DIGITAL PARTICIPATION IN URBAN DESIGN AND PLANNING

Addressing Data Translation Challenges in Urban Policy- and Decision-Making through Visualization Techniques

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

Abstract. Digital technologies and online platforms, such as e-participation and crowdsourcing tools, are revolutionizing citizen engagement in urban design and planning by enabling large-scale, asynchronous, and individual participation processes. This evolution towards more inclusive and representative decision- and policy-making, however, presents a significant challenge: the effective utilization of the vast amounts of textual data generated. This difficulty arises from distilling the most relevant information from the extensive datasets and the lack of suitable methodologies for the visual representation of qualitative data in urban practices. Addressing this, the paper deploys AI-based analysis methods, including Natural Language Processing (NLP), Topic Modeling (TM), and sentiment analysis, to efficiently analyze these datasets and extract relevant information. It then advances into the realm of data representation, proposing innovative approaches for the visual translation of this textual data into multi-layered narratives. These approaches, designed to comply with a comprehensive set of both quantitative and qualitative interpretation criteria, aim to offer deeper insights, thus fostering equitable and inclusive governance. The goal of this research is to harness the power of qualitative textual data derived from online participation platforms to inform and enhance decision- and policy-making processes in urban design and planning, thereby contributing to more informed, inclusive, and effective urban governance.

Keywords. Digital Participation, Textual Big Data, Natural Language Processing, Spatial Data Analytics, Data Visualization

1. Introduction

Participatory methods in policy and decision-making have gained increasing importance, emphasizing diversity, inclusivity, equity, and democratic governance.

ACCELERATED DESIGN, Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2024, Volume 2, 201-210. © 2024 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

Citizen participation aims to reflect the values and goals of residents, which are crucial especially in the early design and planning stages (Ataman et al., 2022). However, traditional offline methods such as face-to-face meetings or focused groups often cause bias or limitation to certain groups, leading to exclusion and inequality in citizen participation (Nabatchi, 2010). Accordingly, the advent of digital technologies, including e-participation tools that provide platforms for online debates in natural languages, offers possibilities to overcome such biases by enabling broader and more geographically diverse citizen participation. Despite this, digital participation poses challenges in analyzing the extensive textual data these platforms generate. Traditional analysis methods, which focus mainly on numeric or categorical data, often neglect textual data, which includes essential perspectives from minorities and marginalized groups. Additionally, digital platforms and AI-based methods like Natural Language Processing (NLP) and Topic Modeling (TM) complicate data interpretation and analysis, since they yield statistical insights that may not be intuitively implementable by decision-makers. To harness the full potential of digital participation for inclusive and equitable governance, it is crucial to develop more advanced and comprehensive approaches for interpreting datasets. These approaches should facilitate the meaningful use of participation data, ensuring that all stakeholders, including traditionally underrepresented groups, are included in decision-making processes.

This study explores interpretation criteria and learning potential from participation data and introduces approaches for visually representing large-scale citizen participation datasets. These approaches are crucial for navigating the complexities of today's accelerated design environment, contributing to the discourse on shaping design and planning practices through the utilization of AI-based NLP tools and visualizations for textual datasets representing inhabitants' perspectives. Specifically, the research aims to provide effective strategies for (a) visualizing participation data from diverse sources and varying scales, (b) identifying significant discussion points and navigating impactful perspectives using hierarchical summaries, and (c) ensuring inclusivity in the analysis process while avoiding reliance on subjective pre-selection methods. The study contributes to the existing literature on participatory decision-making processes by highlighting the importance of advanced data analysis and visualization techniques in promoting equitable and inclusive governance.

2. Current Practices in Participation Data Visualization

Current digital participation practices predominantly focus on analyzing and visualizing demographic and numerical features, frequently overlooking crucial post-participation phases where the data is translated into insights for decision- and policy-makers after participation. While these features can yield valuable general statistical insights, they face limitations in effectively representing content and exploring datasets in detail. The use of such voluminous textual data, therefore, has been recognized as one of the biggest challenges of big data analytics (Siegel, 2013). This shortcoming results in certain datasets remaining partially or entirely underexploited, underscoring the necessity for more comprehensive analysis and representation methods that thoroughly address content-related aspects of participation data and provide deeper, multi-layered insights into the underlying information.

