A Social Citizen Dashboard for Participatory Urban Planning in Berlin: Prototype and Evaluation

Daniel Fürstenau	Flavio Morelli	Kristina Meindl	Mattias Schulte-Althoff	Jochen Rabe
Copenhagen Business School	FU Berlin	FU Berlin	FU Berlin	TU Berlin
dfu.digi@cbs.dk	flavio.morelli@	kristina.meindl@	matthias.schulte-	rabe@tu-berlin.de
FU Berlin	fu-berlin.de	fu-berlin.de	althoff@fu-berlin.de	

Abstract

Participatory urban planning enables citizens to make their voices heard in the urban planning process. The resulting measures are more likely to be accepted by the community. However, the participation process becomes more effortful and time-consuming. New approaches have been developed using digital technologies to facilitate citizen participation, such as topic modeling based on social media. Using Twitter data for the city of Berlin, we explore how social media and topic modeling can be used to classify and analyze citizen opinions. We develop a Social Citizen Dashboard allowing for a better understanding of changes in citizens' priorities and incorporating constant cycles of feedback throughout planning phases. Evaluation interviews indicate the dashboard's potential usefulness and implications as well as point to limitation in data quality and spur further research potentials.

1. Introduction

Urban planning continues to change through pervasive digitalization [1, 2, 3]. Some projects, including Pulse Lab Jakarta, use open source tools to involve the community in shaping public spaces [4]. Sidewalk Labs Toronto has also fostered open discussions using digital tools, such as public displays showing which data is collected or streams of planning meetings [5]. Interactive urban planning tools such as the City Matrix from MIT Media Lab use deep learning to support decision-making processes [6]. Taken together, these examples highlight a diversity of approaches using digital technologies for *participatory urban planning*. However, while social media has become a common way for citizens to express their opinions, the use of social media for participatory citizen planning has been underrepresented in the literature on smart cities [7]. One method that particularly lacks research in the context of urban planning is topic modeling – a group of machine learning algorithms allowing to recognize

thematic clusters in large volumes of texts. These thematic clusters can be used to further analyze citizen opinions and predict new topics [8]. Although topic modeling was used in urban planning [9], its usage differs widely from the classical use case of thematic classification. The lack of participation in the urban planning process shows a definitive need for tools that can bridge the communication gap between city officials and citizens. So far, there is little understanding on how to apply topic modelling in this context in order to address these shortcomings.

The purpose of this paper is to *develop a Social Citizen Dashboard* supporting in citizen participation and urban planning for the case of the city of Berlin. The approach focusses on social media data using a topic modeling approach. This is important because it has valuable implications for the city planning process, as it utilizes the known potentials of social media and topic modeling for the specific needs of the city. We want to give a starting point on how to integrate these tools into the planning cycle and encourage their use as a participatory measure.

In the context of the city of Berlin, a prototypical dashboard has been implemented that visualizes social media data and allows analyses according to three views: a topical, an in-depth, and a spatial view. Demonstrating the approach, a corpus of approximately 250,000 Twitter tweets was compiled over a period of two months and examined using the Latent Dirichlet Allocation (LDA) algorithm [10]. Twitter was chosen because this platform is freely accessible, widely used, and offers citizens untapped potential for participation in urban planning. The Twitter platform also provides a free, although limited, Application Programming Interface for streaming tweets. The results show how the spatial distribution of those different topics in a city can be visualized and analyzed. It is possible to find out which topics are particularly relevant for citizens in a period and district. This brings together different perspectives and offers one integrated solution, which goes beyond existing participatory measures in Berlin and involves planners and citizens more actively. Since

urban demands and conditions are different in each city, the solution presented in this paper was fitted to the context at hand. The dashboard provides an initial overview on informal opinions expressed by twitter users. Another potential is its usage for campaign management by directly targeting and filtering for specific city projects, e.g. bike-lane construction.

Five interdisciplinary evaluation interviews as well as a focus group provide insight into the usefulness of the dashboard and indicate further research opportunities in the future. The implications for participatory urban planning processes are considered as are limitations of the data and methods used.

2. Participation in Urban Planning

A participatory approach means involving citizens in urban planning – either by including them directly in shaping the city, such as in remodeling districts, or by collecting and identifying relevant data for the urban planning process [2]. Participation counteracts the disconnect between citizens, experts, and politicians, in that citizens are not seen as mere test persons or consumers of urban space. Although participation may be time-consuming, it enables a consensus on complex and conflicting opinions. This view may, however, be biased towards groups that make their voice heard the most, or more drastically, exclude certain groups from the planning process altogether [11, 12, 13]. Therefore, the resulting consensus may not reflect the actual opinions of citizens and the achieved consensus can be fragile [14]. What is needed, therefore, are solutions developed through participation that are based on the actual priorities and wishes of the community. Solutions that may then be more widely accepted when implemented, due to their initially higher transparency within the planning process [15]. To achieve this, some researchers have particularly highlighted the "ongoing dispute for words, meaning, discourses, visions" [16], which is also a central issue in social media. In addition, social cohesion is promoted by a common vision, in which all stakeholders are involved [4]. For a successful urban planning project, it is crucial that participation is not only a tool of politicians to placate or manipulate citizens, but that citizens are treated as true partners that are directly affected by the project [17]. In addition, participative measures can be useful in gathering an early understanding of citizen's informal attitudes, which goes beyond institutionalized opinions [18].

