A flexible and lightweight interactive data mining tool to visualize and analyze digital citizen participation content

Sergio Bachiller
s.bachillerrubia@gmail.com
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Madrid, Spain

Lara Quijano-Sánchez*
lara.quijano@uam.es
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Madrid, Spain

Iván Cantador
ivan.cantador@uam.es
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Madrid, Spain

ABSTRACT

Citizen collaboration through current digital participation platforms can entail the generation of large amounts of complex content, which may hide relevant citizens' concerns, requests and initiatives, diluted in isolated individual proposals. Addressing this problem, in this paper, we present an interactive data mining tool for citizen participation data visualization and analysis. The tool implements natural language processing, text similarity, and graph clustering techniques to group proposals with common objectives, identify trends and recurrent topics of interest, and filter and present information according to several criteria. The tool is flexible, able to process different sources of data, and lightweight as it uses simple data structures and dynamic HTML-based visualization and interaction. As a case study, the tool has been instantiated with a dataset obtained from the Decide Madrid e-participatory budgeting platform.

CCS CONCEPTS

• **Applied computing** → **Document management and text processing**; • **Human-centered computing** → **Visualization**; • **Information systems** → **Data mining**.

KEYWORDS

citizen participation, data visualization, data clustering, text similarity, civic technologies

ACM Reference Format:
Sergio Bachiller, Lara Quijano-Sánchez, and Iván Cantador. 2021. A flexible and lightweight interactive data mining tool to visualize and analyze digital citizen participation content. In *Proceedings of ACM SAC Conference (SAC'21).* ACM, New York, NY, USA, Article 4, 9 pages. https://doi.org/xx.xxx/xxx_x

1 INTRODUCTION

With the advent of social media and mobile computing, nowadays there is a plethora of digital citizen participation channels, ranging from general-purpose online social networks to ad hoc e-participation, e-consultation and e-voting platforms.

The huge, ever increasing citizen generated content leads to an information overload problem for both citizens and government stakeholders in decision and policy making tasks. In particular, they may feel overwhelmed by the large amount of data, whose exploration and understanding could result challenging and frustrating. Also, citizens may feel thwarted if their proposals do not reach sufficient visibility and impact. In this sense, we may find several proposals by different authors that address the same problem, but in different ways, for distinct city locations, or entailing distinct initiatives and potential solutions

To address the above problems, there is a need for information systems capable to process and mine citizen generated content, as well as to summarize, visualize and analyze relevant extracted information overtime. Motivated by this need, in this paper we present a flexible, lightweight data mining tool that helps unraveling public deliberation contents to both governments and citizens. On the one hand, local governments are constantly evolving organizations, and when neither long-term growth plans nor daily operations of critical city services are stable, the need for robust, accessible and meaningful community engagement is crucial. The goal here is to develop systems and strategies that enable people to form policies with an impact on their communities and own lives [18]. In this context, our tool makes these activities simpler by interactively analyzing knowledge generated from public initiatives, e.g., detecting trends in the concerns posed to policymakers, persistent demands or particular seasonal problems. On the other hand, considering the citizens' contributions is an important objective for governments, scientists and companies, since it can have a great impact on issues of common interest. Existing civic technologies focus mainly on rewards, only dealing with values such as power, achievement, security and encouragement, and leaving a gap with respect to other values [26]. The presented tool allows unifying objectives by grouping and visually monitoring proposals (that could have been accepted or remain unanswered), thus serving as a stimulus for citizens to increase quality (i.e., more informed and argued proposals) and quantity (more proposals due to ease of use) on their generated content.

In summary, we present an interactive tool for citizen participation data visualization and analysis, which is built upon the *Tableau* interactive data visualization software, and which is lightweight and easy to configure, as well as generic, since it is reusable for other related domains and different languages as it uses data from

*Corresponding author.

SAC'21, March 22-March 26, 2021, Gwangju, South Korea
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8104-8/21/03…$15.00
https://doi.org/xx.xxx/xxx_x

https://www.tableau.com

an external database (provided by Amazon Web Services in our case). The tool, whose code is publicly available, provides several visualization functionalities, allowing both temporal and geographical analysis by means of different diagram bars, heat maps, and time series graphs. These visualization components are dynamic, enabling to filter data according to configurable constraints based on proposal categories and topics, time, and districts and neighborhoods of a city.

Implementing natural language processing, text similarity, and graph theory techniques, groups of proposals related to common topics of interest are created and visualized. This clustering level allows a better and easier extraction of patterns and insights when analyzing the published citizen generated content. In addition, it segregates the proposals in a more exhaustive way, opening the study of citizens' needs in each of the areas of a city. This study would not only detect inequalities in the different districts, but would also serve to inform both the citizenry and the municipal government of what is being demanded. Lastly, the tool has a co-production functionality based on the retrieval of existing similar proposals. Through this functionality, a citizen who is interested in submitting a new proposal can first bring it into the tool and check if there are related ones. As a case study, we instantiated the tool with Open Government Data from Decide Madrid, the electronic platform for citizen participation of the Madrid City Council.

2 RELATED WORK

As stated in [10], there is an increasing demand for data exploration interfaces that support analytical processes, allowing non expert users (referred as citizen scientists) to absorb and derive knowledge from city-related data, by placing high demands in terms of usability. There is thus a need for developing interfaces to support collaborative, community-led inquiries into data. In this context, as shown in [22], geovisualization plays an import role to reinforce the citizens' involvement and engagement in participatory processes.

Revising the research literature on data visualization and analysis in digital government, we can find series of papers devoted to identify and summarize public concerns, proposals and opinions via citizen consultation and participation[7, 9, 13, 16]. Chong et al. [7] present a data analysis framework that, implemented for Denton, a medium-size city in Texas, uses an open-ended survey as a collection instrument for citizen input, and employ concept-based analytics to convert such input into valuable insights. A tag cloud visualization is used to identify major problems and suggested solutions reported by citizens.

Instead of explicitly requesting input from citizens, in [13], Giatsoglou and colleagues present CITYPULSE, a modular tool implemented for Santander (Spain) which automatically gathers citizen feedback from social media (Twitter in particular) and uses temporal and geographical bar charts, as well as tag clouds and visual maps, to aggregate and visualize of those data for revealing and highlighting latent information in terms of the city's emerging topics and trends. Also targeting Twitter as input data source, in [16], Hubert et al. present a range of visualization techniques to analyze the interactions between citizens and government on the issues of public concern. In this case, not only trending topics are considered, but also a number of signals regarding the level of government activity, the intensity of citizen response, the resources shared between both stakeholders, and the sentiment expressed by citizens.