Visualization literacy is crucial in decision-making and policy formulation, involving the adept use of data visualizations to interpret information within a data domain (Boy et al., 2014). Traditional methods like parallel coordinates, treemaps, and Sankey diagrams, often used for demographic or geo-locational data, tend to neglect the importance of textual data in participation datasets, risking the loss of valuable insights in higher-level decisions. To rectify this underutilization of textual data, identifying patterns, trends, and anomalies in large-scale participation is essential, integrating systematic filtering and sorting for insightful exploration (Aparicio, 2017). Additionally, AI techniques like NLP, TM, and sentiment analysis in textual analysis require clear explanations of their results in relation to the datasets. Consequently, the visualization of such datasets and analysis results must be specifically designed to suit the dataset's needs, the project's context, and the objectives.

In order to develop effective visualization methods, understanding the strengths and weaknesses of current approaches is vital. Present techniques offer several benefits, including low computational costs, direct insights into demographic and numeric data, and scalability for larger datasets (Valkanova et al., 2015). These methods convert complex data into visually readable formats, aiding an intuitive understanding of information. However, these techniques exhibit significant deficits in fully representing participation data. A key limitation is their lack of relational and hierarchical representation, which are essential for deciphering complex data configurations (Wang et al., 2020). Moreover, their dependence on predefined categorizations constrains their flexibility in processing dynamically evolving data. Another major drawback is their inability to offer interactive features like filtering, zooming, or selecting data subsets for post-participants, limiting in-depth exploration and analysis of participation data. However, the most significant drawback is their focus primarily on quantitative aspects, often neglecting the qualitative dimensions of participation data. This oversight becomes a critical issue when dealing with complex, multi-layered datasets that are not easily understood through traditional reading and scanning methods. Addressing these challenges requires the development of more advanced and versatile visualization techniques. Such methods should facilitate a comprehensive understanding of participation data, encompassing both its quantitative and qualitative dimensions, and be adaptable to meet the diverse requirements of various stakeholders.

3. Hamburg: Project-based Participation

This paper focuses on Hamburg, employing the city as a case study to underscore its systematic execution of a project-based participation system. Such a system exemplifies a targeted strategy for engaging stakeholders, predominantly at the neighborhood level, to address specific local concerns (Simonofski et al., 2021). Generally initiated by designers, urban planners, or local authorities, this methodology solicits direct feedback from the community, which is indispensable for making informed decisions. Focusing on particular areas or urban issues allows stakeholders to gather precise, context-relevant insights that are critical in the conceptual stages of urban development. By utilizing this specific feedback, stakeholders then strategically allocate budgets and resources, ensuring that the community's needs and preferences are given precedence. This approach cultivates a sense of ownership and collaboration among stakeholders and residents, resulting in more resilient and sustainable urban solutions. In this context, Hamburg, as a notable city in northern Germany, is characterized by its rich historical heritage and a diverse population of approximately 1.85 million (*Statistische Jahrbücher - Statistikamt Nord*, 2023). The city's urban framework, comprising seven boroughs and 104 unique neighborhoods, each presenting its own set of challenges and opportunities, calls for a localized, context-sensitive approach to urban design and planning. This approach is crucial in accommodating its residents' varied needs and preferences, rendering Hamburg an ideal case study for examining the challenges associated with translating diverse, community-sourced data into actionable urban design and planning strategies.

