From Text to Insight: A Natural Language Processing framework for analysing community perceptions for urban climate resilience

Abstract

Urban resilience strategies should reflect and strengthen community perceptions of environmental and climate risks. However, policymakers often struggle to engage diverse groups and incorporate insights from multiple and informal sources into resilience plans. This study presents a data-driven framework using Natural Language Processing to analyse community narratives regarding local risks and urban governance. By combining complementary analytical and visualisation techniques, we identify public concerns and perceptions and translate them into actionable urban resilience strategies. This approach is applied to a publicly owned neighbourhood in Lisbon facing aging infrastructure, limited access to services, energy poverty, and high unemployment. Primary data were collected from web-scraped sources and focus groups involving public authorities, local associations, and elderly inhabitants. Unsupervised topic modelling (BERTopic) was employed to identify key themes in digital narratives, including infrastructure quality, environmental conditions, and relations with local social institutions. Focus group transcripts were analysed through sentiment analysis, combining lexicon-based (VADER) and transformer-based (BART) models with expert validation and comparative assessment. Results show low levels of community awareness of environmental and climate risks and significant divergences between institutional and community perceptions of priorities, highlighting gaps in communication, institutional trust, and the perceived effectiveness of municipal renewal initiatives. In response, we propose soft co-design climate actions and targeted community awareness campaigns to strengthen risk perceptions and adaptive capacity. This framework offers transferable lessons for urban climate resilience by integrating insights from both digital and physical communities that are often underrepresented in formal decision-making, thereby supporting more inclusive local adaptation planning.

Keywords: Natural Language Processing; Community disaster preparedness; Risk perceptions; Vulnerable groups; Climate-related risk analysis; Integrated analysis.

1. Introduction

Climate change is one of the most pressing challenges of the 21st century and in the near future, particularly in urban environments (IPCC, 2023; World Meteorological Organization, 2024). The frequency and severity of extreme events are increasing, with record-breaking heatwaves, prolonged droughts, and devastating wildfires across Southern Europe and Mediterranean. Intense rainfall, rising water levels, and widespread floodings, driven by uncontrolled urbanisation and deforestation have also severely affected central and southern European countries (European Court of Auditors, 2024; World Meteorological Organization, 2024; CEMS, 2024).

To mitigate the impacts of multi-sectoral cascading risks, public authorities should prioritise disaster risk reduction (DRR) planning across housing, critical infrastructure, and community preparedness through whole-of-society approaches (Clark et al., 2025). Strengthening community preparedness is particularly important for protecting vulnerable populations, who are often more exposed to and less able to recover from catastrophic events and climate-related hazards (Cutter, 2024; Fekete et al., 2025). In this context, urban resilience is closely shaped by how local communities perceive climate-related and environmental threats. In fact, community risk perceptions play a key role in how individuals and communities understand, prepare for, and respond to both natural hazards and human-made threats (Yu et al., 2021). When risk perception is not incorporated into DRR planning, policies often fail to achieve their intended impact, either in scope or scale (IPCC, 2022; Zander et al., 2023). Understanding community risk perceptions is essential since the attitudes of people and ways of evaluating risk influence whether and how they adopt risk-mitigation practices (Tang et al., 2022; Dhar et al., 2023). Indeed, low risk perception often leads to insufficient preparedness and weakened system responses, even in areas with high exposure to hazards. Incorporating risk perception across all phases of disaster risk management contributes to inclusive, context-sensitive policies that can enhance resilience by aligning institutional responses with the lived experiences and local needs of residents or visitors (Cutter, 2024).

The literature on community risk perception, covering social and behavioural psychology, human geography, DRR, comparative planning and climate justice, presents different interpretations (Dhar et al., 2023). Over the years, it was analysed using social surveys and statistical modelling techniques, such as multinomial logistic regression (Buchenrieder et al., 2021). Linear regression was used to identify relationships between risk perception and preparedness, as discussed by Kiani et al. (2022) in their study of the post-earthquake context in Pakistan.

In recent years, Natural Language Processing (NLP) and text mining have become increasingly common in different domains, driven by the rapid growth in the volume of text data generated daily across multiple sources and by advances in these computational methods. When legal consent is granted by data owners, the analysis of unstructured data from news media and social media platforms, like Twitter and Sina Weibo (a Chinese microblogging website), offer timely, wide-reaching, and sensitive insights during disasters (Veltri & Atanasova, 2015; Tang et al., 2022; Zhai et al., 2024; Zhang et al., 2024).

Local communities and other key stakeholders working in a specific area can express their views through informal digital platforms, face-to-face interactions, or other public discussion forums. Such primary data can capture dimensions of lived experience that are often not reflected in traditional statistical or official datasets. Therefore, where the available information is sufficiently rich, AI-based computational approaches (including NLP techniques) are highly relevant to urban climate research, although they remain largely exploratory and underutilised in practice. Despite its potential to leverage large volumes of unstructured data for tasks such as clustering and sentiment analysis, the use of NLP in urban planning research, and particularly in the field of urban climate management, remains underdeveloped. This study fills this gap by collecting, clustering, and processing multi-source data to provide insights for urban climate resilience through the integration of perspectives from both digital and physical communities.

More precisely, we analysed the perceptions of key stakeholders in a social housing neighbourhood in Lisbon (*Bairro da Boavista*), selected for its unique social and environmental context and its public ownership, which allows for a potentially greater direct impact of the local public authority on decision-making.

Textual data are collected from different sources across multiple scales (from national to local), including web-scraped digital content and five focus group discussions, involving representatives of public authorities, local associations, and elderly inhabitants. The data are analysed, mapped, and organised into thematic patterns using NLP tools, with results validated by a panel of three experts: a sociologist, a geographer specialising in civil protection and vulnerability risk assessment, and a civil protection official from the Lisbon Fire Brigade.

Since the data came from diverse sources and is rich, the study used a deductive approach as its main method, while also incorporating a reflexive, data-driven component to identify patterns and insights.

The key contributions of this study can be summarised as follows:

- Development of a semantic framework to analyse fine-grained contextual information across diverse data sources and multiple administrative levels. We integrated automated techniques with expert validation to provide a textual and visual analysis of the perceptions and perspectives of diverse communities that are often underrepresented in public decision-making.
- Comparative evaluation of NLP approaches, combining topic modelling using BERTopic with sentiment analysis based on BART (a transformer-based model) and VADER (a lexicon-based model).
- Identification of potential actions for enhancing urban climate resilience.

The identified research gap regards the lack of an integrated framework in urban climate resilience research that combines textual data from both online and offline community sources to inform locally grounded future strategies. Such a framework is needed to better align perceived environmental and climate risks with hazard mapping and urban renewal initiatives, while also incorporating dimensions of community engagement, social inclusion, and climate governance.

This paper presents findings from activities undertaken within RETIME – Urban Adaptation and Alert Solutions for a TIMEly (re)Action, a Horizon Europe project co-funded by the European Union under Grant Agreement No. 101147113 (2024-2028).

It is organised as follows. *Section 2* explains the rationale for selecting the NLP tools, highlights their use for disaster risk reduction, urban governance, and the public sector, and situates the study within the existing literature. *Section 3* presents the research context, detailing the case study and its key characteristics and risks. *Section 4* describes the workflow, data sources, and tools used for web scraping and focus group analyses. The results are discussed in *Section 5*, while *Section 6* examines the potential integration of these insights into public initiatives to enhance urban climate resilience strategies. Research limitations and key challenges are summarised in *Section 7*. Finally, *Section 8* outlines the main findings for policymakers in other contexts and suggests directions for future research.

2. Automated NLP techniques and related work

2.1. Topic Modelling and Sentiment Analysis methods

Topic Modelling is an unsupervised learning technique that automatically uncovers hidden themes within large corpora of text (Hankar et al., 2025). Traditional methods, such as Latent Dirichlet Allocation (LDA), Probabilistic Latent Semantic Analysis (PLSA), and Non-Negative Matrix Factorisation (NMF), are typically applied to bag-of-words representations. While proven relevant, these models overlook context-based meanings and semantic nuances. This limits their ability to capture semantic relationships, making it difficult to distinguish between polysemy and synonymy, often resulting in less coherent topics (Abdelrazek et al., 2022).

BERT (Bidirectional Encoder Representations from Transformers) architecture overcomes this limitation. BERTopic employs BERT embeddings in combination with dimensionality reduction (e.g., UMAP - Uniform Manifold Approximation and Projection) and density-based clustering (e.g., HDBSCAN - Hierarchical Density-Based Spatial Clustering of Applications with Noise) to extract consistent topics. It combines transformer-based embeddings with clustering and frequency-based analysis to generate consistent topic representations (Grootendorst, 2022). This model automatically detects topic structures using unstructured texts. In comparative studies among different NLP-techniques, BERTopic consistently outperformed classical models in terms of coherence, topic diversity, and human interpretability, especially with short-text corpora such as social media posts or news articles (Hankar et al., 2025; Hristova & Netov, 2022). These characteristics make this tool appropriate within public discourse on risk perception, where contextual meaning is essential.

The process begins with SBERT (Sentence BERT) that converts text into dense semantic embeddings. These are then reduced in dimensionality using UMAP to preserve structure and enhance clustering efficiency. HDBSCAN clusters the data without requiring a predefined number of topics, while also identifying noise. Next, CountVectorizer tokenizes the text, and c-TF-IDF (Term Frequency-Inverse Document Frequency) calculates term importance within each cluster, highlighting key topic-defining words. An optional fine-tuning step can further improve coherence by merging or adjusting topics (Grootendorst, 2022). Its architecture is shown in Figure 1.

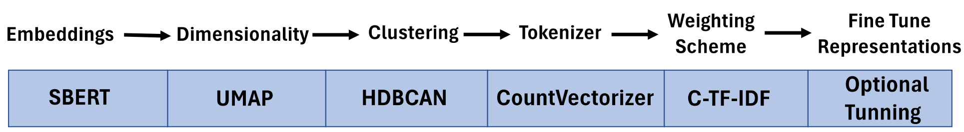

Figure 1. BERTopic algorithm (adapted from Grootendorst, 2022).

Alongside topic modelling, sentiment analysis uncovers textual emotions. It is the process of collecting and examining individuals' views and attitudes on a variety of subjects, using techniques from NLP and text mining to identify the underlying sentiment, typically categorised as positive, negative, or neutral (Liu, 2012). Within the range of existing methods, VADER is a lexicon and rule-based tool that was developed by Hutto & Gilbert (2014). It uses a curated list of words (lexicon), each with associated sentiment scores, and applies grammatical rules to interpret intensity and emphasis. VADER is effective for analysing sentiment in social media contexts due to its sensitivity to informal language, emotive expressions, and slang. Given the informal aspect of the chosen data for this study (focus groups transcripts), VADER can be considered an effective tool. Furthermore, a custom lexicon can be incorporated to train a more accurate algorithm.

In addition to rule-based approaches like VADER, transformer-based models, such as Facebook AI's BART, introduced by Lewis et al. (2019) is also robust. BART is a sequence-to-sequence model that combines the strengths of two foundational architectures in NLP: BERT and GPT (Generative Pre-trained Transformer). During training, BART uses a denoising autoencoding objective where sentences are deliberately corrupted (e.g., by masking or shuffling tokens), and the model learns to reconstruct the original sentence. This dual structure enables BART to perform well in both understanding and generation tasks, including summarisation, machine translation, and zero-shot sentiment classification.

Zero-shot classification is a machine learning technique that enables a model to assign labels to data it has not been explicitly trained on, relying instead on its general understanding of language to make predictions about new and unseen examples (Kyritsis et al., 2023). In zero-shot settings, BART reframes sentiment analysis as a textual entailment problem, comparing the input text with hypothesis statements like 'This text expresses a positive sentiment', without requiring supervised training on labelled data (Lewis et al., 2019). This makes it effective in low-resource or multilingual environments.

2.2. Related work on NLP applications in text analysis

Topic modelling and sentiment analysis are increasingly used to analyse public discourse and disaster risk perception research. Topic modelling techniques are applied through semantic and thematic analysis of the corpus, while sentiment analysis supports the theoretical framework linking expressed sentiment to perceived risk, with more negative sentiment corresponding to higher perceived risk (Chang & Wang, 2018; Zhang et al., 2024).

