MISCELLANEA

Bank i Kredyt 56(5), 2025, 613-642
DOI: 10.5604/01.3001.0055.3046

BANK I KREDYT
BANK & CREDIT

Decoding public sentiment on pension policies in China through natural language processing

Xiaohong Xie*, Magdalena Osińska#

Submitted: 4 November 2024. Accepted: 24 February 2025.

Suggested citation: Xie X., Osińska M. (2025), Decoding public sentiment on pension policies in China through natural language processing, *Bank i Kredyt*, 56(5), 613–642, DOI: 10.5604/01.3001.0055.3046.

Abstract

This study aims to reveal public sentiment toward China's pension policies from January 2018 to August 2023, leveraging over 260,000 Weibo posts to identify key themes and demographic differences. Advanced Natural Language Processing (NLP) techniques, including sentiment analysis and latent Dirichlet allocation, are employed to explore six topics, such as societal impact and policy integrity, while uncovering demographic and regional variations. The findings reveal that policy changes significantly influence public sentiment, with greater negativity observed around institutional and structural aspects of the policies. These results underscore the need for public education on pension reforms and fraud prevention, providing actionable insights for policymakers in an aging society. The study contributes to behavioural finance theory by illustrating how heuristics like availability bias and loss aversion shape public reactions to pension reforms. However, social media data may not fully represent less active groups like older adults, highlighting the need for broader research methods.

Keywords: pension policies, public sentiment, natural language processing, sentiment analysis

JEL: H55, J18, C38

* Department of Econometrics and Statistics, Nicolaus Copernicus University in Torun; e-mail: xiaohong.xie@umk.pl; ORCID: 0000-0003-4296-6287.
# Department of Economics, Nicolaus Copernicus University in Torun; e-mail: emo@umk.pl; ORCID: 0000-0002-9796-2892.

1. Introduction

China stands on the brink of a momentous demographic shift. As projected by WHO (WHO, 2019), by 2040, about 28% of the nation's population will have crossed the age of 60. The inevitable growth of the elderly demographic, coupled with the ramifications of China's one-child policy from 1980 to 2015, presents multifaceted challenges. The policy has led to the formation of smaller family units, placing an augmented burden on them to support an expanding aged population. These factors accentuate the pressures on the present pension system, underscoring the imperative for robust adaptive measures and reforms.

In response, China's pension policy is evolving dynamically. A clear transition is observed from the 13th Five-Year Plan (for the years 2016–2020) to the 14th Five-Year Plan (for the years 2021–2025) in the approach and goals of China's pension strategy. The 13th Five-Year Plan's essence was the creation of a multi-pillar, inclusive, equitable, and enduring social security framework (Gov.cn 2017). Contrastingly, the subsequent 14th Five-Year Plan builds on these foundations, while introducing market-driven institutions and bolstering corporate engagement to mitigate the strains of an aging population. Moreover, the 14th Plan underscores the importance of the elderly's mental well-being and overall contentment. Additionally, a pivotal aspect of this plan is the proposal for a gradual deferment of the statutory retirement age (Gov.cn 2022a). A significant recent update is the gradual extension of the statutory retirement age beginning in 2025, with increases phased in every few months until the age limits reach 63 for men, 58 for those women who currently retire at 55, and 55 for women who currently retire at 50 (Gov.cn 2024).

As China undergoes significant pension policy transitions, these changes are increasingly reflected in public discourse. In today's digitally-dominated era, platforms like Sina Weibo,[1] a major social media site in China, have become pivotal in capturing public sentiment. With its extensive user base, Weibo offers a direct window into diverse public opinions on these policy shifts. However, the complexity and volume of data on Weibo present challenges for traditional analysis methods. This is where Natural Language Processing (NLP) comes into play. NLP is a field of artificial intelligence that focuses on the interaction between computers and human (natural) languages. It involves the development of algorithms and systems that can read, understand, and derive meaning from human languages (Chowdhary 2020). NLP enables the extraction of key insights from large volumes of text by identifying patterns, sentiments, and themes.

The previous research highlights significant advancements in the application of NLP for analysing public sentiment in various domains, such as public health (Abiola et al. 2023; Anoop, Sreelakshmi 2023), political landscapes (Ali et al. 2022; Gorodnichenko, Pham, Talavera 2021), and environmental issues (Sun et al. 2023; Yang et al. 2023). However, there is a conspicuous gap in applying these advanced analytical techniques to financial or economic policies, particularly pension policies in China. Besides, previous studies have focused on general attitudes towards pension systems and welfare state policies, often within a specific national context or under certain demographic settings. However, there is a lack of in-depth analysis that combines the advanced techniques of NLP with a large-scale examination of social media data to unravel the nuances of public opinion on pension policies in China.

[1] https://weibo.com/newlogin?tabtype=weibo&gid=102803&openLoginLayer=0&url=https%3A%2F%2Fweibo.com%2F.

This study aims to understand public sentiment toward China's pension policies from 2018 to 2023 and to identify key themes and demographic differences using NLP techniques and thematic analysis. This study fills a critical gap by leveraging NLP techniques to analyse over 260,000 posts from Sina Weibo, offering unique insights into public sentiment about pension policies in China. This approach is especially pertinent given China's distinct socio-economic landscape and rapidly evolving demographic profile. The research also seeks to explore how these sentiments vary across different demographic groups and regions, a perspective that has been underexplored in the existing literature. Applying NLP techniques such as sentiment analysis and Latent Dirichlet Allocation (LDA), Weibo posts are analysed to decode complex social sentiments. The focus on the post-1980 generation, shaped by the one-child policy and pivotal to the pension structure, underscores the study's significance. The research spans five years (2018–2023) and offers a novel longitudinal perspective on the shifts in public opinion against the backdrop of policy and economic changes. By doing so, the study not only contributes to the existing body of research on public opinion toward pension policies but also extends the application of NLP in social media analysis to a new and critical area of public policy in a significant geopolitical context. Besides, it enables making international comparisons in the area of interest.

2. Literature review

2.1. Pension schemes in the behavioural finance literature

Behavioural finance considers departures from rational thinking based on proven evidence and hypotheses that measure experiences and interactions to determine rational actions and decisions. In contrast, irrational thinking is usually based on emotions, and limited or selective evidence often mixed with those emotions. Rational thinking creates motivation because there is structure and clear evidence, while irrational thoughts cause anxiety (David, Cramer 2009; Rathi, Geetha 2024). Through their experimental studies, Kahneman and Tversky (1979) linked the notion of irrational thinking to the concept of multiple cognitive biases that characterize human thinking processes and decision-making. They hypothesized that people make decisions based not on rational premises and strict rules of reasoning but on heuristics as opposed to rationality. Heuristics should be considered a simplified way of expressing opinions while making decisions. Instead of deliberately using the strict rules of reasoning, the investor frequently uses a simplified pattern, often unaware of it. The most often observed effects of investment activity, including retirement decision-making, are the effect of sunk costs, mental accounting, the disposition effect, the endowment effect, the status quo effect, and the effect of myopia and loss aversion (Osińska, Stawicki 2017). Benartzi and Thaler (2007) summarised the heuristics observed while selecting retirement plans. Their experiments and literature review found that shortcuts (the round number heuristics), the 1/n heuristics, and the naïve portfolio diversification heuristics are among the most common when considering future pension plans. They recognized that when some heuristics act well in a small, non-complicated system, their use in very complex situations is ineffective and leads to non-optimal solutions.

2.2. Applications of NLP in analysing public sentiment

Natural Language Processing (NLP) has advanced sentiment analysis research in various domains. Abiola et al. (2023) used TextBlob and VADER to analyse COVID-19 pandemic responses on Twitter in Nigeria, revealing diverse public perceptions. Anoop and Sreelakshmi (2023) conducted sentiment analysis and topic modelling on Reddit comments, uncovering themes like symptoms, transmission, and societal issues.

In politics, Gorodnichenko, Pham, and Talavera (2021) studied information diffusion and bot influence during the 2016 E.U. Referendum and the U.S. Presidential Election, highlighting the interplay between social media and political discourse. Ali et al. (2022) analysed 7.6 million tweets to understand public sentiment toward presidential candidates in the 2020 U.S. election.

NLP has also been applied to environmental issues. Yang et al. (2023) analysed Sina Weibo data to uncover demographic variations in public attitudes towards environmental issues in China. Sun et al. (2023) analysed Sina Weibo's comments on plastic pollution control in China, emphasizing the importance of public sentiment in shaping environmental policies. These studies demonstrate the utility of NLP in extracting public sentiment from social media data across various domains, including public health, politics, and the environment.

2.3. Public perspectives on pension system policies

Some literature explores public attitudes toward pension systems and related welfare state policies. Svallfors (2011) revealed stable and increasing support for the welfare state in Sweden, indicating robust public backing despite political shifts. Fernandez and Jaime-Castillo (2013) provided a cross-national European perspective, finding that attitudes towards pension policies are significantly influenced by existing policy structures, which aligns with the opposing policy feedback theory. Van Dalen, Henkens, and Oude Mulders (2019) focused on Dutch employers' concerns regarding pension reforms linked to life expectancy, highlighting support for alternative policies like a flexible pension age. Schuetz et al. (2023) explored the German public's support for pension reforms in the context of demographic changes, showing an increased preference for raising the retirement age.

In China's context, Li and Wu (2018) found that the New Rural Pension Scheme (NRPS) increased trust in government at both central and local levels among potential beneficiaries but did not increase expectations of government responsibility for old-age support. Wang and Timonen (2021) identified disparities between individual perceptions and official assumptions underlying the current pension regime, with varying interpretations of retirement income inequality between enterprise workers and public-sector employees. Building on these findings, Long, Niu, and Yi (2024) studied Chinese residents' attitudes toward an individual pension system, identifying key factors such as trust and perceived usefulness influencing participation intentions. These studies collectively illustrate the complexity of public attitudes towards pension policies in China, shaped by factors like demographic trends, existing policy frameworks, societal perceptions, and the specific benefits individuals receive from pension reforms.

In summary, this study aims to fill the research gap by analysing public sentiment towards China's pension policies using a large dataset of Sina Weibo posts and advanced NLP techniques. The research

uncovers changes in key themes and demographic differences in opinions over the period 2018–2023, providing a detailed understanding of the shifting public discourse on pension reforms in China's unique socio-economic context.

3. Methodology

As outlined in Figure 1, the research framework of this study follows the research conducted by Huang et al. (2022). It systematically navigates through various phases, initiating tailored data collection via web crawling. The subsequent data cleaning and preprocessing stages ensure data reliability, setting the stage for an overall analysis focusing on gender and age cohorts. Sentiment analysis, executed through a fine-tuned Bidirectional Encoder Representations from Transformers (BERT) model, classifies posts into positive or negative sentiments and supports detailed time, gender, and region analyses. Sentiment analysis is understood here as defined in Giachanou and Crestani (2017); it involves the detection of sentiment polarity, categorizing the expressed opinion in a text as either positive or negative. The methodology culminates LDA (Blei, Ng, Jordan 2003) for comprehensive topic modelling, ensuring a thorough exploration of public sentiment and thematic trends related to pensions.

