

UAI JOURNAL OF ARTS, HUMANITIES AND SOCIAL SCIENCES

(UAIJAHSS)



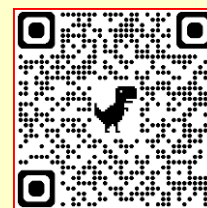
Abbreviated Key Title: UAI J Arts Humanit Soc Sci

ISSN: 3048-7692 (Online)

Journal Homepage: <https://uaipublisher.com/uaijahss/>

Volume- 2 Issue- 10 (October) 2025

Frequency: Monthly



The Mediating Role of Content Interest in the Relationship Between Temporal Persistence Factors and Reading Behavior of Sociological Books

Jiexuan Liu

Nanjing Normal University, Nanjing, Jiangsu, China

Corresponding Author: Jiexuan Liu

ABSTRACT

This study investigates the borrowing behavior of Category C (Sociology, Chinese Library Classification) books at Nanjing Normal University (2016–2024), using a sample of circulation data covering 13,117 book titles, 13,344 readers, and 59,198 loans, via a multi-method approach. For theme analysis, subject terms from the 6XX field (core for CNMARC/UNIMARC subject description) were English-translated, preprocessed (tokenization, word frequency counting), and semantically vectorized. Voyant Tools visualized the top 10 subject categories, 100 high-frequency terms, and cross-period clusters, revealing readers' thematic preferences. Linear regression took "Length of Borrowing Years" and "Category C9" as independent variables, and "Total Borrowings" as the dependent variable; model metrics confirmed these variables as significant direct predictors. Mediation effect analysis verified C9's significant mediating role between "Length of Borrowing Years" and "Total Borrowings." Decision tree analysis used "Renewal" as the dependent variable, and "Total," "Gender," "Persistence" as independent variables, identifying "Total" and "Persistence" as key renewal drivers. This study clarifies sociological book borrowing's temporal patterns, thematic preferences, and influencing factors, providing empirical support for library resource allocation and service optimization.

KEY WORDS: Sociology Books; Theme Analysis; Linear Regression; Mediation Effect Analysis; Decision Tree Analysis; Borrowing Behavior

1. Introduction

Reading behavior of sociological books—defined as sustained engagement with materials exploring themes like social structure, cultural identity, inequality, and community dynamics—does not operate in isolation. Instead, it is shaped by temporal persistence factors: long-term influences that foster consistent interaction with sociological content. These factors include repeated reading interventions, consistent resource access, enduring reading habits, and environmental support.

A burgeoning body of scholarship identifies content interest as a pivotal mediator in the relationship between these temporal persistence factors and reading outcomes—including sustained engagement with sociological literature, mastery of sociological concepts, and lifelong engagement with disciplinary reading. Within this context, content interest denotes readers' subjective affinity for sociological themes (e.g., racial equity, demographic aging, the societal implications of technology), culturally resonant content (e.g., texts examining diasporic identities or Indigenous social systems), and foundational disciplinary topics (e.g., social

stratification, institutional sociology). Even contextual temporal drivers—such as book bans drawing attention to controversial sociological content or pandemic-related concerns about societal resilience—spark content interest, motivating readers to seek out and persist in reading relevant sociological books (Ananthakrishnan et al., 2025; Busetta et al., 2024).

This research advances the field of sociological book borrowing behavior studies through four key innovations: First, it establishes a multi-method mechanistic framework, integrating temporal trend analyses (via “Lengths of Borrowing Years”), linear regressions (for direct predictors of total borrowings), and Bootstrap-based mediation analyses (to reveal C9’s mediating role between borrowing durations and total borrowings), moving beyond descriptions to explain why borrowing patterns emerge. Second, it adopts standardized metadata-driven thematic analyses—using 6XX fields with Voyant Tools’ semantic vectorization and cross-period visualizations—to identify top thematic categories, high-frequency terms, and evolutionary trends (e.g., the 2022–2024 shift to “technology integration”), ensuring objective, replicable insights. Third, it develops interpretable renewal predictions via the CRT algorithm (a “white-box” model) with predictors like “Total Borrowings” and “Persistence,” validating performances across three periods (2016–2018, 2019–2021, 2022–2024) and highlighting the risks of excluding time-varying factors. Fourth, it contextualizes findings with the CLC, detailing the contents of C Category sub-categories (e.g., C0, C9) and linking C9 to borrowing/mediation outcomes, explaining C9’s dominant borrowing volumes and bridging abstract data with library collection logics for actionable resource allocations.

