ASYNCHRONOUS DIGITAL PARTICIPATION IN URBAN DESIGN PROCESSES

Qualitative Data Exploration and Analysis with Natural Language Processing

CEM ATAMAN[1], BIGE TUNÇER[2] and SIMON PERRAULT[3]
[1,2,3] Singapore University of Technology and Design, Singapore
[1]cem_ataman@mymail.sutd.edu.sg, 0000-0001-8355-7119
[2]bige_tuncer@sutd.edu.sg, 0000-0002-1344-9160
[3]simon_perrault@sutd.edu.sg, 0000-0002-3105-9350

Abstract. This paper aims to improve the usability of qualitative urban big data sources by utilizing Natural Language Processing (NLP) as a promising AI-based technique. In this research, we designed a digital participation experiment by deploying an open-source and customizable asynchronous participation tool, “Consul Project”, with 47 participants in the campus transformation process of the Singapore University of Technology and Design (SUTD). At the end of the data collection process with several debate topics and proposals, we analysed the qualitative data in entry scale, topic scale, and module scale. We investigated the impact of sentiment scores of each entry on the overall discussion and the sentiment scores of each introduction text on the ongoing discussions to trace the interaction and engagement. Furthermore, we used Latent Dirichlet Allocation (LDA) topic modelling to visualize the abstract topics that occurred in the participation experiment. The results revealed the links between different debates and proposals, which allow designers and decision makers to identify the most interacted arguments and engaging topics throughout participation processes. Eventually, this research presented the potentials of qualitative data while highlighting the necessity of adopting new methods and techniques, e.g., NLP, sentiment analysis, LDA topic modelling, to analyse and represent the collected qualitative data in asynchronous digital participation processes.

Keywords. Urban Design; Digital Participation; Qualitative Urban Data; Natural Language Processing (NLP); Sentiment Analysis; LDA Topic Modelling; SDG 10; SDG 11.

1. Introduction

Digitalization and new computational technologies have enabled citizen participation in urban design processes to become more inclusive by reaching more residents. Many digital participation tools have been implemented and applied in urban design

POST-CARBON, Proceedings of the 27th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2022, Volume 1, 383-392.

processes, that intend to go beyond receiving directed feedback from residents on a design proposal represented as rendered visualizations or site plans (Alter et al., 2019). Accordingly, some digital or analogue urban participation campaigns utilize map-based or 3D representations when interacting with residents (Tan, 2016; Tomarchio et al., 2019). These campaigns operate through residents participating in the urban design process by directly manipulating program or urban performance-related aspects of a curated urban proposal created by them or offered to them, usually in an engaging and collaborative digital environment, or a physical setting where verbal communication mediates the collaboration on the physical participation artefact. The aim of these processes is in line with the UN Sustainable Development Goals (SDGs): SDG 10, which promotes reducing inequality by encouraging social, economic, and political inclusion with equal opportunities and SDG 11, which enhances inclusive and sustainable urbanization in policy implementations and efficiency.

However, especially when urban designers are in an early stage of design, understanding the prominent values and goals of residents regarding the urban space is essential. This can be facilitated by highly effective means of citizen participation, such as discussions, debates, and proposals, all expressed in natural language in digital mediums. A variety of asynchronous digital participation tools with such functionality are deployed in urban practices, which allow a high number of stakeholders to participate together at each person's own convenience and own schedule in one medium to effectively discuss urban issues (Klein, 2011; Seifert & Rössel, 2019). The output of such participation processes is mostly qualitative data because of the predominant textual interaction between participants. Nevertheless, the qualitative textual data usually remains unused, which contains essential information regarding urban issues from participants' perspectives (Rathore et al., 2018). The reasons for this underutilization are, firstly, the challenges in filtering the collected data to understand the most valuable/useful information, and secondly, the lack of suitable representations and visualizations of the qualitative data in urban design processes. By considering these shortcomings in this research, we would like to facilitate the translation of collected qualitative participation data to be used by designers and decision makers in urban design processes.

