Research and Analysis of Data

**Shruti Varsha Venkatraman**

*VIT University, India*

**ABSTRACT**

Data of different types can be used for various things when analyzed properly and efficiently. To know about the best methods there needs to be existing literature on these, including the stages of analysis, its challenges and outcomes of the methods. There is very less literature on this topic on a general outline. Without proper study, the outcomes that we get would not be precise and useful. Each challenge faced in the different stages of data analysis has to be taken into consideration and worked upon to undertake the analyzing process efficiently. As time goes, different types of data are evolving. Using such daily collected and updated data would prove to be very useful provided that the sources are verifies and the quality is assured. This paper covers the process of analyzing data in a general look and showcasing the challenges faced, the methods used as of now, its advantages and disadvantages.

***Keywords:*** Data, Data discovery, Data collection, Data preparation, Data analysis, LBSN, Data integrity, Data life cycle, Data transformation, Scalability

**Introduction**

Social media dominates in almost every field these days. And it has proved to be very useful in providing data for various fields like research, findings, surveys and audits. Data of several types are used for many purposes. It is a major source for everything.

Data from social media usage, can show us a lot. It can provide information of the aspects and patterns of communication. LBSN data can be recognized as volunteered geographic information —VGI— since the expressed perceptions, interests, needs and behaviors are published online voluntarily by the users and refer to unique and specific places in cities (Campagna, 2016; Jiang, Alves, Rodrigues, Ferreira, & Pereira, 2015; Kitchin, 2013)). The diversity of LBSNs, and the content retrieved from them, offer a multi-perspective approach to the study of cities. There is considerable research using data from Facebook, Twitter and Instagram, some of the most globally-renowned LBSNs, that covers different topics in relation to diverse fields of knowledge (Jiang et al., 2015; Milne, Thomas, & Paris, 2012; Serrano-Estrada, Marti, & Nolasco-Cirugeda, 2016; Van Canneyt, Schockaert, et al., 2012a).

The importance of obtaining meaningful information from data is a huge opportunity for further development of technology and can pose as a challenge too, which can also help in improving the present methods of data collection. It also provides researchers new ways to try doing their research. This means that the method and way to do the research and further studies will be improved which can take data analysis to another level of precision and efficiency.

To extract meaningful information, they will need to carefully plan the way of research and studies on it. Along with that, they have to see a way in which they can include the heterogeneous nature of it and doing it keeping in mind the social and economic analyses and government policies over it. The particular properties and challenges that the current Big Data context opens require specific architectures for information systems particularly designed to retrieve, process, analyze and store such volume and variety of data. Therefore, we are living the constant births of new technologies conceived to be useful in this context such as, to mention some, cloud and exascale computing (Bahrami and Singhal, 2014; Reed and Dongarra, 2015). Given this recent technological and data revolution, research in this topic is in its early stage (Chen et al., 2014).

The critical nature of stages of data analysis is often put aside as a small bug by those who collect this data and use it to retrieve information, prepare it and analyze it. While this saves time, this has the power to change the nature of the information extracted or derived. Such a transformation can make this process useless as it does not achieve its main objective i.e, collecting relevant and precise data. Moreover while it shows what research can be done, it does not show how it can be done. This aspect is much needed for the accuracy of resultant information derived from it. For example, Afzalan, Sanchez, and Evans-Cowley (2017) studied the factors that affect how city and planning organizations select and decide online tools for public engagement in aid of smart city development. The result of the study showed that planning organizations should choose a participation tool on the basis of the following: capacities of organizations, characteristics of communities who will use the tool, user-community norms and rules, and the tool's capabilities.

The objectives of this paper are: 1. Find the existing research in the field of data analysis, 2. Discuss about the paper’s data sets, advantages and disadvantages, 3. Come up with our findings and propose our approach.

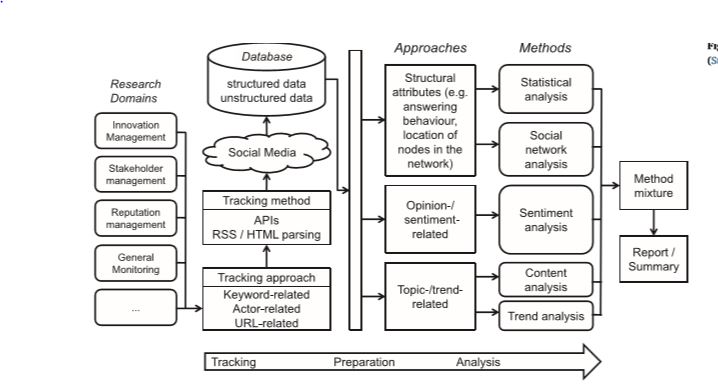
**1.LITERATURE REVIEW**

**1.1 Social media analytics – Challenges in topic discovery, data collection, and data preparation (Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics–Challenges in topic discovery, data collection, and data preparation. *International journal of information management*, *39*, 156-168.)**

The research reveals that in the search query results, only one out of five search results were relevant to the research questions. Moreover, there is a lack of comprehensive discussion of social media analytics. The existing framework doesn’t show how to carry out research.

The research was carried out arguing that the complexity of steps other than data analyses has not yet been adequately covered in research and that there is widely accepted standards on how to proceed with each step.