This study carries theoretical, methodological, and practical value for library and information science, sociology of education, and reader behavior research: theoretically, it enriches the reader behavior research framework in sociology by identifying the dual trend of stable core themes (e.g., research methods, social structure) and evolving emerging topics (e.g., technology integration, aging, feminist studies) in readers’ attention, and through mediation effect analysis, empirically verifies that C9 (Sociology & Demography) partially mediates the relationship between “Length of Borrowing Years (LBY)” and “Total Borrowings (TB)”, revealing the direct and indirect (via preference for high-demand disciplines) impacts of long-term borrowing persistence on total borrowings and overcoming prior overreliance on descriptive correlations; methodologically, it promotes standardization and integration by adopting standardized metadata-driven thematic analysis (ensuring objectivity and replicability) and integrating multi-method approaches (temporal trend analysis, linear regression, mediation analysis, CRT decision trees) to explore different reader behavior dimensions, while its temporal validation of prediction models (2016–2018, 2019–2021, 2022–2024) addresses the oversight of ignoring time-varying factors (e.g., declining predictive role of “Gender” later) in previous renewal prediction studies; practically, it provides actionable guidance—for academic libraries, it aids collection development by identifying C9 as the most borrowed subcategory and mapping high-frequency themes to prioritize high-demand acquisitions and optimize collections, and for higher education institutions and sociological departments, it informs curriculum and teaching resource development by aligning materials with readers’ thematic preferences to connect academic inquiry with real-world applications.

2. Previous Research

2.1 Content Interest as a Mediator of Engagement,

Motivation, and Temporal Persistence

Temporal persistence factors like targeted reading interventions or repeated motivational support rely on content interest to translate into sustained sociological reading behavior. For instance, Chilufya et al. (2025) developed a social robot (“BookBot”) that stimulates fourth-grade students’ reading interest through sociologically relevant interactions—including book content discussions, character analysis tied to social themes, and personalized recommendations. Their findings show that content interest, sparked by the robot’s interventions, enhances reading engagement, indirectly linking this temporal persistence factor (robot-mediated repeated interaction) to consistent reading behavior. This aligns with Wu et al. (2023), who found that for high school students, reading motivation (a temporal driver of persistence) predicts satisfaction with sociological books via self-efficacy—and content interest acts as a key driver of self-efficacy, bridging motivation to sustained reading.

Content interest also addresses barriers to engagement, particularly for readers facing disciplinary or language challenges. Nausa et al. (2024) studied Colombian master’s students struggling with access to sociological research (due to low English proficiency and poor reading skills) and found that matching reading content to students’ disciplinary interests (e.g., social inequality, community development) eliminated these barriers. By aligning materials with pre-existing content interest, temporal factors like repeated exposure to accessible texts translated into consistent engagement. Similarly, Hiebert (2023) emphasized that reading amount (a direct measure of temporal persistence) is critical for mastering sociological disciplinary language and concepts—but content interest reinforces this persistence, ensuring readers maintain engagement long enough to build familiarity with specialized themes.

Cultural relevance emerges as a sub-dimension of content interest that strongly mediates engagement. Jones and Lynch (2023) examined Black students’ after-school reading of African American literature (featuring sociological themes of race, identity, and systemic oppression) and found that content interest in culturally resonant topics mediates sustained reading and comprehension of sociological concepts. For these students, temporal factors like repeated after-school reading time only led to meaningful engagement when materials reflected their cultural experiences—highlighting content interest’s role in validating readers’ identities and sustaining persistence.

2.2 Environmental and Contextual Factors: Home, School, and Digital Platforms

The home literacy environment—defined by factors like book quantity, parental involvement, and consistent access to sociological materials—acts as a temporal persistence factor, with content interest mediating its impact on reading behavior. Mohammed et al. (2023) studied caregiver-child reading dynamics and found that sustained caregiver-child interaction with sociological books correlates with children’s literacy development—but this relationship is fully mediated by content interest in social themes. Specifically, home access to sociological books (a temporal resource) only drives engagement if children find the content (e.g., stories about community, social roles) interesting.

For multilingual populations, HLE’s influence is further refined by content interest in cultural and linguistic contexts. Aleksic and Durus (2025) examined multilingual preschoolers and found that HLE factors (book quantity, parental interest in sociological themes) predict reading skills—but content interest in multilingual social contexts (e.g., stories about cultural diversity, multilingual community life) mediates this link. For these children, temporal

persistence (e.g., daily reading time with caregivers) translates into skill development only when materials align with their linguistic and cultural interests.