Zander et al. (2023) applied both Latent Dirichlet Allocation (LDA) topic modelling and sentiment analysis to assess public perception during the 2021 floods in Germany, while Zhai et al. (2024) used a similar approach in the context of hurricanes, employing the VADER sentiment algorithm alongside LDA. Bidirectional Long Short-Term Memory (BiLSTM)-based models have also been adopted for sentiment analysis in risk contexts (Cheng et al., 2024). BiLSTM was combined with LDA for sentiment classification and keyword extraction from social media during Typhoon Lekima in 2019 (Tang et al., 2022). A similar approach was also applied for health emergencies by Zhang et al. (2024), where a sentiment lexicon and BiLSTM were used to analyse risk perception levels.

By leveraging user-generated content (UGC), hybrid methodologies have also emerged in urban planning and post crisis-management. Stellacci & Moro (2023) assessed the quality of urban public spaces (historic squares and train stations) through the analysis of TripAdvisor reviews using VADER, with the aim of enhancing planning strategies in Italian heritage cities experiencing over-tourism. Twitter data and focus group discussions related to the Canterbury earthquakes in 2011 (in New Zealand) were analysed by Mora et al. (2015) exploring long-term perceptions of seismic risk. Hristova & Netov (2022) applied BERTopic to a large corpus of media and forum data to track public perception of distance learning during the COVID-19 pandemic, focusing on the relationship between media framing and emerging public attitudes. Oliveira et al. (2024) used text mining on news articles to reveal how media narratives frame the Rio Grande do Sul floods, influencing public understanding of vulnerability and climate change.

Across different tools and data sources, all these studies show how NLP can shape and enhance public domains (Kraus et al., 2021; Muhammad et al., 2025; Stellacci and Moro, 2023). In fact, as highlighted by Hristova & Netov (2022) and Zhang et al. (2024), among others, news media not only inform, but also actively shape public discourse, offering a rich dataset for analysing key aspects in contemporary built environments. Since news content often relies on expert and institutional sources, it can play a key role as a trusted medium for information dissemination during emergencies (Wu et al., 2021).

3. Research context

This analysis focuses on *Bairro da Boavista*, a publicly owned neighbourhood in Lisbon, Portugal, selected for its municipal housing stock and sociological profile of its population. The area is characterised by a diverse range of urban and architectural solutions that reflect both the economic challenges and the opportunities associated with publicly managed housing. These include the potential for coordinated spatial planning and the implementation of more inclusive policies, which are particularly effective in this context given the housing stability. Publicly owned buildings with controlled rents often face limited resources but can benefit from integrated renewal plans and urban resilience measures, opportunities rarely available in privately owned or differently managed buildings. In addition, the neighbourhood can serve as a testing ground for broader urban resilience strategies.

Bairro da Boavista was also selected as case study due to its exposure to multiple climate-related risks and its relevance within the current rehabilitation and adaptation agenda in Lisbon. In fact, the neighbourhood has benefited from several public investments (such as the construction of the primary school and new residential building blocks) and is covered by rehabilitation programmes at both district and city levels, including the 2016-2030 flood risk mitigation plan. This makes this neighbourhood an important case study for examining community engagement and promoting sustainable urban environment under conditions of limited resources.

Therefore, the combination of natural hazards, social vulnerability, active public investment, and ongoing adaptation initiatives makes *Bairro da Boavista* a valuable case study for exploring community-centred approaches to urban climate resilience.

Section 3.1 outlines key aspects of the sociological profile of the population and the neighbourhood's construction history, while *Section 3.2* describes local risks and public initiatives aimed at strengthening urban resilience and social cohesion.

3.1. Case study

Bairro da Boavista falls under the jurisdiction of the Benfica Parish Council, the level of government closest to residents and responsible for local governance and community services. It was developed in response to the population boom and housing shortage that affected the city during the 1930s and 1940s (Figure 3). Located in the north-western part of Lisbon, this neighbourhood is home to approximately 1,900 residents within a relatively small area (5.6 km²) and includes people from diverse ethnic backgrounds.

The neighbourhood consists of privately occupied housing that is publicly owned and managed by the municipal company Gebalis,[1] along with local facilities such as a school and other community facilities. Despite the different phases of construction of the housing stock (1960s–1970s, 1976–1977, 1981–1984, 1988–1996, 1997–1999) strong community ties have persisted over time. The building stock is mainly composed of mid-rise residential buildings, typically ranging from 4 to 8, and in some cases up to 12 storeys – most of which without elevator (INE, 2021) – with lower-rise structures (e.g., service buildings or older fabric) generally limited to 2–3 storeys.[2] *Bairro da Boavista* also displays a relatively balanced intergenerational population structure, although the overall

---

[1] 94.5% are tenants, and only 3.3% are building owners, and about 2.1% fall into other categories (source: Gebalis, www.gebalis.pt).

[2] The 2021 Census data (INE) does not provide detailed household composition, but it indicates that 49.2% of households consist of 1-2 people, 31% of 3-4 people, and 19.9% of five or more. This distribution suggests a substantial proportion of small households, which may include single-person and potentially single-parent households. Such a pattern points to a possible presence of social vulnerability linked to household isolation, although further disaggregation is required for confirmation.

level of formal education among residents is relatively low and the unemployment (or informal employment) is high (INE, 2021).
Even with the recent rehabilitation programme (2018–2019) and ongoing infrastructure improvements (IEA EBC, 2020), this neighbourhood continues to face significant challenges at both the building and district level, including accessibility barriers for people with reduced mobility, limited links to Lisbon city centre services, and building energy poverty.

Figure 2. Panoramic view and satellite image of *Bairro da Boavista* (Lisbon).

3.2. Natural hazards and local urban renewal initiatives

Natural hazards relevant to *Bairro da Boavista* have been identified from digital platforms and district urban plans, with potential impacts on various user groups (Figure 3).
The hazard maps were derived from a range of sources, including international and European platforms such as the Risk Data Hub, the European Forest Fire Information System (EFFIS), the European Facilities for Earthquake Hazard and Risk (EFEHR), and Copernicus services, including the Copernicus Emergency Management Service (CEMS) and the European Severe Weather Database (ESWD). These datasets were complemented by high-resolution information from local geoportals. For the Lisbon Metropolitan Area, the *Lisboa Interactiva Data Portal* was used, as it aggregates high-resolution geospatial data collected specifically for the city of Lisbon, while local plans offer more detailed coverage at the district level (CML, 2018, 2019).
These natural hazards, outlined below, reflect both local conditions and broader regional climate and geological risks.

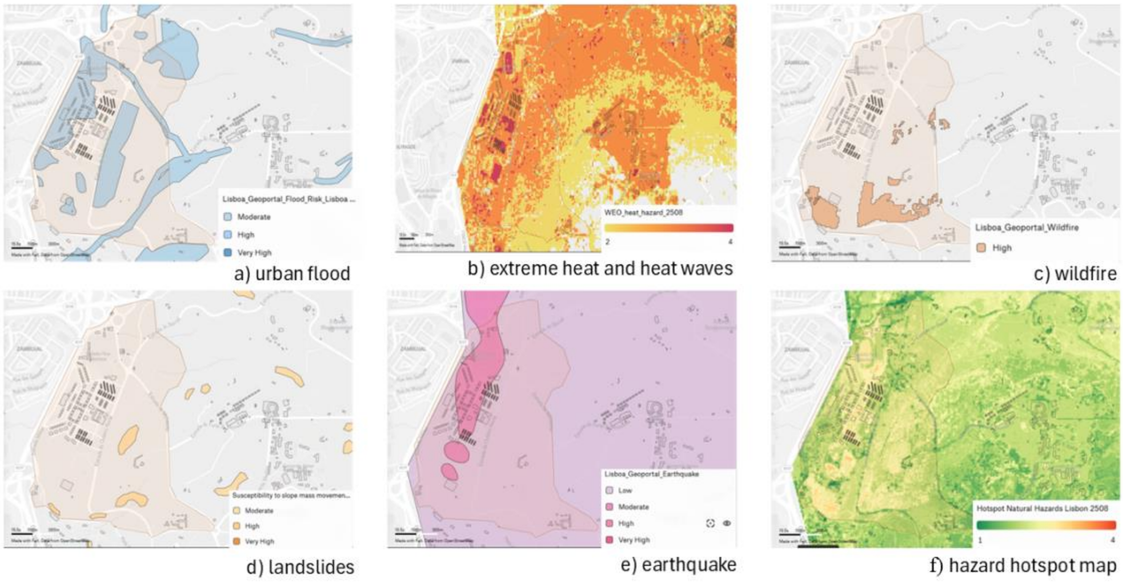

Figure 3. Hazard maps of *Bairro da Boavista*, Lisbon: (a) urban flooding; (b) heat waves; (c) wildfires; (d) landslides; (e) earthquakes; and (f) hazard hotspot map (data sources: OpenStreetMap; map produced by WEO – Environmental Analytics from Satellites).

Urban flood is assessed at a moderate level on the moderate–high–very high-risk scale, but it can still have significant impacts, particularly during periods of intense rainfall associated with severe storm events. In this densely built urban environment, stormwater drainage systems may become overwhelmed throughout the area adjacent to the main road, *Estrada Militar*, increasing the likelihood of flash flooding (Figure 3a). Key vulnerable groups include ground-floor residents, elderly people, and individuals depending on critical infrastructure.

Extreme heat and heat waves are represented on a color-coded scale from low to very high risk (1-4 scale), which constitutes a growing hazard for Lisbon (Figure 3b). Rising temperatures associated with climate change, combined with the urban heat island effect, can intensify heat events and increase risks to public health. In this area, elderly people and low-income households are particularly vulnerable. Other key vulnerable groups include children and individuals working in primary schools and the stadium, as well as residents in the central residential blocks, where risk levels range from 3 to 4.

The heat hazard categories are established using Land Surface Temperature (LST) to monitor urban thermal differences and identify critical hotspots. LST was derived from Sentinel-3 SLSTR thermal bands sharpened to 10 m with Sentinel-2 VIS-NIR using pyDMS (Li et al., 2023). Retrieval utilised split-window algorithms for daytime (S8-S9) and nighttime (S7-S9) based on fused BT, land cover (LC), NDVI, and TCWV inputs (Li et al., 2023). LST pixels were classified into hazard levels using established air temperature thresholds (MeteoSwiss, 2025; OECD, 2025; WHO, 2023) adapted to LST, consistent with the lower bound of the empirically observed LST–Ta difference for mixed urban surfaces in temperate European climates (typically 3–10°C depending on surface type and conditions). The resulting hazard levels are defined by the following specific LST thresholds: Level 1 (< 35°C), Level 2 (> 35°C), Level 3 (> 40°C), and Level 4 (> 45°C).

Other hazards can be considered less significant for this pilot site. For instance, localised *wildfires* may occur in non- built-up areas (Figure 3c), while *landslides* may arise in areas affected by slope instability combined with heavy rainfall (Figure 3d). *Bairro da Boavista* is located at a relatively elevated position and is therefore not considered particularly vulnerable to coastal flooding or tsunami impacts. The *seismic hazards* mapped for this area (e.g., soil response, potential shaking intensity) indicate a moderate to high geological risk, as Lisbon is located in a seismically active region (Figure 3e).

Hazard hotspot map (Figure 3f) identifies critical urban zones by combining multiple high-resolution datasets for six primary hazards in the Lisbon pilot area in a single map: wildfire, windstorm (10-year return period), landslide, extreme heat, urban flood, and seismic activity. These inputs are sourced from a combination of satellite-derived hazard maps, the JRC Risk Data Hub, and the Lisbon Geoportal. To ensure comparability across different measurement units and hazard characteristics, each individual map is first rescaled to a uniform 1–5 scale. The final hotspot index is derived by calculating the unweighted average of all rescaled hazard layers and these calculated mean hazard scores are classified into five descriptive risk levels: Very Low (≤ 1.90), Low (≤ 2.15), Moderate (≤ 2.40), High (≤ 2.75), and Very High (> 2.75).