3.1. Data collection

Building upon the methodologies established by Huang et al. (2022), Sun et al. (2023), and Yang et al. (2023), this study employs web crawler technology – a digital tool designed to automatically navigate the web and collect specific data based on predetermined criteria. Utilizing the Python programming language, we have tailored our crawler to mine relevant data from the social media platform Sina Weibo. The process of web crawler is shown in Figure 2, and the specific steps are described in the following:

1. Environment configuration: A computational setup is established by integrating essential libraries, ensuring efficient data retrieval and parsing. User-agent details are configured to mimic genuine browser interactions. The focus keywords are '养老金', '退休金', and '养老保险', synonymous terms denoting pensions.

2. URL design and query structure: leveraging Weibo's search mechanism, a specific URL pattern is constructed. It targets the selected keywords from 1 January 2018 to 31 August 2023, ensuring daily granularity in data retrieval.

3. Data retrieval and refinement: each keyword is queried against the set date range on Weibo. For every keyword-date combination, the first ten result pages are scrutinized. Relevant content and metrics are extracted, while extraneous elements, such as emojis, are filtered out.

4. User profiling and metadata acquisition: user profiles on Weibo are further analysed to obtain metadata like gender, birth date, and location.

5. Data structuring and storage: post-extraction, data is systematically structured into rows and columns and archived in a CSV format, ensuring compatibility with analytical platforms.

3.2. Data pre-processing

In the data pre-processing phase, we remove duplicate entries based on content from our Weibo dataset. Dates are standardized, and any discrepancies are addressed. User ages are derived from postdates to gauge demographic patterns, and outlier ages are deleted and left in blank. We rigorously clean the content column by eliminating platform-specific notifications, English letters, URLs, mentions, certain hashtags, and non-standard characters.

During the subsequent steps in data pre-processing, we focus on refining the textual content for analysis. Firstly, the text undergoes segmentation with the use of Jieba in Python, a widely recognized Chinese text segmentation tool. This breaks down the textual content into individual terms or phrases, which are pivotal for analysing textual data in Chinese. However, textual data often contains noise – words that appear frequently but don't carry significant meaning, known as stopwords. We deploy an extensive set of Chinese stopwords from: the Python stopwordsiso library, Baidu Stopwords, Harbin Institute of Technology Stopwords, and Machine Intelligence Laboratory of Sichuan University Stopwords (Liu, He, Dai 2021), complemented by manually curated words and common punctuation.

3.3. Sentiment analysis

In sentiment analysis, particularly when deciphering users' sentiments from shared content, traditional techniques relying on sentiment dictionaries and machine learning face inherent challenges. These approaches often require heavy reliance on established sentiment lexicons and demand manual feature engineering, rendering them less adaptive to the dynamic nuances of contemporary text. While deep learning frameworks have shown promise in textual data interpretation, as highlighted by Sagnika, Mishra, and Meher (2021), our research employs the BERT model (Devlin et al. 2019), specifically the 'bert-base-Chinese' variant designed for the Chinese language.

The essence of our methodology lies in leveraging BERT's transformer architecture, which introduces a bidirectional mechanism for context comprehension. Traditional models typically process text linearly, either from left to right or vice versa, thereby overlooking the potential influence of surrounding words. BERT, however, evaluates the context of each word (w_i) by considering both its preceding ($w_{i-n},...,w_{i-1}$) and succeeding ($w_{i+1},...,w_{i+n}$) elements within a sentence, enabling a more nuanced understanding of linguistic meanings essential for precise sentiment categorization. Mathematically, this can be represented as:

$$\text{Context}\left(w_i\right) = f\left(w_{i-n},\ldots,w_{i-1},w_i,w_{i+1},\ldots,w_{i+n}\right) \tag{1}$$

where f denotes the transformation function applied by BERT to derive the contextual embedding of word w_i.

Our method employs the BERT model, fine-tuned with the ChnSentiCorp dataset PaddlePaddle,[2] referenced by Ou et al. (2022). This dataset is designed for PaddlePaddle's ERNIE training, containing labelled reviews of hotels, laptops, and books from online platforms, with reviews averaging 110 words.

[2] https://github.com/PaddlePaddle/ERNIE, original work published 2019.

This critical step significantly improves the model's proficiency in discerning and categorizing sentiments in Chinese texts, as evidenced by an AUC score of 0.98 on our test set. This score underscores the model's enhanced accuracy in identifying sentiment nuances. Subsequently, we utilize this fine-tuned model to classify sentiments in Sina Weibo posts, categorising them as positive or negative.

3.4. LDA topic model

Our analysis leverages the LDA model, a probabilistic framework for uncovering latent thematic structures within a large corpus of text data. Originating from the work of Blei, Ng, and Jordan (2003), LDA posits each document d as a mixture of topics, and each topic t as a distribution over words w. The model's generative process conceives each document as being created through the following steps: for each document d, decide on the number of words N_d it will contain, choose a distribution over topics θ_d based on a Dirichlet distribution parameterized by α, and for each of the N_d words, pick a topic t from θ_d and select a word w from a topic-specific distribution ϕ_t, which is also drawn from a Dirichlet distribution, but parameterized by β.

The cornerstone of our LDA application is the determination of the optimal number of topics K, which guides the granularity of the topics extracted from the data. Following the methodology outlined by Blum and Raviv (2023), we employ the coherence score to systematically gauge the clarity and distinction among topics identified. This measure aids in selecting K that maximizes topic interpretability. Specifically, our approach involves iterating over a range of K values, training an LDA model for each, and computing the coherence score C_v for the resulting topics. The optimal K is thus identified as the one that yields the highest C_v, formalized as:

$$K_{\text{optimal}} = \underset{K}{\operatorname{argmax}}\, C_v\left(K\right) \tag{2}$$

Upon conducting the LDA analysis with the optimal K, we derive a set of topics where each topic is represented by a distribution over words, $P(w|t)$, highlighting the probabilistic association of words with the inferred topics. This analysis, executed in the native Chinese language, ensures the preservation of nuanced meanings and cultural contexts inherent to the dataset.

In the final step of our analysis, we address the language barrier for non-Chinese readers by translating the key terms from the topic visualizations. Utilizing the Google Cloud Translation API and supplementary manual adjustments, we translate the salient Chinese tokens (w) to English, facilitating a broader comprehension of the identified topics. This cross-lingual translation is performed judiciously to maintain the semantic integrity of the topics uncovered by the LDA model.

4. Numerical results

4.1. Demographic analysis

Table 1 delineates the demographic distribution of the users behind the 268,052 posts from the Sina Weibo dataset. On the platform, users can identify their gender as female or male or opt not to specify. Regarding gender distribution, male posts predominate with 148,338 entries, accounting for 55.34%

of the total. In contrast, female users contributed 119,701 posts, comprising 44.66% of the dataset. It is worth noting that 13 users chose not to disclose their gender. Analysing age demographics, 61.05% of the users do not specify their age. However, it is evident that younger individuals primarily engage the platform. Specifically, posts by users aged 25–34 emerge as the most substantial segment, amounting to 14.27% of the total posts. The frequency of posts gradually declines with age, as shown in Table 1.

4.2. Sentiment analysis

To visualize the temporal evolution of sentiments over the years, a stacked bar chart (Figure 3) illustrates the proportion of positive and negative sentiments toward pension schemes from 2018 to 2023. Data showcase a discernible trend in the rise of negative sentiments over these years. In 2018, the negative sentiment was recorded at 41.6%, which marginally decreased to 40.1% in 2019. However, post-2019, there is a consistent surge in the proportion of negative sentiments. By 2020, it rose to 42.2%, followed by a more pronounced increase to 44.5% in 2021. The two subsequent years, 2022 and 2023, witnessed even steeper hikes, culminating at 49.9% and 51.0%, respectively.

Figure 4 presents the monthly proportion of negative sentiment towards pensions, with notable points A, B, C, and D marking significant public opinion events. Point A corresponds with China's nationwide introduction of reverse mortgage in August 2018 (Hanewald et al. 2020), allowing elderly homeowners to draw a regular income stream from their property while retaining ownership. This point sees an initial increase in the proportion of negative sentiment, surpassing 50% in the subsequent month. At Point B, following the policy issued on 22 June 2020, to extend the period for phased reduction and exemption of corporate social insurance premiums to mitigate economic strains caused by the COVID-19 pandemic (Gov.cn 2020), there is a sharp decrease in negative sentiment observed in the month that followed. Point C marks the introduction of the Individual Retirement Account (IRA) system on 4 November 2022 (Gov.cn 2022b), a tax-deferred savings plan permitting individuals to build their retirement funds with deferred tax benefits, which corresponded with a notable decrease in negative sentiment in this month. Conversely, Point D corresponds to the release of the *Outline of the Strategic Plan for Expanding Domestic Demand (2022–2035)*, which underscored the gradual postponement of the mandatory retirement age (Gov.cn 2022c). Following its publication on 14 December 2022, there is a notable increase in negative sentiment, culminating in February 2023. This upsurge suggests growing public concern or discontent with the proposed changes to retirement policy during this interval.

To elucidate the factors influencing both the escalation and mitigation of negative sentiments, this study examines linguistic variations during four pivotal periods: July–August 2018, June–July 2020, October–November 2022, and January–February 2023. Figure 5 illustrates the terms that exhibited the most significant shifts in frequency across these intervals.

For Point A, the increase in negative sentiment is analysed through posts from July to August 2018. The top-left sub-plot of Figure 5 shows a noticeable shift in the frequency of terms such as 'collateral', 'retiree', 'adjustment', 'fund', 'reverse mortgage for seniors', and 'housing'. Notably, 'collateral', 'fund', 'reverse mortgage', and 'housing' experienced pronounced increases. This change implies that public discourse around reverse mortgages and related financial terms intensified alongside growing concerns, as reflected in the uptick of negative sentiment during this period.

Point B reflects a decrease in negative sentiment. Illustrated in the top-right sub-plot of Figure 5, from June to July 2020, there is an increase in terms such as 'retiree', 'adjustment', 'corporate', and 'human resources'. This period coincides with the implementation of policies that reduced enterprise insurance contributions, a response to the economic strain of the COVID-19 pandemic. The dip in negative sentiment suggests the policy was favourably perceived, addressing the economic worries of that time.

Point C indicates a decrease in the share of negative sentiment. As shown in the bottom-left sub-plot of Figure 5, between October and November 2022, terms like 'individual retirement', 'fund', 'participant', 'account' and 'capital' become more frequent. The introduction of the tax-deferred IRA system during this time appears to have been met with a positive reaction, demonstrating the government's commitment to pension reform and its positive reception.

For Point D, there is a consistent rise in negative sentiment. The bottom-right sub-plot of Figure 5, between January and February 2023, highlights an increase in terms such as '65', 'delayed retirement', 'launch', 'work', 'China', and 'basic'. The terms '65 years old' and 'delayed retirement' are particularly prominent, pointing to discussions around potential policy changes to increase the retirement age to 65. The increased frequency of these terms correlates with growing public disapproval or resistance to such pension reforms.

4.3. Sentiment analysis by demographic groups

Life cycle stages, especially concerning financial stability and retirement, can influence public sentiment on topics like pensions. As shown in Table 2, the age cohorts under 24 show apparent variations in their sentiment regarding pensions. The youngest group, those under 18, exhibit the lowest negative sentiment at 38.29%. However, this negative sentiment sees a noticeable rise to 45.82% as individuals transition into early adulthood (18–24). The sentiment remains fairly consistent later on, with the 25–34 and 35–49 age groups registering negative sentiments at 44.28% and 44.10%, respectively. A significant uptick is evident for those approaching 50–59 or in the early stages of retirement (60–69), with figures at 49.00% and 49.46%. However, for those aged 70 and above, the sentiment recedes slightly to 47.89%.