2.3 Digital and Technological Platforms: Gamification and Algorithmic Recommendations

Digital tools represent a modern class of temporal persistence factors, as they enable repeated, scalable access to sociological content—and content interest mediates their effectiveness. Yang et al. (2023) developed gamified portfolios for students, featuring goal-setting and genre rewards focused on sociological genres (e.g., social justice, urban studies). Their results show that content interest mediates gamified activities (a temporal factor of repeated, reward-driven interaction) and persistent exploration of sociological books—with interested students more likely to maintain engagement beyond initial gamification prompts.

Algorithmic recommendation systems, which provide ongoing access to tailored content, also rely on content interest to drive persistence. Tran et al. (2023) studied social relations/rating-based algorithms that recommend sociological books and found that content interest mediates algorithm-driven suggestions (a temporal factor of continuous, personalized exposure) and consistent reading. Readers who received recommendations aligned with their interests (e.g., aging studies for those concerned with community dynamics) were far more likely to sustain engagement than those receiving generic suggestions—confirming content interest's role in translating algorithmic support into long-term behavior.

2.4 Content Interest, Special Populations, and Lifelong Sociological Reading

Early exposure to sociological content and content interest lay the groundwork for lifelong reading behavior, with temporal factors like parental reading or cultural resource access mediated by interest. Shigemasa et al. (2024) studied infant-parent reading and found that parental reading of sociological-themed books (e.g., stories about social structures, community roles) to toddlers improves children's early comprehension of social concepts—and early content interest in these topics predicts lifelong engagement with sociological books. For these children, temporal factors like repeated parental reading only shape long-term habits if they spark interest in social themes during early development.

Indigenous children represent a population where cultural relevance in content interest is particularly critical. Sianturi and Hurit (2024) examined decolonized picture books (integrating Indigenous cultural experiences and sociological themes like cultural identity, community governance) for Indigenous children and found that content relevance (tied to cultural background) mediates repeated reading and sustained engagement. For these children, temporal factors like regular access to culturally specific books translate into persistence only when the content reflects their lived experiences—underscoring the need for culturally responsive content to activate interest.

For older adults, content interest mediates temporal factors like library access or leisure reading time and engagement with sociological books—often with spillover effects on well-being. Otani et al. (2025) studied Japanese older adults and found that library book quantity (a temporal factor of consistent resource access) reduces functional disability risk—but this relationship is mediated by content interest in aging and community themes. Older adults who accessed sociological books aligned with their interests (e.g., studies on intergenerational relationships, community support) maintained more consistent reading habits, driving health benefits.

Srivastava, P. and Srivastava, M. (2023) found that Indian elderly adults who read sociological books on aging and intergenerational dynamics experience reduced loneliness—but only if they have content interest in these topics. For these readers, temporal factors like daily leisure reading time only improve well-being when paired with interest-aligned content.

2.5 Cross-Outcome Links: Content Interest, Temporal Persistence, and Beyond Reading Behavior

Content interest does not only mediate temporal factors and reading behavior—it also bridges sustained sociological reading to broader outcomes like cognitive function, emotional well-being, and cultural participation.

Early content interest in sociological themes mediates the link between childhood temporal factors (e.g., book access) and later cognitive health. Tani et al. (2024) found that a book-rich childhood (a temporal factor of sustained early resource access) fosters content interest in sociological content—and this interest mediates the relationship between childhood book environment and later cognitive function, particularly for older adults with low education. For these individuals, early exposure to sociological books only preserved cognitive function if it sparked interest, ensuring they maintained reading habits into adulthood.

During periods of stress, content interest in sociological themes mediates temporal persistence (e.g., sustained reading) and emotional resilience. Busetta et al. (2024) studied Italian university students during the pandemic and found that sustained reading of sociological books on community resilience mitigates anxiety—but this effect is mediated by content interest. Students who were interested in community-focused themes maintained more consistent reading habits, leading to greater emotional benefits.

Historical and regional studies confirm that content interest has long mediated temporal persistence factors and sociological reading behavior, with context-specific themes shaping engagement. Agafonova (2023) examined early 20th-century Russian tram passengers, who read sociological books (focused on urban issues like worker strikes, community dynamics) during daily commutes. The study found that content interest in these context-relevant themes mediated regular tram time (a temporal factor of daily, repeated reading opportunities) and sociological book choice—passengers consistently selected materials aligned with their interests in urban social issues, translating commute time into sustained engagement.