Differently to previous work, instead of exploiting generic, unstructured content from social media and explicitly requested and limited citizen feedback, we explore the use and analysis of citizen proposals generated in ad hoc electronic participatory platforms. Moreover, instead of simple and fixed visualization functionalities, our tool allows configurable and interactive interfaces to data analysis. Besides, the proposed tool benefits from a flexible instantiation for different data sources, and an easy and lightweight, but efficient deployment and execution.

3 CASE STUDY

In this section, we present Decide Madrid, the electronic participatory budgeting platform of Madrid City Council, and the case study for which our tool has been instantiated. Before, we provide a brief introduction to participatory budgeting that motivates of our work.

3.1 Electronic participatory budgeting

Participatory budgeting (PB) represents one of the most popular mechanisms for involving citizens in decision making[14]. In PB, citizens participate in the processes to allocate part of the public budgets in initiatives and projects for different areas, such as public safety, education, health, and mobility.

The goal of this type of participatory process is to create a functional democracy in which community members influence on the actions of local governments [11]. When citizens see results and impact from their proposals, they feel involved and engaged, and thus tend to participate more. Studies have shown that the level of citizen participation in PB is in general low [32]. This may have arisen from the approaches followed in many PB cases, where the main focus has been put on offering technological solutions instead of understanding the citizens' needs [29].

Hence, several researchers have advocated a turn towards openness in participatory design [21] and increasingly strive to promote empowerment by demonstrating and delivering to people tools and technologies for ownership, reuse and adoption for their situated ends [5]. In this context, reviews of the PB literature suggest a continuous shift in research from a more purely technological approach to a more holistic view, in which other social and technological issues could be integrated to improve citizen participation [2].

These evidences and investigations support the hypothesis of this work about developing an agile and simple tool to enhance the summarization, clustering and visualization of content produced in electronic participatory budgeting (ePB) platforms, facilitating its analysis and understanding by both citizens and governments.

3.2 The Decide Madrid platform

Decide Madrid, active since September 2015 with 420,000 registered users, is the online platform where the Madrid City Council orchestrates the city's annual participatory budgets. In the platform, a series of initiatives and projects are proposed, discussed and supported by registered residents. All the content generated

https://github.com/sbachiller/EParticipationAnalysis-DecideMadrid

https://decide.madrid.es/

in the platform, i.e., proposals text and metadata, comments and votes are publicly available as open data .

Throughout a year, Decide Madrid allows proposals (27,662 as far as June 2020) to be created, discussed and supported. Once a proposal obtains the necessary support (i.e., 1% of the registered population in the city), it is submitted for citizen vote. Those supported proposals that achieve a simple majority of votes are subject to be conducted by Madrid City Council. Part of the municipal budget for this means is allocated. Despite the fact that there are certain proposals that cannot be carried out by the City Council since they are the competence of another organization, there are others that are viable and finally get funding to be implemented, such as the building of a day center, the creation of new subsidies, and the replacement of transportation elements.

4 MINING TOPICS OF INTEREST

The large amount of data published in ePB and other citizen participation platforms makes it difficult to explore and understand the underlying information, and make decisions accordingly. This fact originates many citizens to give up deciding not to participate, and citizens' concerns not to be supported, thus causing discontent among those who participate, since they do not see any progress on their proposals [25].

To alleviate this effect, our aim is to develop a tool capable to extract rich information from ePB platforms, such as particular interests and problems in each neighborhood and district, and issues that represent general concerns of majorities (minorities). In this way, it would be possible to better outline for both citizens who use a platform and local governments what inhabitants are asking for, facilitating the work of decision and policy makers, and ultimately leading to a improvement on the citizens' quality of life.

Having as input documents with the title, abstract and text of citizen proposals , the tool performs the following tasks: i) the content of the proposals is transformed using natural language processing (NLP) techniques, ii) text similarities are computed over the processed documents and used to build a document relatedness graph, and iii) a graph-based clustering method is apply to group duplicate and/or similar proposals.

The final output of these tasks consists of citizen proposal clusters that are analyzed to identify, among other issues, topics of interest. Hence, the tool not only allows visualizing the raw data contained in ePB platforms, but summarize and facilitate the understanding of underlying problems and proposed solutions.

In the next section, we present the text processing techniques carried out prior to visualizing information by the developed tool.

4.1 Text processing

In order to ensure that the computation of document similarity is accurate, it is necessary to treat the textual content appropriately. We accomplish this task using common tools in a NLP pipeline.

The first step of the pre-processing pipeline is the correction of mistakes in the texts. Special characters are first removed, thus avoiding possible problems in the misspelling correction. For this task, we use the Python version of the *Hunspell* spell checker, which is used by some desktop programs such as OpenOffice and Mozilla Firefox. We feed the Hunspell checker with the LibreOffice dictionary and a vocabulary composed of names of streets, places and services in Madrid, obtained from the city open data portal.

Once the texts have been corrected, two more pre-processing tasks are executed: i) *Stopwords* removal and extraction of nouns, adjectives and verbs, which are valuable to find relevant similarities. This task was performed using the *Spacy* library; ii) *Lemmatization*, which consists of replacing each word by its corresponding canonical form.

4.2 Document similarity

Estimating the similarity between two texts is a extensively studied task in the NLP field [17, 19, 27]. In this work, we seek to find a similarity measure with two main characteristics, namely lexical and semantic similarity, that is, the words appearing in the texts are from the same context and have the same meaning, respectively. These characteristics allows differentiating between two texts such as *"I went to the bank to withdraw money"* and *"I sat in a bank and I found money,"* which have a large lexical similarity, since their entities (person, bank, money) are the same, but have a small semantic or contextual similarity, because putting the *bank* word in context, we find that in the first case it refers to a place, and in the second case, it refers to an object.