Looking at gender, the female users, accounting for a total of 148,338, demonstrate a slightly higher negative sentiment regarding pensions, registering at 46.45%. In contrast, the male users, numbering 119,701, exhibit a negative sentiment proportion of 43.77%.

Figure 6 illustrates the regional differences in negative sentiments towards pensions in China, revealing significant disparities. The Heilongjiang province exhibits the highest negative sentiment at 56.98%. In contrast, Tibet shows the lowest at 39.18%. Interestingly, while Tibet had the highest birth rate (14.24%) and natural population growth rate (8.76%) in China in 2022, Heilongjiang recorded the lowest birth rate (3.34%) and a negative population growth rate (-5.75%) (National Bureau of Statistics, 2022). Additionally, Heilongjiang faced a deficit in its basic pension insurance in 2021, the only Chinese province with such a shortfall, amounting to RMB -37.2338 bn (National Bureau of Statistics 2022), underscoring the fiscal implications of an aging population on regional pension schemes.

The discrepancies in pension sentiment are not merely numerical but indicate a deeper crisis of confidence and sustainability in China's pension framework. These variations can be attributed to a mix of regional economic disparities, localized pension policies, and the cumulative effect of personal and collective pension-related experiences. The economic inequalities between provinces, amplified

by inconsistent regional growth, critically challenge the feasibility of maintaining a coherent and equitable national pension system (Zhang et al. 2023). The present study not only calls for urgent policy attention but also demands a reevaluation of the structural integrity of the pension system to ensure its long-term viability and fairness (Zhang et al. 2023).

4.4. LDA analysis

In our study, we employ the coherence value metric to identify the most suitable number of topics for our analysis. Figure 7 clearly demonstrates that the coherence score peaks when the number of topics is set to six. Consequently, we determined that six topics provide the highest level of topic coherence, making it the optimal choice for our LDA model.

Figure 8 illustrates that Topic 1 focuses on the financial pressures and societal impacts associated with an aging demographic in China. Within the Weibo platform, which is primarily utilized by users between 18 and 50 years old, discussions frequently involve terms like 'money', 'salary', 'elder', 'elderly', and 'child'. These terms reflect the financial concerns prevalent among this age group, who often face the challenge of supporting older and younger generations. The analysis of these discussions indicates the significant stress experienced by the working-age population, who must manage the economic implications of extended familial responsibilities. This topic underscores the importance of developing policy measures addressing the financial burdens of individuals supporting multiple generations.

As shown in Figure 9, Topic 2 presents themes closely related to those in Topic 1, evidenced by the overlapping circles representing both topics. Topic 2 centres on familial duties, personal life decisions, and their financial ramifications. This topic frequently addresses life events such as 'retirement', 'marriage', and 'getting old'. The concurrent mention of 'son' and 'daughter' along with 'parents' indicates a dialogue focused on the responsibilities toward both younger and older family members. The combination of terms like 'money', 'work', and 'marry' with emotional expressions such as 'unwilling' and 'feel' suggests that these discussions may involve significant concerns or reluctances related to the financial and emotional strains of familial obligations. This analysis points to the need for a deeper understanding of the interplay between financial stability and family responsibilities within the societal context.

Topics 3, 4, and 5 (Figure 10 to Figure 12) appear to emanate from media sources or government policy propaganda. Topic 3, underscored by terms such as 'urban and rural residents', 'participate in insurance' and 'social security', emphasizes the administrative and procedural dimensions of social welfare. Topic 4, characterized by terms like 'pay the fees', 'retirement age', 'medical insurance' as well as 'staff' and 'unemployment' delves deeper into the mechanics of the employment-based social security system. Meanwhile, Topic 5, highlighted by terms like 'develop', 'fund', 'pension', 'market' and 'reverse mortgage for seniors' focuses on the overarching financial strategies, market innovations, and reforms in the pension and insurance arenas. The structured nature of the terminology and the context indicate a formal tone, reinforcing the notion of their origins in official communications or media narratives on governmental policies.

Topic 6 which is shown in Figure 13 revolves around potential security and legal concerns related to pensions. The terms 'company', 'strike' and 'court' indicate possible disputes or challenges associated

with pension arrangements within corporate entities. Meanwhile, terms such as 'scam', 'criminals', 'illegal law', 'cheat' and 'fraud' highlight fears or instances of deception and fraudulence directly linked to the payment of pension or related financial dealings. This topic provides insights into the public discourse that highlights safety, legality, and potential misconduct issues within the pension systems. The frequency and context of these terms reflect a significant level of public attention and concern regarding the integrity and security of pension schemes.

A deeper insight into the topics shows that Topic 1 corresponds to societal issues while Topic 2 is related to personal and family life. They gave feedback to each other, as social and personal lives are mutually connected. On the other hand, Topic 3 relates to general pension policy at the national level, Topic 4 refers to particular regulations concerning individual payments, participation, and work, and Topic 5 corresponds to propaganda on the development of the pension fund system in China. Topic 6 is the opposite and corresponds to such frameworks as scams, criminals, or illegal rules. Linking them we may state that Topics 3, 4, and 5 constitute an institutional framework for pensions and pensioners in China for the near future. They definitely bring up emotions that are enhanced by specified social, cultural, and economic constraints (one-child policy, responsibility for elderly people, longevity, etc.). Topic 6 is related to the institutional framework but reveals deficiencies in the system, resulting in scams and cheating. The proportion of negative sentiment supports these findings. These are as follows: Topic 3 – 56.21%, Topic 4 – 52.52%, Topic 5 – 52% and Topic 6 – 54.43%. On the other hand, for Topic 1, only 23.24% of sentiments were negative, and for Topic 2 – 21.68%.

The paper's findings correspond to defined biases and heuristics. Firstly, the negativity bias is strongly observed, where people focus more on negative than positive events. This is evident in the steady increase in negative sentiment, particularly regarding government policy announcements, reforms, and the rising statutory retirement age. Negative sentiment (over 54%) is especially prevalent in discussions on pension-related fraud and security concerns. Financial pressures, such as balancing support for both children and elderly family members, give rise to concerns about losing financial security, creating an emotional response shaped by loss aversion heuristics (Topic 1). Similarly, adverse reactions to market-driven reforms and reverse mortgages indicate concerns over losing control or stability in retirement finances (Topic 5).

The anchoring effect is observed as people rely on initial, often biased or incomplete, information. Sentiment analysis reveals that formal language and authoritative tones in pension policy discussions (e.g., regulatory mechanisms, administrative frameworks) may reinforce scepticism toward institutions (Topics 3–5). The status quo bias explains resistance to policy changes, such as increasing the statutory retirement age and disrupting long-standing expectations (Topic 4). Availability heuristics also play a role; the focus on terms like 'fraud', 'criminals', and 'illegal law' highlights how widely publicized incidents amplify fears about the system's integrity (Topic 6).

Regional and cultural variability further shape responses to policies. For instance, lower negativity in Tibet contrasts with higher dissatisfaction in Heilongjiang, reflecting regional economic conditions and local cultural attitudes. Additionally, Chinese social norms emphasize familial obligations, as seen in discussions about financial burdens tied to supporting parents and children. This also explains reluctance toward life decisions like marriage or retirement, reflecting the tension between individual financial goals and societal expectations (Topic 2).

5. Discussions and conclusions

In conclusion, through an extensive NLP analysis of over 260,000 Weibo posts spanning from January 2018 to August 2023, this article has provided a comprehensive understanding of public sentiment toward pension policies in China. The findings indicate that policy changes and announcements significantly influence public opinion on pensions. Structural changes in multi-pillar pension systems, which require greater individual responsibility for retirement saving (Haupt 2023), likely contribute to the heightened concerns observed in public sentiment.

Moreover, the analysis demonstrates that pension policy is critical across various social and demographic characteristics, including gender, age groups, and provinces. Generational differences are a key factor, as Xie, Osińska, and Szczepaniak (2023) emphasize that varying attitudes toward retirement savings across age cohorts reflect diverse financial priorities and life-stage concerns. Regionally, provinces like Heilongjiang, burdened by economic challenges and pension fund deficits, exhibit the highest levels of dissatisfaction. In contrast, Tibet shows relatively lower negativity, likely due to stronger economic stability and adequate social security measures. Institutional factors also significantly influence public sentiment, aligning with Zhao, Qian, and Shan (2021) findings that trust in pension policies is shaped by perceptions of effectiveness and equity, driven by factors such as pension coverage and supportive social policies.

LDA analysis identifies six topics, with Topics 1 and 2 focusing on societal and personal aspects and Topics 3 to 5 addressing institutional pension frameworks. Topic 6 exposes systemic vulnerabilities such as scams. Sentiment analysis shows higher negative sentiments about institutional topics, indicating public concerns about policy and system effectiveness. The higher negative sentiments related to institutional topics align with the concept of loss aversion in behavioural finance theory, which suggests that individuals are more sensitive to potential losses than gains (Kahneman, Tversky 1979).

In the context of pension policies, the public may perceive institutional vulnerabilities and inefficiencies as threats to their financial security, leading to heightened negative sentiments. Moreover, the variation in sentiment across topics underscores the importance of considering the heterogeneity in individuals' behavioural responses to different aspects of pension policies, as emphasized by behavioural finance theory (Mitchell, Utkus 2003).

This study contributes to applying behavioural finance theory by demonstrating how cognitive heuristics, such as availability bias and loss aversion, shape public sentiment toward pension policies. Availability bias offers insights into why widely publicized issues, such as scams and fraud in pension systems, disproportionately influence public perception, amplifying negativity even when such cases may not reflect the overall system's integrity (Tversky, Kahneman 1974). Loss aversion further explains the heightened negative sentiments observed in response to institutional vulnerabilities and policy changes perceived as financial risks, such as delayed retirement (Kahneman, Tversky 1979).

Additionally, the demographic and regional variations in sentiment highlighted in this study align with behavioural finance theory's emphasis on heterogeneity in behavioural responses (Mitchell, Utkus 2003). For instance, older adults nearing retirement are more prone to loss aversion, reacting strongly to changes threatening their financial stability. Similarly, regions facing economic challenges, such as Heilongjiang, exhibit heightened sensitivity to perceived risks, likely influenced by availability bias (Zhao, Qian, Shan 2021). By integrating these theoretical insights with advanced NLP techniques, the present research provides a novel framework for understanding public sentiment and its cognitive underpinnings in complex policy environments.

The implementation of these insights is pivotal for effective pension reform in China. Policymakers should be conscious of the nuanced findings of this study, particularly the varied sentiments across different demographics and regions. This involves rectifying systemic weaknesses and enhancing public trust and understanding through transparent communication and inclusive policymaking. Engaging with the public to educate it about pension reforms and actively addressing their concerns can help bridge the gap between policy intentions and public perception. Also, preventing scams by educating the most threatened group of the population (the elderly people) is required to build trust in the public system.

As China's demographic landscape continues to evolve, these adaptive and responsive measures in pension policy implementation will be key to ensuring both the system's sustainability and its acceptance by the populace. Implementing the NLP and LDA methodology has definitely enabled the uncovering of topics that troubled the Chinese society and sentiments that permeated it in the considered time frame, confirming their wide applications for key social and financial issues.