Most of the existing papers are isolated case studies that collect a large set of data set during a specific time duration on a specific topic and analyze it quantatively. The steps followed in all the research papers are common, therefore the field of “Social Media Ananlytics” aims to combine, extend, and adapt methods for the analyses of social media data. While frameworks by Aral, Dellacrcas, and Godes(2013) and van Osch and Coursaris(2013) are useful to decide the topic of the research and to locate individual projects within a larger context, they do not oﬀer guidance on how to carry out the research, and which challenges might arise.



*Fig.1. Social media analytics framework*

To ensure the success of a new social media analytics project, researchers and practitioners should plan ahead and carefully consider how they will address each of these challenges well before they arise. Not all of the data is relevant to the research topic, and thus irrelevant topics need to be ﬁltered out. Advanced topic detection algorithms promise to solve this problem. The approach of drawing four V’s from big data literature to categories challenges found in the analysis of the literature.

**1.2 Spatial data analytics of mobility with consumer data (Tranos, E., & Mack, E. (2018). Big data: A new opportunity for transport geography? *Journal of Transport Geography*)**

In the fields of transport and geography, obtaining past data, analyzing them and using those results is a very efficient way of improving the transport and geography services that are already available.

In the research papers that are on this same topic, the authors have used census data to undertake the spatial analysis in many countries. For example, Rae, 2016; Lima et al., 2017; Parolin and Rostami, 2017) used census data for spatial analysis and shows that the data is typically accurate and are with high spatial granularity details. It is easily accessible and presented with high standards of documentation. However, census data is also limited in significant ways – collection is infrequent and publication of the data can be slow. Moreover, the outputs are aggregated and do not permit longitudinal analysis for individuals; and limited insights are provided into consumption, health, lifestyle and wealth.

Consumer data in the commercial context has not received the attention in research that it has to get. If we take telephone or mobile telephone records, from large operators like Vodaphone or Airtel, we would get huge reliable data of individual movement. We, humans generate billions of calls everyday. The geolocation of these calls can be collected, prepared and analyzed for more useful information is the transport and geolocation fields (de Montjoye et al., 2014).

The proposed approach was using consumer data instead of census data. With this approach we could get data that is continually collected and updated. This way would be more useful as some of the most detailed outputs that we would need from census data will take several years to be released from the collection.

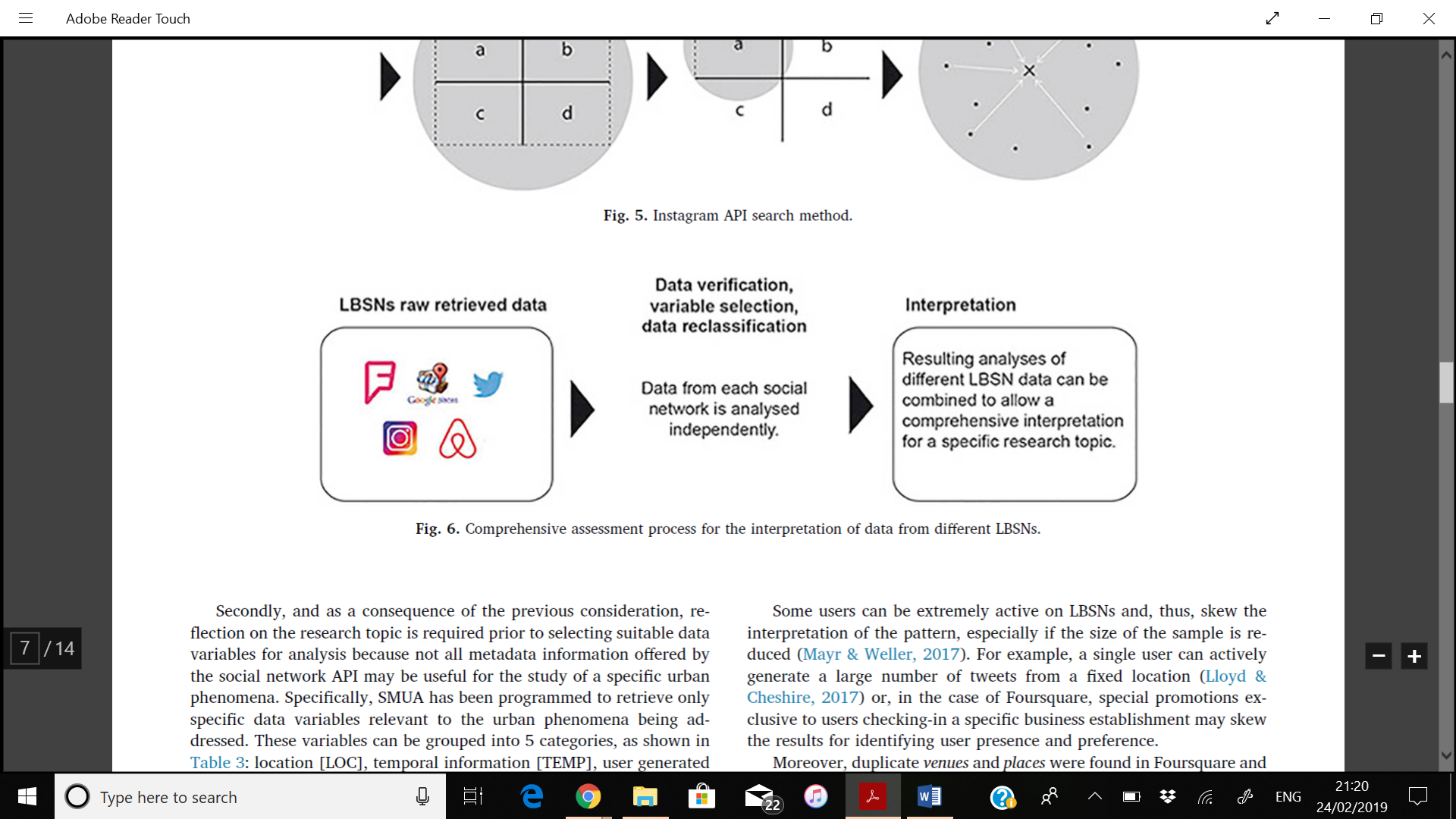
**Data collection:** Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes.