The main stakeholders in the urban planning process are citizens, (communal and private) companies, NGO's, and administrations (Figure 1). Because

administrations have a democratic and legal mandate, they have clear guidelines when it comes to planning projects. Therefore, the mode of interaction between them and other stakeholders is largely predetermined. On the other hand, there is potential for conflict between citizens and companies, due to often differing priorities. While companies are very sensitive to issues that might affect their long-term profitability, citizens may be more concerned with inclusiveness and developing a sense of belonging. Achieving social cohesion within their communities and neighborhoods can be of great interest, as it impacts citizen's individual well-being [19]. Participatory planning can, if done correctly, foster trust between the stakeholders, thus avoiding conflicts at an early stage in the planning process.
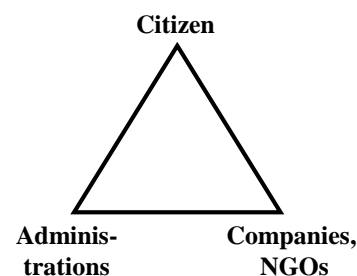
Figure 1. Stakeholders in urban planning

The traditional urban planning process has few participatory elements. It often lacks transparency and open discussion since there may not be sufficient channels for interaction between the public and official actors [18]. However, cities are beginning to incorporate different technologies to increase citizen participation. For example, the city of Glasgow has started multiple smart city initiatives to increase public engagement [20]. This included a so-called Open City Dashboard, which provides users with real-time information about Glasgow. In general, city dashboards can be used to monitor different activities within cities, such as traffic, housing, cultural life or citizen's opinions. They receive data from multiple sources and use information with the purpose of visualization, analysis or control [21]. Social media represents an important source of data in this context because it holds information about the location, behavior, and sentiment of its users. It has shown to be a promising tool for timely and cost-effective citizen engagement. Use-cases that incorporate social media data have been researched for different themes and cities around the world. Zhou et al. [22] analyze data from the online social network WeChat to detect cultural demand patterns in the city of Beijing. Ye et al. [23] explore the spread of rumors on social media in relation to the Ebola outbreak in two Chinese cities.

The degree to which urban dashboards facilitate citizen participation varies and depends on different factors. Some projects support a two-way flow of information between citizens and administrators. They can be open to the public and accessible via a website. Others are solely used for the decision making of one group, e.g. city planners [21]. The application of such dashboards is highly contextual and must be fitted to the demands of a city. For the case of Berlin, there has not yet been a sufficient approach to incorporate informal citizen attitudes into the urban planning process – e.g. as reflected in social media data. Thus, there seems to be a lack of understanding on how to make this kind of information accessible to stakeholders like city administrators. Further insights into topics and emotions regarding urban life would likely promote social-cohesion and citizens' individual well-being.

In this regard, our project aims to develop a *Social Citizen Dashboard* for the case of the city Berlin. The aim is to use data from social media as a novel form of citizen participation. This can be a complement to traditional participation formats like citizen workshops. Be it in the form of advanced visualizations and decision support tools or continuing the discussions in the virtual realm. To this end, we focus on data from social media platforms. Such platforms have been heralded to foster social exchange and communication, promoting social contagions [24], but have also been criticized as giving a platform to populists or promoting echo chambers [25]. Nevertheless, these platforms produce large amounts of data, which can be potentially employed towards meaningful ends in the city planning process. This requires additional analytical capabilities regarding the thematic, spatial, and temporal aggregation of data to which this paper contributes. It is concerned with the question of whether topic modelling of social media data can be used to identify citizen's priorities and to present the results in a dashboard as an additional source of information in the urban planning process. We will discuss the implications of this approach. The user group consists primarily of city planners, including engineers, and other administrative personnel. In a next step, it should be opened to citizens as well as private and communal companies.

3. Methodological Approach

3.1. Research Context

Our research emerged within the context of interdisciplinary research involving urban planning, digitalization and sociological/ psychological

scholars in Berlin. It was spurred by a lack in opportunities for citizens to participate digitally in the planning processes. First, one of the authors, an urban planning and digitalization scholar, was involved in developing a digital end-to-end solution for participatory urban planning within a research project. This author was interested in exploring the use of social media for participatory planning, which sparked first explorations into the topic. The solution developed in this paper was initially supporting his project directly and was co-developed with the project's team. Building on the results, another author, an information systems scholar and digitalization scholar, took the approach further by preparing its use in an interdisciplinary project conceptualizing an urban transformation map. In this context, the approach was evaluated through interviews and a focus group regarding its potential and limitations. Together, these two projects present a case in which a solution to extract social media data and synthesize insights from it through topic modeling emerged, which used exemplary data from Twitter as the backdrop of the approach.