To minimise the impacts of climate-related events and enhance urban resilience and social cohesion, several initiatives have been implemented in *Bairro da Boavista* over the years, alongside citywide programs such as Lisboa Cidade Resiliente, ReSist, the LIFE LUNGS project, and DECA. *Eco-Bairro Project* (2007–2013), within the National Strategic Reference Framework,[3] is one of the key initiatives, which involved the renovation of residential buildings, creation of eco-facilities (Eco-Centre and Eco-Gardens), installation of solar thermal systems, free neighbourhood Wi-Fi (Net-Verde), pedestrian pathways (PediBus), and community engagement campaigns. Additionally, *Eco-Bairro Boavista Ambiente+* (2012) focused on upgrading public spaces, improving building energy performance, and integrating renewable energy sources (solar, photovoltaic, wind), along with new pedestrian areas. *Morar Melhor Programme* (2023–2026) is currently underway, targeting 14 building lots (168 homes, 471 residents), with roof and façade interventions to enhance comfort, safety, and energy efficiency. Other programmes, promoted by the public building manager institution (Gebalis) more focused on improving social cohesion and urban identity, are: i) *Lotes Com Vida* (2019, 2021) encouraged resident involvement in lot management and decision-making; ii) *Community Champions League* (since 2019) engaged local youth in sustainability initiatives through sport; and iii) *Memories Project* (2024) aimed to preserve the heritage and stories of relocated residents following demolitions.

4. Methodology

To understand the perceptions of the stakeholders relevant to this case study, the methodology for processing and analysing textual data using NLP techniques is illustrated in Figure 4.

[3] https://ec.europa.eu/regional_policy/en/information/publications/brochures/2008/cohesion-policy-2007-13-national-strategic-reference-frameworks

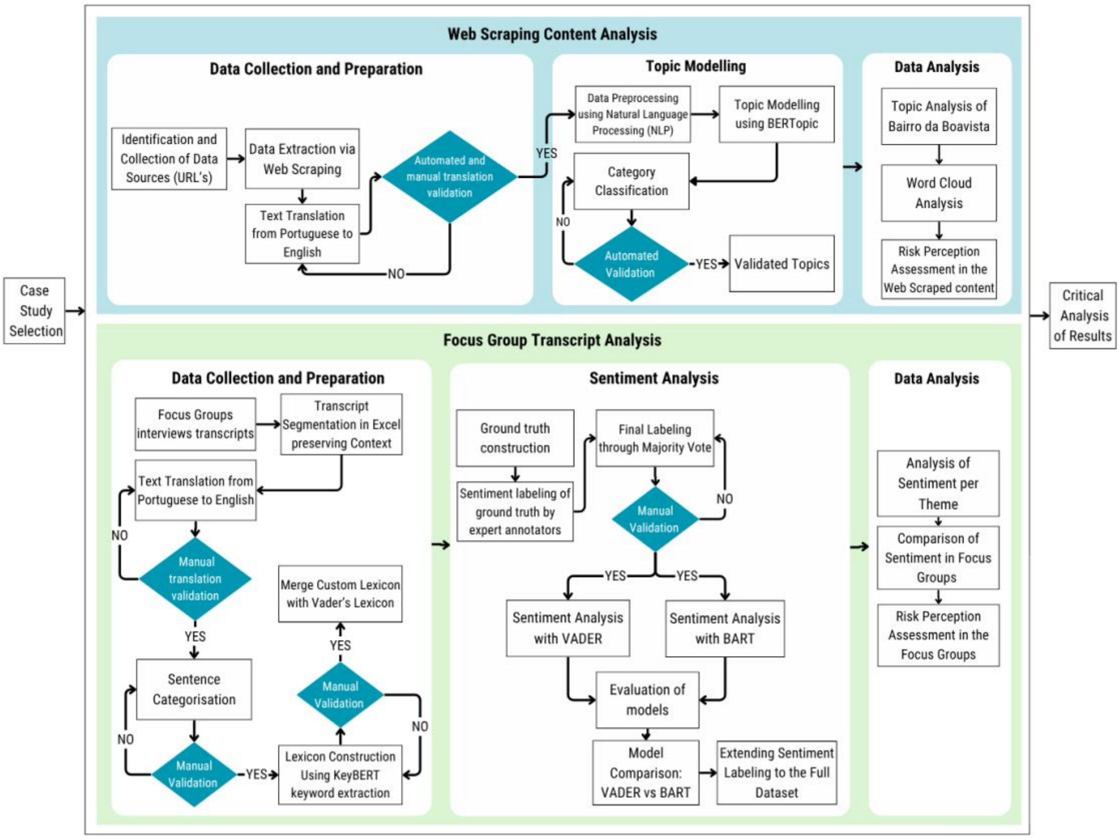

Figure 4. Natural Language Processing framework.

Data analysis is conducted during two parallel steps, with their respective sources described in *Sections 4.1* and *4.2*. One step focuses on web-scraped content from various digital sources, and the second step regards the analysis of focus groups data. As shown in Figure 4, each step involves three phases: i) data collection and preparation; ii) NLP-based modelling (topic modelling and sentiment analysis); and iii) data analysis.

4.1. Web scraping content analysis: data sources and NLP tools

We analysed 25 data sources across national, city, and neighbourhood levels, comprising news outlets, stakeholder websites, and community blogs. These sources are all in Portuguese, media outlets, including print, digital, radio, and television (Figure 5). Blogs from resident associations were also included, reflecting active community voices.

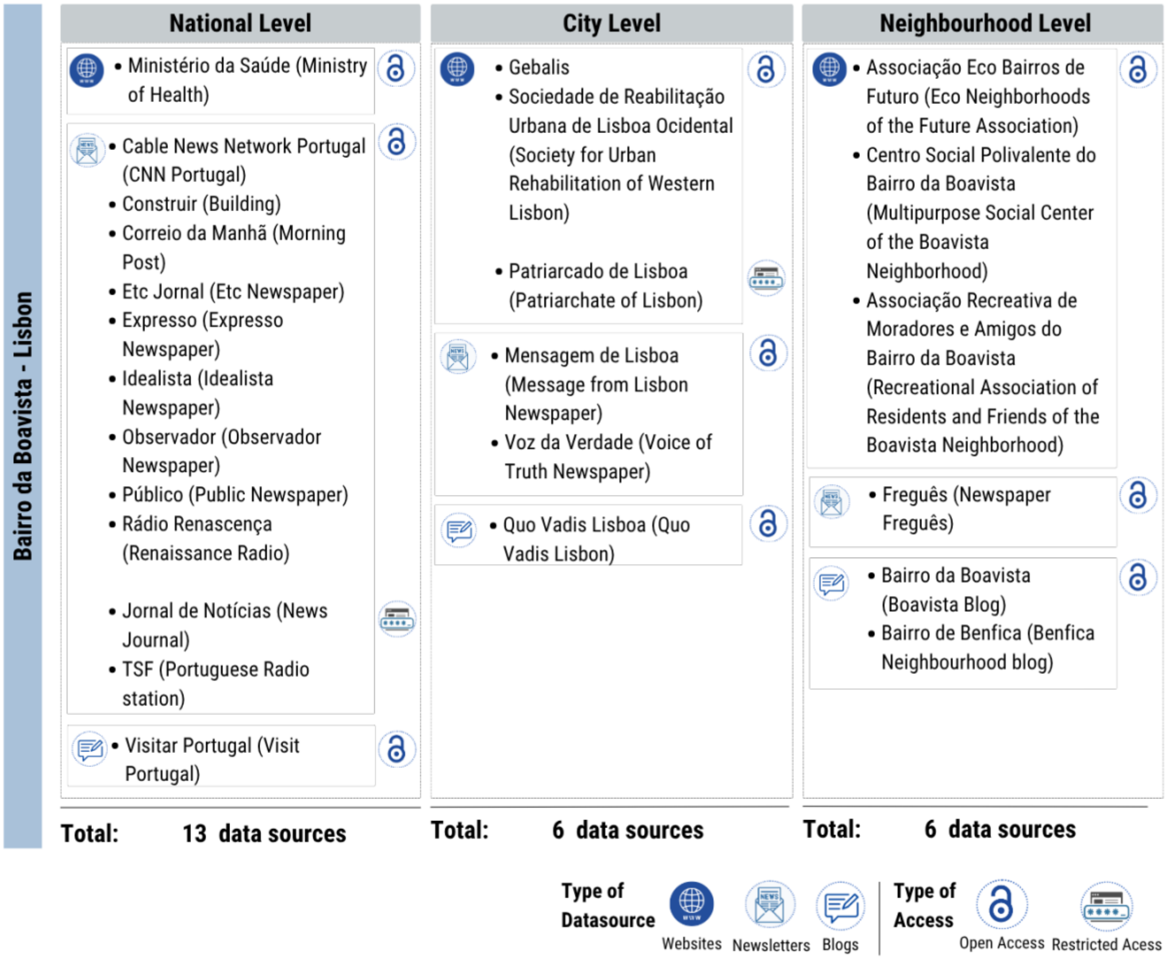

Figure 5. Overview of stakeholders and online data collected within this research.

The analysis of digital contents covered 33 websites, with a total corpus size of 24 kilobytes. The content was collected from July 2024 to November 2024, as a part of the broader stakeholder mapping analysis conducted within the RETIME project, using a custom web scraping tool that uses python's libraries *BeautifulSoup* and *Selenium*.[4] Compliance with legal requirements including the General Data Protection Regulation (European Parliament and Council of the European Union, 2016), were ensured. Regarding website authorisation, robots.txt files were checked for each domain (by appending /robots.txt to the URL) to determine the permissions for data access and crawling, as these files specify the terms governing automated data extraction. Where directives were absent or unclear, website owners were contacted, and consent was obtained. Data sources that did not permit web scraping (e.g., Facebook) were excluded from collection and analysis. In cases where prior access was restricted (e.g., Patriarcado de Lisboa and TSF), explicit permission was requested and granted before data collection. Under approval of the participants, each session was recorded, transcribed using Transkriptor. The content was automatically translated from Portuguese to English using the deep translation library powered by Google Translate API (Application Programming Interface). To reduced translation biases, including those arising from polysemic terms, idiomatic expressions, irony, sarcasm, and context-dependent meanings, all translated content was manually verified by two bilingual moderators of the focus groups to ensure

[4] *BeautifulSoup* (https://pypi.org/project/beautifulsoup4/); *Selenium* (https://www.selenium.dev/)

consistency in meaning, tone, and intent. This process helped avoid decontextualised interpretation and ensured semantic equivalence with the original text. A CSV file containing each website's URL, original content, and translated content was automatically generated for the topic modelling analysis. Other details are provided in *Section 7*.

4.2. Focus groups transcript analysis: data sources and NLP tools

Five focus group discussion sessions were conducted between November 2024 and February 2025, involving 25 Portuguese-speaking participants, and moderated by academicians who are experts in urban planning and risk management (Table 1). These participants were selected based on their administrative, political, and organisational relevance (Table 1, 'PuA1' and 'PuA2'), their role in supporting daily social activities (Table 1, 'LA' and 'PrA'), and their representation of a vulnerable group within the neighbourhood[5] (Table 1, 'LI').

Table 1. Focus groups related to *Bairro da Boavista* (Lisbon).

Focus Group ID	**Stakeholders involved in the *Bairro da Boavista* (Lisbon)**	**Date and Format**	**No. of Participants (P) and Moderators (M)**
PuA1	Members of Benfica Parish Council	27-11-2024 In person	5 P 4 M
PuA2	Representatives of Lisbon City Council, of a public housing management company in Lisbon (Gebalis), and of Municipal civil protection service	19-12-2024 Online	6 P 2 M
LA	Members of *Bairro da Boavista* childcare centre, President of Cape Verdean Association, and the Coordinator of Association *ReTrocas*	20-12-2024 Online	3 P 2 M
PrA	Internal staff of a public charitable institution in Lisbon (i.e., *Santa Casa da Misericórdia de Lisboa*, a private legal entity with public utility status)	19-02-2025 In person	6 P 2 M
LI	Sample of elderly inhabitants in *Bairro da Boavista*	19-02-2025 In person	5 P 3 M

***PuA:** Public Authorities: **PuA1**: Parish officials and **PuA2**: Representatives of Lisbon City Council; members of a public housing management company in Lisbon; members of the municipal civil protection service in Lisbon; **LA:** Local Associations; **PrA:** Private Authorities; **LI:** Local elderly Inhabitants.*

The focus group discussions followed an open format around four key themes: 1) Role of local authorities, 2) Socio-architectural identity, 3) Climate-related risks and mitigation

[5] *Bairro da Boavista* exhibits a relatively balanced age structure: children (0–14) and elderly residents (65+) are similar in number, with youth-to-adult and elderly-to-adult ratios of 31.8% and 32.1%, respectively (INE, 2021). Elderly inhabitants were selected because they represent a key vulnerable group.

measures, and 4) Vulnerabilities and strengths. Table 2 details the set of indicators for each theme that are discussed during the focus groups. These themes reflect relevant dimensions of community life and urban experience.