**1.3 Social Media data: Challenges, opportunities and limitations in urban studies (Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2019). Social Media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, *74*, 161-174.)**

The critical nature of data collection is often put aside as a small bug by those who collect this data and use it to retrieve information. While this saves time, this has the power to change the nature of the information extracted or derived. Such a transformation can make this process useless as it does not achieve its main objective i.e, collecting relevant and precise data.

The importance of obtaining meaningful information from data is a huge opportunity for further development of technology and can pose as a challenge too, which can also help in improving the present methods of data collection. It also provides researchers new ways to try doing their research. This means that the method and way to do the research and further studies will be improved which can take data analysis to another level of precision and efficiency.

Although closer scrutiny of the data is necessary for more effective data filtering and mining, crowdsourcing technologies —including social networks— provide great opportunities for researchers and designers involved in the analysis of urban environments (Al-Ghamdi & Al-Harigi,2015; Lee et al., 2013; Liu, Zhou, Zhao, & Ryan, 2016; Quercia, Aiello, Schifanella, & Davies, 2015a, 2015b).



*Fig.2. Comprehensive assessment process for the interpretation of data from different LBSNs.*

SMUA's functionality and interface has been specifically designed to collect geolocated social network data.

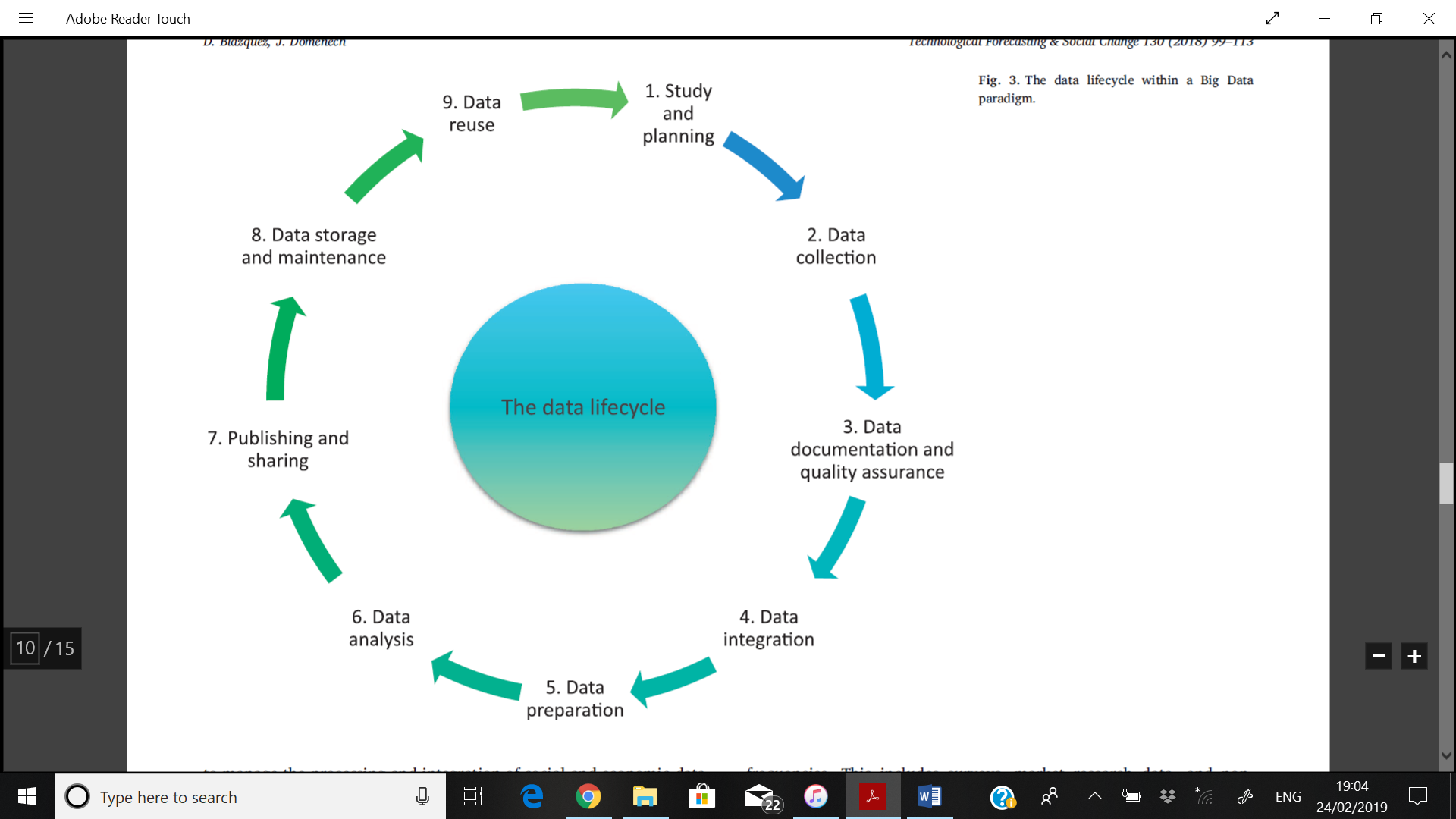
**1.4 Big Data sources and methods for social and economic analyses (Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, *130*, 99-113.)**

Information and communication technologies has brought a lot of development in the field of data analysis. People who deal with analysis this data will have to deal with a huge amount of data which will be heterogeneous. To extract meaningful information, in other words to transform data, they will need to carefully plan the way of research and studies on it. Along with that, they have to see a way in which they can include the heterogeneous nature of it and doing it keeping in mind the social and economic analyses and government policies over it.