3.1. PARTICIPATION PROCESS

The City of Hamburg established the Digital Participation System (DIPAS) in 2016, which exemplifies a proactive response to the increasing need for effective public engagement in urban development (Thoneick, 2021). DIPAS, a comprehensive digital platform, facilitates citizen participation in planning and design processes by enabling them to share opinions and feedback and engage in discussions on various urban development initiatives. Its interactive interface, featuring map integrations, allows users to access spatial information and understand the impact of proposed projects. The platform categorizes discussion topics, simplifying navigation and contribution to relevant conversations. The study focuses on analyzing the structure of a participation dataset from DIPAS, comprising 28 project-based datasets with diverse components like comments, timestamps, likes, geolocations, and categorical information. This diversity is reflected in the range of projects covered, from bicycle rental systems to climate-friendly neighborhoods. DIPAS is open to all Hamburg residents without pre-selection or registration, offering a digital, asynchronous, and anonymous platform for individual contributions. The data collection process involves authorities initiating projects on DIPAS, setting timelines, and inviting public contributions for three to six weeks, after which the content remains visible but closed for further submissions.

3.2. DATA ANALYSIS

Data Pre-Processing: Textual big data analysis requires a corpus, which is a large and structured set of texts. However, raw data often contains irregularities and unnecessary text such as URLs, user mentions (@), abbreviations, and advertisements. In our research, we leveraged the "NLTK" Python library to clean the data by converting all texts to lowercase, tokenizing, removing punctuations, special characters, stopwords, and finally using stemming and lemmatization. We then removed the infrequent words (below 10 occurrences) and generated bigrams and trigrams from frequently co-occurring phrases. Subsequently, the "Gensim" TF-IDF algorithm was implemented to filter out ubiquitous words (low value = 0.03). This comprehensive data cleaning and preparation yielded a refined corpus, structured for effective topic modeling.

Topic Modeling (TM): We used TM algorithms to extract the topics from our corpus. TM is an unsupervised Machine Learning technique, as the labels are retrieved directly from the text without any prior labeling (Blei, 2012). TM algorithms determine the probability of a word or phrase belonging to a specific topic, i.e. a collection of words, based on their similarity. This study deployed the BERTopic model to capture complex relationships between words and phrases in a more accurate way.

Sentiment Analysis: It assesses the balance of positive and negative responses about a particular subject by analyzing diverse textual sources to understand people's attitudes, and feelings toward various entities like services, locations, events, or topics (Liu, 2012). In this study, the SentimentIntensityAnalyzer was utilized to produce sentiment scores ranging from -1.0 to 1.0, reflecting the overall emotional inclination.

Geospatial Data Analysis: Geospatial data enhances the understanding of participation data by providing spatial context and visualizing geographic patterns and relationships. In this study, we used leaflet.js as an open-source JavaScript library to create an interactive map based on the GIS information of each comment and reply.

3.3. DATA INTERPRETATION

After applying TM algorithms to datasets and determining sentiment scores for each comment and reply, a mere tabular presentation falls short in elucidating correlations and hierarchical relationships. To bridge this gap, we developed a set of data interpretation criteria, divided into quantitative and interpretative categories. The quantitative criteria include metrics and numerical results, instrumental in depicting the general attributes of the dataset, dominant topics, and choosing suitable TM algorithms for data visualization. These criteria encompass a variety of topics, such as engagement levels indicated by the number of replies, popularity, and quality reflected in support or like counts, comment clarity and structure through coherence scores, dominant themes, and trends via frequent words or phrases, individual participation levels, and insights into the discussion's pace and flow from response times.

However, to fully grasp the details of large-scale participation data, it is imperative to integrate both quantitative and qualitative measures. This approach captures the essential contextual information. Our interpretative criteria involve content analysis to discourse quality, depth, and perspectives; examination of discussion hierarchies to identify patterns and themes; network analysis to understand community dynamics and participant relationships; time analysis for insights into engagement patterns over time; sentiment analysis to gauge the overall tone and mood; assessment of argument quality and logical structuring; and evaluation of the inclusiveness and range of perspectives in the discourse. Incorporating both sets of criteria fosters a holistic understanding of participation data, thereby refining data visualization techniques and ensuring comprehensive accessibility of individual data for post-participants.