In [28, 31], the main trends, examples, limitations and successes of the most popular methods of text similarity are reported. They show how the Jaccard index is not appropriate for the proposed problem, as it does not deal with different elements with the same semantic meaning. With respect to techniques in which *embeddings* (vectors that represent words) are used, they highlight how the K-Means algorithm to establish similarities is very sensitive to the number of vector features and requires knowing a priori the number of clusters, and how the cosine similarity can be improved depending on the followed word vector generation method. Other described techniques include *Latent Semantic Indexing* (LSI), which seeks to reduce the size of *embeddings* by assuming that there is a space of smaller dimensions in which representing all the words of a text with certain loss of information, and *Word Mover's Distance* (WMD) [20], which aims to find the minimum distance between two documents within a vector space. WMD allows establishing a high similarity between two sentences without common words, such as *'Obama speaks to the media in Illinois'* and *'The president greets the press in Chicago'*, as it manages to capture the semantics from whole documents and corpora.

In order to use a method that captures both lexical and semantic similarities, in this work we advocate for the WMD similarity, since it stands out over the simplest methods and at the same time does not require a pre-labeled dataset to be executed, facilitating thus its reusability of the developed tool for different domains and languages.

The WMD similarity is inspired by the *Earth Mover Distance* transportation problem, aiming to find similarity (distance) between two texts even if they have no words in common. WMD leverages the results of advanced embedding techniques like word2vec and Glove [15]. It treats text documents as weighted point clouds of embedded words. The distance between two text documents A and

https://datos.madrid.es

We have employed wor2vec on a corpus in Spanish https://github.com/dccuchile/spanish-word-embeddings#word2vec-embeddings-from-sbwc

B is calculated by the minimum cumulative distance that words from the text document A need to travel to match exactly the point cloud of text document B. Hence, the distance measures the dissimilarity between two text documents as the minimum amount of distance that one document's embedded words need to "travel" for reaching another document's embedded words.

This measure computes distance and not similarity. For this reason, all the values of the distance matrix WMD of dimension NxN (where N denotes the number of documents) and the maximum distance are used to compute a similarity matrix as follows: $Similarity(i, j) = 1 - WMD(i, j)/max\_distance$.

4.3 Document clustering

Instead of using classic clustering techniques such as K-Means and agglomerative clustering, we propose to use recent approaches applied to detecting communities of interest in urban contexts [1, 12, 23]. For such purpose, we build a non directed, weighted graph whose nodes represent the citizen proposal documents and whose edges are assigned with the computed document similarity values.

On the built graph, we apply the Louvain method [4] , which locally optimizes the modularity of the graph and associates nodes until convergence, with a good execution time of $O(n \cdot log(n))$. This clustering method, in contrast to others like K-Means, it does not need a fixed number of clusters, but rather it adapts to the problem.

Applying this algorithm directly on the graph, which we can assume is totally connected, results in a single community that represents the entire graph. To fix this, we removed edges with weights lower than certain value (representing the level of desired similarity within the community). In this work, all edges with weights lower than $min\_weight = 0.55$ (motivated by the range in which considerable similarities were observed) were removed. Some results are shown in Table 1. In this table, where the 10 largest communities are shown, we can verify the concern of citizens regarding our case study platform, being this the community with the most proposals, such as '*An option of NO SUPPORT in Madrid Decide*', '*Advertising Madrid Decide Platform*' or '*Group similar proposals in Decide Madrid*'. Finding such proposals in such a large community verifies our hypotheses and reinforces the indications of the referred preliminary studies, summarized in [25]. In other communities, we find documents similar to each other, thus achieving the desired unification and summarization objectives when it comes to understanding the large volume of proposals.

5 CITIZEN PARTICIPATION ANALYSIS TOOL

To address the outlined information overload problem, we have developed a lightweight web application consisting of a simple HTML-based data panel that, through the use of date, location and category based filters, and several interactive graphs, allows visualizing, exploring and analyzing the data obtained from public deliberation platforms (the Decide Madrid platform in our case) in an easy and clear way. The tool thus serves as a decision support system for the municipal government, and contribute to information transparency, engaging citizens into what is happening in their city. In this section, we present the developed tool, instantiated with citizen generated content from the Decide Madrid ePB platform.

5.1 Data sources and structures

The Madrid City Council, through its open data portal, provides data collections related to the city. Among these collections, we focus on the proposals created by 24,482 users of Decide Madrid until September 11, 2019, and assigned with one or more categories such as Urban Planning, Animals, Mobility, and Security and Emergencies. The obtained database contains 21,746 proposals and their associated 86,102 comments. Each proposal has a title, a summary, a description, social tags, publication date and the number of received supports. Also, proposals are tagged with one or several of 30 existing general categories. Downloaded data also includes a geographic repository with almost 1,500 streets and POIs of Madrid with their corresponding districts and neighborhoods. A proposal was also assigned with it corresponding geographic area, i.e., a street, a neighbour (among the existing 129), a district (among the existing 21) or the whole city. Using web crawling and scrapping techniques, thematic taxonomies and semantic annotations, and ad hoc developed metrics as described in [6], the retrieved data has been automatically extended with the following information for each proposal: topics that refine the assigned categories (among 325 different options), measures of popularity and controversy. The database was uploaded to a MySQL database in Amazon Web Services. Later, the database schema was automatically imported into the tool, hence as we next described, others that present similar schemes (with equivalent columns for the later described filters) can be easily incorporated into the tool.

We note that for reusability purposes, if the data structure is maintained, other use cases with similar scopes or in alternative languages could be loaded into the tool. Furthermore, the data filters of the tool could be easily adjusted to the characteristics of a new given database. Also, given that the techniques used for text similarity and clustering described in Section 4 are not based on particular topics, languages and corpora, the tool is extremely flexible.

5.2 Data analysis functionalities

According to [3, 8, 24], a data visualization tool, without too many simultaneous variables, helps on the interpretation of underlying information. In the context of citizen participation, it also may encourage the co-production of new proposals, and may engage others to participate in the process.