Similarly, Garcia-Gonzalez et al. (2023) studied historical CRA (Reading Promotion Centers) programs in Chilean school libraries, which historically focused on social justice (a core sociological theme). They found that content interest mediates library access (a temporal factor of sustained resource availability) and student engagement—though this mediation has diminished in modern times, as programs now prioritize academic outcomes over social justice themes. This historical shift highlights how content interest's mediating role depends on alignment between temporal factors (e.g., library programming) and readers' interests in contextually relevant sociological themes.

Existing studies on sociological book borrowing behaviors in libraries have notable limitations. First, they overly rely on single-dimensional analyses (e.g., descriptive statistics, simple correlations), failing to explore the temporal dynamics of borrowings (e.g., the persistence of long-term reading interests) or the mechanistic relationships between influencing factors (e.g., the indirect effects of borrowing durations via specific book categories).

Second, they inadequately integrate metadata and thematic analyses—rarely using standardized metadata (e.g., CNMARC/UNIMARC 6XX subject term fields) and instead relying on subjective title classifications or free-text keywords, which lead to inconsistent subject delineations and the inability to capture fine-grained thematic clusters or cross-period interest trends. Third, they pay limited attention to mediation mechanisms and interpretable renewal predictions; few studies examine indirect variable pathways, and renewal predictions often use opaque “black-box” algorithms or exclude time-varying factors, reducing generalizability. Fourth, they lack contextualization with standardized classification systems (e.g., the Chinese Library Classification, CLC), failing to clarify how sub-categories (e.g., C9) differ in influencing borrowings or why specific sub-categories dominate borrowing volumes.

The primary purpose of this study is to systematically explore the characteristics, dynamics, and influencing mechanisms of sociological book borrowing behavior at Nanjing Normal University, with four specific objectives:

First, to identify the thematic characteristics of sociological books favored by readers, and analyze the temporal evolution of high-frequency subject terms across three periods (2016–2018, 2019–2021, 2022–2024)—clarifying both the stable core themes and phased emerging topics in readers’ disciplinary interests.

Second, to unpack the key factors driving readers’ total sociological book borrowings, and verify the mediating role of C9 (CLC subcategory “Sociology and Demography”) in the relationship between “Length of Borrowing Years” and “Total Borrowings”, revealing the direct and indirect mechanisms shaping borrowing outcomes.

Third, to develop an interpretable prediction model for readers’ book renewal behavior using the Classification and Regression Trees (CRT) algorithm, and validate the model’s performance across different time periods—assessing how factors like “Total Borrowings,” “Gender,” and “Persistence” influence renewal decisions and the risks of excluding time-varying predictors.

Fourth, to translate the above analytical findings into actionable insights for academic libraries and higher education institutions—providing evidence-based guidance for optimizing sociological book collection allocation, refining reader services, and aligning

teaching/learning resources with readers’ evolving disciplinary needs.

3. Data and Method

The research sample consists of the circulation data of category C (Sociology) books from Nanjing Normal University between 2016 and 2024, covering 13,344 readers, 13,117 types of books, and a total loan volume of 59,198.

3.1 Book Theme Analysis

Field 6XX constitutes the core field group for describing the subject content of documents in both CNMARC (China Machine-Readable Cataloging Format) and UNIMARC (Universal Machine-Readable Cataloging Format). Its primary function is to serve as a “subject access point,” enabling three key capabilities: (1) accurately revealing the document’s subject matter with greater granularity than classification fields; (2) supporting multi-dimensional retrieval (encompassing disciplinary subjects, personal names, geographical locations, institutions, and document titles); and (3) facilitating cross-classification association (aggregating documents on the same subject that may fall under different classification numbers). Users can search through library catalog systems via “subject terms”. Publishers and libraries use Field 6XX to unify subject descriptions during cataloging, avoiding metadata inconsistencies caused by free-text vocabulary.

The visualization is generated via Voyant Tools, a web-based platform dedicated to digital text analysis. Its development relies on a series of interrelated analytical procedures. First, the subject terms of books in Category C borrowed by readers are translated into English and input into the text box. Initially, the input digital texts undergo preprocessing, which involves tokenization (to segment the text into individual words) and word frequency counting. Subsequently, semantic vectorization techniques—such as word embedding or topic modeling—are applied to convert these words into high-dimensional semantic vectors. These vectors capture the semantic relationships between words, with words of similar meanings positioned in closer proximity within the vector space. Through text analysis, the top 10 subject term categories are mined, and the top 100 high-frequency subject terms as well as subject term clusters (including the top 50 high-frequency terms) across three time periods are visually presented.