4. Learning Potentials from Participation Datasets

Understanding the insights embedded in participation data, termed learning potentials, is crucial for guiding decision-making and policy formulation. In this process, visual representations offer invaluable opportunities for navigating the diverse perspectives of participants. This section introduces a variety of visualization methods designed to support both qualitative and quantitative interpretations comprehensively.

4.1. TOPICS & CLUSTERS

To systematically explore patterns, trends, hierarchical links, and similarities within massive textual datasets, this study employs NLP methodologies that enable stakeholders to effectively navigate the data and provide crucial insights for decision-makers and policy formulators. These methods facilitate a preliminary comprehension of the participation data, laying the groundwork for further detailed investigations.

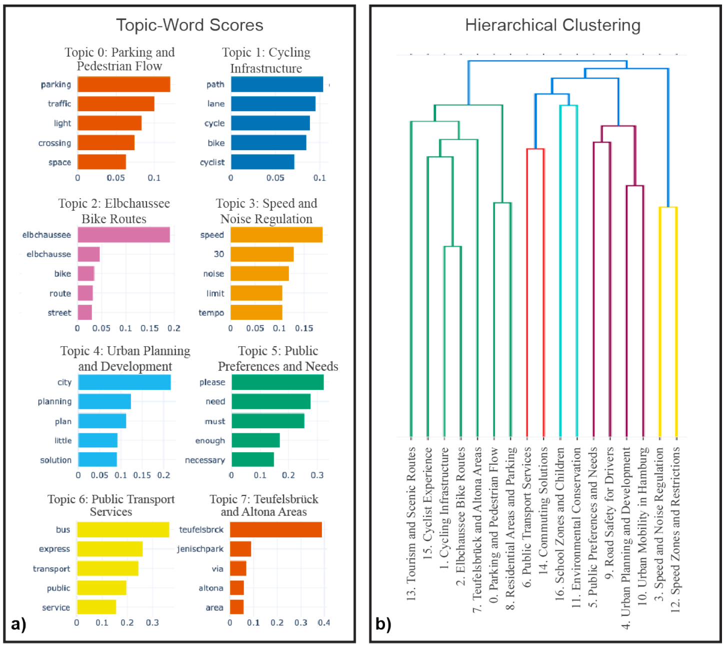

Figure 1: Illustration of topics and clusters based on TM results (a) Topic word scores for the first 8 topics and (b) Hierarchical clustering graph.

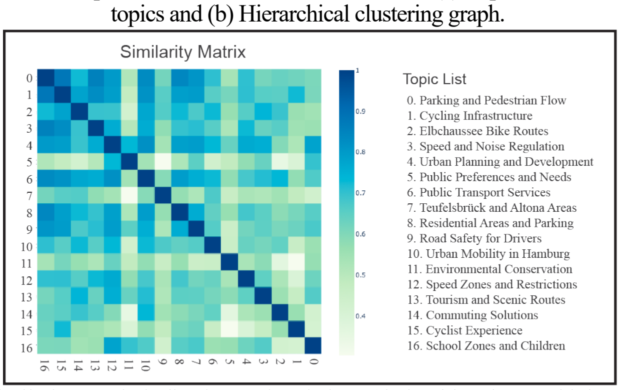

Figure 2: Similarity matrix indicating topic relations (darker blue denotes higher similarity).

In the analysis process, stakeholders initially gain control over the topic number parameter. Upon setting this parameter to 17 topics, three visualizations of the TM results are developed. Among these, bar charts emerge as a simple yet effective tool for illustrating topic-word scores (Figure 1a), displaying topics with their respective keywords and significance scores for a clear overview. Further, the hierarchical relationships among these topics are depicted through a hierarchical clustering graph (Figure 1b), delineating root topics and their connections to reveal the underlying themes. For example, at a higher level of the hierarchy, one might find overarching themes such as “Cycling,” which could further be subdivided into more focused categories like “Experience,” “Bike Routes” and “Infrastructure.”