3.2. Developing the Social Citizen Dashboard

Figure 2 shows the methodological steps for developing our Social Citizen Dashboard based on social media data and topic modeling in the context of participatory urban planning. As explained above, the **conceptualization** took place in the context of a research project on participatory urban planning. After the initial idea emerged, one of the authors conducted interviews and requirements analysis workshops with the project team, over a period of approximately 6 months in 2017. As part of the requirements analysis, the author also took part in a smart city hackathon to generate ideas and test initial design solutions. This informed the **development** phase, mainly in 2018, focusing on collecting data from a pilot case study, modeling data, and building visualizations and analyses, especially in the form of the Social Citizen Dashboard.
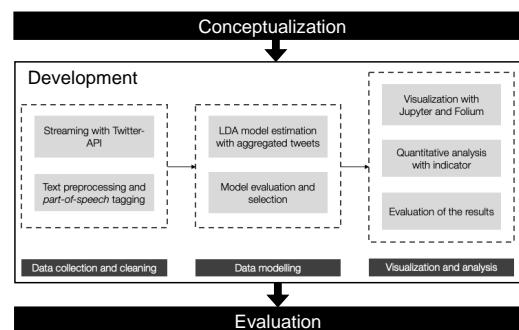

Figure 2. Methodological steps

Data collection and cleaning. Building the Corpus with the Twitter API, a corpus of tweets was compiled from August 14, 2018 to October 15, 2018 using the Twitter streaming API and it contains a total of 250,028 tweets, of which approximately 100,000 are in German. The collection of tweets is limited to the city of Berlin. This way, the methods developed in this paper can be tested directly on data from a specific city in Germany. In addition, there is a broad spectrum of Twitter users in Berlin. The individual tweets are transmitted by the API in JSON format. Each tweet object equivalent to one posted tweet has numerous attributes [26]. Since participatory planning is always related to the geographical level, attributes such as place and coordinates are particularly relevant for citizen participation. Although the free API provides only a fraction of all tweets, the sample is still representative of the population [27, 28].

Then, texts were preprocessed, and part-of-speech tagging was conducted. Text cleaning is necessary for the application of text mining algorithms. First, the individual messages are processed so that they contain only characters and spaces. URLs are not considered to be interpretable and are deleted, tweets with only image or video are ignored and user handles are removed. All tweets under 30 characters are ignored because they have a low information content. Very frequent words in the German language (stop words) are ignored. To avoid overfitting, only words that occur at least four times in the corpus but not more than 60% of all documents in the corpus are considered. The quality of the learned topics can also be improved by filtering out word categories with little or no information content. It was shown that the restriction of selected words to nouns increases the coherence of the learned topics [29]. In this paper, all models will therefore be trained with nouns only. A part-of-speech tagger is used for the German language, so that words can be identified as nouns, verbs, adjectives, adverbs. etc. [30]. No lemmatization or stemming was used.

Data modeling. Topic Modeling is a category of algorithms that can be used to recognize themes in large volumes of documents [31]. One of the most common topic modeling algorithms is LDA [10]. Applying Topic Modeling to Twitter data poses certain challenges. These can be traced back to the short length of the tweets. Initial approaches to solving the problem of short texts were based on aggregation. Two forms of aggregation were tested: tweet aggregation by user and tweet aggregation by a particular word. In both cases, it was shown that aggregation can lead to a better LDA model [32]. In this paper, the implementation of LDA of the GenSim library is used together with the aggregation by user. This is a natural choice of aggregation, as the citizen is the focus in this participatory approach. Alternative tools for topic modeling are, for example, Java-based Mallet or BigARTM (Python) [33].

Visualization and analysis. Dashboards are a suitable means to communicate results through visualization and interaction [34]. The technical aspect of topic modeling becomes accessible for laymen in the context of urban planning. The creation of a dashboard should fulfil three criteria. First, the results should be interactive to facilitate working with the data. Second, not only the structure of the topics should be visible, but also filtering options with respect to relevance and period should be provided. Third, the spatial component should be considered in the visualization, since citizen participation primarily takes local interests into account.

Evaluation of the LDA models and interpretation of the learned topics are two challenging aspects of topic modeling. Topics are ultimately evaluated by the human sense language perception, even though the topics themselves are generated by purely statistical methods. The first indicator for the evaluation of topic models is perplexity. A lower perplexity indicates a better model [10]. Other metrics try to quantify the coherence of the learned model: Röder et al. [35] propose a new indicator called CV. It estimates the coherence of topics and model, which is why it is also called Coherence Score. Higher values of CV indicate a higher coherence.

The evaluation of the results is not limited to the visual level. By employing topic modeling, quantitative statements can be made about the collected corpus. For this purpose, an indicator that represents the time change of the individual topics in different parts of the city is developed. Thus, trends in the priorities of citizens can be uncovered, which can serve as a basis for decision-making in urban planning. Based on the achieved results, the quality of the Twitter data and their suitability for the urban planning process can be discussed. This question arises because of the wide range of topics and user groups represented on Twitter.

The added value of the developed approach to urban planning was then considered within the **evaluation phase**, which took place in 2020. This phase drew on 5 interviews with experts from urban planning, digitalization, and psychology. The interview structure was based on Meuser and Nagel [36]. The interviews proceeded by (1) asking about possible users and usage scenarios for the dashboard. Next (2), the interview partners addressed current shortcomings in the urban planning process. Furthermore (3), we asked how the dashboard can be

incorporated in the urban planning process. We then (4) explored suggestions to increase the practical usefulness of our dashboard. Finally (5), we asked for limitations of the dashboard. The insights showed that the dashboard is a promising source of information for the city of Berlin. Concurrently, the evaluation sparked interesting discussions surrounding its practicability and possible limitations. This offered us valuable starting points for possible improvement, which were further elaborated in a focus group at the City Lab Berlin in July 2020. We presented the dashboard to a focus group and gathered insights with an open discussion and questionnaires. Table 1 contains a description of the five interview partners and ten focus group participants.