Table 2. Key themes of the focus group interviews and their sub-themes.

Theme	Sub-themes
Theme 1 - Role of local entities and authorities in the neighbourhood	i) Cooperation between different organisations ii) Governance and decision-making process iii) Public services and infrastructure iv) Community engagement v) Security and legal framework vi) Partnerships and collaboration
Theme 2 - Socio-architectural features and urban identity	i) Sense of belonging and community identity ii) External perceptions iii) Urban regeneration and its impact on residents' well-being iv) Cultural features
Theme 3 - Climate-related risks and potential Mitigation measures	i) Climate-related vulnerabilities ii) Energy efficiency and urban sustainability iii) Resilience strategies and emergency panning
Theme 4 - Vulnerabilities and strengths	i) Socio-economic vulnerabilities ii) Housing conditions and public health iii) Education and prospects iv) Use and maintenance of public spaces v) Social support and community-based networks vi) Local initiatives and success stories vii) Criminality and other local risks

The transcript texts of the focus groups were analysed through two-steps: topic modelling (step 1), detailed in *Section 4.2.1*, and sentiment analysis (step 2) detailed in *Section 4.2.2*, while the discussion of the results is described in *Section 5.2*.

4.2.1. Topic Modelling (*step 1*)

Text embeddings were generated using the paraphrase-MiniLM-L6-v2 model from the Sentence-Transformers library. Dimensionality reduction is performed using UMAP (n_neighbours = 15, min_dist = 0.2). Then, the reduced embedding was clustered using HDBSCAN (min_cluster_size = 2, min_samples = 1). Topic term representation was obtained using class-based c-TF-IDF, with Maximal Marginal Relevance (MMR) set to 0.35 for topic diversity. During the vectorisation pre-processing stage, digits, symbols, and special characters (except apostrophes) were removed. Additional steps included lemmatisation using spaCy,[6] the application of a custom stop word filtering, and the use of n-gram range (1, 2) to capture both unigrams and bigrams. The topics were based on the Social Vulnerability Index (SoVI) (Cutter, 2024), that includes social, economic, and demographic factors, such as income, housing quality, health, and access to social services. The topics were identified using BERTopic, with manual adjustments to the model parameters to optimise the outcomes. Following the method undertaken by Farea

[6] https://spacy.io/

et al. (2024), this study focused on a small, targeted corpora, i.e., five topics shown in Table 3.
First, the most representative words for each topic were examined, following the approach by Farea et al. (2024). Second, BERTopic's inter-topic distance map was employed to evaluate the degree of overlap between topics, which is subsequently verified manually. Once extracted, the topics were labelled according to the SoVI categories (Cutter, 2024).

Table 3. Topics identified in the web scraped content of *Bairro da Boavista* (Lisbon).

Topics	**Description**
Topic 1 - Environmental and Neighbourhood Concerns	Environmental and neighbourhood issues that matter most to local communities, including housing conditions, access to services and green spaces, infrastructure quality, and other factors that affect residents' day-to-day lives.
Topic 2 - Community and Social Institutions	Community-focused institutions (e.g., churches, schools, and community spaces), reflecting social support structures, religious involvement, civic engagement, and shared cultural values.
Topic 3 - Infrastructures and Development	Infrastructural and urban development, particularly related to housing projects and urban or building improvement. It also focuses on institution-led initiatives.
Topic 4 - Public Health and Safety	Public health and safety, intersecting with community resilience, mobility, and interaction.
Topic 5 - Housing Safety and Quality	Housing and urban issues, related to challenges with access, density, and affordability.

4.2.2. Sentiment Analysis (*step 2*)

The focus group transcripts were segmented into smaller units, while preserving the contextual meaning within the corpus, to improve the performance of sentiment models designed for short-form text. This pre-processing step aligns with the models' input constraints and enhances sentiment detection accuracy by minimising contextual noise and ensuring each textual unit conveys a coherent sentiment expression.
Two models were employed for the sentiment classification:
1) VADER, a rule-based sentiment classifier with both the default and custom lexicon;
2) BART (facebook/bart-large-mnli), a zero-shot classifier implemented using the Hugging Face Transformers library. Texts were classified into 'positive', 'neutral,' or 'negative' with confidence scores.
A customised lexicon, that improves sentiment analysis, as proven in literature (Deng et al., 2017; Stellacci and Moro, 2023), was developed to enhance VADER in this study. KeyBERT (model: all-mpnet-base-v2) was used to extract domain-specific keywords per theme and per group. Terms are manually filtered, validated, and annotated with sentiment scores on a three-point scale [-1, 0, +1]. The lexicon size was expanded from 501 to 926 terms using spaCy for inflections, following the approach defined by Gatti et al. (2015) to increase coverage and improve results.
Since the custom lexicon used sentiment scores ranging from -1 to 1, but VADER's lexicon ranges from -4 to 4, scores were scaled as follows: -1 → −2.0, 0 → 0.0, +1 → +2.0. This scaling is essential to preserve the original sentiment meaning of the custom

lexicon while aligning it with VADER's scoring system. It also prevents the custom terms from having an unfairly large influence compared to VADER's existing terms.
Finally, the custom lexicon was merged with VADER's default lexicon to incorporate its existing annotations. To assess and compare the models, a ground truth dataset comprising 30% of the focus groups corpus (n = 134) was manually annotated by a panel of three experts: i) a sociologist specialised in civil protection; ii) a geographer expert in social vulnerability assessment; and iii) an official from civil protection service. Inter-annotator agreement is evaluated using Krippendorff's Alpha (α = 0.6387), representing a moderate and satisfied agreement. In cases where some disagreements of the three experts were noted, the final label was assigned based on a majority vote.
Table 4 presents selected excerpts of sentences from the focus groups for each theme (listed in Table 2), a sample of associated equivalents, along with their respective VADER and BART scores based on a [-1, 0, +1] threshold.

Table 4. Samples of focus group sentences and related scores.

Focus Group ID (detailed in Table 1) **Theme** (detailed in Table 2)	**Excerpts of sentences**	**Equivalents**	**Score [-1, 0, +1]**	
			VADER	**BART**
LA, **Theme 1 –** Role of Local Entities and Authorities in the Neighbourhood	*'The Residents' Association and, I don't know, Os Putos* traquinas [an association whose name means 'Cheeky kids'], *the Pensioners' Association, each worked on their own and then it's funny to see that over time, for a variety of reasons, these organisations have started to work together and they began to realise that things worked better, at least this is the perception I have.'*	'residents', 'association', 'pensioners', 'organisations', 'better'	1	1

PuA2, **Theme 2** – Socio-architectural Features and Urban Identity	*'It's good for the neighbourhood, in the sense that the fact that we're improving and giving people new homes is helping to improve their quality of life and the impact, even in terms of energy issues and energy efficiency, because these new homes already have these features, they already have electricity panels, even the materials used in the construction are materials that help to improve the thermal comfort of the homes.'*	'good', 'help', 'improve', 'issues', 'energy', 'comfort', 'quality of life', 'impact', 'efficiency', 'electricity', 'new', 'thermal comfort'	1	1
PuA2, **Theme 3** – Climate-Related Risks and Potential Mitigation Measures	*'Yes, so head, is the risk of urban fire, risk of forest fire. I think it's the biggest risk, I would say.*	'urban fire', 'risk', 'forest fire'	-1	-1
LI, **Theme 4** – Vulnerabilities and Strengths	*'People who can't go home have medical and home care support.'*	'medical', 'support', 'care'	1	-1

PuA: *Public Authorities:* ***PuA1****: Parish officials and* ***PuA2****: Representatives of Lisbon City Council; members of a public housing management company in Lisbon; members of the municipal civil protection service in Lisbon;* ***LA:*** *Local Associations;* ***PrA:*** *Private Authorities;* ***LI:*** *Local Elderly Inhabitants.*

5. Discussion of results

5.1. Web scraping content analysis

The thematic analysis on topic modelling of blogs and social media reveals five dominant topics, shown in Figure 6, whose frequency is based on a [0, 0.05, 0.1] threshold.

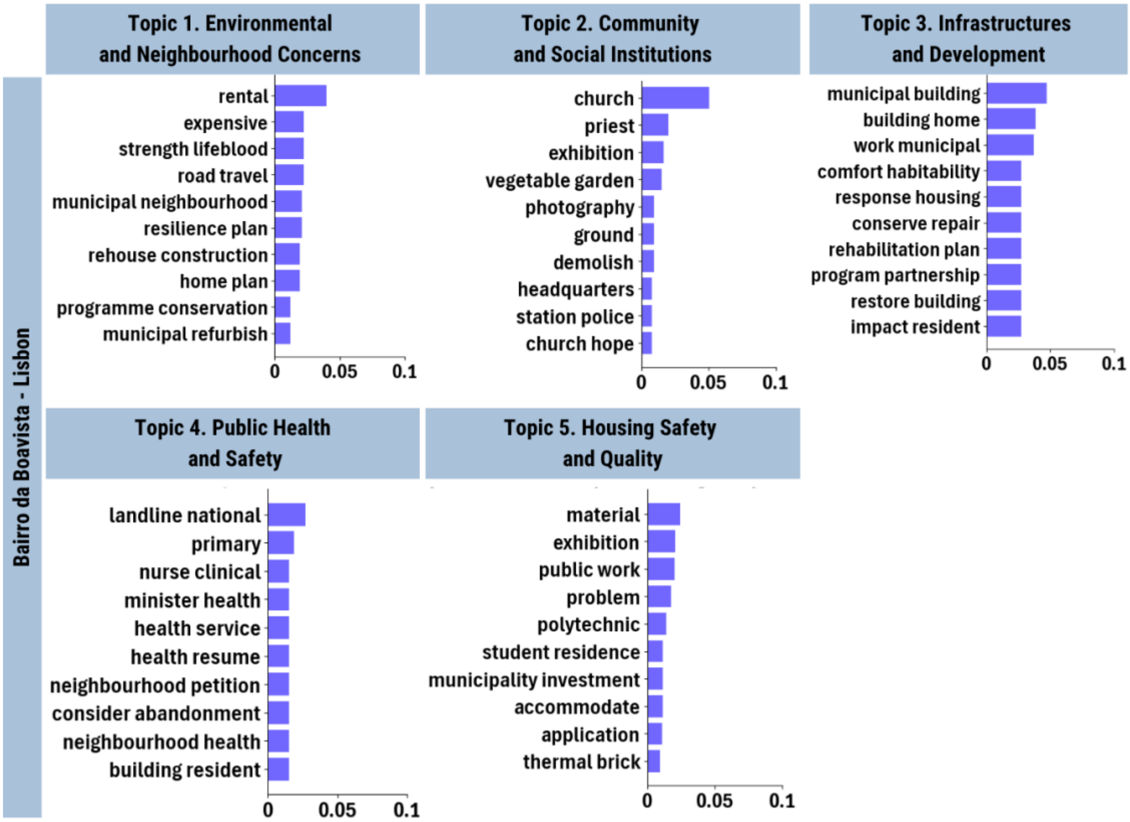

Figure 6. Five topics identified in blogs and social media related to *Bairro da Boavista* (Lisbon).

Regarding *Topic 1 - Environmental and Neighbourhood Concerns*, on one hand, terms like ‘rental’ and ‘expensive’ exemplify concerns about the cost of living, suggesting that housing affordability is a critical factor influencing neighbourhood quality. At the same time, expressions such as ‘resilience plan’ and ‘house reconstruction’ indicate community awareness of recent initiatives by public housing authorities aimed at addressing local needs through energy upgrades and building retrofitting. ‘Home plan' and 'programme conservation' reflect efforts to preserve and enhance citizens’ living conditions, demonstrating residents’ awareness of current public-led design interventions.

Topic 2 - Community and Social Institutions highlights the role of community spaces and institutions. Terms like ‘church’ and ‘vegetable garden’ suggest that shared public spaces are important for social interaction and mutual support. Terms like ‘exhibition’, ‘priest’, and ‘photography’ point to cultural and religious activities that helped build identity and attractivity within the neighbourhood.