To quantitatively evaluate semantic similarities, a Similarity Matrix (SM) is generated, employing cosine similarity calculations for each pair of topic vectors (Figure 2). This technique aids in identifying topic clusters with similar semantic meanings by quantifying a similarity score on a scale from 0 to 1. For instance, the "Elbchaussee" dataset (Figure 2) reveals a substantial similarity score of 0.88 between the topics of ‘Parking and Pedestrian Flow’ (T0) and ‘Cycling Infrastructure’ (T1). Such a high degree of similarity indicates that these issues are frequently discussed together and are conceptually interconnected in the public discourse.

4.2. GEO-SPATIAL ANALYSIS

The geographic information embedded in discussion topics of participation datasets is critical, enabling location-specific analysis and correlation at the neighborhood scale. This study has devised a map interface to provide insights into the geographical and temporal distribution of contributions across various projects in the DIPAS dataset.

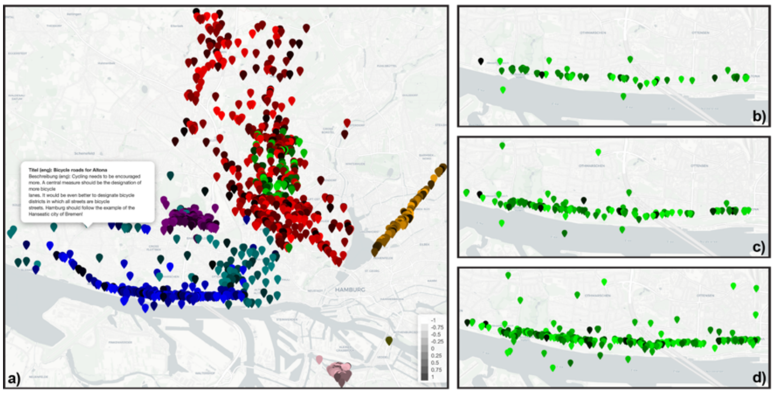

Figure 3: Map interface displaying eight project-based participation datasets in Hamburg (a), and the temporal evolution of the Elbchaussee project on March 3 (b), April 4 (c), and April 17 (d), 2018.

This interface facilitates three analytical approaches:

(i) It displays comment locations on a map to identify engagement patterns and trends in different neighborhoods. This feature aids in pinpointing outlier engagements and mapping areas of conflict and consensus as depicted in Figure 3a.

(ii) The temporal evolution of contributions is analyzed to track the engagement levels over time. This analysis allows stakeholders to uncover new themes or escalating conflicts by identifying key moments affecting participation, correlating events with temporal patterns, and assessing the continuity of participation (Figures 3b, 3c, and 3d).

(iii) Emotional resonances in community engagement are integrated to combine sentimental, spatial, and temporal dimensions. This approach is vital for evaluating argument quality and perspective diversity on a location-specific basis. It informs strategies for promoting cross-community collaboration and dialogue (Figure 3).

In this interface, actual comments and replies are accessible by hovering over the respective location pins to enable content analysis as displayed in Figure 3a. Different colors represent various projects, while shades indicate the sentiments of each contribution, providing a comprehensive view of participation data.

4.3. MULTI-LAYERED INFORMATION ANALYSIS

To enhance the understanding and navigation of massive textual datasets, this study lastly introduces Sunburst Diagrams (SD) and Text-Network Diagrams (TND). These techniques dissect the structure and content of online discussions, highlighting the hierarchy of topics, comments, and responses along with embedded numerical data. They offer stakeholders an intuitive way to analyze the content and analytical depth.