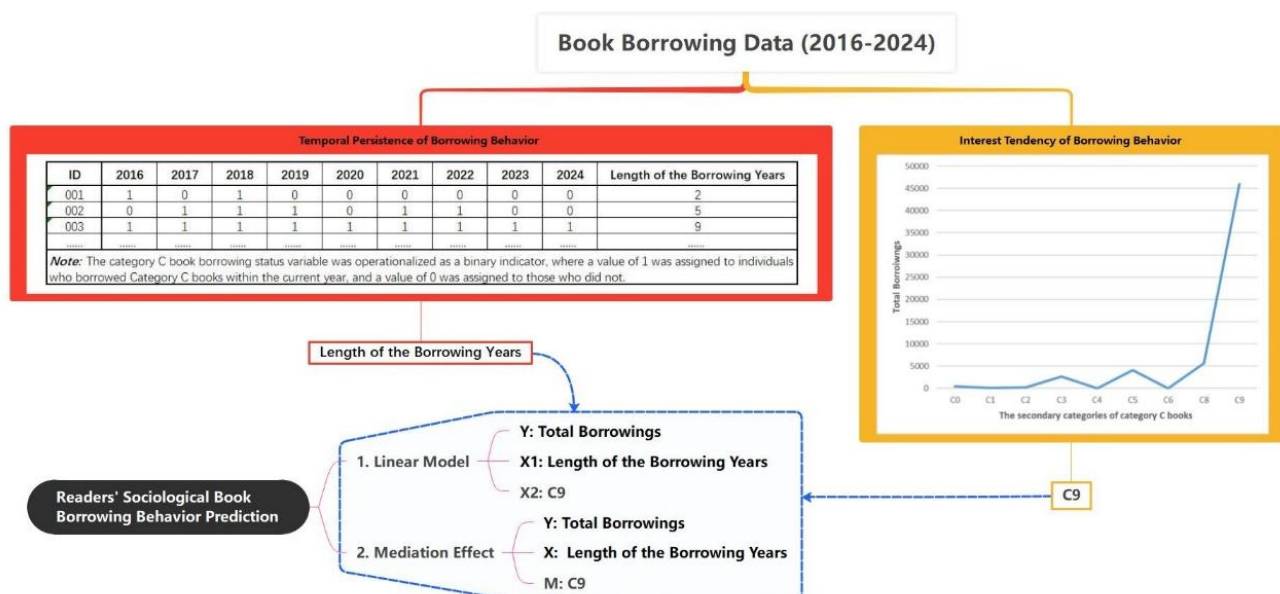


Figure 1. Research design

Figure 1 illustrates the research design centered on book borrowing data from 2016 to 2024. It first delves into the temporal persistence of borrowing behavior, utilizing a binary indicator (assigning 1 for individual book borrowing and 0 otherwise) and examining the "Length of the Borrowing Years". Then, it analyzes the interest tendency of borrowing behavior via a line chart. Finally, for predicting students' sociological book borrowing behavior, two approaches are employed: a linear model with "Total Borrowings" as the dependent variable and "Length of the Borrowing Years" and "C9" as independent variables; and a mediation effect analysis where "C9" mediates the relationship between "Length of the Borrowing Years" and "Total Borrowings".

Comparing these two algorithms, the linear model offers a straightforward way to assess the direct impacts of borrowing duration and "C9" on total borrowings, being simple and interpretable. In contrast, the mediation effect analysis uncovers the indirect pathway through which "C9" transmits the influence of borrowing years on total borrowings, providing a more nuanced understanding. The significance lies in that the linear model helps identify key direct predictors, while the mediation analysis enriches the comprehension of the underlying mechanisms, together offering a comprehensive view of factors shaping students' book - borrowing behavior and aiding in more targeted interventions or resource allocation in educational settings.

Category C in the Chinese Library Classification (CLC), titled "General Works of Social Sciences", is a classification summary of basic and universal theories and methods in the field of social sciences. Its secondary categories follow a logical system of "from basic theory to practical application, and from the whole to the part". The specific categories and their core contents are as follows:

The secondary category with the classification code C0 is "Theories and Methodologies of Social Sciences", which covers the basic theories of social sciences (such as the objects, nature, and system of social sciences), research methods (such as survey methods and statistical methods), and the history of academic thought and its schools.