**Data transformation:** it is the process of converting data from one format or structure into another format or structure. Data transformation is critical to activities such as data integration and data management.

The particular properties and challenges that the current Big Data context opens require specific architectures for information systems particularly designed to retrieve, process, analyze and store such volume and variety of data. Therefore, we are living the constant births of new technologies conceived to be useful in this context such as, to mention some, cloud and exascale computing (Bahrami and Singhal, 2014; Reed and Dongarra, 2015). Given this recent technological and data revolution, research in this topic is in its early stage (Chen et al., 2014).

In the domain of healthcare, a complete and specific Big Data analytics architecture was developed by Wang et al. (2016a). This architecture was based on the experiences about best practices in implementing Big Data systems in the industry. A number of functionalities are expected to be considered when designing a Big Data architecture for a specific knowledge field, business or industrial process. These include: Data sources, data extraction, data loading and preprocessing, data processing, data analysis, data transformation, interfacing and visualization, data storage and model specification.



*Fig. 3. Data life cycle*

**Data life cycle:** The data life cycle is the sequence of stages that a particular unit of data goes through from its initial generation or capture to its eventual archival and/or deletion at the end of its useful life. Although specifics vary, data management experts often identify six or more stages in the data life cycle.

The proposed work includes layers and modules to manage the processing and integration of data. It is organized in three layers:

Data analysis layer: the main processes are getting knowledge from the data that is input and having an overall process of collecting it from the source, preparing and publishing the data.

Governance layer: its role is to apply the policies and regulation on the data and managing licenses of the data sets.

Persistence layer: it deals with storing and managing the data to make it easy to access and use for data analysis.

**1.5 Smart cities with big data: Reference models, challenges, and considerations (Lim, C., Kim, K. J., & Maglio, P. P. (2018). Smart cities with big data: Reference models, challenges, and considerations. *Cities*, *82*, 86-99.)**

The concept of smart cities is spreading worldwide gaining popularity. It has become so as it can improve economy, mobility, environment, people, living standards and city governance(Abella, Ortiz-de-Urbina-Criado, & De-Pablos- Heredero, 2017; Angelidou, 2015; Caragliu, Del Bo, & Nijkamp, 2011; Vanolo, 2014). Seeing such improvement in big sectors of not only companies but also countries, we can tell that research on methods for data transformation and extraction has significant urgency for development.

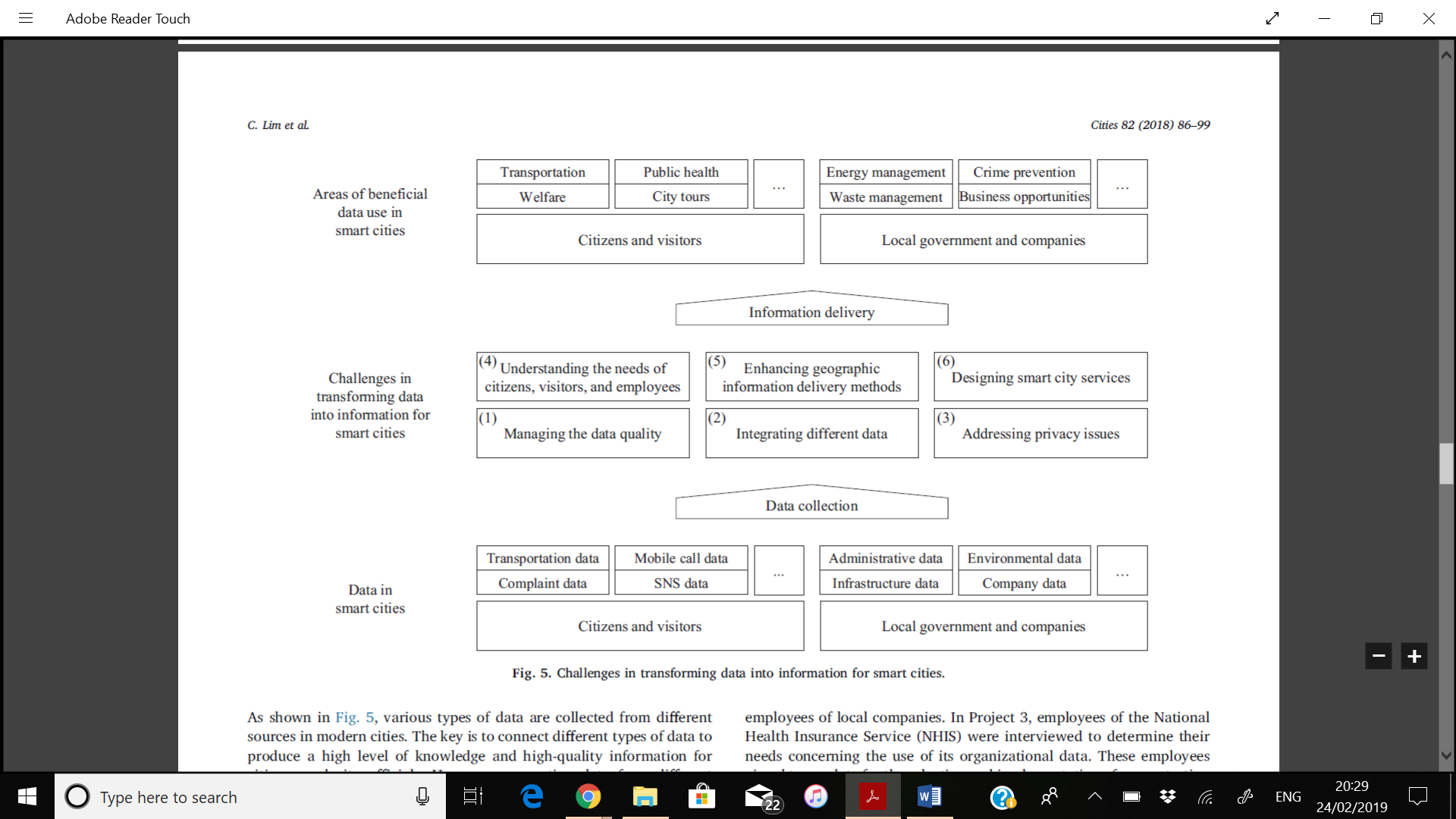
The Rio de Janeiro government used data from cameras and sensors to address various concerns of the city such as weather, energy, and safety (Kitchin, 2014). Other cases include Santander in Spain (Díaz-Díaz, Muñoz, & Pérez-González, 2017) and Cosenza in Italy (Cicirelli, Guerrieri, Spezzano, & Vinci, 2017).