While this study offers extensive insights, it is limited by its reliance on social media data, potentially overlooking the perspectives of demographics less active online, such as older adults who are significantly affected by pension policies. Moreover, the methodology used does not establish causal relationships, focusing instead on correlational insights between public sentiment and policy changes. Future research should expand to include empirical surveys and focus groups to gather a more comprehensive array of viewpoints and explore methodologies that can test for causality.

References

Abiola O., Abayomi-Alli A., Tale O.A., Misra S., Abayomi-Alli O. (2023), Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser, *Journal of Electrical Systems and Information Technology*, 10(1), 5, DOI: 10.1186/s43067-023-00070-9.

Ali R.H., Pinto G., Lawrie E., Linstead E.J. (2022), A large-scale sentiment analysis of tweets pertaining to the 2020 US presidential election, *Journal of Big Data*, 9(1), 79, DOI: 10.1186/s40537-022-00633-z.

Anoop V.S., Sreelakshmi S. (2023), Public discourse and sentiment during Mpox outbreak: an analysis using natural language processing, *Public Health*, 218, 114–120, DOI: 10.1016/j.puhe.2023.02.018.

Benartzi S., Thaler R. (2007), Heuristics and biases in retirement savings behavior, *Journal of Economic Perspectives*, 21(3), 81–104, DOI: 10.1257/jep.21.3.81.

Blei D., Ng A., Jordan M. (2003), Latent Dirichlet allocation, *Journal of Machine Learning Research*, 3, 993–1022, DOI: 10.1162/jmlr.2003.3.4-5.993.

Blum A., Raviv A. (2023), The effects of the financial crisis and Basel III on banks' risk disclosure: a textual analysis, *Finance Research Letters*, 53, 103680, DOI: 10.1016/j.frl.2023.103680.

Chowdhary K.R. (2020), Natural language processing, in: K.R. Chowdhary (ed.), *Fundamentals of Artificial Intelligence*, Springer India, DOI: 10.1007/978-81-322-3972-7_19.

David D., Cramer D. (2009), 6 rational and irrational beliefs in human feelings and psychophysiology, in: D. David, S. Lynn, A. Ellis (eds.), *Rational and Irrational Beliefs: Research, Theory, and Clinical Practice,* Oxford Academic, DOI: 10.1093/acprof:oso/9780195182231.003.0006.

Devlin J., Chang M.-W., Lee K., Toutanova K. (2019), *BERT: Pre-training of deep bidirectional transformers for language understanding*, arXiv, 1810.04805, DOI: 10.48550/arXiv.1810.04805.

Fernandez J.J., Jaime-Castillo A.M. (2013), Positive or negative policy feedbacks? Explaining popular attitudes towards pragmatic pension policy reforms, *European Sociological Review*, 29(4), 803–815, DOI: 10.1093/esr/jcs059.

Gorodnichenko Y., Pham T., Talavera O. (2021), Social media, sentiment and public opinions: evidence from #Brexit and #USElection, *European Economic Review*, 136, 103772, DOI: 10.1016/j.euroecorev.2021.103772.

Gov.cn (2017), *State Council on the issuance of the scheme for the aging population and the construction of the elderly care system within the Thirteenth Five Year Plan*, https://www.gov.cn/zhengce/content/2017-03/06/content_5173930.htm?cid=357.

Gov.cn (2020), *Ministry of Human Resources and Social Security, Ministry of Finance and General Administration of Taxation Circular on issues concerning the extension of the implementation period of the policy of phased reduction and exemption of enterprise social insurance premiums, etc.*, https://www.gov.cn/zhengce/zhengceku/2020-06/25/content_5521984.htm.

Gov.cn (2022a), *State Council on the issuance of the scheme for the Development of a Retirement Benefit System within the Fourteenth Five-Year Plan*, https://www.gov.cn/zhengce/zhengceku/2022-02/21/content_5674844.htm.

Gov.cn (2022b), *Ministry of Human Resources and Social Affairs, Ministry of Finance, State Administration of Taxation, China Banking and Insurance Regulatory Commission, and China Securities Regulatory Commission, jointly issued the implementation measures for individual pensions*, https://www.gov.cn/xinwen/2022-11/05/content_5724785.htm.

Gov.cn (2022c), *Central Committee of the Chinese Communist Party and the State Council issued the "Outline of the Strategic Plan for Expanding Domestic Demand (2022–2035)"*, https://www.gov.cn/zhengce/2022-12/14/content_5732067.htm.

Gov.cn (2024), *Delaying the mandatory retirement age in a steady and orderly manner*, https://www.gov.cn/zhengce/202409/content_6974483.htm.

Hanewald K., Bateman H., Fang H., Wu S. (2020), Is there a demand for reverse mortgages in China? Evidence from two online surveys, *Journal of Economic Behavior & Organization*, 169, 19–37, DOI: 10.1016/j.jebo.2019.10.023.

Huang H., Long R., Chen H., Sun K., Li Q. (2022), Exploring public attention about green consumption on Sina Weibo: using text mining and deep learning, *Sustainable Production and Consumption*, 30, 674–685, DOI: 10.1016/j.spc.2021.12.017.

Kahneman D., Tversky A. (1979), Prospect theory: an analysis of decision under risk, *Econometrica*, 47(2), 263–291, DOI: 10.2307/1914185.

Li Z., Wu X. (2018), Social policy and political trust: evidence from the new rural pension scheme in China, *The China Quarterly*, 235, 644–668, DOI: 10.1017/S0305741018000942.

Liu S., He T., Dai J. (2021), A survey of CRF algorithm based knowledge extraction of elementary mathematics in Chinese, *Mobile Networks and Applications*, 26(5), 1891–1903, DOI: 10.1007/s11036-020-01725-x.

Long Z., Niu J., Yi F. (2024), Driving factors and micro-mechanism of residents' participation intention in individual pension system, *Heliyon*, 10(2), e24488, DOI: 10.1016/j.heliyon.2024.e24488.

Mitchell O., Utkus S. (2003), *Pension Design and Structure. New Lessons from Behavioral Finance*, Oxford University Press, DOI: 10.2139/ssrn.464640.

National Bureau of Statistics (2022), *China Statistical Yearbook – 2022*, http://www.stats.gov.cn/sj/ndsj/2022/indexch.htm.

Osińska M., Stawicki J. (2017), Identification of heuristics in the process of decision making on financial markets, in: K. Nermend, M. Łatuszyńska (eds.), *Neuroeconomic and Behavioral Aspects of Decision Making*, Springer International Publishing, DOI: 10.1007/978-3-319-62938-4_8.

Ou H., Yu L., Tian S., Chen X. (2022), Chinese adversarial examples generation approach with multi--strategy based on semantic, *Knowledge and Information Systems*, 64(4), 1101–1119, DOI: 10.1007/s10115-022-01652-1.

Rathi K.N., Geetha D. (2024), Relationship between rational and irrational investment decisions and multiple intelligence of investors, *International Journal of Economics and Financial Issues*, 14(4), 282–289, DOI: 10.32479/ijefi.16118.

Sagnika S., Mishra B.S.P., Meher S.K. (2021), An attention-based CNN-LSTM model for subjectivity detection in opinion-mining, *Neural Computing and Applications*, 33(24), 17425–17438, DOI: 10.1007/s00521-021-06328-5.

Schuetz J., Uebelmesser S., Baginski R., Aprea C. (2023), Pension reform preferences in Germany: Does information matter?, *European Journal of Political Economy*, 80, DOI: 10.1016/j.ejpoleco.2023.102476.

Sun Y., Wang D., Li X., Chen Y., Guo H. (2023), Public attitudes toward the whole life cycle management of plastics: a text-mining study in China, *Science of The Total Environment*, 859, 159981, DOI: 10.1016/j.scitotenv.2022.159981.

Svallfors S. (2011), A bedrock of support? Trends in welfare state attitudes in Sweden, 1981–2010, *Social Policy & Administration*, 45(7), 806–825, DOI: 10.1111/j.1467-9515.2011.00796.x.

Tversky A., Kahneman D. (1974), Judgment under uncertainty: heuristics and biases, *Science*, 185(4157), 1124–1131, DOI: 10.1126/science.185.4157.1124.

Van Dalen H.P., Henkens K., Oude Mulders J. (2019), Increasing the public pension age: employers' concerns and policy preferences, *Work, Aging and Retirement*, 5(3), 255–263, DOI: 10.1093/workar/waz004.

Wang Q., Timonen V. (2021), Retirement pathways and pension inequality in China: a grounded theory study, *International Journal of Sociology and Social Policy*, 41(13/14), 96–111, DOI: 10.1108/IJSSP-09-2020-0454.

WHO (2019), *Building a better world for an ageing population*, https://www.who.int/westernpacific/articles/item/chinas-ageing-population-will-feature-in-global-project.

Xie X., Osińska M., Szczepaniak M. (2023), Do young generations save for retirement? Ensuring financial security of Gen Z and Gen Y, *Journal of Policy Modeling*, 45(3), 644–668, DOI: 10.1016/j.jpolmod.2023.05.003.

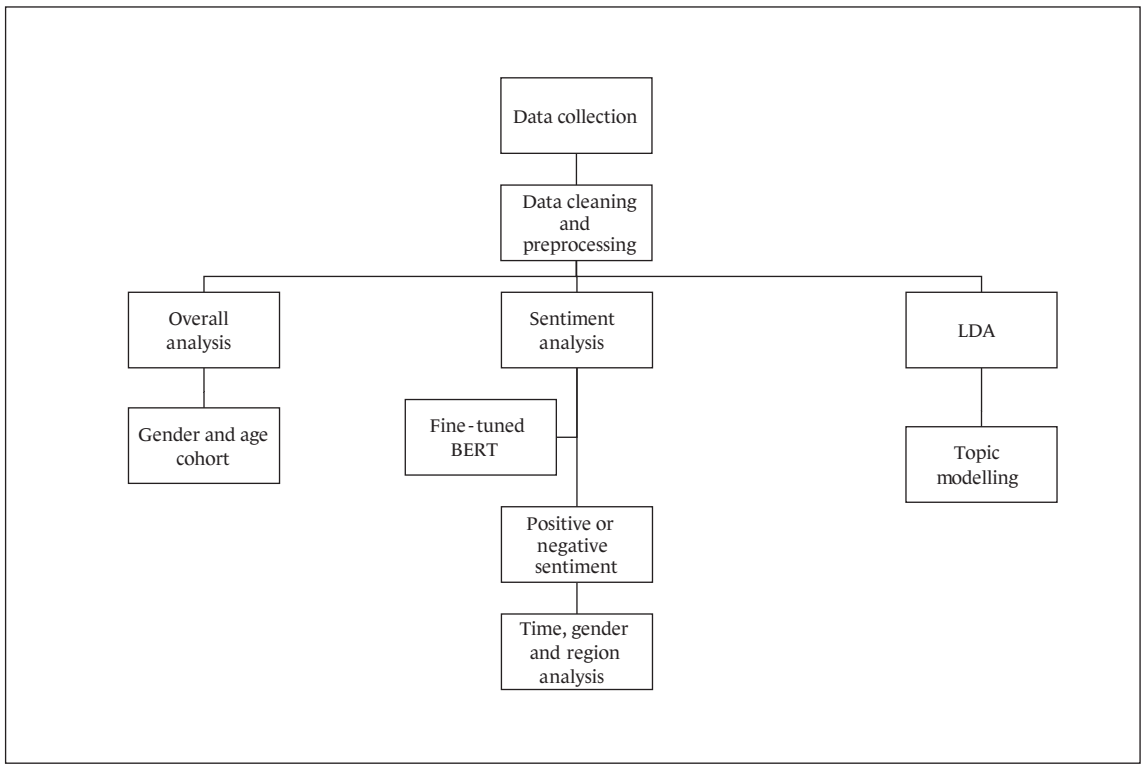
Yang L., Wu S., Li G., Yuan Y. (2023), Explore public concerns about environmental protection on Sina Weibo: evidence from text mining, *Environmental Science and Pollution Research*, 104067–104085, DOI: 10.1007/s11356-023-29757-2.