Our tool presents a number of tabs that could be easily expanded to show more aspects of the database: analysis by categories (and topics), communities, and districts (and neighborhoods), temporal evolution and influence distribution. For each tab, there are several filters: date, category, topic, district, neighbourhood, community and number of proposals, which allows the user to interactively explore information. Proposals are also displayed in real time as filters are adjusted in the interface. The tool interface leads to two types of analysis: temporal and geographical. As for the temporal analysis, by means of bar charts, users can analyze the number of proposals presented in each district over time. This information is grouped by months and divided in slots of trimesters per year. Users can further unravel information by selecting concrete districts or time periods. In the evolution over time tab, users can study the

https://datos.madrid.es/portal/site/egob

Community	Main category	Proposal examples
#105 Decide Madrid	Citizen participation	An option of NO SUPPORT in Madrid Decide Advertising Madrid Decide Platform Group similar proposals in Decide Madrid
#34 BiciMad	Mobility	Creation of safe bike lanes and improvement of the Madrid asphalt Bike lane in all the streets of Madrid Expand bicimadrid to more neighborhoods
#34 Blue parking zone (SER)	Mobility	Flexible allocation of SER areas for residents Delete reserved areas in official buildings More blue parking areas for people with reduced mobility
#12 Transport card	Mobility	Include BiciMAD in the Transport Pass Free transfer between metro, bus Free transport pass for the unemployed
#9 Night transport	Mobility	Introduce a public transport night schedule METRO open all night on weekends Reestablish night bus service
#12 Free public transport for minorities	Mobility	Senior citizen transport card Free transport for people with minimum resources Free transport for children under 12 and unemployed
#24 Dog poop	Animals	Significant penalties for dog owners Fines to people who deposit garbage on the street Fine dog owners for not removing their feces
#6 Clean city	Environment	Clean the streets Increase cleaning in Madrid More trees and clean air in Madrid
#24 Sanctions for dirtying streets	Environment	Prohibit the use of public roads as a bar counter Tax for those who have a dog Treat all the neighborhoods as if they were Serrano street
#2 public transport reach	Mobility	Pavones Goya direct EMT line Expansion of line 102 EMT to Cibeles That Bicimad reaches more districts

Table 1: The 10 largest communities formed by the Louvain clustering method, together with their main category, and some example proposals.

trends of detected topics of interest (i.e., the communities described in Section 4.3). To do so, the topics are presented as a series of bars for the months in the year, grouped by trimester. The height of the bars represent the number of proposals belonging to the communities for each month. Users can also filter by the number of communities displayed vertically, a chosen time span, and involved districts. An example of this analysis is exemplified in Figure 3. The counterpart graph is presented; in this case, analyzing the evolution of proposals belonging to the existing categories.

Regarding the geographical analysis, similarly to the participation by district temporal graph displayed in the participation tab, bar charts representing participation by district are presented. Again, filters for desired span of time and districts to study are available. Also to facilitate visualization, the tool presents a topographical map of Madrid divided by districts, where the influence of communities and categories for each district are highlighted (darker colours representing high production volume). This tab, presented later in Figure 4, aids users to study possible geographical correlations and phenomena, as districts that are actually close are painted according to their GPS coordinates.

To assist users in information exploration, three general tabs are introduced with the aim of identifying proposals with desired characteristics through a chain of filters. The goal of the three tabs is changing the filter order and the type of graphics that adapt to the specific temporal or geographical analysis. The first one, finding through category, is presented in Figure 1. It allows users to first select a desired time spam, then select categories presented in horizontal bar charts, and set filters by topics, districts and neighbours. We note that inside each filter, the number of proposals to be shown can also be established. In the last graph, a temporal analysis of the retrieved proposals is presented. The second tab, finding through category, is analogous to the previous, exchanging categories by identified communities. The last tab, finding through district, is presented in Figure 2; here, after selecting a desired time span, the volume of proposals in districts is represented with the corresponding bigger or smaller bubbles. This selection allows for further tuning the search indicating neighbours to visualize inside the prior selection, and then categories and topics related to the chained filters. Again, a final temporal analysis graph of the selected information is presented.

Lastly, the tool allows, given certain (new) proposal, finding similar proposals, which enable avoiding duplication and merging of proposals from the platform. Space limitation prevents the authors from showing this functionality as a screenshot figure.

6 ANALYSIS INSIGHTS

In this section, we analyze the influence of the topics of interests identified in the different districts of the city of Madrid, and their evolution over time. An in-depth social study could be carried out, in which socio-demographic variables would be used to know and understand the motivations underlying the proposals [6]. In particular, we next present a sample of insights hidden within the large volume of data, which were easily identified thanks to the data filtering and visualization functionalities of our tool.

As mentioned before, we can analyze the participation by district and its evolution, where for instance districts with the highest incomes tend to be those that have participated the least, and districts with the lowest incomes are those that present more proposals. Through the temporal visualization component (Figure 3), a deeper analysis of the evolution of participation is possible by adding two