Intelligent transportation systems can contribute to the energy saving of smart cities (Chen, Ardila-Gomez, and Frame (2017)). This study identified the four main steps of smart mobility solutions as a means to achieve energy savings, and then discussed the institutional, technical, and physical conditions required by each step. The factors that affect how city and planning organizations select and decide online tools for public engagement in aid of smart city development. The result of the study showed that planning organizations should choose a participation tool on the basis of the following: capacities of organizations, characteristics of communities who will use the tool, user-community norms and rules, and the tool's capabilities (Afzalan, Sanchez, and Evans-Cowley (2017)).

The proposed approach is of two parts: using case analysis and action research. They are used for identifying reference models and understanding the challenges and considerations. The challenge faced include integrity of data and scalability.

**Data integrity:** Data integrity is the maintenance of, and the assurance of the accuracy and consistency of, dataover its entire life-cycle, and is a critical aspect to the design, implementation and usage of any system which stores, processes, or retrieves data.

**Scalability:** Scalability is the capability of a system, network, or process to handle a growing amount of work, or its potential to be enlarged to accommodate that growth. A system whose performance improves after adding hardware, proportionally to the capacity added, is said to be a scalable system.



*Fig.4.Challenges in transforming data into information for smart cities.*

**2. COMPARISON OF EXISTING WORK AND DISCUSSION**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameters | Paper 1 | Paper 2 | Paper 3 | Paper 4 | Paper 5 |
| Name | Social media analytics – Challenges in topic discovery, data collection, and data preparation | Spatial data analytics of mobility with consumer data | Social Media data: Challenges, opportunities and limitations in urban studies | Big Data sources and methods for social and economic analyses | Smart cities with big data: Reference models, challenges, and considerations |
| Citation | Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics–Challenges in topic discovery, data collection, and data preparation. *International journal of information management*, *39*, 156-168. | Tranos, E., & Mack, E. (2018). Big data: A new opportunity for transport geography?. *Journal of Transport Geography*. | Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2019). Social Media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, *74*, 161-174. | Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, *130*, 99-113. | Lim, C., Kim, K. J., & Maglio, P. P. (2018). Smart cities with big data: Reference models, challenges, and considerations. *Cities*, *82*, 86-99. |
| Problem | What challenges do researchers face when discovering topics, collecting and preparing social media data for further analyses | Consumer data will provide valuable ammunition to augment academic investigations in transport geography | The collection of data from LSBN and other social medias are seldom addressed when they are taken to retrieve information. It has its own challenges which can affect the data derived from it. Along with that, it has its limitation and obliqueness which can are often overseen by the people who handle that type of data. | The challenges that are associated with preparation of huge data for efficient data analysis. | Research on the ways and methods for transforming data into useful and meaningful information is very rare. |
| Proposed approach | To ensure the success of a new social media analytics project, researchers and practitioners should plan ahead and carefully consider how they will address each of these challenges well before they arise. | There should be a limit to which consumer data can be accessed and used. | SMUA-Social Media Urban Analyser | Data analysis layer, Governance layer  Governance layer | The proposed approach is of two parts: using case analysis and action research. They are used for identifying reference models and understanding the challenges and considerations |
| How the results of proposed approach are compared with other similar approaches? | Conventional tools failed when used to compile several million rows whereas programming lets millions of rows compile together easily. | Consumer data have distinctive characteristics which oﬀer great potential when used as a complement or as an alternative to census data.  Consumer data are contemporaneous whereas census data are legacy  Consumer data are dynamic and longitudinal whereas census data is cross-sectional and static. | Even though most common social media APIs use the Rest method (Brown, Soto-Corominas, Suárez, & de la Rosa, 2017), an approach to Twitter Streaming API request method is rather similar not in terms of quantity but in terms of data representability. | A more exhaustive data lifecycle to date was proposed by Rüegg et al. (2014), who included up to eight stages: the first four stages (planning, data collection, data quality control, and analysis) correspond to managing data in a traditional project which is new (no previously results or data exist). | The six challenges addressed in this paper were consistent with those discussed in literature papers such as Al  Nuaimi et al., 2015; Ben Sta, 2017. This showed that their research confirmed and augmented existing knowledge and arguments in the literature for further research. |
| Datasets used | It allowed to verify the major problem faced while researching | Census data  Consumer data  Diary entries  Wearable technologies | Quadtree decomposition method | relational databases managing, external, searches, comments | Preventive local administration, Local operations management, Local network development, Local information diffusion |
| Advantages | Visual analytics involves visualization and automatic processing of data. Data visualization is considered as a research output, a way of communication results, and an integral part of research.  This search allowed to note that the volume of data is the most frequently mentioned challenge. | Consumer data are continually collected and updated.  Consumer data is wide ranging in terms of socio-economic and demographic characteristics of populations.  Consumer data are continuously changing and hence can be used for dynamic research. | The process of retrieving data and grouping them into specific entities is made easier by implementing the particular methods accordingly, in this case it is SMUA. | Unnecessary information ,useless data, duplicates or noise is cleaned beforehand to make data analysis easier. While analyzing data, due to unnecessary data or noise, the process may be delayed or it may use extra storage to story the data to be analyzed and there may be data redundancy too. | The proposed solution helped improve the process of transforming the data into meaningful information. This lead to accurate and very useful information which helped get better ideas for the smart cities in many cities across the world. |
| disadvantages | The research is completely relied on papers specifically mentioning the challenges they solve.  Some papers provide solutions without examples.  The problems/challenges aren’t universal. | Consumer data are collected by business organization and so a gap exists.  Consumer data usage can lead to invasion of privacy.  Reliability of consumer data is very challenging | Reliance on data accessibility makes the retrieval process vulnerable to the changes in access conditions. | The integration of the architecture in the existing organizational information systems is a critical process to ensure the smooth generation of forecasts and nowcasts. The implementation of the modules in a proper cloud computing environment so that the system can scale easily is also crucial. | This process which requires enhanced connectivity needs new equipment and organizations and increases the complexity of human jobs. |

**3. OUR FINDINGS**

These papers provided an insight into the present studies and application of the stages of data analysis namely, data discovery, data collection, data preparation and data analysis. From these works we learnt the challenges and difficulties faced in those stages of data analysis which are often not found or taken into consideration or just simply put aside.

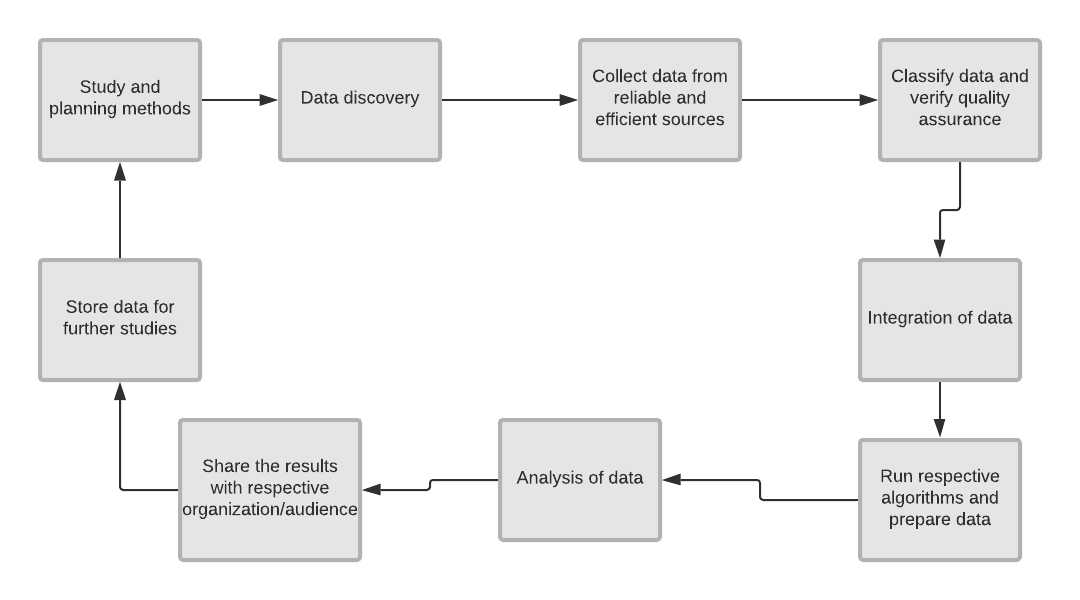
The authors used various methods to find, collect, prepare and analyze data. Some of those methods were: the four v’s approach ---volume, velocity, variety, and veracity—which was used to filter out the necessary ones from the unnecessary data for better and efficient data collection

The main challenges faced when searching for data is finding a reliable source and the SMUA which has the functionality and interface that has been specifically designed to collect geolocated social network data.