Regarding *Topic 3 - Infrastructures and Development*, ‘municipal building’, ‘comfort habitability’, and ‘rehabilitation plan’ are related to official development projects aimed at upgrading buildings and living conditions. ‘Program partnership’ and ‘restore building’ point to cooperation between public bodies and other actors.

In *Topic 4 - Public Health and Safety*, terms like ‘nurse clinical’ and ‘health service’ show that health care access is a major concern. ‘Neighbourhood petition’ and ‘consider abandonment’ reflect moments in which residents feel neglected and respond through collective actions. ‘Minister health’ and ‘landline national’ reflect how local concerns are framed in relation to national-level authorities.

Regarding *Topic 5 - Housing Safety and Quality,* terms like ‘material’ and ‘thermal brick’ point to performance-based aspects of construction, such as the energy efficiency. ‘Public work’ and ‘municipality investment’ signal efforts by governmental authorities, while

‘problem’ and ‘application’ reflect the challenges residents face in understanding the process. The terms ‘student residence’ and ‘polytechnic’ highlight positive transformations in this neighbourhood associated with the introduction of new education infrastructure.
Figure 7 shows the distribution of the topics across the various source documents. The bubble’s size is proportional to the number of documents for a given topic within a particular source type, with the numerical value indicated inside the circle. Larger bubbles indicate greater prominence of the topic in the respective source.

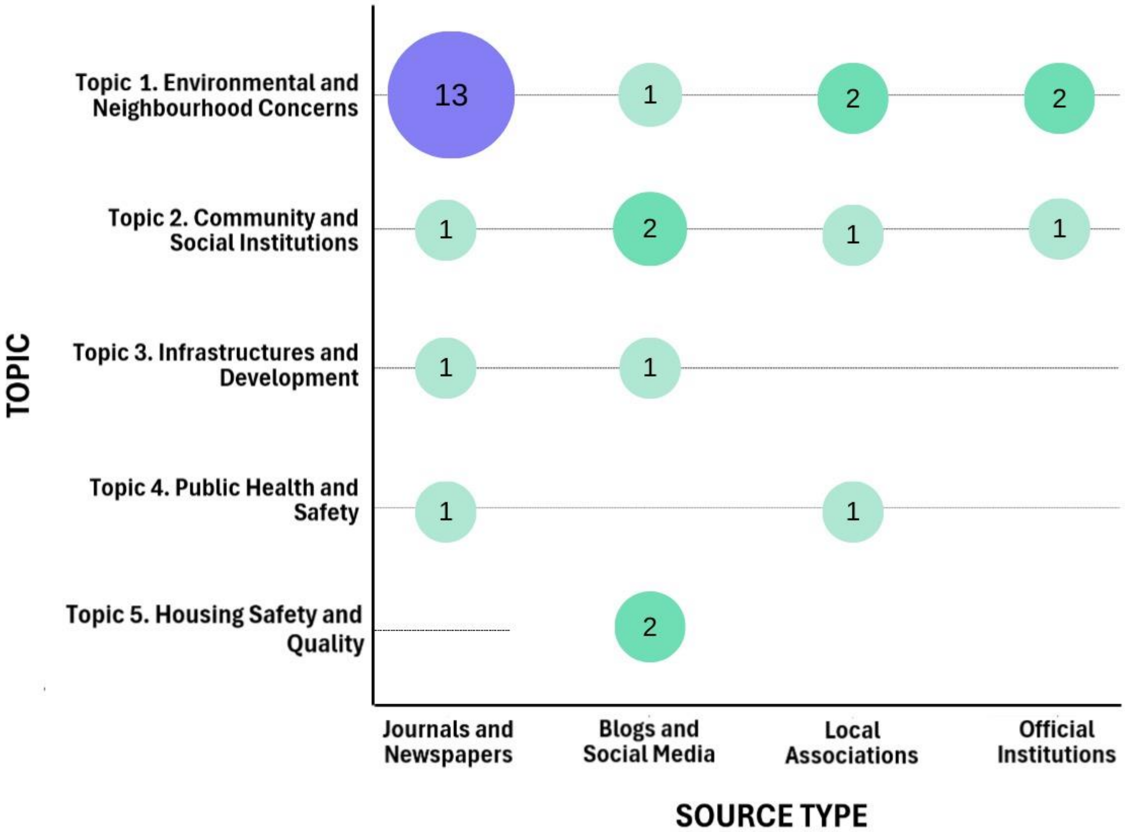

Figure 7. Influence of websites/journals news related to *Bairro da Boavista* (Lisbon) by source type.

Topic 1 - Environmental and Neighbourhood Concerns as the most frequently addressed theme in the dataset, appearing in 13 of the 33 analysed datasets. This topic is widely covered in journals and newspapers, yet in other source type is less discussed. The remaining topics receive low attention, with only one or two mentions across different source types. While less prominent, these topics are evenly discussed in blogs and social media, local associations, and official institutions.
Notably, *Topic 5 - Housing Safety and Quality* appears exclusively in blogs and social media, suggesting that housing concerns are primarily voiced through informal, community-led channels rather than official or media outlets. *Official Institutions* show limited engagement overall, covering just four of the five topics and with low frequency. This may indicate a narrower communication focus or limited public discussion of these issues. *Local Associations* engage broadly, addressing most topics except for *Topic 3 - Infrastructures and Development*. This reflects their role in responding to diverse community needs, even though key terms such as ‘hospital’, ‘road’, and other infrastructure-related words are absent. *Blogs and social media* stand out as key platforms for public dialogue on community, infrastructure, and housing.

Overall, these findings suggest that both informal and institutional channels should prioritize interconnected topics related to urban climate resilience.

5.2. Focus groups transcript analysis

Sentiment scores within the [–1, 0, 1] threshold are presented across the four core themes (detailed in Table 2) in Figure 8. Analysing sentiment polarity through this framework provides insights into both general attitudes and detailed perceptions of each neighbourhood aspect.

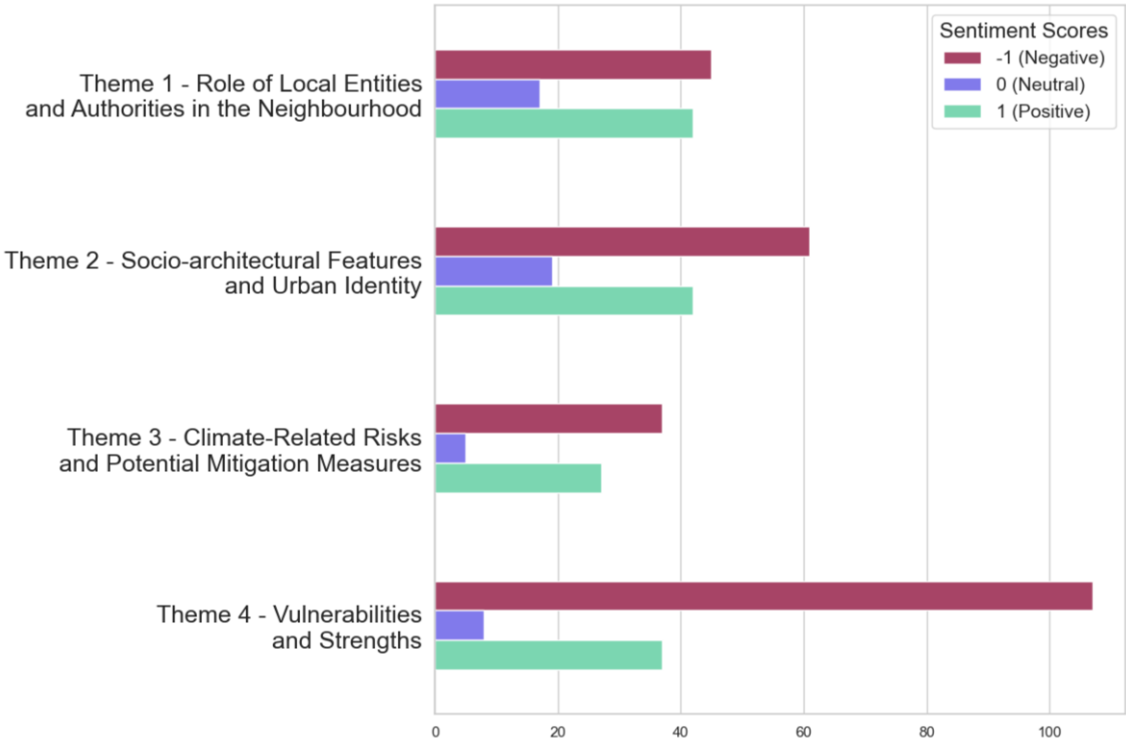

Figure 8. Sentiment distribution across the transcripts of the focus groups for each theme.

As shown in Figure 8, the perceptions of the participants of the five focus groups are predominantly negative across all four themes, particularly for *Theme 4 - Vulnerabilities and Strengths*. In contrast, sentiment scores are more balanced for *Theme 1 - Role of Local Entities and Authorities in the Neighbourhood*. Positive scores reflect trust in ongoing or past partnerships and community engagement activities, while negative sentiment indicates dissatisfaction with governance or service provision. Concerns regarding housing conditions, unemployment, crime, and access to social support are also common. *Theme 2 - Socio-Architectural Features and Urban Identity* also shows high negative sentiment scores, suggesting dissatisfaction among participants related to urban regeneration processes and cultural identity. Although lower than the negative scores, positive sentiment reflects a shared sense of belonging to the neighborhood's character. *Theme 3 - Climate-Related Risks and Potential Mitigation Measures* receives the least feedback overall but is relatively balanced, indicating both concerns over environmental vulnerabilities and recognition of the importance of adopting new sustainability and resilience strategies.

To better understand these results, the average sentiment scores for each sentence transcript was then analysed in all focus groups. Figure 9 shows how perceptions vary across four thematic categories among the key stakeholder groups.

Figure 9. Sentiment score distribution across the transcripts of the focus groups per themes and participants.

Institutional actors such the Benfica Parish Council ('PuA1') and other entities like Gebalis, the Lisbon City Council, and the Municipal Civil Protection ('PuA2') attributed positive sentiments in relation to *Theme 1*, on the role of local governance and *Theme 2* regarding urban identity (e.g., members of public authorities show a positive sentiment of 0.57 regarding socio-architectural features). In contrast, the sample of inhabitants ('LI') and the working staff from *Santa Casa da Misericórdia de Lisboa* ('PrA') exhibit negative sentiment, especially when discussing *Theme 3 - Climate-Related Risks* and *Theme 4 - Vulnerabilities and Strengths* (with scores of -0.75 and -0.66, respectively).
These patterns suggest a divergence of perceptions among institutional stakeholders, community organisations, and residents, potentially highlighting gaps in communication, trust, or perceived effectiveness of local policies.

5.2.1. VADER vs. BART

Classification performance was then assessed using the following standard metrics: accuracy, precision, recall, F1 score (macro average), while confusion matrices are computed using scikit-learn. As shown in Table 5, the VADER default model performs the worst overall, with a macro F1 score of 0.38 and 42% accuracy.

Table 5. Model performance: VADER vs BART.

Metric	**VADER default**	**VADER custom**	**BART**
Accuracy	0.42	0.46	0.56
Precision	0.40	0.53	0.54
Recall	0.41	0.45	0.53
F1 Score	0.38	0.39	0.48
Positive F1 Score	0.52	0.55	0.61
Neutral F1 Score	0.18	0.14	0.14
Negative F1 Score	0.44	0.47	0.68

It performs moderately well on positive sentiment (F1 = 0.52) but struggles significantly with neutral sentiment (F1 = 0.18). This shows a general-purpose lexicon has trouble capturing the subtle, less emotional expressions common in the dataset. The VADER custom model, enhanced with a domain-specific lexicon, improves notably. It raises positive class recall from 0.70 to 0.86, increases macro precision to 0.53, slightly improves macro F1 to 0.39, and boosts accuracy to 46%. These results confirm the findings by Deng et al. (2017) that domain-specific lexicons handle context-dependent sentiment better than general ones. The BART model performs best overall, achieving 56% accuracy, a macro F1 score of 0.48, and the highest F1 for negative sentiment (0.68). Sentiment patterns can be defined thanks to the contextual and generative nature of BART model. However, the model still performs poorly on neutral sentiment (F1 = 0.14), indicating that even advanced models struggle to accurately identify neutral or ambiguous content. Analysing the confusion matrices of each model (Figure 10) allows a closer look into each model performance.