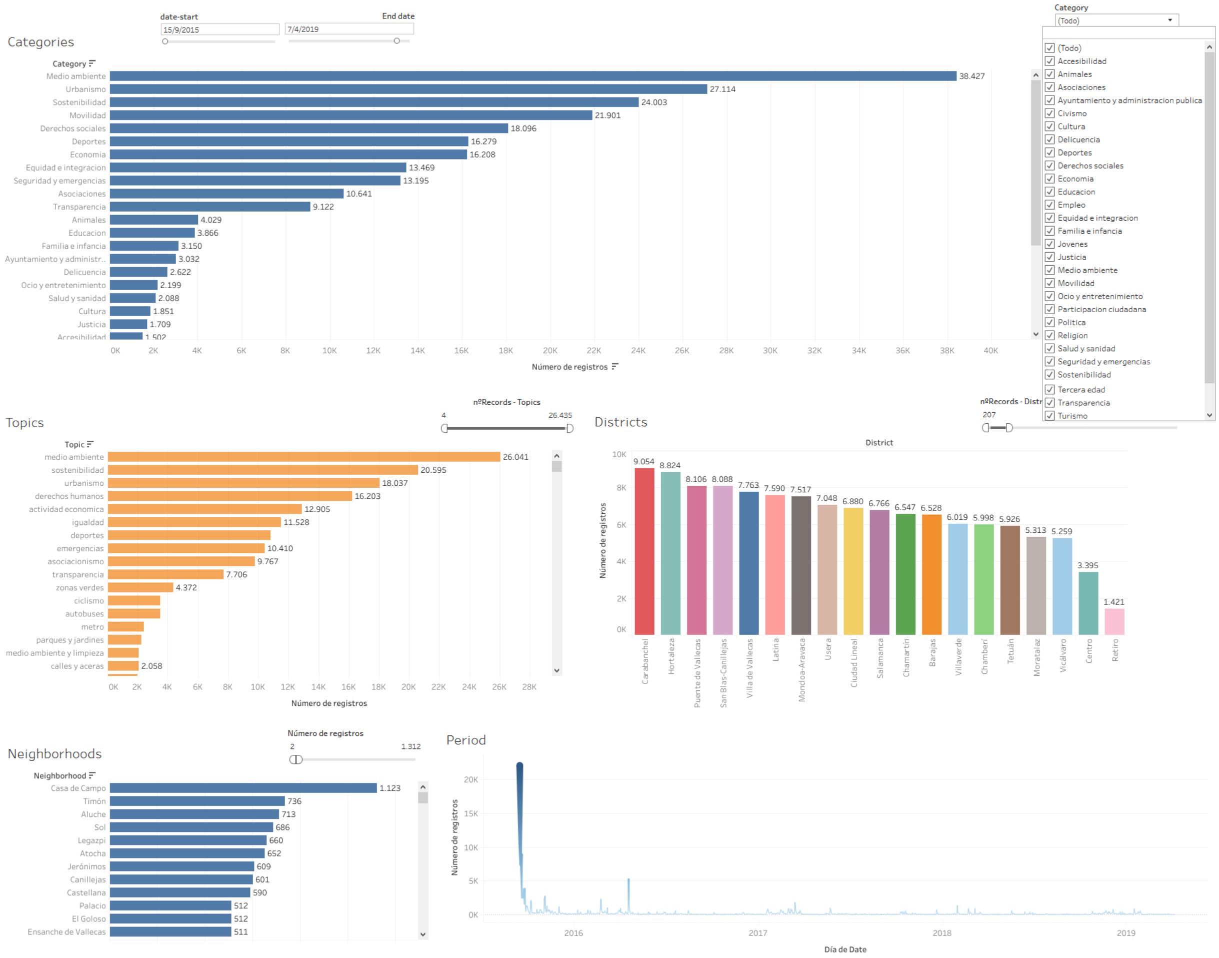


Figure 1: Bar charts to visualize proposal popularity and controversy information among categories, topics, districts, neighborhoods and time. Proposals are displayed in real time as filters are adjusted in the interface.

new variables: the community (at the top) and the category (at the bottom). In both cases, we can see a significant decrease on the number of proposals during the summer months. With these bar graphs, it is possible to estimate the seasonality of a problem over time, for example, in the Animals category, which increases considerably during the last months of the year, surely caused by the abandonment of pets or animals that are given away during Christmas. It is also possible to find a certain seasonality in the evolution of the communities and categories. Regarding the community that *Scope of public transport* represents, we can see that it is in the first trimester of each year (or even the fourth of the previous year) 3, when it has the highest impact, probably with the goal of being included in the objectives of the organizations responsible for that year. Thanks to this data panel, it is easy to discover the seasonality of groups of proposals, and even to detect unusual peaks, such as the one shown in Figure 3 (in the middle) for the transparency community in April 2016. In that month, the General Director of Economy of Madrid City Council presented her resignation, after documents that linked her to a illegal company in Panama were leaked.

In Figure 4, which shows the thematic influence panel of our tool, we can observe the impact of categories and communities on each district and area in the city. Specifically, in the map on the left, we can analyze the impact of the selected community, *Scope of public transport*, and find out that there are clear needs in the periphery districts. In fact, there are experts who indicate that the M-30 road to the east of the city is a great barrier that split the population and leads to the creation of initiatives such as *Park-30*. In the map on the right, we can see that in the *Tourism* category, the majority of proposals come from downtown districts, where the number of places of interest is much higher.

Using the communities extracted as explained in Section 4.3, we can, in a much more concrete way than the category, identify the needs that are most demanded in each district. Also, analyzing

https://www.eldiario.es/politica/directora-economia-ayuntamiento-madrid-panama_1_4025672.html

https://www.elespanol.com/espana/madrid/ 20200527/parque-30-radical-ideal-quiere-eliminar-kilometros-m-30