This research aims to support urban designers with data analysis and informed decision-makings in the use of large-scale asynchronous digital participation tools that predominantly collect textual data. In this paper, we aim to investigate the way in which we can facilitate qualitative urban big data sources by utilizing AI-based techniques, i.e., Natural Language Processing (NLP), sentiment analysis, and LDA topic modelling to provide insights into qualitative participation data, which are mostly neglected in participatory urban design practices.

2. Background

The increasing development in data processing and information technologies provides promising opportunities in urban design practices (Milz & Gervich, 2021). They encourage authorities to work on processes for residents and engage with them in important decisions and policies, which deepens the relationship between democracy and the vibrancy of civic life (Goldsmith & Crawford, 2014). In this context, digital participation tools and processes enable residents to collaborate with local governments and provide the advantage of massive amounts of new information for effective and efficient solutions to urban problems (Ataman & Tuncer, 2022). As a general rule, these tools offer greater flexibility in time and location of data collection while promoting social learning for better collective decisions (Tekler et al., 2020). The use of mixed methods for quantitative and qualitative data collection and analysis has been particularly important in such participatory design practices (Lobe et al., 2020). Yet, especially for the qualitative data, the utilization of digital participation data is not fully covered in the urban design domain (Rathore et al., 2018).

It remains a challenge to enable the use of massive digital participation data and utilize qualitative data for informed urban design processes. Several researchers already focused on the possible ways of qualitative data analysis, e.g., mind mapping (Burgess-Allen & Owen-Smith, 2010), correspondence analysis (Habib et al., 2012), and very recently topic modelling (Lock & Pettit, 2020). Topic modelling is especially useful to organize and offer insights for stakeholders to understand large datasets consisting of unstructured text bodies. Yet, it becomes a challenge to interpret the semantic meaningfulness of topics by using such statistical measures of topical coherence (Chang et al., 2009). Therefore, the operationalization of a topic model according to a specific context and research setting is essential.

This study is unique as it deploys and combines various methods from other domains for analysing and visualizing participation data in different levels specifically for urban practices, e.g., NLP, sentiment analysis, LDA topic modelling. In the end, this paper contributes to the use of qualitative big data in asynchronous digital participation processes by providing information about:

- the interaction between and involvement of participants on different levels (i.e., module level, topic level, and entry-level) in a digital participation process,
- AI-based approaches for qualitative participation data to utilize and interpret participation outputs,
- new visualizations and representations to translate qualitative data into informed decision-makings.

Based on the above, this study concludes with several recommendations for the analysis of digital qualitative data in large-scale, individual, and asynchronous participation processes in urban design practices.

3. Methodology

This section presents the experiment conducted for collecting and analysing participation data in a campus transformation context. The process is explained in participation phases, i.e., pre-participation, participation, and post-participation.

3.1. PRE-PARTICIPATION: TOOL SELECTION & CASE STUDY

Within the pre-participation phase, we chose and modified an open-source digital asynchronous digital participation tool, Consul Project, to implement it for a case study. The tool is written in Ruby on Rails using a PostgreSQL database for data storing. Consul Project provides several participation modules with different features and interfaces, i.e., debates, proposals, voting, to discuss the case study (Figure 1). The case study in this experiment is the transformation project of SUTD campus. The campus is accessible by the subway, buses, and private cars and nearby several urban hubs such as a EXPO Convention & Exhibition Centre, Changi Business District, and an urban coastal park (ECP) while hosting a variety of facilities, such as a library, student housing, staff housing, sports and recreation facilities, and F&B and services to fulfil the needs of its daily users. The campus is also a part of a large-scale transformation project aiming at increasing the vibrancy of the East Coast area due to its location.