The next big challenge would come from using consumer data like mobile telephone records, or from their search results, etc. As these are consumer data, the availability is tricky and could be taken by business organizations or other third parties who might not give it for research. As per the proposed approach in Tranos, E., & Mack, E. (2018). Big data: A new opportunity for transport geography?. *Journal of Transport Geography* , we could use consumer data. Our finding in this is that we should not let business organizations take this data which could be a valuable resource for further development not only in spatial analysis, but also for data analysis.

**Data analysis:** Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data.

In some papers listed above, a few authors used user interfaces for evaluating data. This method is very efficient but it could be manipulated or duplicated. We would suggest to use algorithms that would clearly lay out the various conditions and procedures so that this data would not be manipulated or duplicated.



*Fig.5. Generic architecture*

**CONCLUSION**

When it comes to studying the stages of data analysis, the objective has to be clearly identified. Among these papers, the objective has been identified properly, although not clearly. The previous researches on analysis of data the objectives have been taken in a general perspective rather than specific. There are significant improvements in the fields of data analysis which yet needs more study.

To bring about a better, reliable and efficient method of data analysis, we would have to incorporate not just one method, but many methods together so that it can encompass all types of data available. And for this to happen, more researches should be done specifically in this field which could be used for improving it. As discussed above, one of the main challenges were that in literature there no methods or ways to undertake data analysis. In this paper we have done anm analysis of different literature papers on the same base topic data analysis and discussed about the methods used, its pros and cons and which methods is the best.

A for the future proceedings, we would like to bring about more data in literature that could be used by the personnel attending to the issue of data analysis and its challenges such as scalability, availability of data, integrity of data , transformation of data, data quality, data provenance, management of huge volumes of information, heterogeneity of data, integration of data from different sources, data matching, bias, availability of tools for properly analyzing such kind of data, processing complexity, privacy and legal issues, and data governance are slowly but steadily striken out leading to efficient and better method for this process (Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, *130*, 99-113.).