Table 1. Description of interviewees and focus group participants

N=5	Interviewee field	Expertise
I-1	Psychology	>10 years
I-2	City Planning & Governance	> 5 years
I-3	Urban Planning & Design	> 5 years
I-4	Business & Data Analytics	> 2 years
I-5	Sociology of Technology	> 2 years
N=10	**Represented disciplines**	
Focus Group	Director Commercial Properties (1), Digital Sales Manager (1), Digital Urban Designer (1), Project Manager Urban Planning (1), Business students (4), Citizen Volunteers (2)	

4. Results

4.1. Elements of the Social Citizen Dashboard

We develop a Social Citizen Dashboard for the city of Berlin. To begin, we explore the question of an intuitive interpretation and communication of the results. The front end was created in the form of a Jupyter notebook and consists of three parts: (1) an overview of the topics, (2) a detailed view into the topics, and (3) a geographical heat map.

The *topic overview* is visualized using the Python module LDAvis [37]. It offers an intuitive visualization of the topics of the selected model, such as politics or leisure (see Figure 3). First, it displays the frequency of the most important words for a topic and in relation to the whole corpus. Second, the representation of the proximity of the different topic clusters to each other (Jensen-Shannon divergence) and the number of associated tweets through the circumference of the circle is particularly useful to understand the underlying structure. The reduction to two dimensions is performed by principal component analysis (PCA). In addition, there is the option of

clicking on individual words, as opposed to topics, on the right-hand side. The frequency of a word in a topic is represented in this case by the circumference of the circles of the individual topics on the left side of Figure 3.

The *detail view* of the tweets on a certain topic can be filtered according to criteria like relevance and period. The relevance refers to the distribution of the various topics of each document (in this case, each aggregated user profile) and indicates the minimum probability that the respective topic must have in a document in order to be displayed. The time period is important because topics change over time – for example, before and after a state election. In addition, further statistics are provided, such as the number of tweets and the number of users.

The *geographical view* locates the tweets that contain exact coordinates with the library Folium (see Figure 4). The map of the city of Berlin is based on OpenStreetMaps. As there are several thousand tweets with coordinates, simultaneously displaying all tweets would be overwhelming. Folium offers an elegant solution: the tweets are clustered by region and replaced by a small circle with the number of tweets in each region. Granularity increases with zoom until only individual tweets are displayed. The small circles indicating the number of tweets act as a heat map: this allows to quickly identify those districts in which people are tweeting the most.

4.2. Demonstration

The Social Citizen Dashboard is now demonstrated using data collected for Berlin from the period from August 14, 2018 to October 15, 2018. Figure 3 gives an overview of collected topics for this period.

Figure 3. Topic overview visualization

As shown on the right side of Figure 3, a topic contains a top list of terms most relevant to itself. Here, Topic 17 includes terms like "Merkel", "SPD", "Bayern", "CSU", or "Bundestag". These words are associated with the general theme politics. Another

topic containing, for example, words such as "weekend", "Friday", "night", or "festival" could be associated with leisure. For most topics, one can quickly decide on an appropriate title by looking at the corresponding top terms. For others, further interpretation is necessary.

Next, we demonstrate how the approach can be used to analyze the time and spatial distribution of tweets quantitatively. For this purpose, all tweets with exact coordinates (approximately 17% of the 100,000 German tweets) were grouped by locality.

Figure 4. Geographic visualization with Folium for Berlin-Kreuzberg

The selected localities were: Kreuzberg, Charlottenburg and Schoeneberg. Figure 4 shows the heat map for a location in Berlin-Kreuzberg. The tweets were then divided into four two-week periods, starting August 20, 2018. The following indicator was calculated for each topic in each subgroup:

$$t_k = \frac{1}{n_v} \sum_{i=1}^{n_v} p_{i,k}$$

Where n_v is the number of tweets in subgroup U and $p_{i,k}$ is the probability estimated by the LDA algorithm that the i-th tweet will be assigned to topic k. The indicator describes how a topic k is represented in a subgroup. Only the tweets of users posting between three and 300 tweets during the entire period (mid-August to mid-October) were considered. Because of the aggregation, the results refer to users, not to individual tweets, and are summarized in Table 1. They are exemplified by topic 7 (the Jewish Museum, a site located in Kreuzberg) and 17 (politics).

First, the number of tweets in each subgroup should be noticed. In localities with a younger demography and a more active cultural life such as Kreuzberg, there is much more tweeting than in more familial localities such as Charlottenburg and Schoeneberg. This highlights one challenge in collecting enough tweets for all districts. Due to the relatively small number of tweets in each subgroup, caution regarding changes is advised when interpreting the table. In addition, a lower number of tweets

during the second period can be observed in all three districts. There are two possible explanations for this anomaly. The first being, that fewer tweets with exact coordinates were provided by the streaming API at this time. Second, it could be that most of the tweets with coordinates were posted during these two weeks mainly those identified to be more active users, such as news organizations that were excluded due to filtering of users with over 300 tweets.