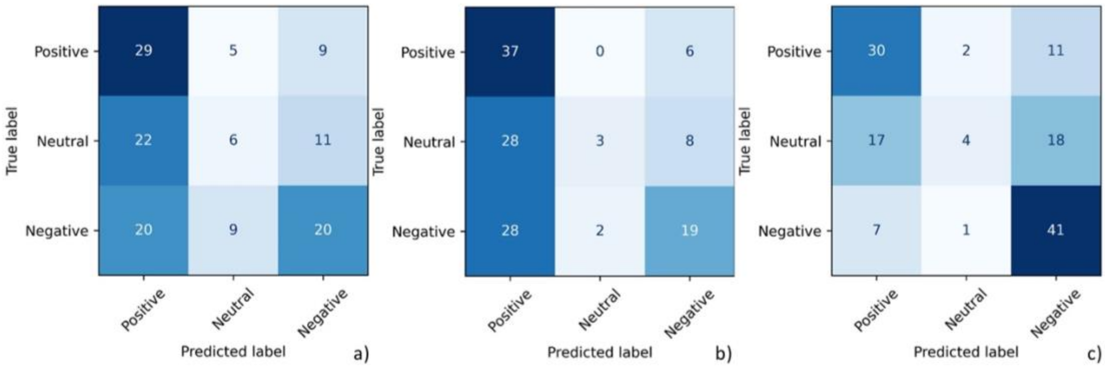

Figure 10. Confusion matrices: VADER default (a), VADER custom (b) and BART (c).

VADER Default often mislabels neutral and negative sentiments as positive, causing low neutral F1 scores. The VADER Custom model reduces this confusion, improving positive precision and lowering misclassifications, though some neutral-negative errors remain. The BART model separates classes best, especially negative vs. positive, but still struggles with neutral cases. Overall, domain-specific lexicons increase VADER's accuracy, yet BART's context-aware approach delivers stronger performance, especially in handling nuanced sentiments.

5.2.2. Word frequency analysis

A word frequency analysis was conducted on the five group transcripts, identifying the 200 most frequently used words. The results indicate that the terms related to climate change risks are not frequently mentioned. Then, it was checked whether the terms related to local risks, described in *Section 3.2*, are mentioned during the focus groups.

The most frequently terms mentioned by participants across the different focus groups are highlighted in bold in Figure 11, whereas local high-impact vulnerabilities (such as floods, wildfire risks, and seismic soil instability - expressed through terms such as 'ground', 'earthquake', 'soil', or 'terrain') appear sporadically. Extreme heat, drought and windstorms, another high-impact hazards in *Bairro da Boavista*, are not mentioned during the focus groups.

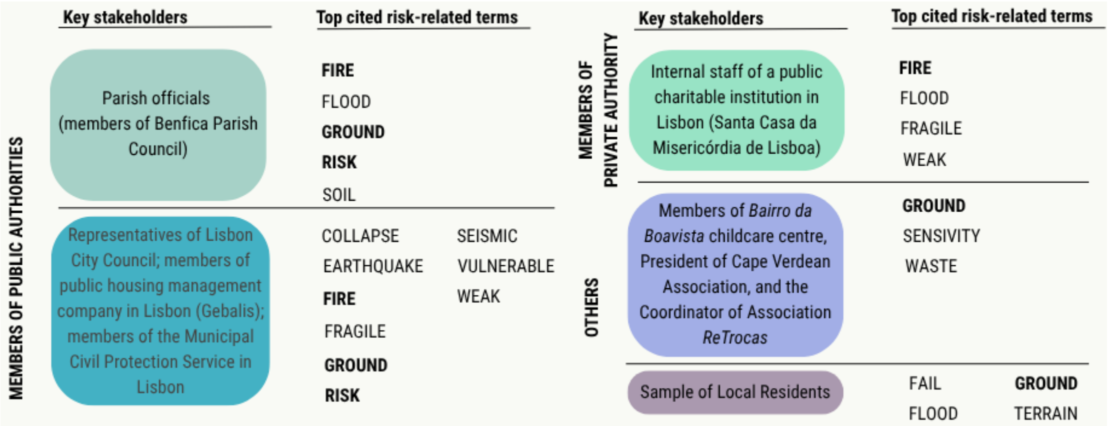

Figure 11. Top cited risk-related terms during the focus groups.

The observed divergence between local high physical exposure and low community risk perception has important implications. The tendency for risk to be perceived as abstract or secondary to more immediate concerns suggests limitations in the risk communication implemented across governance levels, as well as a broader disconnect between formal risk governance frameworks and everyday lived experience. As noted by Wachinger et al. (2013), while high risk perception does not automatically translate into protective action, since individuals aware of natural hazards may still choose not to respond, greater perceived risk is generally associated with increased preparedness and risk-mitigating behaviour, as also clearly stated by IPCC (2022).

5.3. Integrated insights from NLP analysis

5.3.1. Cross-methods analysis of focus group transcripts

The findings of this research are diverse and significant. The analysis of digital content reveals a strong focus on public investment, and efforts to enhance infrastructure and housing quality. There are significant community concerns about rising housing costs and population displacement, alongside deficiencies in public health services (e.g., shortages of hospital staff), which strain residents' well-being and social cohesion. Housing affordability and urban planning are identified as critical factors influencing the quality of life in the neighbourhood. The uncertainty of community to stay in this place can reduce the attachment to the place and the interest in collaborating.

At the same time, the preservation of community institutions and social spaces is recognised as a key factor in maintaining local connections. Overall, *Bairro de Boavista* is seen to be undergoing substantial changes. While there is optimism regarding urban renewal and development, concerns exist about increasing housing costs, displacement, and the preservation of community identity.

The sentiment analysis of the focus group transcripts reveals contrasting perspectives, which is both common and expected in urban analysis (Kim & Feng, 2024). Such mismatches can influence trust in climate governance and its perceived legitimacy, particularly when municipal policies are viewed as top-down or insufficiently participatory (Aylett, 2015).

In this context, institutional actors generally express positive sentiments, particularly regarding urban governance and identity. Conversely, residents and community-based organisations tend to voice negative sentiments, especially when underling the low preparedness to face climate-related risks and the long-standing socio-economic vulnerabilities. These findings highlight the contrast in how institutional and community actors perceive the challenges facing the neighbourhood. The disparities in sentiment

underscore the need for inclusive, participatory, and context-sensitive strategies in urban planning and risk communication. Bridging these gaps through a dialogue between authorities and residents is essential for building trust and enhancing legitimacy and effectiveness (Castán Broto & Bulkeley, 2013; IPCC, 2022).
A comparison between the results of VADER and BART shows that detecting neutral sentiment remains a challenge for both models. A domain-specific lexicon in VADER outperforms its default version, highlighting the value of customised lexicons for domain-specific applications. However, this approach is time-consuming. While KeyBERT facilitated more automated keyword extraction, manual selection and validation still required substantial effort. To address this, BART was tested as an efficient alternative, with better results with less labour compared to VADER.

5.3.2. Word cloud analysis: web-scraped content *vs* focus groups transcripts

Although comparing web-scraped content with focus group transcripts is challenging due to differences in scale and stakeholders, a word cloud analysis was conducted on a combined dataset of both sources to highlight dominant vocabulary.
Similarities and shared cultural values were observed among the e-community and the stakeholders involved in the focus groups (Figure 12). Overall, the analysis highlighted a shared emphasis on the public sphere and community-related issues, particularly housing, public infrastructure, and community support. Community cohesion and other positive values were reflected through words like ‘belonging’, ‘community’, ‘associations’, and ‘social’.

Figure 12. Word clouds from web scraping (a) and focus groups (b) related to *Bairro da Boavista* (Lisbon).

The word cloud based on blogs and social media web scraping (Figure 12a), with the top terms ‘church’, ‘project’, and ‘building’, ‘people’, ‘rehabilitation’ emphasised collective effort, planning, and shared purposes. The word cloud from focus groups transcripts (Figure 12b), highlighting ‘school’, ‘work’, and ‘house’, ‘population’ also reflected environments and activities that support learning, cooperation, and community involvement.
The word cloud analysis also revealed relevant differences but also drew attention to the absence or relatively low prominence of key topics. Web-based content reflected urban and building improvements, as indicated by terms such as ‘investment’, ‘efficiency’, ‘intervention’, and ‘program’. In contrast, focus group data emphasised everyday challenges related to living conditions, safety, sanitation, and access to services, reflected in terms such as ‘food bank’, ‘rubbish’, ‘problem’, ‘vandalise’, ‘trafficking’, and ‘drug’. In the focus groups data, references to ‘mental health’ and ‘poor’, as well as to risks such as ‘flooding’, ‘flood’, ‘hot’, ‘fire’, ‘earthquake’, and ‘tsunami’ pointed to a shared and

multi-dimensional perception of vulnerability. These terms were mentioned infrequently and therefore appeared very small in Figure 12b (highlighted with a rectangular box). Words related to risks or vulnerabilities that appear infrequently may reflect limited awareness of these issues within the e-community or focus groups' participants, their absence from everyday experience, or social discomfort and reluctance to disclose such information (as observed in other research contexts, e.g., Skafida, Morrison, and Devaney, 2022), rather than indicating their absence.
Taken together, these preliminary findings portrayed a community facing both changes in the physical environment and socio-economic challenges, while remaining anchored by a sense of belonging and supported by robust local public structures, including associations, facilities, and initiatives. As described in the word frequency count (*Section 5.2*), community climate-risk awareness is low, being overshadowed by other socioeconomic concerns.

6. How to improve urban climate resilience in *Bairro da Boavista?*

The required actions to improve urban climate resilience in *Bairro da Boavista* are diverse and complementary, ranging from soft co-design climate actions and targeted community awareness campaigns to Nature-based solutions (NbS). These are below detailed according to the findings from each data source and data analysis employed in this research context.
From web-scraped content: The key findings regard the uncertainty concerning the use of publicly owned buildings by local inhabitants, the importance of cultural and shared public spaces, inadequate thermal insulation in buildings, and increased community awareness of recently implemented or ongoing community facilities. These findings point to the need to strengthen institutional trust, communication, participatory actions, and ensure housing stability. From a climate adaptation perspective, digital communities also emphasise the importance of enhancing green and shared areas to mitigate urban heat island effects. Beyond the outdoor space, building retrofitting is essential to improve thermal performance and energy efficiency to withstand winter conditions and heatwaves. Additional strategies include the integration of passive cooling measures (e.g., green roofs, reflective materials, or natural ventilation), the expansion of urban tree canopies, and the implementation of NbS.
From focus group transcripts: The need for targeted strategies to strengthen social capital and community networks, enhance economic resilience and social cohesion, and increase risk awareness, particularly among local stakeholders, who expressed more negative sentiment than representatives of public authorities and local associations.
The low level of interest or attention to natural hazards of focus groups' participants (*Sections 5.2.2* and *5.3.2*) highlights the need to reinforce community participation in urban resilience planning and risk preparedness, by leveraging the role of non-state actors in local governance. This aligns with the broader literature on global environmental governance that emphasises the growing involvement of non-state actors such as NGOs, community groups, population associations and other organisations in responding to climate change (Bulkeley & Newell, 2010).
Drawing on a critical analysis of the web-scraped content and focus group data, the following practical recommendations are presented to address the concerns and priorities identified in this research (Figure 13):

- *Sustainable economy & livelihoods:* Develop green jobs initiatives targeting low-income and unemployed residents, linked to building retrofits and circular economy projects. Provide incentives for local small and medium enterprises to implement climate-resilient solutions, such as NbS approaches and thermal retrofitting (OECD, 2025).

- *Education & skills development:* Integrate topics on resilience, climate adaptation, and civil protection into school curricula and apprenticeship. Provide financial support for courses for residents, vocational training in green construction and emergency response for accelerating the green transition and address local needs (OECD, 2025).
- *Mapping and communication tools for inclusion and social resilience:* Create interactive visualisation tools to show vulnerable areas and track resilience strategies. Use visual storytelling and participatory workshops to foster shared understanding of local risk factors. To further raise local risk awareness, stakeholder engagement should be systematically integrated into local urban strategies, since engaging the public and fostering justice in all planning processes is one of the key principles outlined by Meerrow and Woodruff (2020).