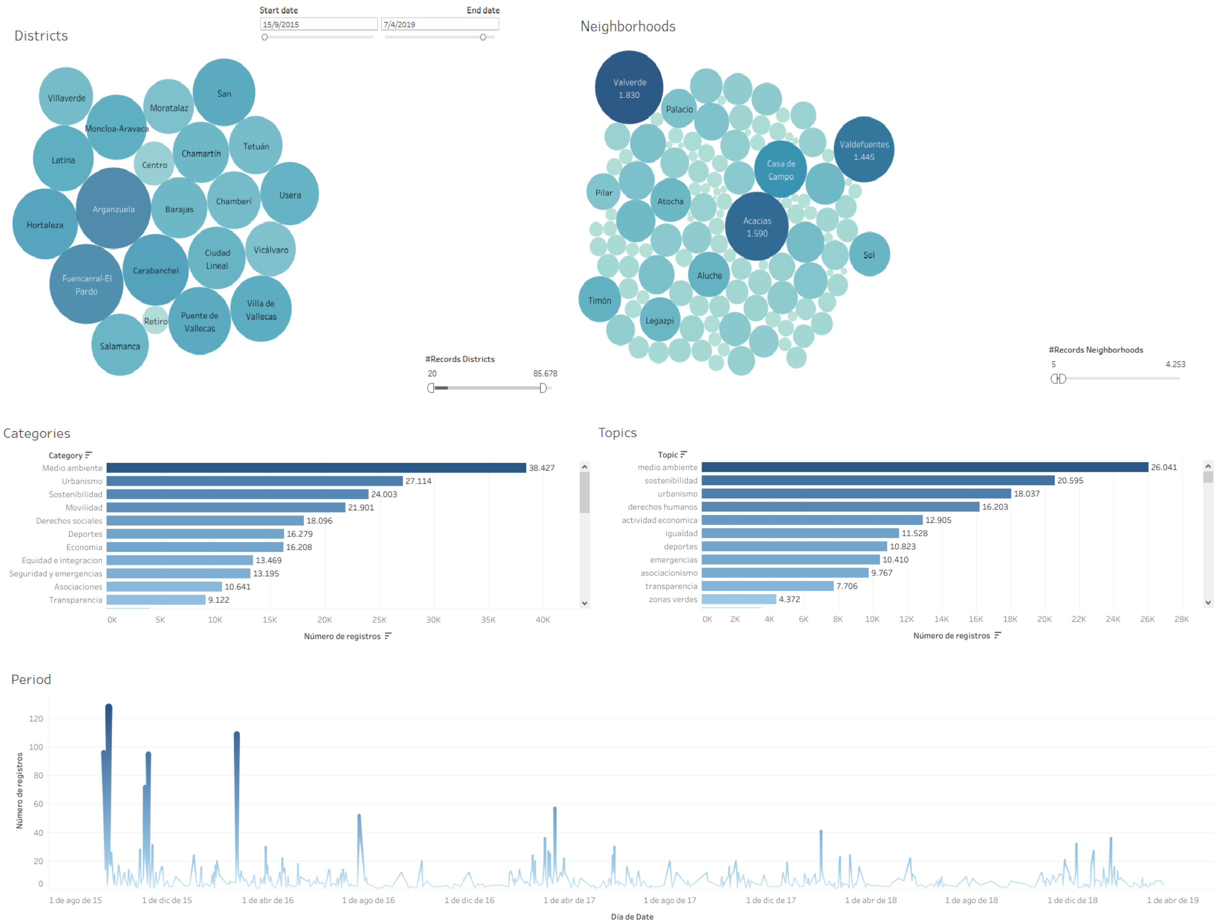


Figure 2: Interactive bubble, bar and time series charts to visualize and explore category and topic distributions over districts, neighborhoods and time. Proposals are displayed in real time as filters are adjusted in the interface.

the communities exemplified in Table 1, the districts where each one has the most influence can be checked. For instance, regarding community #9 Nigh transport, the most affected is the university district, Moncloa-Aravaca, which should have enough resources to cover the great transport demand that exists and that despite the large number of stations on the main metro lines, it should be bared in mind that the Madrid Metro does not have night service, with all the burden falling (taking into account what this means in a university neighborhood, where it is mostly populated by young people who need this kind of service) on night buses. The district only has four night city buses, all originating from Plaza de Cibeles, and also taking into account the large time lapses between buses, it is logical that Moncloa-Aravaca calls for an improvement in its night transport services. This fact is highlighted for example in in Figure 4 left.

7 CONCLUSIONS

In recent years, many cities have implemented online platforms for citizen participation, in which inhabitants participate in municipal decisions and actions by means of proposals and debates. To date, these platforms are suffering problems related to the citizens' frustration for the lack of visibility and impact of their proposals, and to a low participation due to information overload and exploration difficulties. To address these problems, researchers have proposed the development of technological solutions to summarize and contextualize citizen feedback, and visualize individual and community needs and concerns [6, 22, 30]. Following this direction, we have presented a flexible interactive tool that provides a variety of visualization and analysis functionalities for generic citizen generated content, and facilitates to both citizens and local governments the understanding of the underlying problems in a city and proposed solutions.

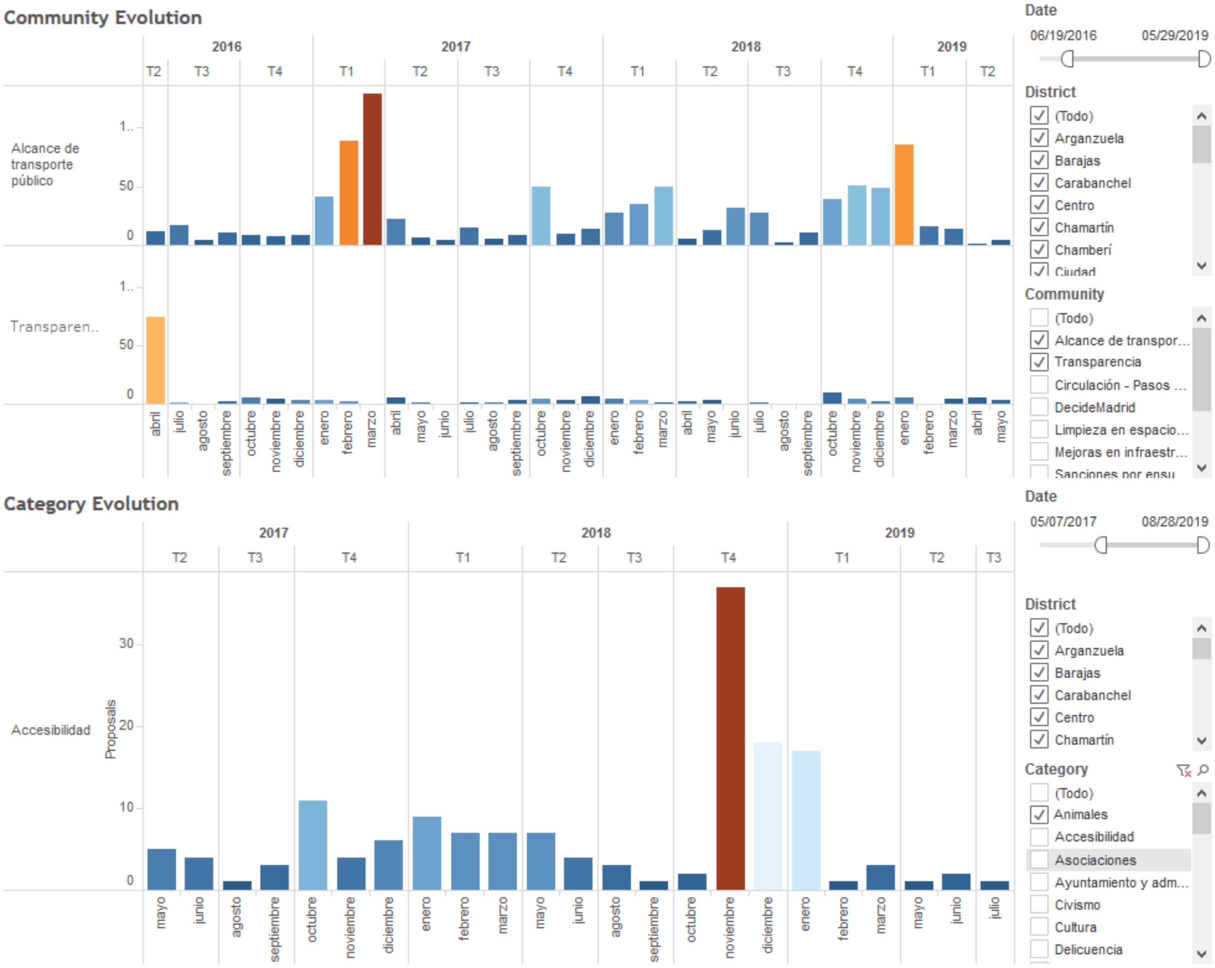
8 ACKNOWLEDGEMENTS

This work was supported by the Spanish Ministry of Science and Innovation (PID2019-108965GB-I00) and the Centre of Andalusian Studies (PR137/19).