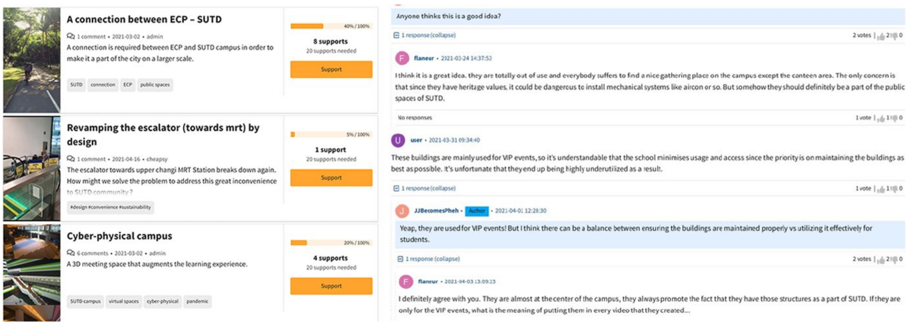

Figure 1. Consul Project Modules: Left: Proposal interface. Right: A discussion on a debate topic.

3.2. PARTICIPATION: PROCEDURE AND PARTICIPANTS

The 47 participants (40% female, 45% graduate and 24% architecture students) participated in the experiment voluntarily, and no incentives were used in this experiment. The selection criteria for the participants were, firstly, to be a student in the university, and secondly, to have completed their first mandatory design course to ensure a foundation in design theory and thinking. Prospective participants were sent an email that included their username, password, and user instructions, which allowed them to anonymously access the participation tool. After receiving the user account information and instructions, the participants were given two months to discuss and propose their ideas through the tool about the transformation of the university campus.

3.3. POST-PARTICIPATION: ANALYSIS METHODS

After the participation phase, the collected participation data was pre-processed to record timestamps of entries, parent arguments, involved participants, and support numbers. This process provided a hierarchical data structure to analyse the qualitative data. By using this pre-processed data, we investigated possible ways to visualize the interaction between participants and the evolvement of debate topics and proposals. Then, we implemented AI-based NLP techniques, i.e., sentiment analysis and LDA topic modelling. We used these NLP methods as powerful computational techniques for data mining and formulating relationships among data and textual documents.

Sentiment Analysis determines the ratio of positive to negative engagements about a specific topic by analysing bodies of text, such as written discussions, comments, and reviews. It is used to obtain insights from opinions, attitudes, and emotions of individuals towards entities, e.g., topics, services, locations, or events (Liu, 2012).

Although sentiment analysis has been used in many fields such as politics, economics, business, and urban planning, the implementation of sentiment analysis in the participatory design domain is not fully discovered yet (Roberts et al., 2018). In this experiment, we used SentimentIntensityAnalyser from the NLTK library in Python, which generates sentiment scores ranging between -1.0 (negative) and 1.0 (positive) corresponding to the overall emotional leaning of the text.

LDA topic modelling, on the other hand, was deployed to classify texts in debates and proposals to a particular topic. It is an intuitive approach that calculates the similarity between source data to reveal their respective distributions of each cluster over topics (Jelodar et al., 2019). It provides the examination of multiple topics by generating a probabilistic distribution of words under a topic to extract thematic content from text-specific data (Wang & Taylor, 2018). In this experiment, the NLTK Python library was implemented to pre-process the data as it is one of the most common and powerful libraries for NLP and computational linguistics.

Before analysing the data, a basic data cleaning was conducted. As Consul Project allows its users to use words, URLs, mentions, abbreviations, etc., we cleaned the texts by removing stopwords, URL links, user mentions (@), and special characters (punctuations and numeric numbers) as they are unnecessary for further analysis. Then, we tokenized the entries and convert them to lowercase. The final data consists of 139 sentences with 1182 words.

4. Results and Discussion

The analysis and interpretation of the collected data are discussed according to the methods that we described in the analysis methods for the post-participation phase. The goal is to provide effective and meaningful ways to analyse digital qualitative participation data to be used by other stakeholders.

4.1. USER INVOLVEMENT & INTERACTION IN PARTICIPATION

After the data pre-processing, the user involvement in different modules is compared through the numbers of involved participants, entries, and supports. In the debate module, engagement is correlated with the number of entries and supports in each individual debate topic. In the proposal topics, on the other hand, only the support numbers represent the engagement as the participants mostly used the support feature of the module, and active participation cannot be measured through the entry numbers (Figure 2). The results confirmed the proper use of the participation modules by the participants according to their features, e.g., deliberation, feedback, supports, and likes.