The secondary category C1 is "Current Situation and Overview of Social Sciences", encompassing the overall development status of social sciences in various countries/regions, research trends, academic institutions (e.g., academies of social sciences, university research institutes), academic organizations, academic journals, and literature reviews.

The secondary category C2 is "Institutions and Organizations of Social Sciences", which includes the organizational structure, functions, activities, and directories of specialized institutions in the field of social sciences (e.g., social science research institutes, policy research centers) and academic organizations (e.g., sociology societies, economics associations).

The secondary category C3 is "Research Methods of Social Sciences", focusing on specific methods and tools for social science research, including quantitative research (e.g., data analysis, model construction), qualitative research (e.g., interviews, case studies), interdisciplinary research methods, and research design.

The secondary category C4 is "Education and Popularization of Social Sciences", covering the education system in the field of social sciences (e.g., the setup of social science majors in universities, curriculum and teaching materials), popular science activities (e.g., the promotion of social science knowledge, public education programs), and related resources.

The secondary category C5 is "Series, Collected Works, and Serial Publications of Social Sciences", which includes collected literature in the field of social sciences, such as series (e.g., "Frontier Series of Social Sciences"), collected papers, yearbooks (e.g., Yearbook of Chinese Social Sciences), and serial publications like journals.

The secondary category C6 is "Reference Books of Social Sciences", referring to reference tools in the field of social sciences, such as dictionaries (e.g., Dictionary of Social Sciences), encyclopedias (e.g., Encyclopedia of China: Social Sciences Volume), handbooks, indexes, and catalogs for retrieval and reference purposes.

The secondary category C7 is "Information Retrieval Tools of Social Sciences", specifically tools used for retrieving social science literature, including guides to literature databases, retrieval strategies, citation indexes (e.g., tools related to the Social Sciences Citation Index - SSCI), and handbooks on literature classification and indexing.

The secondary category C8 is "Statistics", which involves statistical theories and methods applicable in the field of social sciences, including descriptive statistics, inferential statistics, regression analysis, sampling surveys, the application of statistical software (e.g., SPSS, Stata), and specialized statistical techniques in various social science disciplines (e.g., economics, sociology).

The secondary category C9 is "Sociology and Demography". Sociology here covers areas such as social structure, social stratification, social mobility, social organizations, and social issues; demography includes population size, structure, distribution, fertility and aging, and population policies, as well as related sub-fields (e.g., medical sociology, rural sociology). It can be seen from Figure 1 that the borrowing volume of C9 books is significantly higher than that of books in other secondary categories.

3.2 Linear Regression Analysis

In the top menu bar of SPSS, select Analyze → Regression → Linear. This opens the "Linear Regression" main dialog box, where variables and model parameters are configured. "Total Borrowings" is the dependent variable being predicted. "Length of the Borrowing Years" and "C9" are the independent variables. "Forward Stepwise" method is used. This method starts with no predictors and iteratively adds the most statistically significant one at each step until no more significant predictors remain, aiming to build an optimal model efficiently.

The "Model Summary" table serves as the core output of this forward stepwise linear regression analysis, and its indicators must be systematically interpreted to evaluate the validity, explanatory power, and predictive accuracy of the optimal model generated. First, R (the multiple correlation coefficient) quantifies the strength of the linear relationship between the predicted values of the dependent variable "Total Borrowings (TB)" and its observed values; a value closer to 1 indicates a stronger consistency between the model's predictions and actual data, reflecting a more reliable predictive trend.

Second, R Square (R^2) represents the proportion of variance in TB that can be explained by the independent variables (IVs)—"Length of the Borrowing Years (LBY)" and/or "C9"—included in the final model. For instance, an R^2 value of 0.35 means 35% of the variability in TB is accounted for by the IVs, while the remaining 65% is attributed to unmeasured factors or random variation; a higher R^2 generally indicates that the model captures more of the key drivers of TB.

Third, Adjusted R Square is a corrected version of R^2 that adjusts for

the number of IVs in the model. Unlike R^2 , which inherently increases (or remains unchanged) when additional variables are added—even if those variables are non-significant and do not meaningfully improve model fit—Adjusted R Square penalizes the inclusion of redundant predictors. This makes it a more conservative and accurate measure of model performance, especially when comparing models with different numbers of IVs, as it avoids overestimating the model's explanatory power.