Table 2. Topics 7 and 17 in selected city districts in relation to total amount of tweets and changes in relative importance over time

Time period	8/20 - 9/3	9/3 - 9/17	9/17 - 10/1	10/1 - 10/15
Kreuzberg District				
Tweets	**150**	**36**	**129**	**128**
Topic 7	0.067	0.087	0.078	0.152
Topic 17	0.037	0.029	0.041	0.03
Charlottenburg District				
Tweets	**90**	**43**	**70**	**82**
Topic 7	0.057	0.12	0.017	0.068
Topic 17	0.036	0.05	0.042	0.034
Schoeneberg District				
Tweets	**65**	**20**	**33**	**49**
Topic 7	0.044	0.045	0.084	0.061
Topic 17	0.044	0.05	0.051	0.027

To identify significant changes, a threshold of 0.05 was defined for the difference between the smallest and the largest value for a topic in a district. If the fluctuation remains within this range, this is interpreted as no change. The first thing to notice is that, generally speaking, the ratios of the topics remain constant with some exceptions. Most significant changes can be observed in Kreuzberg. Here, one topic with a significant increase in the fourth period is topic 7. The increase in the proportion could be due to an increased number of visitors or increased interest in the museum. One possible explanation may be the holding of two symposia during this period [38, 39]. Apart from this, the distribution of tweets across different districts and time periods is relatively stable. This means that the districts show similar patterns regarding their tweeting behavior.

The indicator t_k developed in conjunction with the subgroups is relevant because it combines three central dimensions of urban planning: space, time and topic. By using topic modeling and coordinates, developments in a district can be analyzed in a targeted manner. The thematic development is important as it reflects the changing priorities of citizens. This information can be used to develop targeted urban planning measures or serve as a basis for interaction with the citizens of a neighborhood.

4.3. Evaluation

Results of the evaluation interviews and focus group confirm that in its *current state of development*, the dashboard's main promise is that it acts like a "fever curve" (Interviewee 1). It shows the rise and fall of certain topics and emotional states. This can be a useful tool in bridging the communication gap between citizens and administrators. It is important to know about citizen's subjective attitudes regarding a planning project since they have an impact on its successful execution (Interviewee 3). However, this kind of information often does not reach authorities. One reason being that channels for citizen participation in Berlin tend to be rather formal and tedious (Interviewee 2). Our dashboard can help to achieve timely updates about public opinion. Relatedly, it offers a detailed look into the city's districts while giving the possibility to filter for specific tweets and hashtags. This way, one can detect core activities for a location of interest. How many people frequent "Boxhagener Platz" at different times in the day (Interviewee 3)? What do they think about a new construction project? Or what do they think about charging points for electric cars in their neighborhood (Interviewee 2)? Is there a district that is completely overcrowded with moviegoers (Interviewee 4)? How do people enjoy bike lanes in Kreuzberg that have been reopened (Interviewee 1)? Answers to such questions can give valuable insight into the dynamics of a neighborhood and by doing so inform the approach to its remodeling. As one interviewee (3) summarized: "Here the city planner gets a tool where he can see certain issues in a neighborhood. So, it has more of a *scouting* function. I can take a quick look at how important a square or street actually is. Is it on people's radar, i.e. how relevant is it?"

Regarding *shortcomings in the current urban planning process*, interviewees and focus group participants noted that informal voices are underrepresented in the process of urban planning. As one interviewee (3) put it: "Urban planners are already trained to be aware of what is happening outside, in other words, social life. But if you want to be close to people, you always rely on neighborhood management methods, meaning you reach out to registered associations and societies. But these are all institutionalized opinions. So, we need tools that can show us the more informal topics, the status quo. Social media can bring this to light." One interviewee (2) noted: "There are a lot of participative initiatives, but they are fragmented into different communities. Left-wing activists, district offices, civic activists and so on … a lot happens in the informal sector".

Interviewees further noted that what's especially beneficial is identifying new needs of the younger generations (focus group participant), and that traditional participation elements, such as workshops often fall short in providing active participation (focus group participant).

Regarding *potentials for incorporating our dashboard into the urban planning process*, interviewees noted the potential to align needs of citizens with priorities and measures taken (focus group participant). As one focus group participant noted, "On Twitter you can also recognize those who reveal something without being asked", meaning informal and unexpressed needs. Interviewee 1 expressed that, "it would be good for reaching digital communities. […] People that tweeted on the same topic, the same political message. A case where apparently the same tweet is running through the network. With topics like Corona lockdown, opening of daycare centers, etc., which are being tweeted there in the city. One can say, this seems to be a topic that interests the different neighborhoods. Then you could argue that you don't define neighborhood by spatial demarcation, but by ideas or shared common interests." In addition, topics can also be positioned in a targeted manner, e.g. by announcing hashtags within campaigns, which can be used for tweeting. This allows, "as in the example of Sidewalk Labs" (Focus Group Participant), for a broader involvement of the public to enable greater participation in decision-making. As one interviewee (3) extended this," it is not so much a supporting system for legally mandated forms of participation. […] Strength of this tool is rather the city analysis and participation of citizens during this analysis." Furthermore, the approach can also be extended to include, for example, data on purchasing power (focus group participant). This shows the potential to incorporate the dashboard as a barometer and as a campaigning tool into the urban planning process.