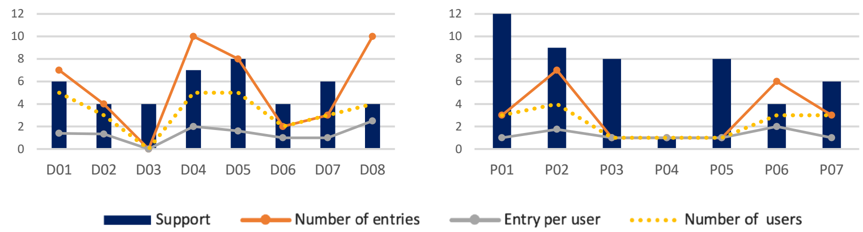

Figure 2. Number of entries and support for debate topics (Left) and proposals (Right).

After the identification of the most engaging debates and proposals, each entry in these modules is mapped according to their timestamps, their hierarchical relationship with other entries, like & dislike numbers that they received, and their sentiment scores on the entry level. The impact of each entry on the overall discussion is revealed by visualizing each discussion topic. Figure 3 presents the entry maps of Debate 4 (as an engaging discussion) and Debate 6 (as a rather inactive debate). As seen in Debate 4, the entries with positive sentiments led to an increased engagement and participation activity while the ones with negative sentiment values as in Debate 6 result in shorter discussions with limited interaction. The analysis results revealed that the initial attitude of an entry is determinant in the engagement of and the interaction between the participants. In order to identify such entries, entry maps and visualizations are essential to provide mediums to stakeholders to trace the most engaging and interacted arguments within the data set and evaluate them accordingly.

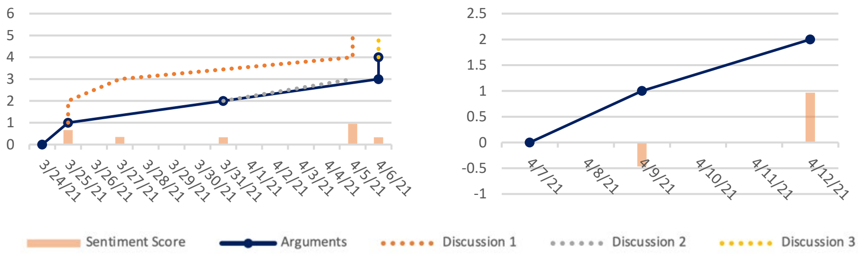

Figure 3. The analysis of Debate Topics: Arguments, discussions, and their sentiment scores (Left: Debate Topic #4, right: Debate Topic #6)

4.2. THE IMPACT OF PARTICIPANTS' ATTITUDE ON PARTICIPATION: SENTIMENT ANALYSIS

In this participation experiment, debate and proposal topics contain introduction texts explaining the target issue to other participants to discuss and evaluate. Although the structure of these introductions is mostly very straightforward with texts and some images, some topics attracted more participants with longer discussions. In order to understand possible underlying reasons, we analysed the sentiment scores of introduction texts and the discussions under each debate and proposal separately. According to the results of the sentiment analysis, the introduction texts of the debate topics with positive sentiment scores (or positive attitudes) result in a higher number of entries; and therefore, more engagement (D1, D4, D5, D7, and D8 in Figure 4).