Fourth, R Square Change reflects the incremental increase in R^2 when a new IV is added to the model during each step of the forward selection process. A statistically significant R Square Change (typically determined by a p-value < 0.05) at a specific step indicates that the newly added IV contributes unique, meaningful explanatory power to the model. For example, a significant R Square Change at Step 1 would mean the first IV (either LBY or C9, depending on statistical significance) explains a substantial amount of additional variance in TB that was not captured by the intercept-only model.

Finally, Standard Error of the Estimate (SEE) measures the average distance between the observed TB values and the values predicted by the model. It quantifies the magnitude of prediction error: a smaller SEE indicates that the model's predictions are more closely aligned with actual TB values, signifying higher predictive accuracy. Conversely, a larger SEE suggests greater variability between predicted and observed values, indicating potential limitations in the model's ability to forecast TB reliably.

3.3 Mediation Effect Analysis

In the Process dialog box, specify "Length of Borrowing Years (LBY)" as the independent variable (X), "Mediator Variable C9" as the mediator variable (M), and "Total Borrowings (TB)" as the dependent variable (Y). Use the Bootstrap method to estimate the indirect effect because it does not rely on the normal distribution assumption. In the "Bootstrap Samples" option, set the number of resampling times 5000. In the "Confidence Level" option, set the confidence interval, usually 95%. After completing the above settings, click "OK" to run the analysis. SPSS will output a series of results, including path coefficients, significance levels, R-squared values, and Bootstrap confidence intervals.

Evaluate the overall fit of the model through the R-squared value. The R-squared value represents the proportion of the variance of the dependent variable explained by the independent variable and the mediator variable. A larger R-squared value indicates a better fit of the model. Check the significance levels (p-values) of the path coefficients of $X \rightarrow M$, $M \rightarrow Y$, and $X \rightarrow Y$. If the p-value is less than 0.05, it indicates that the corresponding path is significant, that is, there is a significant linear relationship between the variables. Focus on the point estimate and 95% confidence interval of the indirect effect. If the 95% confidence interval does not include 0, it indicates that the mediation effect is significant, that is, the independent variable X has an impact on the dependent variable Y through the mediator variable M.

3.4 Decision Tree Analysis

The dependent variable is "Renewal", and the independent variables include "Total", "Gender", and "Persistence". Cross-validation is used, with the maximum tree depth set to 5, the minimum number of cases in the parent node being 100, and the minimum number of cases in the child node being 50. SPSS uses the "CRT" (Classification and Regression Trees) algorithm, which is suitable for both categorical and continuous dependent variables. The decision tree model generates core analytical data across model

performance, node structure, and variable importance.

The analytical data generated by the aforementioned decision tree model collectively address three critical research questions that underpin the model's validity and practical utility for "Renewal" prediction. First, the model's performance metrics—including overall accuracy, cross-validation accuracy, and class-specific indicators such as sensitivity and precision—provide empirical evidence to answer whether the model can reliably predict "Renewal" outcomes and generalize to unseen data. Second, the structural data of the tree's nodes (e.g., root node benchmarks, split rules for intermediate nodes, and "Renewal" rate distributions in terminal nodes) elucidate the sequential and conditional mechanisms through which independent variables (i.e., "Total", "Gender", and "Persistence") drive "Renewal" behavior, rendering the model's predictive logic interpretable rather than opaque. Third, the relative importance scores assigned to each independent variable clarify which factors exert the most substantial influence on "Renewal" (e.g., "Total" typically emerging as the primary driver, followed by "Persistence"), thereby prioritizing variables for actionable analysis. Together, these attributes elevate the model beyond a mere predictive tool: by resolving the three core questions, it serves as a robust guide for data-driven decision-making in practical scenarios—for instance, enabling the design of targeted marketing strategies that focus on users with high "Total" values and strong "Persistence", who are empirically more likely to exhibit "Renewal" behavior.

4. Result

4.1 Top 10 Topics

4.1.1 Sociology Core Disciplines

As the foundational branches of sociology, these core disciplines construct the theoretical framework and research paradigm of the entire discipline, and provide essential methodological and analytical tools for exploring social phenomena: Institutional sociology, sociological detection procedures research, American social studies, rural studies, organizational differences, etc.