REFERENCES

[1] Monira N Aldelaimi, M Anwar Hossain, and Mohammed F Alhamid. 2020. Building dynamic communities of interest for Internet of Things in smart cities. *Sensors* 20, 10 (2020), 2986.

[2] Mariam Asad, Christopher A Le Dantec, Becky Nielsen, and Kate Diedrick. 2017. Creating a sociotechnical API: Designing city-scale community engagement. In



Date
06/19/2016 05/29/2019

District
☑ (Todo)
☑ Arganzuela
☑ Barajas
☑ Carabanchel
☑ Centro
☑ Chamartín
☑ Chamberí
☑ Ciudad

Community
□ (Todo)
☑ Alcance de transpor...
☑ Transparencia
□ Circulación - Pasos ...
□ DecideMadrid
□ Limpieza en espacio...
□ Mejoras en infraestr...
□ Sanciones por ensu

Date
05/07/2017 08/28/2019

District
☑ (Todo)
☑ Arganzuela
☑ Barajas
☑ Carabanchel
☑ Centro
☑ Chamartín

Category
□ (Todo)
☑ Animales
□ Accesibilidad
□ Asociaciones
□ Ayuntamiento y adm...
□ Civismo
□ Cultura
□ Delicuencia

Figure 3: Temporal evolution of proposals by district, category and community.

Conf. on Human Factors in Computing Systems. ACM, 2295–2306.

[3] Raissa Barcellos, José Viterbo, Leandro Miranda, Flávia Bernardini, Cristiano Maciel, and Daniela Trevisan. 2017. Transparency in practice: using visualization to enhance the interpretability of open data. In *18th International Conference on Digital Government Research*. 139–148.

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, 10 (2008), P10008.

[5] Glenda Amayo Caldwell and Marcus Foth. 2014. DIY media architecture: open and participatory approaches to community engagement. In *2nd Media Architecture Biennale Conference: World Cities*. ACM, 1–10.

[6] Iván Cantador, María E Cortés-Cediel, and Miriam Fernández. 2020. Exploiting Open Data to analyze discussion and controversy in online citizen participation. *Information Processing & Management* 57, 5 (2020), 102301.

[7] Miyoung Chong, Abdulrahman Habib, Nicholas Evangelopoulos, and Han Woo Park. 2018. Dynamic capabilities of a smart city: An innovative approach to discovering urban problems and solutions. *Government Information Quarterly* 35, 4 (2018), 682–692.

[8] André Eberhardt and Milene Selbach Silveira. 2018. Show me the data! A systematic mapping on open government data visualization. In *19th Annual International Conference on Digital Government Research*. ACM, 1–10.

[9] Francesca Fallucchi, Michele Petito, and Ernesto William De Luca. 2018. Analysing and visualising open data within the data and analytics framework. In *Research Conference on Metadata and Semantics Research*. Springer, 135–146.

[10] Daniel Filonik, Tomasz Bednarz, Markus Rittenbruch, and Marcus Foth. 2015. Collaborative data exploration interfaces: From participatory sensing to participatory sensemaking. In *2015 Big Data Visual Analytics*. IEEE, 1–2.

[11] Frank Fischer. 2006. Participatory governance as deliberative empowerment: The cultural politics of discursive space. *The American review of public administration* 36, 1 (2006), 19–40.

[12] Nigel Franciscus, Xuguang Ren, Junhu Wang, and Bela Stantic. 2019. Word Mover's Distance for Agglomerative Short Text Clustering. In *Proceedings of the 2019 Conference on Intelligent Information and Database Systems*. 128–139.

[13] Maria Giatsoglou, Despoina Chatzakou, Vasiliki Gkatziaki, Athena Vakali, and Leonidas Anthopoulos. 2016. CityPulse: A platform prototype for smart city social data mining. *Journal of the Knowledge Economy* 7, 2 (2016), 344–372.

[14] Hollie Russon Gilman. 2016. *Participatory budgeting and civic tech: The revival of citizen engagement*. Georgetown University Press.

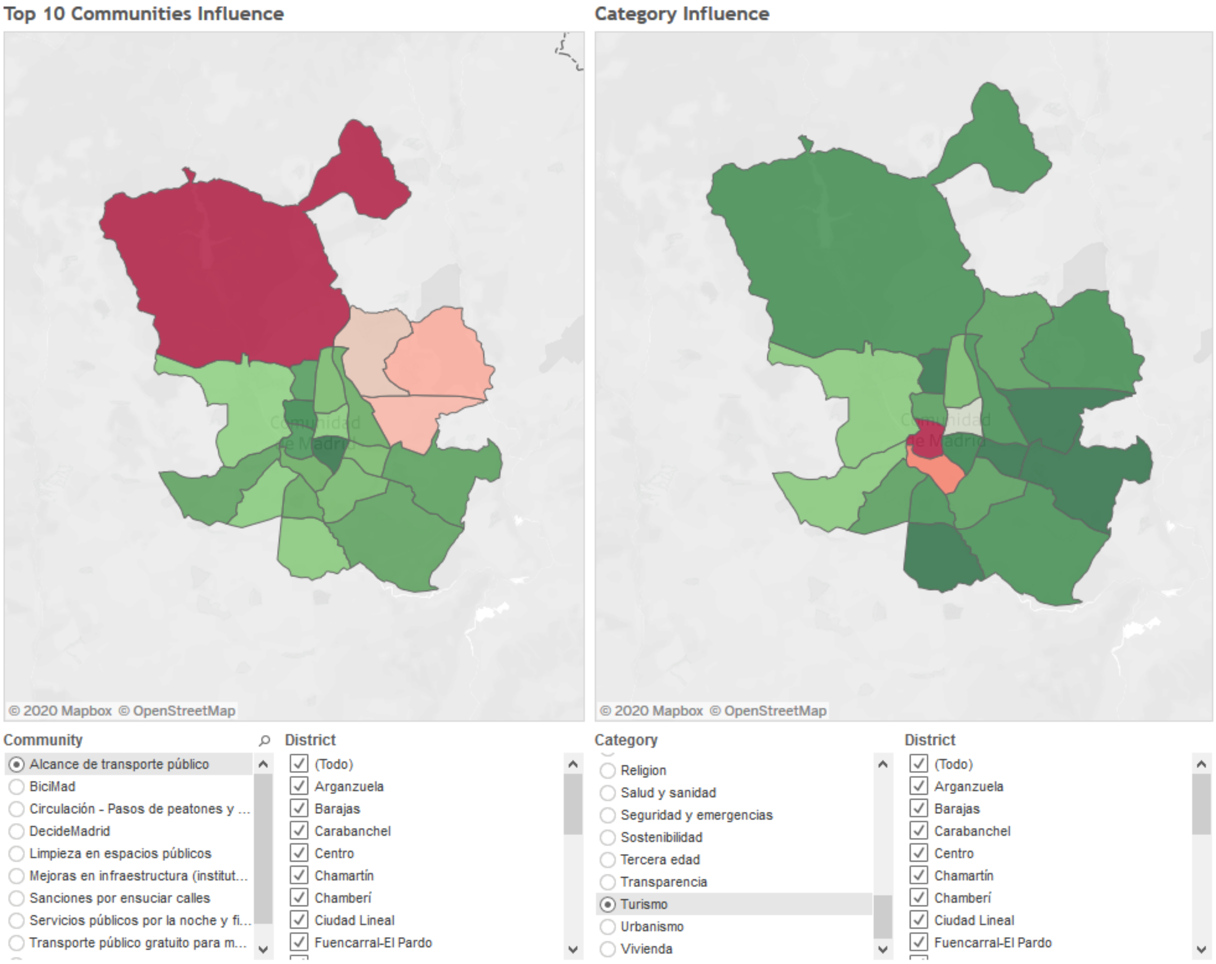
[15] E Hindocha, V Yazhiny, A Arunkumar, and P Boobalan. 2019. Short-text Semantic Similarity using GloVe word embedding. *International Research Journal of Engineering and Technology* 6, 4 (2019).