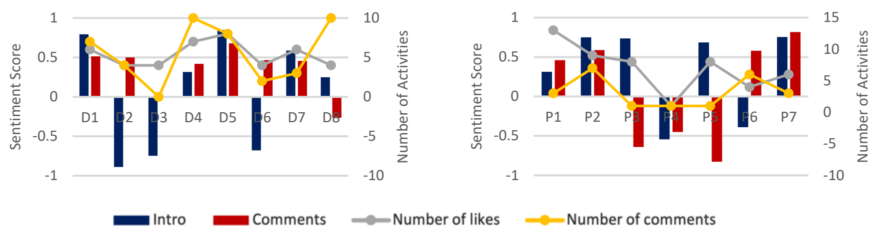

Figure 4. Sentiment analysis of textual data (Left: Debate module, right: Proposal module)

In the proposal module, on the other hand, the attitudes in the responses of the participants are more determinant. The most engaging proposal topics contain more responses with positive sentiment scores, which indicate the willingness of the participants when they like the proposal. In other cases, in which the engagement is relatively lower, negative sentiment scores are observed in the responses (P3, P4, and P5 in Figure 4). This analysis reveals the correlation between positive feedback and increased engagement in proposal topics as the support and comment numbers are in line with the attitude of the participants. As a result, the sentiment analysis of the qualitative data provided several key insights regarding digital asynchronous participation as presented below:

- Each participation module needs to be analysed separately as they contain different data types and interfaces.
- Sentiment analysis results indicate the need for triggering engagement with positive attitudes.
- Higher sentiment scores may indicate future engagement in digital participation, and therefore, these topics require further attention.

4.3. VISUALIZATION OF QUALITATIVE PARTICIPATION DATA: LDA TOPIC MODELLING

In order to enable the use of qualitative data, a huge amount of textual data need to be visualized for decision-makers. In this experiment, the collected qualitative data is visualized through the LDA topic modelling algorithm, which estimates the topic prevalence distributions per entry and the term prevalence distributions per topic. The algorithm uses the corpus (i.e., the collected data consisting of the entries and replies) and the dictionary (i.e., an embedded wordlist in the NLTK library) to build the trained topics with frequent words and their weights. The weighting schema is a numerical statistic which reflects how important a term is to the discussion from the corpus of the experiment (Truica et al., 2016). In other words, the weights of the frequent words indicate their importance in a topic while analysing the data. This statistical measure primarily uses the frequency of a term in a discussion and the frequency of a term in the entire corpus.

The number of topics is the most important parameter to define. It is an ever-present concern to choose the best number of topics in topic modelling as well as in other latent variable methodologies. In this experiment, different numbers are tested in content generation by analysing the keywords of topics to see if they formulate meaningful topics. Eventually, the number 8 is adopted as the topic parameter by combining both statistical measures (weights) and manual interpretation (content). Eventually, we obtain 8 topics that resulted in the LDA topic model. Table 1 presents the LDA model built for one of these 8 topics (i.e., Topic 2) in our data set.

Table I. Most frequent words in Topic 2.

Topic # 2	0.038*"marketing" + 0.030*"budget" + 0.029*"think" + 0.028*"word" + 0.026*"spent" + 0.026*"year" + 0.019*"student" + 0.019*"faculty" + 0.017*"mil" + 0.015*"expense"

As seen in Table I, the most frequent words for the entries in topic 2 are "marketing" and "budget". Many of the entries also indicate the expenses of the university for marketing. Therefore, the topic can be named “financial management”. However, the decision-makers need to observe all the topics and relevant terms together to interpret the collected data. Several visualization methods are deployed for topic modelling, including basic graphs for dominant words in word clouds, or more complex visualizations such as t-Distributed Stochastic Neighbour Embedding (t-SNE) clustering or intertopic distance map as the most referred visualization technique for LDA topic modelling. Figure 5 presents the intertopic distance map of the collected data (relevance metric (λ) = 0.6) based on a multidimensional scaling algorithm. On the left, the area of circles is proportional to the number of words that belong to each topic across the dictionary. Meanwhile, the shorter distance between circles indicates topics that have more words in common. On the right, a bar chart presents the 30 most salient terms, and each bar indicates the total frequency of the term across the entire corpus. When a specific topic is selected as seen in Figure 5, the bar chart displays the most salient words included in that selected topic with a second (darker) bar that presents the topic-specific frequency of words that belong to the selected topic.