Regarding the *increase of its practical usefulness*, the interview partners gave useful suggestions, such as a deeper inclusion of a "time perspective" bypassing the perception that the dashboard "so far, is more focused on spatial perspective" (Interviewee 1). Secondly, the interviewees suggested a classification based on sentiments. For example, asking "is it more positive or negative based? Expressing praise, pain, or indignation? Basically, adding a multidimensional hierarchical system." (Interviewee 1). Thirdly, they recommended high-level overview graphs, such as "word clouds" (Interviewee 3). Fourth, some interviewees called for more interpretable context of the topics, such as with news media articles (focus group participant) or an

automatic labeling of topics in place of numbers, instead of "trusting" manual labels (Interviewee 5).

Finally, the experts offered insight into the limitations of the dashboard. First, interviewees were clear that it would serve for information and opinion gathering but were skeptical about its usefulness in decision making. Second, the experts noted limits in representativeness, and in the different communities that would be addressed by such an approach. Third, some interviewees expressed the need for cross validation with other data sources to circumvent biases in sample selection and representativeness.

5. Discussion

This paper aimed to conceptualize and prototype a Social Citizen Dashboard usable in participatory urban planning. The concept is based on social media data analyzed via topic modeling. In the context of participatory urban planning, it is used to reveal and classify citizen views. Based on the city of Berlin case, we described vignettes to evaluate the importance of topics and their changes over time. A measure was developed to support this analysis quantitatively. The following section will discuss how this approach can be integrated into the city planning process and scrutinize the quality of the data underlying this approach.

5.1. Implications for Urban Planning

In this chapter, we discuss the relevance of the developed approach for the urban planning process and how it can be integrated into such a process. Participatory planning always accompanies a long-term process such as an urban development or construction project. However, it does not circumvent the function of representative democracy or of investors and developers. Today's form of participation does not adequately represent a society within a city that has diverse interests and is under continuous change. Politics has to make decisions to balance conflicting interests. Digitalization is not an end in itself, but a tool for uncovering systematic connections in the city that supports the decision-making process. Nevertheless, it is easy to make the misleading claim that new technologies can represent the "true voice of the people." Topic modelling thus carries some populist potential, as do social media platforms such as Twitter. By having in-formation about which topics are most relevant in the public discussion, decision-makers can focus on current sensitivities and interests, and exploit them for their own benefit. On the other hand, topic modelling as a

clustering method can be a useful tool to uncover and visualize echo chambers [25].

The presentation and analysis of time and spatial trends using an indicator can enable innovative approaches such as participatory budgets [40]. Resources are used in a more targeted way, prioritizing the more discussed topics. The trends shown can serve as orientation for citizens and decision-makers. However, it should be emphasized that some decisions (e.g., a new subway line) are never made in a purely bottom-up process.

Topic modelling can also identify social influences during a project [24]. Through coordinates, it is possible to distinguish relevant topics in different areas of the city or neighborhood. It is then possible to identify the areas in which certain topics – particularly positive or critical of the project – are concentrated and provide a rough idea of differing opinions in each area. Presenting the topics to the stakeholders can also have an impact on the participation process. By showing which topics reflect priorities the most at a given point in time, the results from the topic model can be a starting point and a complement in the discussion of an urban planning projects. For administrators and companies, the topics are a basis for understanding the position of the citizens. This understanding is the precondition for finding a compromise between the different stakeholders in controversial issues.

A possible planning cycle supported by this newly developed Social Citizen Dashboard is displayed in Figure 5. The dashboard bridges between planning and participation, and the information in the chain is used as the city planning project progresses from one phase to the next.
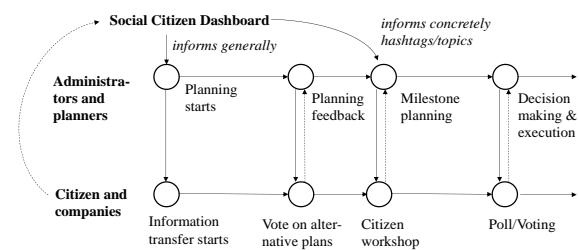
Figure 5. Social Citizen Dashboard in the participatory urban planning process

Along the top lane of the figure run the planning activities from start, planning feedback, milestone planning through decision making in planning processes. On the bottom lane, we see participatory activities from the information transfer start, vote on alternative plans, citizen workshops, through polls/voting. The approach developed in this paper supports both early as well as later phases. During

early phases, citizen priorities can be considered in the form of "fever curves". Later, it can be used within more specific participatory activities such as citizen workshops or voting on alternative plans. Because a broad spectrum of topics and user groups is represented in tweets, it is advisable to proceed in a targeted manner when collecting the data. If citizens discuss on social media and want their opinions to be stored digitally, they should use a predefined hash tag. Thus, a large part of the irrelevant posts is already sorted out during the collection.

5.2. Quality of Twitter Data

Turning to data quality, it should be noted that Twitter poses limits to representativeness. Mostly younger people use Twitter. 44% of twitter users in Germany are between 18 and 34 years old, mainly with a higher average income and degree of education [41]. In the context of Berlin, tweeting is more frequent in inner-city districts which are preferred by this certain demographic. Some users use social media as a way of expressing opinions and exchanging ideas. Another group uses it as a means of self-promotion – whether for professional or personal reasons. Moreover, some users mainly use Twitter to express anger or spread misinformation. Therefore, the information that can be used for the urban planning process varies with each user group.