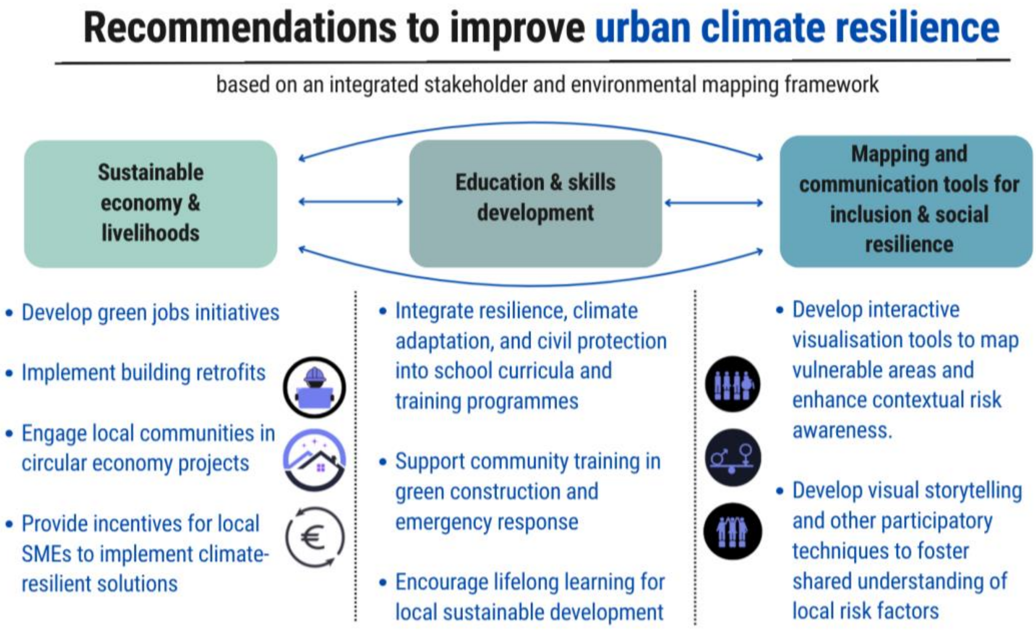

Figure 13. Recommendations to improve urban climate resilience in the case study.

Table 5 presents several stakeholder engagement activities, though not exhaustive, approaches. They are informed by local cultural behaviours (Ferreira et al., 2021), targeting vulnerable groups using effective tools (e.g., Le Dé et al., 2021; Schiele et al., 2022; Pueyo-Ros et al., 2023), while the different roles, power levels, and interests of stakeholders should be considered at the neighbourhood level.

Table 5. List of potential effective stakeholder engagement activities to increase risk awareness in *Bairro da Boavista* (Lisbon).

Stakeholder engagement activities	Expected Impact
Knowledge Café	Facilitates open dialogue and promotes shared understanding of risks, priorities, and guidelines.

Role-Playing simulation and serious games	Explores scenarios to foster empathy and informed decision-making and the adoption of NbS.
Roles Framework	Supports discussion of roles, responsibilities, and needs of different target groups.
Gender Mapping	Visualises gender roles, relationships, and power dynamics.
Stakeholder Influence Mapping	Identifies and analyses key stakeholders, defines engagement levels, and develops climate-related management strategies.
Dot Voting	Enables collective prioritisation of solutions and supports participatory decision-making.

Aside from these potential activities to increase risk awareness and community involvement, it is also evident that the organisation of focus groups was effective not only in deepening understanding of key community stakeholders, but also in fostering dialogue and building trust. This process helped ensure that diverse perspectives were captured, thereby strengthening the connections between scientific research and resilience policymaking.

7. Key limitations and challenges

The study has some limitations, including issues related to idiomatic language, the inherent constraints of automated semantic analysis, the specificity of the domain context, and the limited timeframe of this analysis. To address these challenges, several mitigation measures were implemented.

First, the deep translation using Google Translate API from Portuguese into English can lead to the loss of cultural nuances, irony, or language-specific references, which could impact the accuracy of the findings. This limitation is overcome by a manually validation step by Portuguese native speakers. The lexicon developed using KeyBERT and then enriched manually could be also expanded, as a more comprehensive set of terms would improve the accuracy of sentiment analysis.

Second, the NLP framework is useful for identifying latent patterns in large text corpora (topic modelling) and for comparing sentiment across different groups and sources (sentiment analysis). However, these findings require fine-grained qualitative interpretation and triangulation with official data sources, as well as a deep understanding of the pilot site, to be meaningfully interpreted. Additionally, the complexity of analysis in this specific domain adds further challenges, as community perceptions vary significantly across different groups of communities and this topic is often not represented in social media networks. Extending the duration of the analysis would allow for a more robust assessment of changes in public discourse and risk perceptions. Short-term data may fail to capture gradual shifts that are influenced by policy changes or specific events.

Regardless of these limitations, this study offers the advantage of combining and considering heterogeneous and unstructured data, thus incorporating voices of individuals, through focus groups, often overlooked within traditional data sources, achieving this at minimal cost. Following the implementation of any urban improvement plan aimed at enhancing quality of life and the built environment at neighbourhood level, this analysis can be repeated to assess its impact on the community. While perceptions are inherently subjective and ever evolving, they provide valuable insights into the lived experiences of different user groups.

The proposed integrated framework is exploratory and emerges from the convergence of methods and knowledge across social sciences and humanities, geography, environmental and climate risks, data science, and artificial intelligence. It is grounded in empirical research enabling data collection that supports the theoretical development and the formulation of scientifically grounded hypotheses. The framework is validated through the case study of *Bairro da Boavista*, where richer and multi-source datasets were used, thereby capturing complementary perspectives.
Figure 14 illustrates how public decision-makers and other key stakeholders in urban risk management can leverage the proposed multi-source stakeholder mapping, informed by NLP techniques applied to both primary (research-generated, collected through social surveys) and naturally occurring data, within a feedback-loop framework. This mapping is integrated with environmental insights derived from geospatial risk data platforms and local urban renewal initiatives to support local or municipal informed decision-making.

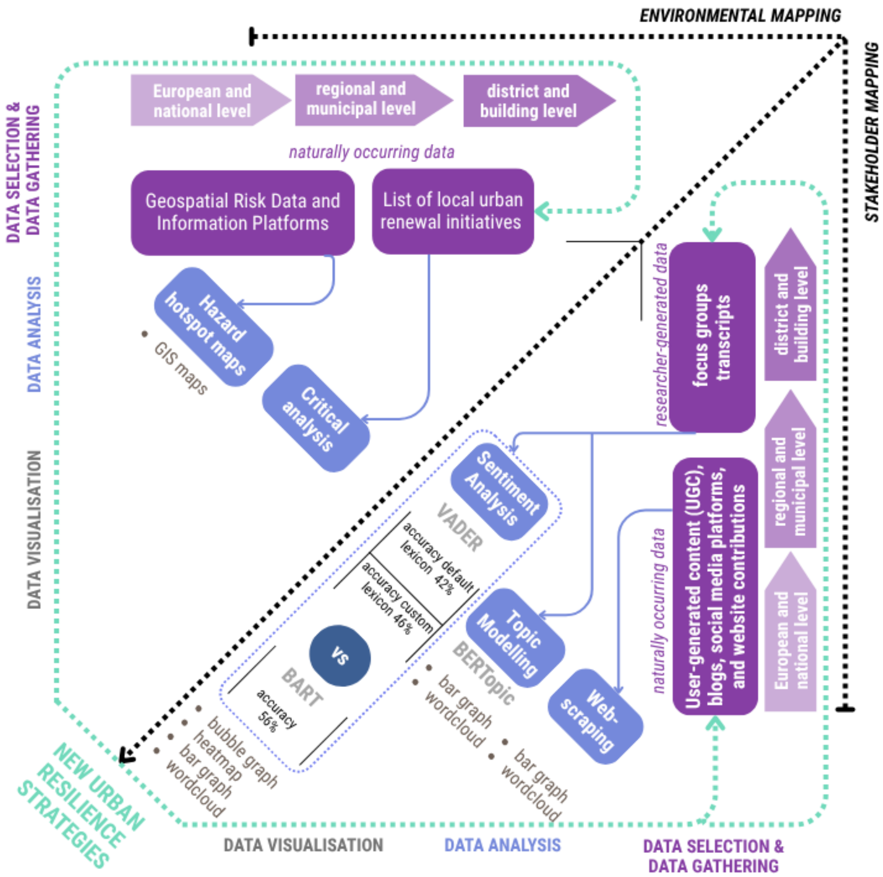

Figure 14. Stakeholder mapping using NLP and environmental analysis to inform urban resilience policymaking.

NLP findings are visualised and compared to environmental analysis based on secondary, naturally occurring data. This methodology can be replicated across other urban environments that share similar urban and socio-demographic characteristics, specifically areas owned and managed by public authorities with controlled rents, multiple hazards exposure, and vulnerable populations. The latter include, for instance, in addition to immigrant or local communities with low levels of formal education and limited access to digital tools or literacy, factors that may shape both their perceptions of environmental and climate risk and their capacity to access to information and other resources.

8. Conclusions and future perspectives

Aligned with the more advanced practices in data processing analysis, we develop a replicable and semantically rich framework that integrates automated techniques with expert-driven validation using both supervised and unsupervised NLP techniques. By analysing multi-source and multi-actor textual data, this approach enables an in-depth analysis of the perceptions and perspectives of diverse communities that may be overlooked in urban planning and environmental and climate risk reduction strategies. These include e-communities engaged through blogs and news media, and local key stakeholders involved through both in situ and online focus groups.

We demonstrate how NLP-based analysis of textual data can serve as an effective tool for understanding stakeholder perspectives and informing tailored actions to enhance urban climate resilience planning, particularly when integrated with hazard map analysis to assess whether environmental and climate risk perception aligns with site-specific knowledge. By capturing a diverse range of experiences, viewpoints, and concerns, this analysis provides an understanding of the local context through the lens of community perspectives, stakeholder priorities, and lived experiences. The insights derived from NLP techniques are particularly valuable when integrated into ongoing and future local strategies and initiatives, some of which were identified in this study, contributing to the creation of a more adaptive and resilient urban environment. This methodological NLP framework, including its thematic classification and data visualisation components, has the potential to be applied in diverse urban contexts, provided that key transferability considerations identified in this empirical research are adequately addressed.

A primary requirement is access to a large volume of data collected with the appropriate ethical approvals and consent procedures. The framework is likely to be most effective in urban areas characterised by mature digital ecosystems, where rich user-generated online content can provide information sources for the web-scraping and data mining components. In contrast, in smaller cities or contexts with lower levels of digital engagement, the availability of online data may be more limited, necessitating a greater effort to collect other type of primary data, such as from focus groups, stakeholder consultations, and other participatory data collection methods. The transferability of the framework may be limited in contexts with low digital participation or where online discourse is not representative of the wider population.

Another important consideration for transferability concerns linguistic and cultural factors. This framework requires manual validation and the involvement of native-speaking experts to ensure the accuracy and reliability of thematic classifications. In multilingual contexts, this would necessitate adapting the translation pipeline and developing context-specific lexicons tailored to local languages and communication practices. Furthermore, local expert validation is essential for preserving cultural and semantic nuances. Other relevant aspects that can limit the use of this framework regard contextual differences in patterns of social media use and digital participation.

The reiterative use of this approach can be effective if at the governance level, translating insights into policy or planning actions can strengthen trust through community involvement and support a feedback-based approach. In this model, community-generated data informs adaptation measures, which are then monitored and refined through iterative cycles of analysis and improvement.

In future research, additional statistical sources and social surveys could enhance interpretative depth, by enabling triangulation with analyses of trends and contextual factors. Future work could also extend the timeframe of both topic modelling and sentiment analysis to capture how multi-actor perceptions and conditions evolve over time, particularly in response to policy changes, infrastructural interventions, or environmental events. When web-scraping and focus group analyses are repeated following the implementation of municipal renewal initiatives, they can be used to measure their impact and assess the perceptions of local communities.