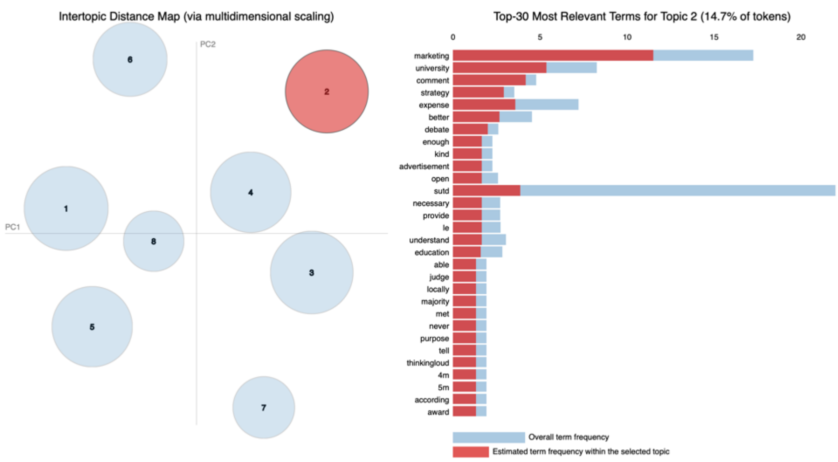

Figure 5. The intertopic distance map of the LDA topic modelling for this experiment (λ=0.6).

The computational analysis and visualization methods are important to convert a large number of dimensions resulting from statistical methods, i.e., LDA topic modelling with an intertopic distance map in this case, to a reasonable number of dimensions. This study presents a valuable and novel way to filter the collected data and visualize textual information to enable decision makers to organize, understand and summarize large collections of qualitative textual participation data.

5. Conclusion

The consequences of the post-carbon cities require major digital shifts towards data collection in participatory urban design practices. In order to achieve a better collaboration for addressing global issues effectively, new data analysis methods are necessary. These methods should support designers with informed decision making in the use of large-scale digital participation data, which contains a huge amount of qualitative data through discussions, debates, and proposals. In this study, we presented the results of an asynchronous digital participation experiment by analysing the textual data obtained from the participants. In order to facilitate the translation of textual data into decision-making processes, we explored several AI-based data analysis methods, i.e., NLP, sentiment analysis, and LDA topic modelling. Eventually, the long-term goal of this study is to provide systematic qualitative data processing and visualization on the entry level, module level, and topic level to support decision-makers in informed urban design practices. This study's focus on underutilized qualitative participation data might also be able to tackle the deployment of new computational methods and technologies in the analysis and visualization of qualitative data in digital asynchronous participation. This preliminary study already promises to transform the ways architects and designers utilize digital participation data in urban design contexts to increase the communication, interoperability, accountability, and quality of representations.

The next step of this study is to implement these techniques in larger data sets to further investigate the analysis of digital qualitative participation data. It will be beneficial to introduce the analysis and visualization methods to designers and decision makers to further study the use of qualitative participation data in early design phases in future works.

Acknowledgements

This research is supported by Singapore University of Technology and Design (SGPCTRS1804) and Singapore International Graduate Award (SINGA). The authors would like to express their highest gratitude to all the participants, who participated in the experiment, for their collaboration and valuable opinions. The authors would also like to sincerely thank Doganalp Ergenc (University of Hamburg) for his contributions especially in the installation of the participation tool in pre-participation phases.

References

Alter, H., Whitham, R., Dawes, F., & Cooper, R. (2019). Learning by design. How engagement practitioners use tools to stretch the creative potential of their citizen participation practice. *13th International Conference of the EAD* (pp. 1387–1397). https://doi.org/10.1080/14606925.2019.1594964

Ataman, C., & Tuncer, B. (2022). Urban Interventions and Participation Tools in Urban Design Processes: A Systematic Review and Thematic Analysis (1995 – 2021). *Sustainable Cities and Society*, 76, 103462. https://doi.org/10.1016/j.scs.2021.103462

Burgess-Allen, J., & Owen-Smith, V. (2010). Using mind mapping techniques for rapid qualitative data analysis in public participation processes. *Health Expectations*, 13(4), 406–415. https://doi.org/10.1111/j.1369-7625.2010.00594.x

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. *Advances in Neural Information Processing Systems 22 - Pro- Ceedings of the 2009 Conference* (pp. 288–296).

