnologia*, XI.
- [Souza et al. 2015] Souza, R., Oliveira, I. P., Junior, F., Sales, L., and Ferraz, F. (2015). Beyond efficiency: How to use geolocation applications to improve citizens well-being. In *SMART'2015*, pages 37 – 40.
- [Vila et al. 2016] Vila, J. J. R., Kozievitch, N. P., Gadda, T. M. C., Fonseca, K., Rosa, M. O., Gomes-jr, L. C., and Akbar, M. (2016). Urban mobility challenges – an exploratory analysis of public transportation data in curitiba. *Revista de Informática Aplicada*, 12:1–14.
- [Weigang et al. 2001] Weigang, L., Yamashita, Y., da Silva, O. Q., XiJun, D., dos Prazeres, M. a. T., and de Oliveira, D. C. S. (2001). Implementação do sistema de mapeamento de uma linha de ônibus para um sistema de transporte inteligente. In *SEMISH'2001*, pages 72–85.
- [Yang et al. 2000] Yang, H., Bell, M. G., and Meng, Q. (2000). Modeling the capacity and level of service of urban transportation networks. *Transportation Research Part B: Methodological*, 34(4):255 – 275.