**REFERENCES**

* Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics–Challenges in topic discovery, data collection, and data preparation. *International journal of information management*, *39*, 156-168
* Tranos, E., & Mack, E. (2018). Big data: A new opportunity for transport geography?. *Journal of Transport Geography*.
* Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2019). Social Media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, *74*, 161-174.
* Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, *130*, 99-113.
* Lim, C., Kim, K. J., & Maglio, P. P. (2018). Smart cities with big data: Reference models, challenges, and considerations. *Cities*, *82*, 86-99.
* Afzalan, N., Sanchez, T. W., & Evans-Cowley, J. (2017). Creating smarter cities: Considerations for selecting online participatory tools. *Cities*, *67*, 21-30.
* Urbieta, A., González-Beltrán, A., Mokhtar, S. B., Hossain, M. A., & Capra, L. (2017). Adaptive and context-aware service composition for IoT-based smart cities. *Future Generation Computer Systems*, *76*, 262-274.
* Frame, G., Ardila-Gomez, A., & Chen, Y. (2017). The kingdom of the bicycle: what Wuhan can learn from Amsterdam. *Transportation research procedia*, *25*, 5040-5058.
* Paroutis, S., Bennett, M., & Heracleous, L. (2014). A strategic view on smart city technology: The case of IBM Smarter Cities during a recession. *Technological Forecasting and Social Change*, *89*, 262-272.
* Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*, 3-13.
* Assunção, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., & Buyya, R. (2015). Big Data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, *79*, 3-15.
* Pääkkönen, P., & Pakkala, D. (2015). Reference architecture and classification of technologies, products and services for big data systems. *Big data research*, *2*(4), 166-186.
* Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, *130*, 99-113.
* Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, *49*, 45-53.
* Alam, S., Albareti, F. D., Prieto, C. A., Anders, F., Anderson, S. F., Anderton, T., ... & Basu, S. (2015). The eleventh and twelfth data releases of the Sloan Digital Sky Survey: final data from SDSS-III. *The Astrophysical Journal Supplement Series*, *219*(1), 12.
* Milne, D., Thomas, P., & Paris, C. (2012). *Finding, weighting and describing venues: Csiro at the 2012 trec contextual suggestion track*. COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANIZATION MARSFIELD (AUSTRALIA) INFORMATION AND COMMUNICATION TECHNOLOGIES CENTER.
* Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A. (2017). Using locative social media and urban cartographies to identify and locate successful urban plazas. *Cities*, *64*, 66-78.
* Van Canneyt, S., Van Laere, O., Schockaert, S., & Dhoedt, B. (2012, November). Using social media to find places of interest: a case study. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information* (pp. 2-8). ACM.
* Jiang, S., Alves, A., Rodrigues, F., Ferreira Jr, J., & Pereira, F. C. (2015). Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, *53*, 36-46.
* Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in human geography*, *3*(3), 262-267.
* Al-ghamdi, S. A., & Al-Harigi, F. (2015). Rethinking image of the City in the Information Age. *Procedia Computer Science*, *65*, 734-743.
* Kabir, H. E., Kim, E. M., Lee, H. Y., Nam, B., & Maeng, J. H. (2013). *Human development report 2013* (Doctoral dissertation, Doctoral dissertation, Ewha Womans University).
* Liu, L., Zhou, B., Zhao, J., & Ryan, B. D. (2016). C-IMAGE: city cognitive mapping through geo-tagged photos. *GeoJournal*, *81*(6), 817-861.
* Wang, P., Xu, B., Wu, Y., & Zhou, X. (2015). Link prediction in social networks: the state-of-the-art. *Science China Information Sciences*, *58*(1), 1-38.
* Abbasi, A., Fu, T., Zeng, D., & Adjeroh, D. (2013, September). Crawling credible online medical sentiments for social intelligence. In *2013 International Conference on Social Computing* (pp. 254-263). IEEE.
* Al-Qurishi, M., Al-Rakhami, M., Alrubaian, M., Alarifi, A., Rahman, S. M. M., & Alamri, A. (2015, March). Selecting the best open source tools for collecting and visualzing social media content. In *2015 2nd world symposium on web applications and networking (WSWAN)* (pp. 1-6). IEEE.
* Alsubaiee, S., Carey, M. J., & Li, C. (2015, May). LSM-Based Storage and Indexing: An old idea with timely benefits. In *Second international ACM workshop on managing and mining enriched geo-spatial data* (pp. 1-6). ACM.