[16] Rocío B Hubert, Elsa Estevez, Ana Maguitman, and Tomasz Janowski. 2018. Examining government-citizen interactions on Twitter using visual and sentiment analysis. In *Proceedings of the 19th Annual International Conference on Digital Government Research*. ACM, 1–10.

[17] R Ibrahim, S Zeebaree, and K Jacksi. 2019. Survey on semantic similarity based on Document Clustering. *Advances in Science and Technology. Research Journal* 4, 5 (2019), 115–122.

[18] David Rios Insua, Gregory E Kersten, Jesus Rios, and Carlos Grima. 2008. Towards decision support for participatory democracy. In *Handbook on Decision Support Systems 2*. 651–685.

[19] Supavit Kongwudhikunakorn and Kitsana Waiyamai. 2020. Combining Distributed Word Representation and Document Distance for Short Text Document Clustering. *Journal of Information Processing Systems* 16, 2 (2020).



Top 10 Communities Influence

Category Influence

Community

● Alcance de transporte público
○ BiciMad
○ Circulación - Pasos de peatones y ...
○ DecideMadrid
○ Limpieza en espacios públicos
○ Mejoras en infraestructura (institut...
○ Sanciones por ensuciar calles
○ Servicios públicos por la noche y fi...
○ Transporte público gratuito para m...

District

☑ (Todo)
☑ Arganzuela
☑ Barajas
☑ Carabanchel
☑ Centro
☑ Chamartín
☑ Chamberí
☑ Ciudad Lineal
☑ Fuencarral-El Pardo

Category

○ Religion
○ Salud y sanidad
○ Seguridad y emergencias
○ Sostenibilidad
○ Tercera edad
○ Transparencia
● Turismo
○ Urbanismo
○ Vivienda

District

☑ (Todo)
☑ Arganzuela
☑ Barajas
☑ Carabanchel
☑ Centro
☑ Chamartín
☑ Chamberí
☑ Ciudad Lineal
☑ Fuencarral-El Pardo

Figure 4: Heat map showing the relevance of proposal topics (communities) and categories on each district

[20] Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. 2015. From word embeddings to document distances. In *Proceedings of the 32nd Conference on Machine Learning*. 957–966.

[21] Sanna Marttila and Andrea Botero. 2013. The Openness Turn'in co-design. From usability, sociability and designability towards openness. *CO-CREATE* (2013), 99–111.

[22] Amal Marzouki, F Lafrance, Sylvie Daniel, and Sehl Mellouli. 2017. The relevance of geovisualization in citizen participation processes. In *Proceedings of the 18th Annual International Conference on Digital Government Research*. 397–406.

[23] Mazdak Nik-Bakht and Tamer E El-diraby. 2016. Communities of interest-interest of communities: Social and semantic analysis of communities in infrastructure discussion networks. *Computer-Aided Civil and Infrastructure Engineering* 31, 1 (2016), 34–49.

[24] Sora Park and J Ramon Gil-Garcia. 2017. Understanding transparency and accountability in open government ecosystems: The case of health data visualizations in a state government. In *18th International Conference on Digital Government Research*. 39–47.

[25] Directorate-General For Internal Policies. 2012. Potential And Challenges of E-Participation In The European Union. (2012).

[26] Francesco Restuccia, Sajal K Das, and Jamie Payton. 2016. Incentive mechanisms for participatory sensing: Survey and research challenges. *ACM Transactions on Sensor Networks* 12, 2 (2016), 1–40.

[27] Omid Shahmirzadi, Adam Lugowski, and Kenneth Younge. 2019. Text similarity in vector space models: a comparative study. In *Proceedings of the 18th International Conference On Machine Learning And Applications*. 659–666.

[28] Adrien Sieg. 2018. Text Similarities: Estimate the degree of similarity between two texts. *Medium* 5 (2018).

[29] Iryna Susha and Åke Grönlund. 2014. Context clues for the stall of the Citizens' Initiative: lessons for opening up e-participation development practice. *Government Information Quarterly* 31, 3 (2014), 454–465.

[30] Nina Valkanova, Sergi Jorda, and Andrew Vande Moere. 2015. Public visualization displays of citizen data: Design, impact and implications. *International Journal of Human-Computer Studies* 81 (2015), 4–16.

[31] Shuiqiao Yang, Guangyan Huang, Bahadorreza Ofoghi, and John Yearwood. 2020. Short text similarity measurement using context-aware weighted biterms. *Concurrency and Computation: Practice and Experience* (2020), e5765.

[32] Yueping Zheng and Hindy Lauer Schachter. 2017. Explaining citizens' E-participation use: The role of perceived advantages. *Public Organization Review* 17, 3 (2017), 409–428.