Zhang S., Wang X., Li H., Wang H. (2023), The geographical pension gap: understanding the causes of inequality in China's pension funds, *PLOS ONE*, 18(7), e0288754, DOI: 10.1371/journal.pone.0288754.

Zhao L., Qian J., Shan W. (2021), Filial piety, pension policy and changing perceptions of elderly care responsibility: evidence from China, *China: An International Journal*, 19(3), 112–131, DOI: 10.1353/chn.2021.0031.

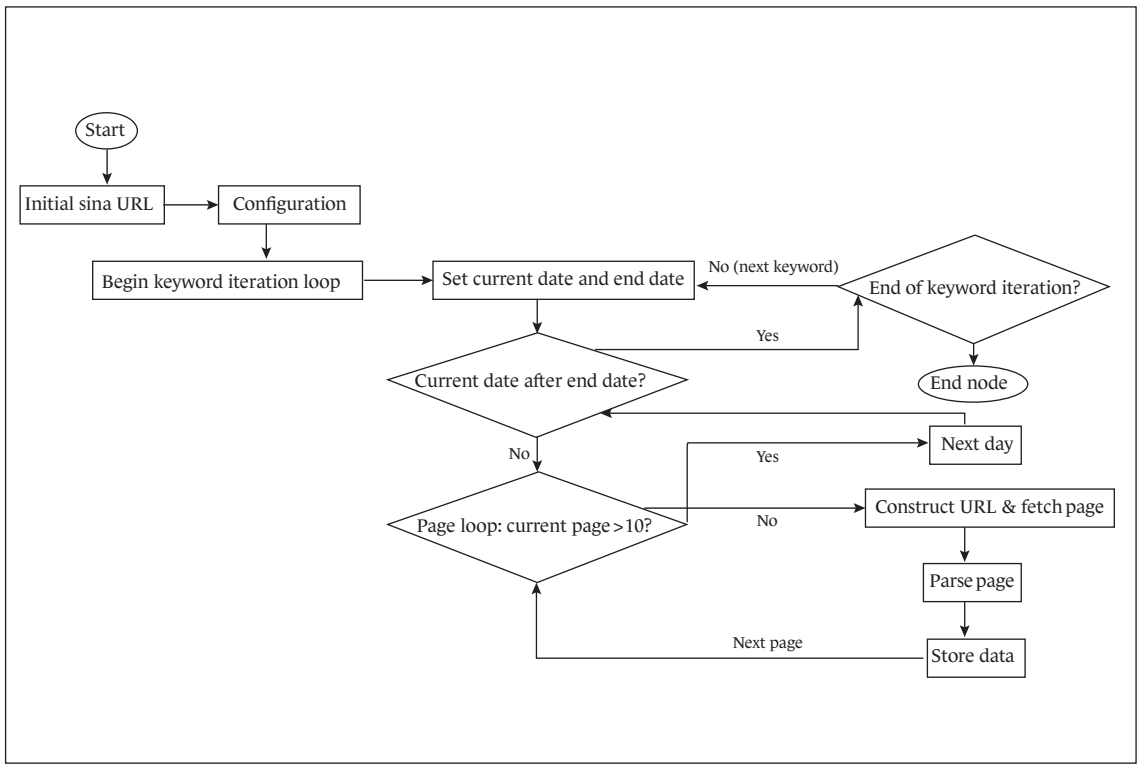
Appendix

Figure 1
Research framework of the study



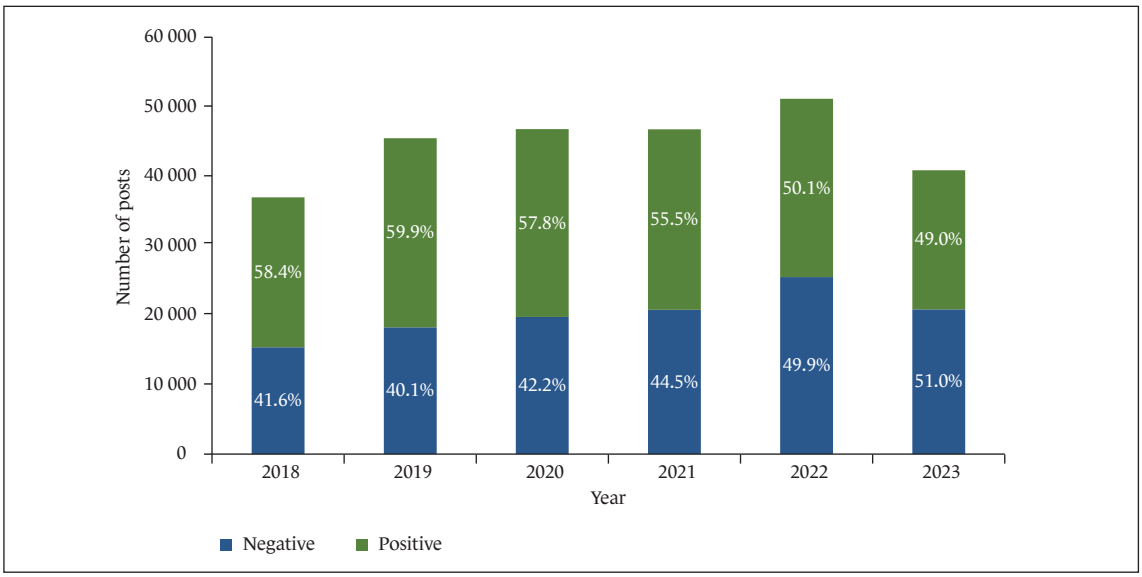
Source: own elaboration.

Figure 2
The process of web crawler



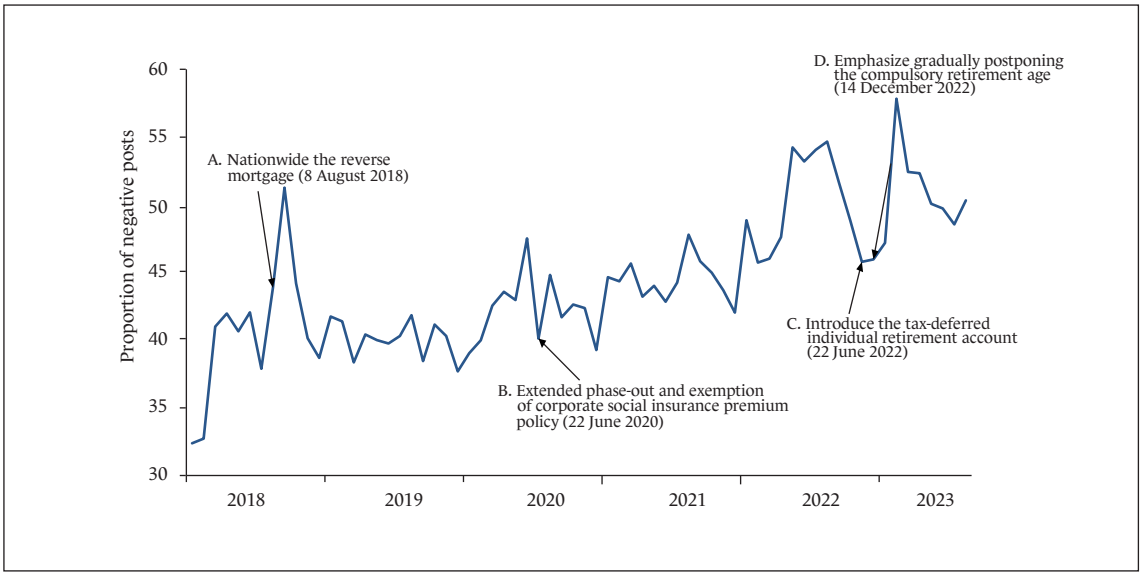
Source: own elaboration.

Figure 3
Annual proportion of negative sentiment posts



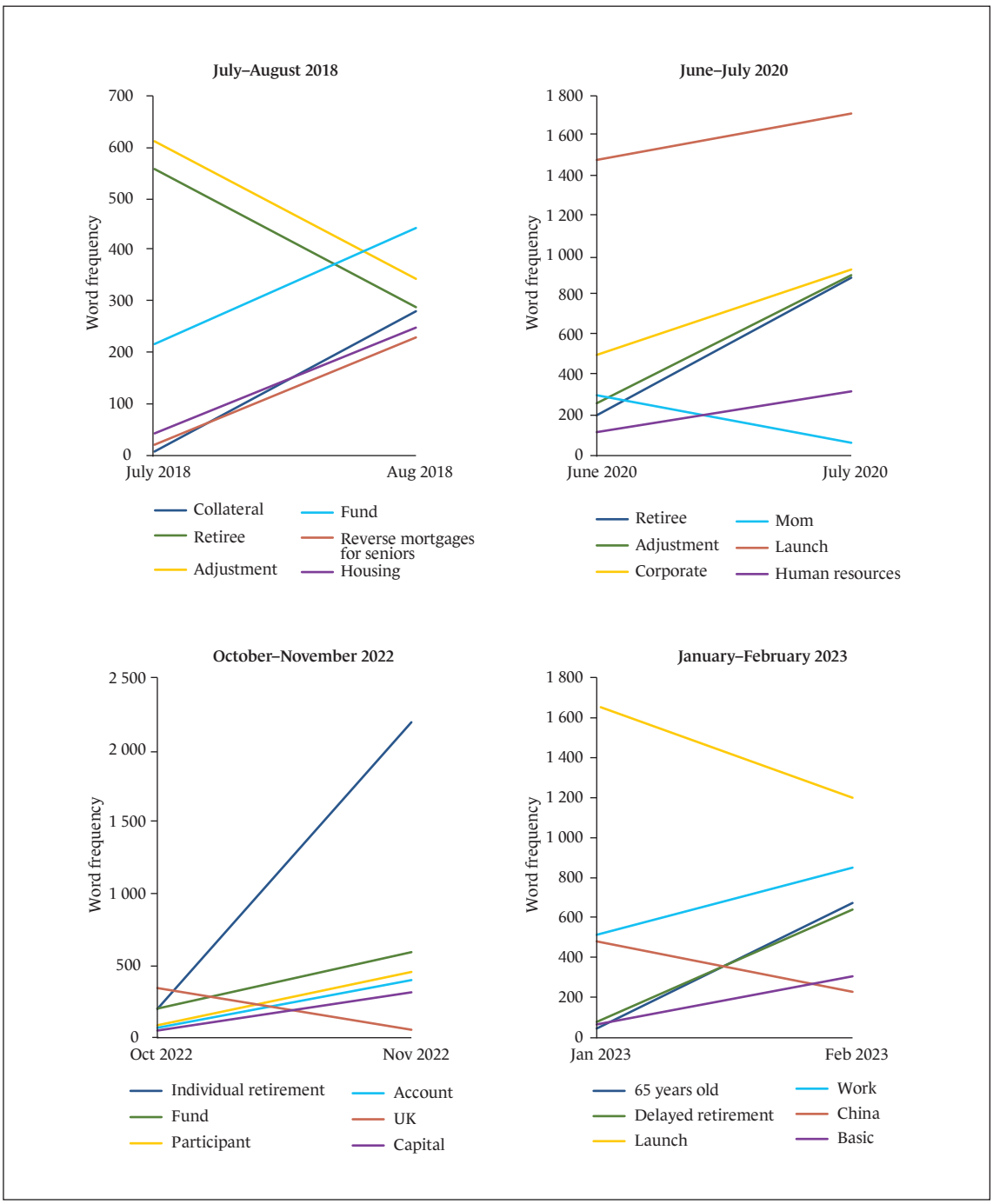
Source: own elaboration.