The brevity of tweets also poses a challenge as no procedure for thematic modeling has become a standard. The method of aggregation by user chosen for this paper provides a good overview of the topics in the Twitter corpus. At the same time, aggregation poses a major problem: the structure of the data is changed. To maintain a consistent methodology, all further data that is examined with the help of the trained model must be processed in the same way – aggregating by user. This results in two challenges. First, the results of the LDA algorithm apply only to each aggregated pseudo-document and it is not possible to automatically deduce the distribution of the individual tweets. Second, it is not possible to classify tweets from the Twitter Streaming API directly with the trained model, requiring a certain quantity of tweets per user for aggregation.

In the literature, the use of abbreviations, colloquial language, and non-standard spelling are often cited as obstacles to the use of topic modeling [42]. Furthermore, there are many accounts operated by organizations. Therefore, large amounts of tweets from news agencies, companies, political parties and bots mix into the represented tweets from Berlin citizens. This is problematic due to not all of them being able to be systematically filtered out.

6. Conclusion

In this study, a Social Citizen Dashboard was developed for participatory urban planning extracting Twitter data and visualizing it via an interactive dashboard. To this end, topic modeling was used as well as spatial and temporal distributions of tweets. We identified three aspects that played an important role: the possibility to interact with the data, filtering by relevance and time period and the geographical representation of the tweets. In addition, the developed key indicator t_k with its division into subgroups provides an overview of the temporal development in different parts of the city. Evaluation interviews show that this could be used to explore the sentiments of citizens and changes in citizens' views over time as well as spur specific planning projects through hashtag-based campaigns. Finally, the relevance of Twitter data for the urban planning process was discussed. To address the limitations of the proposed dashboard, e.g. data quality, future work should consider complementary participatory elements like citizen workshops, as means to include a broader spectrum of Berlin citizens.

7. References

[1] Salter, J. D., Campbell, C., Journeay, M., and Sheppard, S. R. J. "The Digital Workshop: Exploring the Use of Interactive and Immersive Visualisation Tools in Participatory Planning", Journal of Environmental Management 90, 6 (2009), pp. 2090–2101.

[2] Khan, Z., Ludlow, D., Loibl, W., and Soomro, K. "ICT Enabled Participatory Urban Planning and Policy Development: The UrbanAPI Project", Transforming Government: People, Process and Policy 8, 2 (2014), pp. 205–229.

[3] Mukhtarov, F., Dieperink, C., and Driessen, P. "The Influence of Information and Communication Technologies on Public Participation in Urban Water Governance: A Review of Place-Based Research", Environmental Science and Policy 89, (2018), pp. 430–438.

[4] Pulse Lab Jakarta. From Urban Data Collection to Urban Design. A Guide to Participatory Approaches Around the Globe, 2017 [Available at: http://unglobalpulse.org/sites/default/files/From Urban Data Collection to Urban Design.pdf].

[5] Sidewalk Labs. Toronto Tomorrow: A New Approach for Urban Growth, 2019 [Available at: https://quaysideto.ca/ Wp-content/uploads/2019/06/MIDP_Volume0.pdf]

[6] Zhang, Y., Grignard, A., Lyons, K., Aubuchon, A., and Larson, K. "Real-time Machine Learning Prediction of an Agent-Based Model for Urban Decision-making", in Proceedings of AAMAS, (2018), pp. 2171–2173.

[7] Marrone, M. and Hammerle, M. "Smart Cities: A Review and Analysis of Stakeholders' Literature", Business & Information Systems Engineering 60, 3 (2018), pp. 197–213.

[8] Stieglitz, S., Dang-Xuan, L., Bruns, A., and Neuberger, C. Social Media Analytics. "An Interdisciplinary Approach and Its Implications for Information Systems", Business & Information Systems Engineering 6, 2 (2014), pp. 89–96.

[9] Yuan, J., Zheng, Y., and Xie, X. "Discovering Regions of Different Functions in a City Using Human Mobility and POIs", in Proceedings of the 18th ACM SIGKDD, (2012), pp. 186–194.

[10] Blei, D. M., Ng, A. Y., and Jordan, M. I. "Latent Dirichlet Allocation", Journal of Machine Learning Research 3, (2003), pp. 993–1022.

[11] Abram, S. A. "Planning the Public: Some Comments on Empirical Problems for Planning Theory", Journal of Planning Education and Research 19, 4 (2000), pp. 351–357.

[12] Huxley, M. "The Limits to Communicative Planning", Journal of Planning Education and Research 19, 4 (2000), pp. 369–377.

[13] Neuman, M. "Communicate This! Does Consensus Lead to Advocacy and Pluralism?", Journal of Planning Education and Research 19, 4 (2000), pp. 343–350.

[14] Kaza, N. "Tyranny of the Median and Costly Consent: A Reflection on the Justification for Participatory Urban Planning Processes", Planning Theory 5, 3 (2006), pp. 255–270.

[15] Johannessen, M. R. and Berntzen, L. The Transparent Smart City. In Rodríguez Bolívar, M. P. (ed.) Smart Technologies for Smart Governments. Springer, (2018), pp. 67–94.

[16] Pløger, J. "Strife: Urban Planning and Agonism", Planning Theory 3, 1 (2004), pp. 71–92.

[17] Arnstein, S. R. A. "Ladder of Citizen Participation", Journal of the American Institute of Planners 35, 4 (1969), pp. 216–224.