Goldsmith, S., & Crawford, S. (2014). *The responsive city: Engaging communities through data-smart governance.* Jossey-Bass. https://books.google.com/books?id=NpI_BAAAQBAJ&pgis=1

Habib, F., Etesam, I., Ghoddusifar, S. H., & Mohajeri, N. (2012). Correspondence Analysis: A New Method for Analyzing Qualitative Data in Architecture. *In Digital Fabrication* (pp. 517–538). Springer Basel. https://doi.org/10.1007/978-3-0348-0582-7_9

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Klein, M. (2011). The MIT deliberatorium: Enabling large-scale deliberation about complex systemic problems. *ICAART 2011 - Proceedings of the 3rd International Conference on Agents and Artificial Intelligence*, 1(1). https://doi.org/10.1109/cts.2011.5928678

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Lobe, B., Morgan, D., & Hoffman, K. A. (2020). Qualitative Data Collection in an Era of Social Distancing. *International Journal of Qualitative Methods*, 19, 160940692093787. https://doi.org/10.1177/1609406920937875

Lock, O., & Pettit, C. (2020). Social media as passive geo-participation in transportation planning – how effective are topic modeling & sentiment analysis in comparison with citizen surveys? *Geo-Spatial Information Science*, 23(4), 275–292. https://doi.org/10.1080/10095020.2020.1815596

Milz, D., & Gervich, C. D. (2021). Participation and the pandemic: how planners are keeping democracy alive, online. *Town Planning Review*, 92(3), 335–341. https://doi.org/10.3828/tpr.2020.81

Rathore, M. M., Paul, A., Hong, W.-H., Seo, H., Awan, I., & Saeed, S. (2018). Exploiting IoT and big data analytics: Defining Smart Digital City using real-time urban data. *Sustainable Cities and Society*, 40, 600–610. https://doi.org/10.1016/j.scs.2017.12.022

Roberts, H., Resch, B., Sadler, J., Chapman, L., Petutschnig, A., & Zimmer, S. (2018). Investigating the Emotional Responses of Individuals to Urban Green Space Using Twitter Data: A Critical Comparison of Three Different Methods of Sentiment Analysis. *Urban Planning*, 3(1), 21–33. https://doi.org/10.17645/up.v3i1.1231

Seifert, A., & Rössel, J. (2019). Digital Participation. In D. Gu & M. E. Dupre (Eds.), *Encyclopedia of Gerontology and Population Aging* (pp. 1–5). Springer International Publishing. https://doi.org/10.1007/978-3-319-69892-2_1017-1

Tan, E. (2016). *The Evolution of City Gaming* (pp. 271–292). https://doi.org/10.1007/978-3-319-32653-5_15

Tekler, Z. D., Low, R., Zhou, Y., Yuen, C., Blessing, L., & Spanos, C. (2020). Near-real-time plug load identification using low-frequency power data in office spaces: Experiments and applications. *Applied Energy*, 275, 115391. https://doi.org/10.1016/j.apenergy.2020.115391

Tomarchio, L., Hasler, S., Herthogs, P., Müller, J., & Tunçer, B. (2019). Using an Online Participation Tool To Collect Relevant Data for Urban Design. *Proceedings of CAADRIA 2019 Intelligent & Informed* (pp. 747–756).

Truica, C.-O., Radulescu, F., & Boicea, A. (2016). Comparing Different Term Weighting Schemas for Topic Modeling. *2016 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, 307–310. https://doi.org/10.1109/SYNASC.2016.055

Wang, Y., & Taylor, J. E. (2018). Urban Crisis Detection Technique: A Spatial and Data Driven Approach Based on Latent Dirichlet Allocation (LDA) Topic Modeling. *Construction Research Congress 2018*, 250–259. https://doi.org/10.1061/9780784481271.025