* Anderson, K. M., Aydin, A. A., Barrenechea, M., Cardenas, A., Hakeem, M., & Jambi, S. (2015, January). Design challenges/solutions for environments supporting the analysis of social media data in crisis informatics research. In *2015 48th Hawaii International Conference on System Sciences*(pp. 163-172). IEEE.
* Bindra, G. S., Kandwal, K. K., Singh, P. K., & Khanna, S. (2012, July). Tracing information flow and analyzing the effects of incomplete data in social media. In *2012 Fourth international conference on computational intelligence, communication systems and networks* (pp. 235-240). IEEE.
* Cao, G., Wang, S., Hwang, M., Padmanabhan, A., Zhang, Z., & Soltani, K. (2015). A scalable framework for spatiotemporal analysis of location-based social media data. *Computers, Environment and Urban Systems*, *51*, 70-82.
* Carr, J., Decreton, L., Qin, W., Rojas, B., Rossochacki, T., & wen Yang, Y. (2015). Social media in product development. *Food quality and preference*, *40*, 354-364.
* Chae, J., Thom, D., Jang, Y., Kim, S., Ertl, T., & Ebert, D. S. (2014). Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers & Graphics*, *38*, 51-60.
* Chang, Y., Yamada, M., Ortega, A., & Liu, Y. (2014, December). Ups and downs in buzzes: Life cycle modeling for temporal pattern discovery. In *2014 IEEE International Conference on Data Mining* (pp. 749-754). IEEE.
* Chen, S., Guo, C., Yuan, X., Zhang, J., & Zhang, X. L. (2014, October). MovementFinder: Visual analytics of origin-destination patterns from geo-tagged social media. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)* (pp. 239-240). IEEE.
* Chen, S., Yuan, X., Wang, Z., Guo, C., Liang, J., Wang, Z., ... & Zhang, J. (2016). Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE transactions on visualization and computer graphics*, *22*(1), 270-279.
* Weiler, A., Grossniklaus, M., & Scholl, M. H. (2016). Situation monitoring of urban areas using social media data streams. *Information Systems*, *57*, 129-141.
* Cao, G., Wang, S., Hwang, M., Padmanabhan, A., Zhang, Z., & Soltani, K. (2015). A scalable framework for spatiotemporal analysis of location-based social media data. *Computers, Environment and Urban Systems*, *51*, 70-82.
* Tinati, R., Phillipe, O., Pope, C., Carr, L., & Halford, S. (2014, June). Challenging social media analytics: Web Science perspectives. In *Proceedings of the 2014 ACM conference on Web science* (pp. 177-181). ACM.
* Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics–Challenges in topic discovery, data collection, and data preparation. *International journal of information management*, *39*, 156-168.
* Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2017). A model for theanalysis of data-driven innovation and value generation in smart cities' ecosystems Cities, 64, 47–53.
* Afzalan, N., Sanchez, T. W., & Evans-Cowley, J. (2017). Creating smarter cities: Considerations for selecting online participatory tools. Cities, 67, 21–30.
* Artikis, A., Etzion, O., Feldman, Z., & Fournier, F. (2012). Event processing under uncertainty. Proceedings of the 6th ACM international conference on distributed event-based systems (DEBS ’12), 32–43. http://dx.doi.org/10.1145/2335484.2335488.
* Baars, H., & Kemper, H.-G. (2008). Management support with structured and unstructured data – An integrated business intelligence framework. Information Systems Management, 25(2), 132–148. <http://dx.doi.org/10.1080/10580530801941058>
* Bell, G., Hey, T., Szalay, A., 2009. Computer science: beyond the data deluge. Science 323, 1297–1298.
* Birkin, M., Clarke, G., Clarke, M., Wilson, A., 2002. Retail Geography and Intelligent Network Planning. Wiley, Chichester.
* Deng, X., & Newsam, S. (2017). Quantitative Comparison of Open-Source Data for Fine- rain Mapping of Land Use. Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics - UrbanGIS. Vol. 17. Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics - UrbanGIS (pp. 1–8). . https://doi.org/10.1145/3152178.3152182.
* Dunkel, A. (2015). Visualizing the perceived environment using crowdsourced photo geodata. Landscape and Urban Planning, 142, 173–186. <https://doi.org/10.1016/j>. landurbplan.2015.02.022
* Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M., 2011. Election forecasts with Twitter: how 140 characters reflect the political landscape. Soc. Sci. Comput. Rev. 29 (4), 402–418. http://dx.doi.org/10.1177/0894439310386557.
* Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., Baesens, B., 2015. APATE: a novel approach for automated credit card transaction fraud detection using network-based extensions. Decis. Support. Syst. 75, 38–48. <http://dx>. doi.org/10.1016/j.dss.2015.04.013