Figure 4
Monthly proportion of negative sentiment posts as a result of retirement policy changes (A, B, C and D)



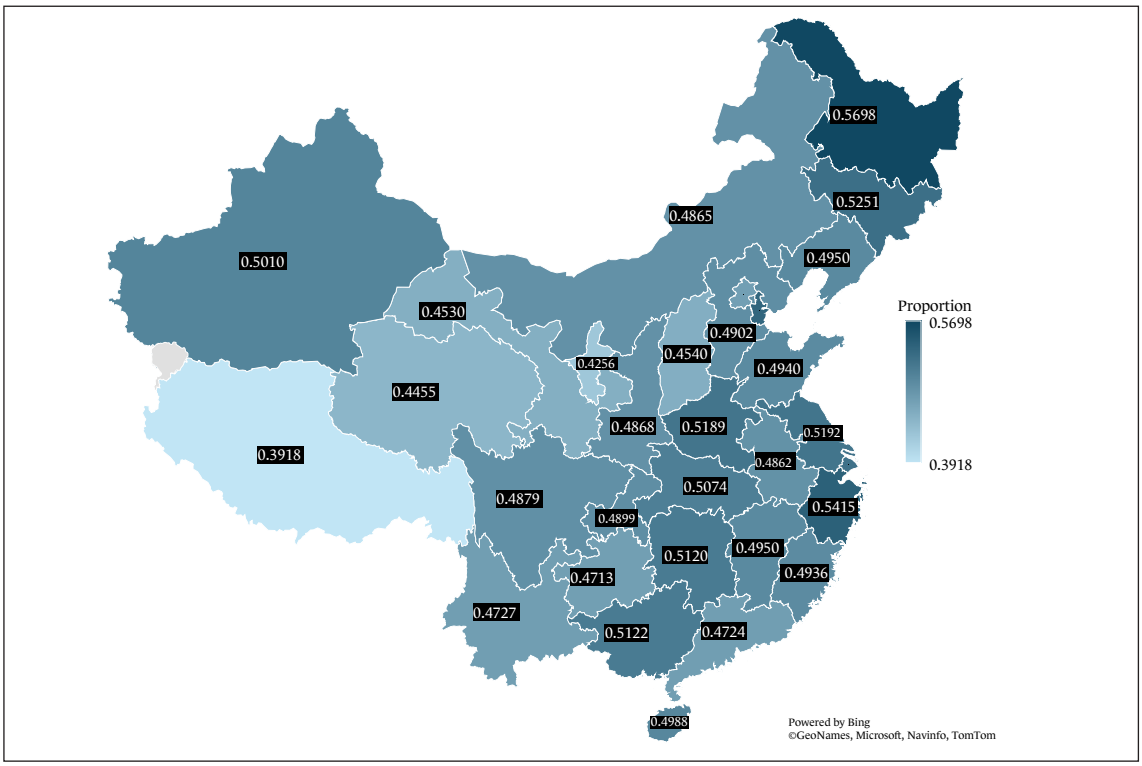
Source: own elaboration.

Figure 5
Top trending words for July 2018 – August 2018, June 2020 – July 2020, October 2022 – November 2022 and January 2023 – February 2023



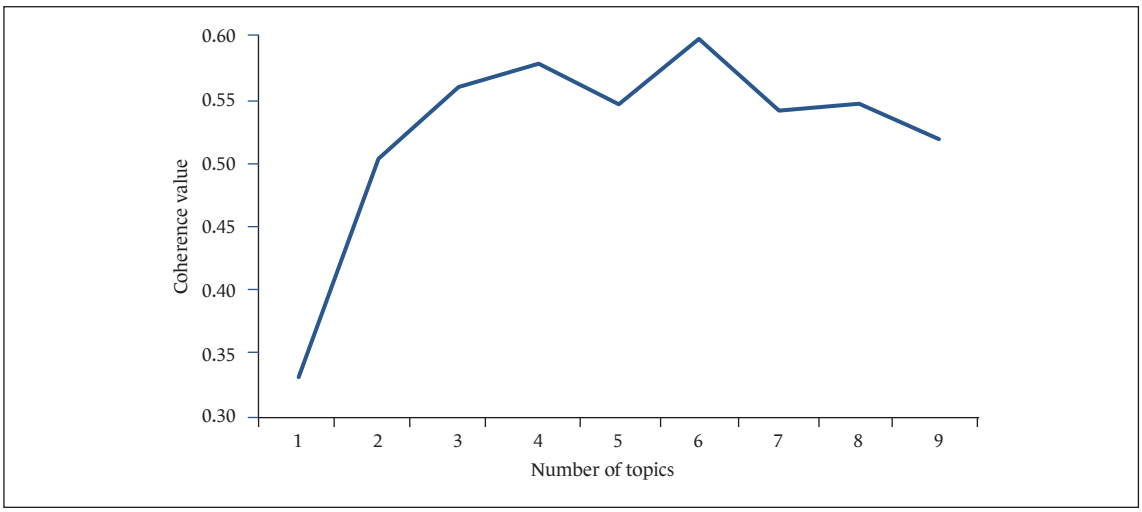
Source: own elaboration.

Figure 6
Proportion of negative sentiment posts by regions (mainland China)



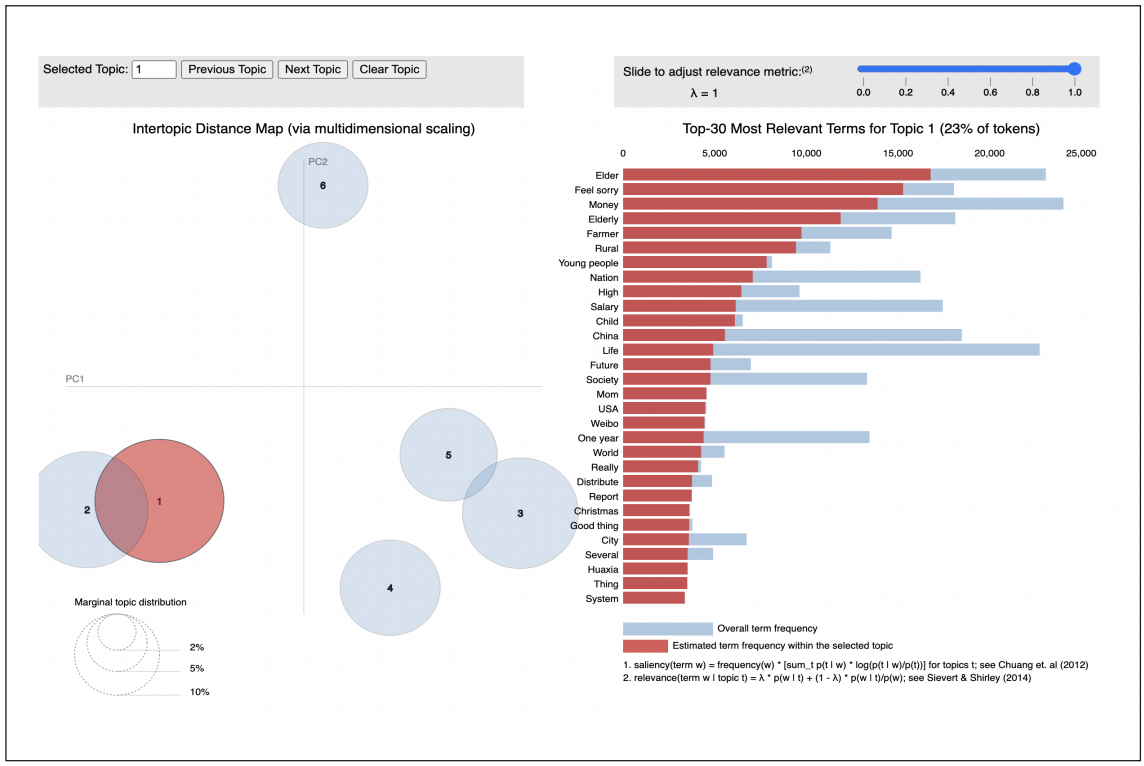
Source: own elaboration.

Figure 7
Coherent score for different numbers of topics



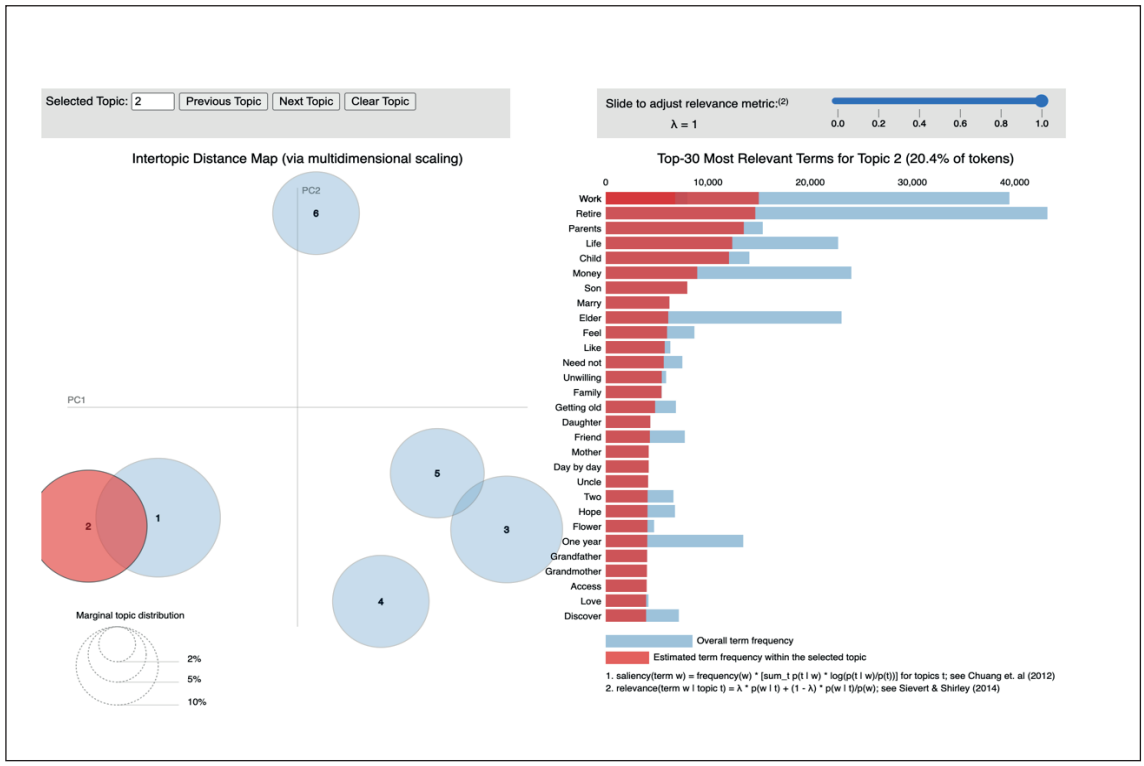
Source: own elaboration.