[18] Kovács-Győri, A., Ristea, A., Havas, C., Resch, B., and Cabrera-Barona, P. "#London2012: Towards Citizen-Contributed Urban Planning through Sentiment Analysis of Twitter Data", Urban Planning, Cogitatio Press 3, 1 (2018), pp. 75–99.

[19] Schiefer, D. and Van Der Noll, J. "The Essentials of Social Cohesion: A Literature Review", Social Indicators Research 132, (2017), pp. 579-603.

[20] Leleux, C. and Webster, W. "Delivering Smart Governance in a Future City: The Case of Glasgow", Media and Communication, Cogitatio Press 6, 4 (2018), pp. 163–174.

[21] Lock, O., Bednarz, T., Leao, S. Z., and Pettit, C. "A Review and Reframing of Participatory Urban Dashboards", City, Culture and Society, Elsevier, (2020), Article 100294.

[22] Zhou, X., Noulas, A., Mascoloo, C., and Zhao, Z. "Discovering Latent Patterns of Urban Cultural Interactions in WeChat for Modern City Planning", in Proceedings of the ACM SIGKDD, Association for Computing Machinery, (2018), pp. 1069–1078.

[23] Ye, X., Li, S., Yang, X., Lee, J., and Wu, L. "The Fear of Ebola: A Tale of Two Cities in China. Big Data Support of Urban Planning and Management", Nature Publishing Group, (2018), pp. 113–132.

[24] Aral, S., Muchnik, L., and Sundararajan, A. "Engineering Social Contagions: Optimal Network Seeding in the Presence of Homophily", Network Science 1, 2 (2013), pp. 125–153.

[25] Colleoni, E., Rozza, A., and Arvidsson, A. "Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data", Journal of Communication 64, 2 (2014), pp. 317–332.

[26] Twitter. Twitter API Documentation, 2018 [Available at: https://developer.twitter.com/en/docs].

[27] González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., and Moreno, Y. "Assessing the Bias in Samples of Large Online Networks", Social Networks, Elsevier B.V. 38, (2014), pp. 16–27.

[28] Morstatter, F., Pfeffer, J., Liu, H., and Carley, K. M. "Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose", in Proceedings of the 7th AAAI Conference on Weblogs and Social Media, (2013), pp. 400–408.

[29] Martin, F. and Johnson, M. "More Efficient Topic Modelling Through a Noun Only Approach", in Proceedings of Australasian Language Technology Association Workshop, (2015), pp. 111–115.

[30] Konrad, M. Accurate Part-of-Speech Tagging of German Texts with NLTK, WZB Data Science Blog, 2016 [Available at: https://datascience.blog.wzb.eu/2016/07/13/accurate-part-of-speech-tagging-of-german-texts-with-nltk/]

[31] Blei, D. M. "Probabilistic Topic Models", Communications of the ACM 55, 4 (2012), pp. 77–84.

[32] Hong, L. and Davison, B. D. "Empirical Study of Topic Modeling in Twitter", in Proceedings of the First Workshop on Social Media Analytics, (2010), pp. 80–88.

[33] Barde, B.V., and Bainwad, A.M. "An Overview of Topic Modeling Methods and Tools". 2017 ICICCS Conference, (2017), pp. 745-750.

[34] Yigitbasioglu, O. M. and Velcu, O. "A Review of Dashboards in Performance Management: Implications for Design and Research", International Journal of Accounting Information Systems, Elsevier Inc. 13, 1 (2012), pp. 41–59.

[35] Röder, M., Both, A., and Hinneburg, A. "Exploring the Space of Topic Coherence Measures", in Proceedings of the 8th ACM International Conference, (2015), pp. 399–408.

[36] Meuser, M. and Nagel, U. The Expert Interview and Changes in Knowledge Production. In A Bogner, B Littig, W Menz (eds.) Interviewing Experts. Palgrave Macmillan UK, Basingstoke, Hampshire, England., (2009), pp. 17–42.

[37] Sievert, C. and Shirley, K. E. "LDAvis: A Method for Visualizing and Interpreting Topics", in Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, (2014), pp. 63–70.

[38] Jüdisches Museum Berlin. Die jüdische Sportbewegung im nationalsozialistischen Deutschland (Symposium), 2018 [Available at: https://www.jmberlin.de Symposium-juedische-sportbewegung-in-ns-deutschland]

[39] Jüdisches Museum Berlin Living with Islamophobia (internationale Konferenz), 2018 [Available at: https://www.jmberlin.de/konferenz-living-with-islamophobia].

[40] Wampler, B. A Guide to Participatory Budgeting. In Shah, A. (ed.) Participatory Budgeting. Washington D.C.: The World Bank, (2007), pp. 21–54.

[41] Rondinella, G. Deutsche Twitterer sind jung, einkommensstark und markenaffin, Horizont, 2019 [Available at: https://www.horizont.net/tech/nachrichten/studie-deutsche-twitterer-sind-jung-einkommensstark-und-markenaffin-174405].

[42] Mehrotra, R., Sanner, S., Buntine, W., and Xie, L. "Improving LDA Topic Models for Microblogs via Tweet Pooling and Automatic Labeling", in Proceedings of the 36th international ACM SIGIR Conference, (2013), pp. 889–892.