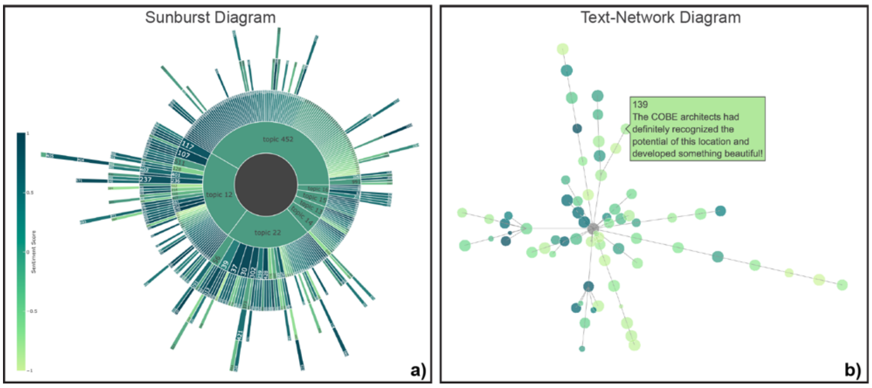

Figure 4: Hierarchical Data Analysis in Digital Participation (a) SD displays sentiment-driven engagement and (b) TND shows discussion dynamics with varied edge lengths and node sizes.

Both SD and TND are designed to depict hierarchical interactions in online debates across projects, enabling analysis of discussion clusters and facilitating comparisons at project and topic levels as in Figure 4. These diagrams graphically represent engagement levels by visualizing the initial (root) comments and subsequent responses (branches). Complex graphs indicate active communities while simpler ones suggest less engagement. They also help identify the breadth or specificity of discussion topics: broader topics garner more responses while niche subjects attract fewer but more detailed responses. Another distinctive feature is their color-coded sentiment scores for each comment and response, mapping the emotional tone of discussions. Positive scores indicate agreement or support while negative scores denote disagreement or criticism. This feature aids in identifying sentiment shifts and their impact.

Furthermore, each diagram type brings unique advantages. TNDs utilize edge lengths to indicate time gaps between comments and responses to reveal the discussion dynamics (Figure 4b). Short edges suggest rapid engagement while longer ones indicate slower interaction. Similarly, the node sizes in TNDs correlate with comment length, indicating the depth and quality of contributions. In contrast, SDs excel in displaying sentiment scores' impact on engagement levels (Figure 4a). They align comments according to sentiment scores at the top hierarchical level, enabling easy identification of the most positively received comments in a discussion.

In conclusion, SDs and TNDs are effective tools for stakeholders to analyze and interpret datasets at project and topic levels. They represent multilayered quantitative and qualitative information and provide insights into patterns of online discussions. Such insights guide communication and collaboration among community members, policymakers, and decision-makers in diverse and complex urban environments.

5. Discussion

So far, this paper has focused on the advantages of project-based digital participation, highlighting specific benefits like precise community feedback and enhanced public trust. Concurrently, it acknowledges challenges such as the potential overshadowing of broader development objectives and the emergence of stakeholder fatigue. To mitigate these challenges, the paper advocates for the application of visualization methodologies that amalgamate project-based participation within an expansive urban context. These methodologies aim to transform human-generated data into actionable insights, thereby harmonizing the advantages of project-based participation with overarching objectives.

Table 1. An overview of the recommended visualization techniques, categorizing them by qualitative and quantitative interpretation criteria and specifying the scale at which each is most effective.

		Thematic Analysis			Geospatial Analysis (Map-based)	Multi-Layered Analysis	
		Clustering	Similarity	Hierarchy		Sunburst	Text-Network
Quantitative Criteria	Topic Number	x	x	x			
	Reply Number					x	x
	Support/Like Number					x	x
	Sentiment Scores				x	x	x
	Frequent Words	x					
	User Activity				x		x
Qualitative Criteria	Content Analysis				x	x	x
	Discussion Hierarchies			x		x	x
	Network Analysis						x
	Time Analysis				x		x
	Quality of Arguments				x	x	x
	Diversity of Perspectives	x			x	x	x
Analysis Scale	Participation Dataset	x	x	x	x		
	Participation Project	x	x	x	x	x	
	Discussion Topic					x	x

Subsequently, the paper delineates the development of visualization systems, emphasizing the necessity of integrating quantitative and qualitative interpretation criteria in digital participation. Quantitative metrics offer a foundation of objectivity and facilitate the analysis of large data sets, while qualitative metrics contribute depth, focusing on the nuances of social dynamics and public sentiment. The synthesis of these metrics within data visualization techniques enables accelerated interpretation and supports evidence-based decision-making in urban development.