Figure 8
LDA topic 1 result



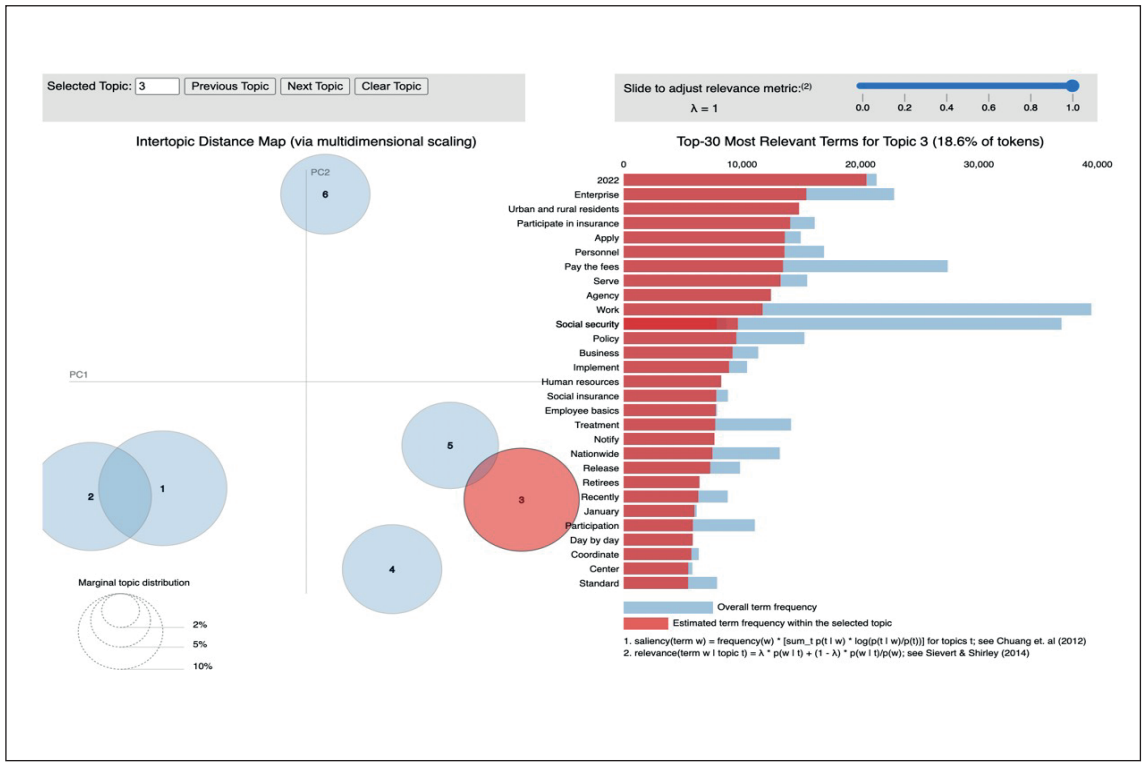
Source: own elaboration.

Figure 9
LDA topic 2 result



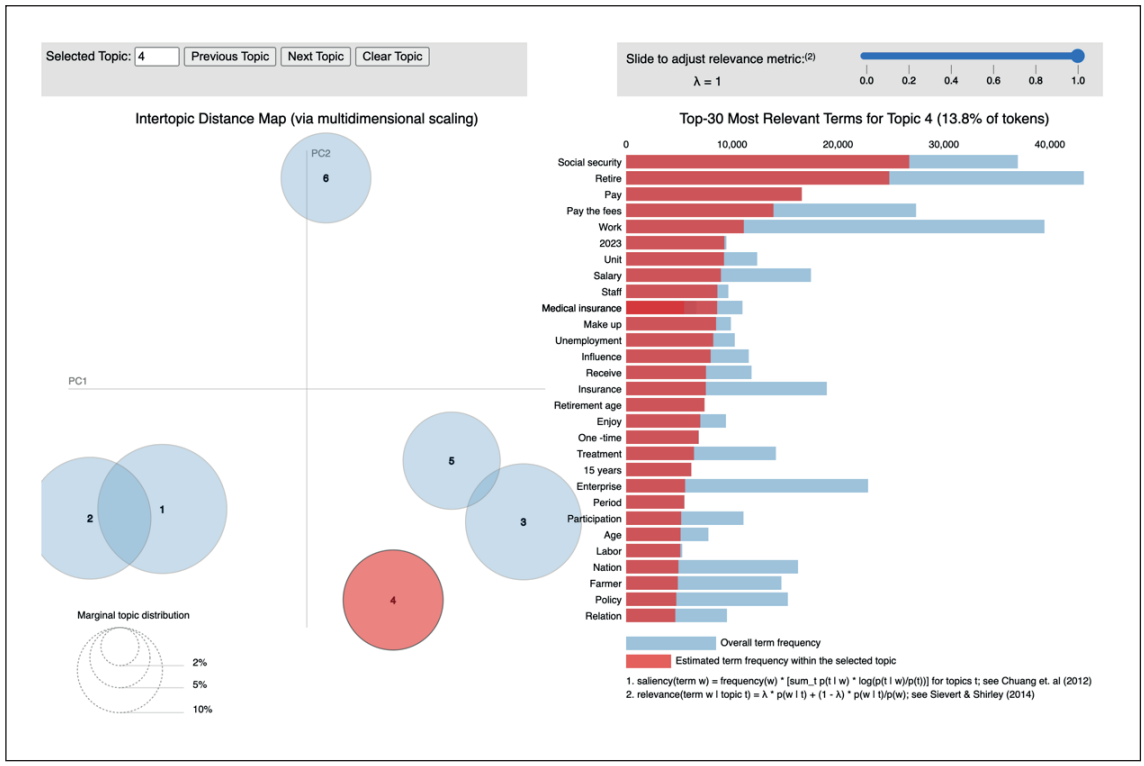
Source: own elaboration.

Figure 10
LDA topic 3 result



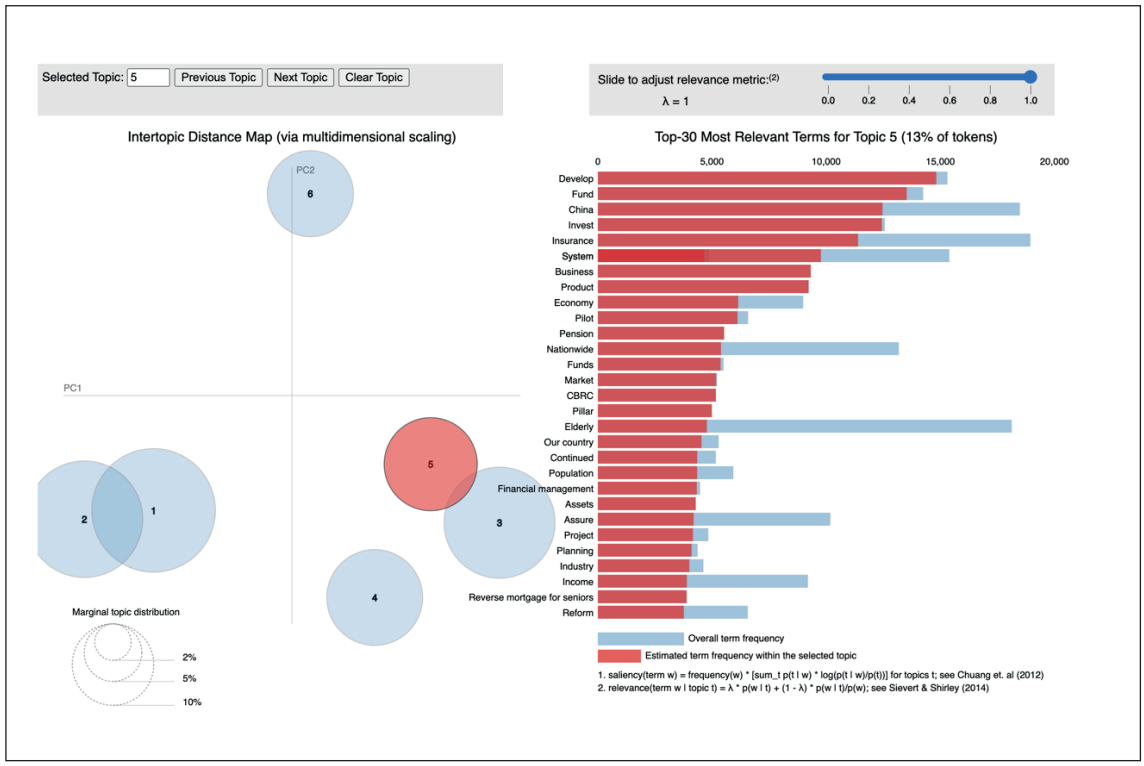
Source: own elaboration.

Figure 11
LDA topic 4 result



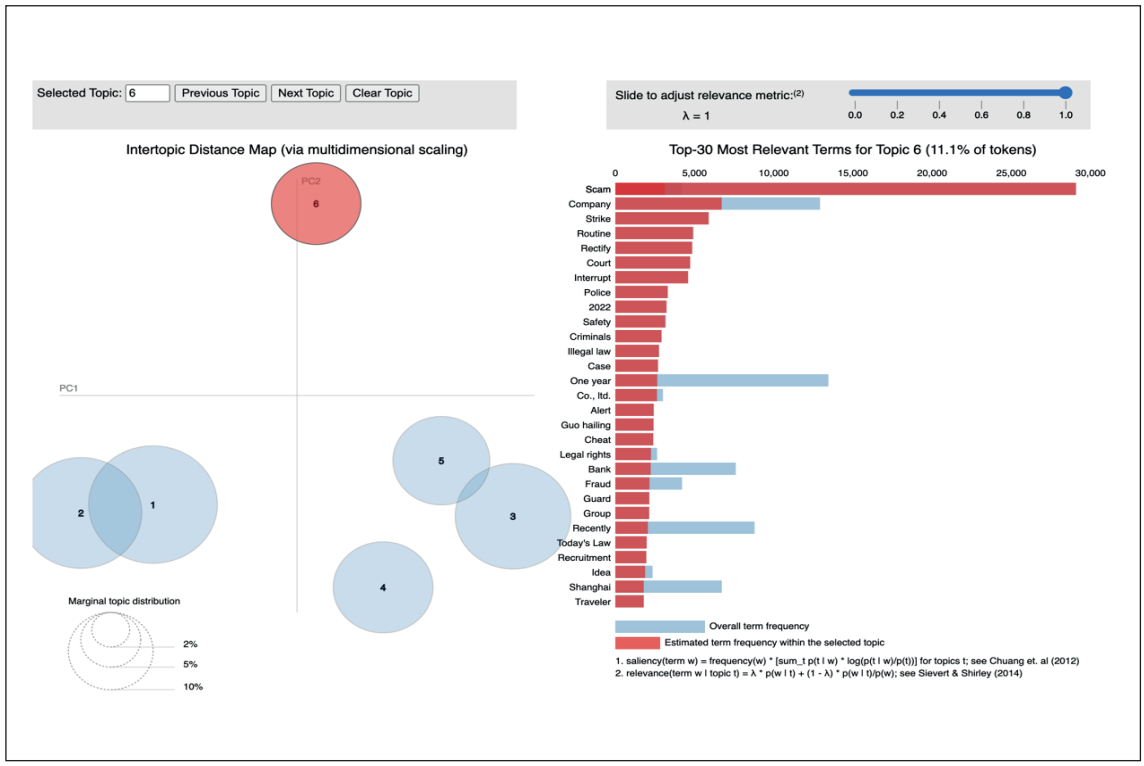
Source: own elaboration.