List of References

Abdelrazek, A., Eid, Y., Gawish, E., Medhat, W., & Hassan Yousef, A. (2022). Topic modeling algorithms and applications: A survey. *Information Systems*, *112*, 102131. https://doi.org/10.1016/j.is.2022.102131

Aylett, A. (2015). *Institutionalizing the urban governance of climate change adaptation*. Urban Climate, 14, 4–16.

Buchenrieder, G., Brandl, J., & Balgah, A. R. (2021). The Perception of Flood Risks: A Case Study of Babessi in Rural Cameroon. *International Journal of Disaster Risk Science*, *12*(4). https://doi.org/10.1007/s13753-021-00345-7

Bulkeley, H., Newell, P., 2010. Governing Climate Change (Global Institutions). Routledge, London.

Câmara Municipal de Lisboa. (2019). *Riscos naturais e antrópicos no Bairro da Boavista*.

Câmara Municipal de Lisboa. (n.d.). *LXi – Lisboa Interactiva*. https://websig.cm-lisboa.pt/MuniSIG/visualizador/index.html?viewer=LxInterativa.LXi

Chang, W.-L., & Wang, J.-Y. (2018). Mine is yours? Using sentiment analysis to explore the degree of risk in the sharing economy. *Electronic Commerce Research and Applications*, *28*, 141–158.

Cheng, G., You, Q., Li, G., Li, Y., Yang, D., Wu, J., & Wu, Y. (2024). Research on the Application of Topic Models Based on Geological Disaster Information Mining. *Information (Switzerland)*, *15*(12). https://doi.org/10.3390/info15120795

Clark, N., Boersma, K., Raju, E., Opromolla, A., Orru, K., Hansson, S., Russo, R., Gargiulo, M. V., Duca, G., Capuano, P., Schweizer, P.-J., Cumiskey, L., Steinhausen, M., Branlat, M., Olson, A., Andersen, N. B., Larruina, R., Atun, F., van Westen, C., Tamborrino, R., & Vollmer, M. (2025). *Strengthening all-of-society approaches for disaster resilient societies through competency building: A European research agenda*. *International Journal of Disaster Risk Reduction, 121*, Article 105345. https://doi.org/10.1016/j.ijdrr.2025.105345

Cologna, V., & Siegrist, M. (2020). *The role of trust for climate change mitigation and adaptation behaviour: A meta-analysis*. *Journal of Environmental Psychology*, 69, 101428. https://doi.org/10.1016/j.jenvp.2020.101428.

Cutter, S. L. (2024). The origin and diffusion of the social vulnerability index (SoVI). *International Journal of Disaster Risk Reduction*, *109*. https://doi.org/10.1016/j.ijdrr.2024.104576

Deng, S., Sinha, A. P., & Zhao, H. (2017). Adapting sentiment lexicons to domain-specific social media texts. *Decision Support Systems*, *94*, 65–76.

Dhar, T., Bornostein, L., Lizarralde, G., & Nazimuddin, S.M. (2023). Risk perception- A lens for understanding adaptive behaviour in the age of climate change? Narratives from the Global South. *International Journal of Disaster Risk Reduction*, https://doi.org/10.1016/j.ijdrr.2023.103886

European Court of Auditors. (2024). *Climate adaptation: EU and Member State planning not consistent with climate risks*. https://www.eca.europa.eu/ECAPublications/SR-2024-15/SR-2024-15_EN.pdf

European Parliament and Council of the European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*.

Fekete, A., Neuner, S., Munschauer, C., Buzna, L., Chovanec, D., Holla, K., Kockar, S., Leopold, U., Stellacci, S., Yilmaz, D., & Ferreira da Silva, C. (2025). An upscaling multi-level and multi-hazard risk assessment for heat and other natural hazards concerning vulnerable groups in Žilina, Slovakia. *Natural Hazards*. https://doi.org/10.1007/s11069-025-07722-1

Ferreira, J., Ferreira, C., & Bos, E. (2021). Spaces of consumption, connection, and community: Exploring the role of the coffee shop in urban lives. *Geoforum, 119*, 21–29.

Gatti, L., Guerini, M., & Turchi, M. (2015). SentiWords: Deriving a high precision and high coverage lexicon for sentiment analysis. *IEEE Transactions on Affective Computing*, *7*(4), 409–421.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *ArXiv Preprint ArXiv:2203.05794*.

Hankar, M., Kasri, M., & Beni-Hssane, A. (2025). A Comprehensive Overview of Topic Modeling: Techniques, Applications and Challenges. *Neurocomputing*, 129638.

Hristova, G., & Netov, N. (2022). *Media coverage and public perception of distance learning during the COVID-19 pandemic: a topic modeling approach based on BERTopic*. 2259–2264.

IEA EBC. (2020). *Annex 75: Cost-effective building renovation at district level – Combining energy efficiency & renewables*. [Publisher if available]. https://www.iea-ebc.org/projects/project?Annex=75

INE-Instituto Nacional de Estatística. (2021). *Censos 2021: Resultados definitivos*.

Intergovernmental Panel on Climate Change (2022). *AR6 WGII: Impacts, Adaptation and Vulnerability*. https://doi.org/10.1177/0963662515613702

Intergovernmental Panel on Climate Change (2023). *Climate change 2023: Synthesis report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Core Writing Team: H. Lee & J. Romero, Eds.). IPCC. https://www.ipcc.ch/report/ar6/syr/

Le Dé, L., Gaillard, J. C., Gampell, A., Loodin, N., & Hinchliffe, G. (2021). Fostering children's participation in disaster risk reduction through play: A case study of LEGO and Minecraft. *International Journal of Disaster Risk Science, 12*(6), 867–878.

Li, R., Li, H., Hu, T., Bian, Z., Liu, F., Cao, B. (2023). Land surface temperature retrieval from Sentinel-3A SLSTR data: Comparison among split-window, dual-window, three-channel, and dual-angle algorithms. *IEEE Transactions on Geoscience and Remote Sensing, 61*, 1–14. https://doi.org/10.1109/TGRS.2023.3288584

Kiani, U. B. N., Najam, F. A., & Rana, I. A. (2022). The impact of risk perception on earthquake preparedness: An empirical study from Rawalakot, Pakistan. *International Journal of Disaster Risk Reduction*, *76*. https://doi.org/10.1016/j.ijdrr.2022.10298

Kim, J. S., & Feng, Y. (2024). Understanding complex viewpoints in smart sustainable cities: The experience of Suzhou, China. *Cities, 147*, 104832. https://doi.org/10.1016/j.cities.2024.104832

Kyritsis, K., Perikos, I., & Paraskevas, M. (2023). *Zero-Shot Sentiment Analysis Exploring BART Models*. 192–197. https://doi.org/10.1109/BCD57833.2023.10466289

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *ArXiv Preprint ArXiv:1910.13461*.

Liu, B. (2012). Sentiment analysis and opinion mining. *Morgan & Claypool Publishers*, *168*.

Meerow, S., & Woodruff, S. C. (2020). Seven principles of strong climate change planning. *Journal of the American Planning Association, 86*(1), 39–46. https://doi.org/10.1080/01944363.2019.1652108

MeteoSwiss. (2025). Urban heat. Federal Office of Meteorology and Climatology MeteoSwiss. https://www.meteoswiss.admin.ch/climate/climate-change/more-extreme-heat/urban-heat.html

Mora, K., Chang, J., Beatson, A., & Morahan, C. (2015). Public perceptions of building seismic safety following the Canterbury earthquakes: A qualitative analysis using Twitter and focus groups. *International Journal of Disaster Risk Reduction*, *13*, 1–9. https://doi.org/10.1016/j.ijdrr.2015.03.008

Organisation for Economic Co-operation and Development (2025). *Employment and Skills Policies for the Green Transition: Review of International Good Practices*, OECD Publishing, Paris, https://doi.org/10.1787/f0c558fa-en

Oliveira, F. M., Holanda, T. C., de Oliveira Silva Ramalho, A. L., da Rocha Kloeckner, N. V., de Moura, I. E. M. O., Ricarte, T. L., & Ribeiro, K. V. (2024). Rains, Tragedies and Media Coverage: Analysis of Floods In Rio Grande Do Sul. *Revista de Gestão Social e Ambiental*, *18*(4). https://doi.org/10.24857/rgsa.v18n4-114

Organisation for Economic Co-operation and Development. (2025). Monitoring exposure to future climate-related hazards: Forward-looking indicator results and methods using climate scenarios. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/07/monitoring-exposure-to-future-climate-related-hazards_2d4a78d2/b9ba6ee0-en.pdf

Pueyo-Ros, J., Comas, J., Säumel, I., Castellar, J. A. C., Popartan, L. A., Acuña, V., & Corominas, L. (2023). Design of a serious game for participatory planning of nature-based solutions: The experience of the Edible City Game. *Nature-Based Solutions, 3*, 100059. https://doi.org/10.1016/j.nbsj.2023.100059

Slovic, P. (2016). Perception of risk. In *The perception of risk* (pp. 220–231). Routledge

Skafida, V., Morrison, F., & Devaney, J. (2022). Answer refused: Exploring how item non-response on domestic abuse questions in a social survey affects analysis. *Survey Research Methods, 16*(2), 227–240. https://doi.org/10.18148/srm/2022.v16i2.7823

Schiele, H., Krummaker, S., Hoffmann, P., & Kowalski, R. (2022). The "research world café" as method of scientific enquiry: Combining rigor with relevance and speed. *Journal of Business Research, 140*, 280–296. https://doi.org/10.1016/j.jbusres.2021.10.075

Stellacci, S., & Moro, S. (2023). Travellers' perspectives on historic squares and railway stations in Italian heritage cities revealed through sentiment analysis. *Journal of Urban Design*, *28*(2), 174–198. https://doi.org/10.1080/13574809.2022.2097903

Tang, J., Yang, S., Liu, Y., Yao, K., & Wang, G. (2022). Typhoon Risk Perception: A Case Study of Typhoon Lekima in China. *International Journal of Disaster Risk Science*, *13*(2), 261–274. https://doi.org/10.1007/s13753-022-00405-6

Veltri, G. A., & Atanasova, D. (2017). *Climate change on Twitter: Content, media ecology and information sharing behaviour*. Public Understanding of Science, 26(6), 721–737.

Wachinger, G., Renn, O., Begg, C., Kuhlicke, C. The risk perception paradox-implications for governance and communication of natural hazards. Risk Anal. 2013 Jun;33(6):1049-65. doi: 10.1111/j.1539-6924.2012.01942.x. Epub 2012 Dec 24. PMID: 23278120.

World Health Organization. (2023). Heat and health. https://www.who.int/news-room/fact-sheets/detail/climate-change-heat-and-health

World Meteorological Organization. (2024). *State of the Global Climate 2024*. https://library.wmo.int/viewer/69455/download?file=WMO-1368-2024_en.pdf&type=pdf&navigator=1

Wu, K., Wu, J., Ding, W., & Tang, R. (2021). Extracting disaster information based on Sina Weibo in China: A case study of the 2019 Typhoon Lekima. *International Journal of Disaster Risk Reduction*, *60*. https://doi.org/10.1016/j.ijdrr.2021.102304

Yu, T., Yang, H., Luo, X., Jiang, Y., Wu, X., & Gao, J. (2021). Scientometric analysis of disaster risk perception: 2000–2020. *International Journal of Environmental Research and Public Health*, *18*(24), 13003.

Zander, K. K., Nguyen, D., Mirbabaie, M., & Garnett, S. T. (2023). Aware but not prepared: understanding situational awareness during the century flood in Germany in 2021. *International Journal of Disaster Risk Reduction*, *96*. https://doi.org/10.1016/j.ijdrr.2023.103936

Zhai, W., Hu, W., Yuan, Z., & Li, Y. (2024). Examining disaster resilience perception of social media users during the billion-dollar hurricanes. *Natural Hazards*, *120*(1), 701–727. https://doi.org/10.1007/s11069-023-06206-4

Zhang, P., Zhang, H., & Kong, F. (2024). Research on online public opinion in the investigation of the "7–20" extraordinary rainstorm and flooding disaster in Zhengzhou, China. *International Journal of Disaster Risk Reduction*, *105*. https://doi.org/10.1016/j.ijdrr.2024.104422

Zhang, Y., Li, X., Liu, Q., Qiu, Z., & Fa, Z. (2024). How to identify public risk perception in public health emergencies and explore its driving mechanism? An empirical analysis in the Xi'an epidemic. *Sustainable Cities and Society*, *108*.