In order to achieve this interpretation, this study introduces three analytical techniques accompanied by six visualization methods, each encompassing a variety of interpretative criteria. The first technique employs NLP and TM to analyze textual data, utilizing charts and matrices for semantic analysis. The second technique integrates spatial, temporal, and emotional dimensions, utilizing geolocation metadata and sentiment scores to enhance the understanding of public engagement. The third method uses SD and TND to dissect complex narratives, integrating structural, emotive, and temporal factors into the visual analysis. Table 1 provides a comprehensive guide, elucidating the types of information captured by these visualization systems. This guide aids stakeholders in navigating vast textual datasets, facilitating informed decision-making and policy formulation following digital citizen participation.

6. Conclusion

Consequently, this paper highlights the crucial role of advanced visualization tools in project-based digital citizen participation. The main objective is to equip stakeholders engaged in data-centric decision-making with a system that enables efficient navigation within datasets, yielding practical insights. The next phase of this research involves integrating these visualizations into a cohesive data visualization system. This system will be linked with a participation framework to enhance the interconnectivity between various elements. Such tools will enable democratized access to intricate data, improve interpretability, and promote informed, inclusive decision-making. Further improvements in visualization methodologies will significantly contribute to digital participation processes, supporting a responsive approach to urban transformation.

Acknowledgments

This project is partially funded by the Singapore International Graduate Award (SINGA), the CGH-SUTD grant CGH-SUTD-HTIF-2022-001, and the Singapore Ministry of Education (Tier 2) grant T2EP20221-0020.

References

Aparicio, G. (2017). Data-Insight-Driven Project Delivery: Approach to Accelerated Project Delivery Using Data Analytics, Data Mining and Data Visualization. *Disciplines & Disruption: Proceedings of the 37th ACADIA* Conference, 102–109.

Ataman, C., Tunçer, B., & Perrault, S. T. (2022). Asynchronous Digital Participation in Urban Design Processes: Qualitative Data Exploration and Analysis with Natural Language Processing. *POST-CARBON - Proceedings of the 27th CAADRIA Conference*, 383–392.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

Boy, J., Rensink, R. A., Bertini, E., & Fekete, J. D. (2014). A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1963–1972.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167.

Nabatchi, T. (2010). Addressing the Citizenship and Democratic Deficits: The Potential of Deliberative Democracy for Public Administration. *The American Review of Public Administration*, *40*(4), 376–399.

Siegel, E. (2013). *Predictive analytics: The power to predict who will click, buy, lie, or die*. Wiley.

Simonofski, A., Vallé, T., Serral, E., & Wautelet, Y. (2021). Investigating context factors in citizen participation strategies: A comparative analysis of Swedish and Belgian smart cities. *Int. J. of Information Management*, *56*, 102011.

Statistische Jahrbücher—Statistikamt Nord. (2023). https://www.statistik-nord.de/presse-veroeffentlichungen/statistische-jahrbuecher

Thoneick, R. (2021). Integrating online and onsite participation in urban planning: Assessment of a digital participation system. *Int. J. of E-Planning Research*, *10*(1), 1–20.

Valkanova, N., Jorda, S., & Vande Moere, A. (2015). Public visualization displays of citizen data: Design, impact and implications. *Int. J. of Human-Computer Studies*, *81*, 4–16.

Wang, Z., Sundin, L., Murray-Rust, D., & Bach, B. (2020). Cheat Sheets for Data Visualization Techniques. *Conf. on Human Factors in Computing Systems - Proceedings*.