Figure 12
LDA topic 5 result



Source: own elaboration.

Figure 13
LDA topic 6 result



Source: own elaboration.

Table 1
Demographic of users

Variable	Label	Count	Proportion (in %)
Gender	male	148,338	55.34
	female	119,701	44.66
	unknown	13	0.00
Age group	< 18	5,435	2.03
	18–24	22,452	8.38
	25–34	38,264	14.27
	35–49	23,481	8.76
	50–59	9,104	3.40
	60–69	3,963	1.48
	≥ 70	1,706	0.64
	unknown	163,647	61.05
Total		268,052	100.00

Note: The 'unknown' age group represents the posts from users who have not specified their age. This group is included in the overall analysis but is excluded from the age-specific sentiment analysis.

Source: own elaboration.

Table 2
Proportion of negative sentiment by age cohort and gender

Age group	Total post count	Negative sentiment proportion (in %)
< 18	5,435	38.29
18–24	22,452	45.82
25–34	38,264	44.28
35–49	23,481	44.10
50–59	9,104	49.00
60–69	3,963	49.46
≥ 70	1,706	47.89
Gender		
Female	148,338	46.45
Male	119,701	43.77

Source: own elaboration.

Dekodowanie opinii publicznej na temat polityki emerytalnej w Chinach za pomocą przetwarzania języka naturalnego

Streszczenie

W starzejącej się populacji Chin odsetek osób powyżej 60. roku życia ma osiągnąć 28% do 2040 r. (WHO 2019). Wywiera to silny nacisk na system emerytalny i wymaga pilnych reform zapewniających jego długoterminową stabilność. Obecne badania dotyczące chińskiej polityki emerytalnej koncentrują się na trendach demograficznych oraz ramach regulacyjnych i pomijają analizę bieżących nastrojów społecznych. Podczas gdy techniki przetwarzania języka naturalnego (NLP) znalazły zastosowanie w takich obszarach jak zdrowie publiczne czy polityka środowiskowa (Abiola i in. 2023; Yang i in. 2023), w niewielu badaniach wykorzystano NLP do eksploracji opinii publicznej na temat polityki gospodarczej, w szczególności w Chinach. Ogranicza to możliwość opracowywania reform, które skutecznie odpowiadałyby na obawy społeczne i budowały zaufanie do systemu emerytalnego.

Celem niniejszego badania jest analiza nastrojów społecznych w reakcji na chińską politykę emerytalną w okresie od stycznia 2018 do sierpnia 2023 r., z wykorzystaniem ponad 260 tys. postów z platformy Sina Weibo. Zastosowanie zaawansowanych technik NLP, w tym analizy sentymentu oraz algorytmu ukrytego przydziału tematów (Latent Dirichlet Allocation, LDA), pozwoliło zidentyfikować kluczowe wątki, monitorować zmiany nastrojów w czasie oraz zbadać zróżnicowanie demograficzne i regionalne opinii publicznej. Analiza sentymentu umożliwia podział postów na pozytywne lub negatywne i śledzenie ich zmian w czasie, natomiast metoda LDA identyfikuje główne tematy odzwierciedlające zainteresowania społeczeństwa. Proces przetwarzania danych obejmował usuwanie szumów, a także ekstrakcję danych demograficznych, takich jak wiek, płeć oraz lokalizacja, co umożliwiło ujawnienie demograficznych i regionalnych wzorców nastrojów. Prezentowane podejście dostarcza cennych informacji dla decydentów, którzy muszą odpowiadać na zróżnicowane obawy społeczne w dynamicznie zmieniającym się systemie emerytalnym Chin.

Analiza ujawnia systematyczny wzrost negatywnych nastrojów od 2018 r., przy czym kwestie instytucjonalne, takie jak ogłoszenia w sprawie polityki rządowej i reformy, często powodowały nasilenie tendencji negatywnych. Istotne wydarzenia, na przykład stopniowe podnoszenie ustawowego wieku emerytalnego oraz wprowadzenie systemu Indywidualnych Kont Emerytalnych, spowodowały wyraźny wzrost negatywnego sentymentu. Trendy demograficzne wskazują, że wśród użytkowników zbliżających się do wieku emerytalnego nastroje negatywne są częstsze niż w młodszych grupach. Niektóre prowincje, np. Heilongjiang, charakteryzują się wyższym poziomem niezadowolenia, podczas gdy w Tybecie nastawienie jest mniej negatywne, co potwierdza, że warunki ekonomiczne oraz trwałość funduszy emerytalnych istotnie wpływają na nastroje społeczne. Wyniki te podkreślają złożoność czynników kształtujących opinie na temat polityki emerytalnej powiązane ze zróżnicowaną strukturą demograficzną i regionalną Chin.

Zastosowanie technik NLP umożliwiło przeprowadzenie szczegółowej analizy tematycznej przy użyciu LDA, która ujawniła sześć głównych tematów występujących w zbiorze danych. Temat 1 odnosi się do presji finansowej, w szczególności obciążeń ekonomicznych osób w wieku produkcyjnym,

zmuszonych do wspierania zarówno dzieci, jak i starszych członków rodziny. Powszechnie występujące w postach wyrazy, takie jak „wynagrodzenie", „pieniądze" czy „dziecko", odzwierciedlają finansowe niepokoje młodych dorosłych, balansujących między wydatkami osobistymi a obowiązkami rodzinnymi. Temat 2 dotyczy decyzji życiowych związanych z rolami rodzinnymi oraz stabilnością finansową i uwzględnia takie wyrazy, jak „małżeństwo", „emerytura" i „rodzice". W dyskusjach z tego obszaru często widać niechęć lub obawy dotyczące ciężarów finansowych wynikających z tradycyjnych obowiązków rodzinnych, co świadczy o szerszych oczekiwaniach społecznych.

Tematy 3–5 odnoszą się do instytucjonalnych aspektów polityki emerytalnej, m.in. systemu opieki społecznej, regulacji oraz reform napędzanych mechanizmami rynkowymi. Temat 3 dotyczy administracyjnych aspektów uczestnictwa w systemie emerytalnym zarówno w obszarach miejskich, jak i wiejskich. Z kolei temat 4 skupia się na systemie zabezpieczenia społecznego opartym na zatrudnieniu. Pojawiają się tu takie sformułowania jak „wiek emerytalny", „ubezpieczenie zdrowotne" oraz „bezrobocie". Temat 5 odnosi się do szerszych strategii finansowych i obejmuje takie pojęcia jak „fundusz", „rynek" oraz „odwrócona hipoteka dla seniorów". Dyskusje w tych obszarach cechuje formalny język, najprawdopodobniej zaczerpnięty z oficjalnych oświadczeń lub mediów, co sugeruje autorytatywny ton dyskursu publicznego w sprawach emerytalnych.

Szósty temat ujawnia systemową podatność, szczególnie w kontekście bezpieczeństwa oraz zgodności z prawem. Wyrazy takie jak „oszustwo", „przestępcy", „nielegalność" oraz „fraud" odzwierciedlają obawy społeczne przed oszustwami związanymi z emeryturami i podkreślają istotne zastrzeżenia dotyczące bezpieczeństwa inwestycji osobistych oraz wiarygodności ram instytucjonalnych. Częsty negatywny sentyment związany z tym tematem (ponad 54%) wskazuje na powszechne obawy dotyczące oszustw oraz integralności systemu emerytalnego. Łącznie tematy 3–6 sygnalizują silny sceptycyzm społeczeństwa wobec instytucjonalnych aspektów systemu emerytalnego, o czym świadczy konsekwentnie wysoki poziom negatywnego nastawienia w dyskusjach dotyczących przejrzystości, regulacji oraz zapobiegania oszustwom.

Wyniki badania podkreślają konieczność włączenia do reform emerytalnych skutecznej komunikacji publicznej oraz inicjatyw edukacyjnych, szczególnie w zakresie zapewnienia przejrzystości i bezpieczeństwa. Analiza wykazuje, że tematy związane z obowiązkami społecznymi i rodzinnymi (tematy 1 i 2) cechują się niższym odsetkiem negatywnych wypowiedzi, co sugeruje, że kwestie dotyczące odpowiedzialności osobistej spotykają się z większym zrozumieniem. Bardziej negatywne nastawienie w dyskusjach dotyczących rozwiązań instytucjonalnych świadczy o rozbieżności między oczekiwaniami społecznymi a komunikacją rządu, zwłaszcza w obszarach wymagających jasności i odpowiedzialności oraz skutecznych mechanizmów zapobiegania oszustwom.

Przeprowadzone badanie wnosi istotny wkład do finansów behawioralnych. Ilustruje, w jaki sposób heurystyki poznawcze, takie jak błąd dostępności oraz awersja do strat, wpływają na nastroje społeczne wobec reform emerytalnych. Błąd dostępności tłumaczy nasilone obawy związane z nagłaśnianiem takich kwestii jak oszustwa, natomiast awersja do strat wyjaśnia silne negatywne reakcje na zmiany polityczne postrzegane jako zagrożenie dla bezpieczeństwa finansowego, na przykład wydłużenie wieku emerytalnego. Integracja tych koncepcji z metodami NLP umożliwia ukazanie, jak zachowania napędzane heurystyką różnią się w zależności od czynników demograficznych i regionalnych, oraz oferuje nowatorskie podejście do analizy dyskursu publicznego na temat polityki społeczno-ekonomicznej.

Badanie to dostarcza również praktycznych wskazówek dla decydentów zarządzających reformą systemu emerytalnego w starzejącym się społeczeństwie. Zwalczanie regionalnych nierówności

wymaga podejścia odpowiedniego dla danego regionu, uwzględniającego zróżnicowanie ekonomiczne i demograficzne. Budowanie zaufania społecznego wymaga natomiast podejmowania aktywizujących działań, takich jak programy edukacyjne podkreślające korzyści płynące z reform emerytalnych oraz zapobieganie oszustwom. Integracja technik NLP i LDA w analizie mediów społecznościowych umożliwia uzyskanie szczegółowego obrazu nastrojów społecznych i identyfikuje główne obawy w rozległej i dynamicznej przestrzeni cyfrowej.

Kluczowym ograniczeniem niniejszego badania jest oparcie się na danych z platformy Weibo, co może wpływać na wyniki przez niedostateczną reprezentację starszych lub wiejskich populacji, które są mniej aktywne online. Ponadto choć metody NLP umożliwiają uchwycenie ogólnych trendów sentymentu, mogą nie oddawać subtelności kulturowych. Obserwacyjny charakter badania ogranicza analizę przyczynową, gdyż opiera się głównie na korelacjach; przyszłe badania powinny uwzględniać tradycyjne metody ankietowe oraz podejście podłużne, aby uzyskać bardziej reprezentatywny i przyczynowo-informacyjny obraz opinii publicznej.

Słowa kluczowe: polityka emerytalna, nastroje społeczne, przetwarzanie języka naturalnego, analiza sentymentu