4.1.2 Literature, Art & Philosophy

This sub-topic integrates three interrelated fields—literature, art, and philosophy—exploring the expression of human thoughts, emotions, and values through literary works, artistic practices, and philosophical reflections. It covers both theoretical research and practical forms, reflecting the diversity of human cultural creation and spiritual exploration, including philosophical analysis of novels, world literature, social literature, prose, philosophy, calligraphy binding, legal stories, etc.

4.1.3 Higher & Vocational Education

This theme focuses on the development, reform, and practical application of higher education and vocational education systems, aiming to bridge the gap between education and social needs, and cultivate talents with professional competence and comprehensive quality. It covers both theoretical exploration of educational models and practical research on teaching methods, including higher education in anthropology, vocational studies, modern education, secondary education, secretarial studies, etc.

4.1.4 Academic Exchange

This theme emphasizes the importance of academic communication and knowledge sharing in promoting academic progress, focusing on the forms, channels, and impacts of academic exchange, as well as the dissemination and innovation of academic achievements. It covers both traditional and modern academic exchange methods, reflecting the global and interdisciplinary trends of current academic

research, including academic conferences, academic works, contemporary thought, postmodernist studies, application technologies, etc.

4.1.5 Specialized Disciplines

This theme covers a variety of specialized disciplines that play important roles in social development and academic research, each with its unique theoretical system and research methods, providing professional support for solving specific social problems and promoting interdisciplinary development, including economic geography, population theory, management science, developmental design, development policy, institutional time management, etc.

4.1.6 Cultural & Ethnic Research

This theme focuses on the exploration of cultural diversity and ethnic characteristics, studying the formation, evolution, and mutual influence of different cultures and ethnic groups, aiming to promote cultural heritage, ethnic harmony, and cross-cultural understanding. It covers both theoretical research on culture and ethnicity and practical research on cultural protection, including Chinese cultural studies, ethnic studies, culture, ancient issues, national studies, etc.

4.1.7 National & Regional Studies

This theme focuses on the study of different countries and regions around the world, exploring their social, economic, cultural, and political characteristics, as well as the differences and connections between them. It aims to provide a comprehensive understanding of global regional development patterns and promote international cooperation and regional stability, including China, United States, Western countries, Japan, South Korea, province, urban, southern regions, etc.

4.1.8 Research Methods & Statistical Tools

This theme provides essential methodological support for academic research, focusing on the development, application, and innovation of research methods and statistical tools, aiming to improve the scientific, accuracy, and reliability of research results. It covers both qualitative and quantitative research methods, as well as the practical application of statistical software, including survey statistics, research statistics, sampling and order control, SPSS, consumer questionnaires, evaluation studies, software, etc.

4.1.9 Psychology & Social Security

This theme combines the study of psychology and social security,

exploring the psychological factors affecting individual well-being and social stability, as well as the role of social security systems in protecting individual rights and promoting social equity. It covers both theoretical research and practical policy analysis, including political psychology, research on relationships, public, historical networks, construction studies, campus, security, risk, lie, welfare history, etc.

4.1.10 Sexual & Gender Studies

This theme focuses on the exploration of sexual behavior, sexual culture, gender identity, and gender equality, challenging traditional gender norms and promoting the understanding and respect of diverse gender and sexual orientations. It covers both theoretical research on gender and sexuality and practical advocacy for gender equity, reflecting the progress of social concepts and the pursuit of human rights, including sexual studies, introduction to sexology, homosexuality studies, women's changes, etc.

4.2 Top 100 Subject Words

Among the top 100 high-frequency subject words in the three time periods of 2016-2018, 2019-2021, and 2022-2024 (see Figures 2, 3, and 4), there are 86 words that appear in all three time periods (accounting for 86%). Among these 86 words:

There are 27 words with a total frequency of ≥ 1000 , which are: research, social, china, sociology, education, higher, institutions, management, science, studies, statistical, analysis, psychology, methods, history, modern, anthropology, organizational, public, relations, work, statistics, national, urban, leadership, application, sociological.

There are 59 words with a total frequency of < 1000 , which are: world, human, cultural, behavior, theory, population, survey, software, issues, culture, countries, thought, relationship, family, interpersonal, western, philosophy, rural, literature, women's, sexual, introduction, group, works, knowledge, procedures, life, Chinese, gender, development, impact, psychological, sexology, marriage, choice, basic, career, policy, opinion, relationships, United States, internet, humanities, love, decision, time, Japan, SPSS, literary, talent, ancient, community, models, province, art, methodology, anthologies, France, nation, behavioral, systems, Qing.

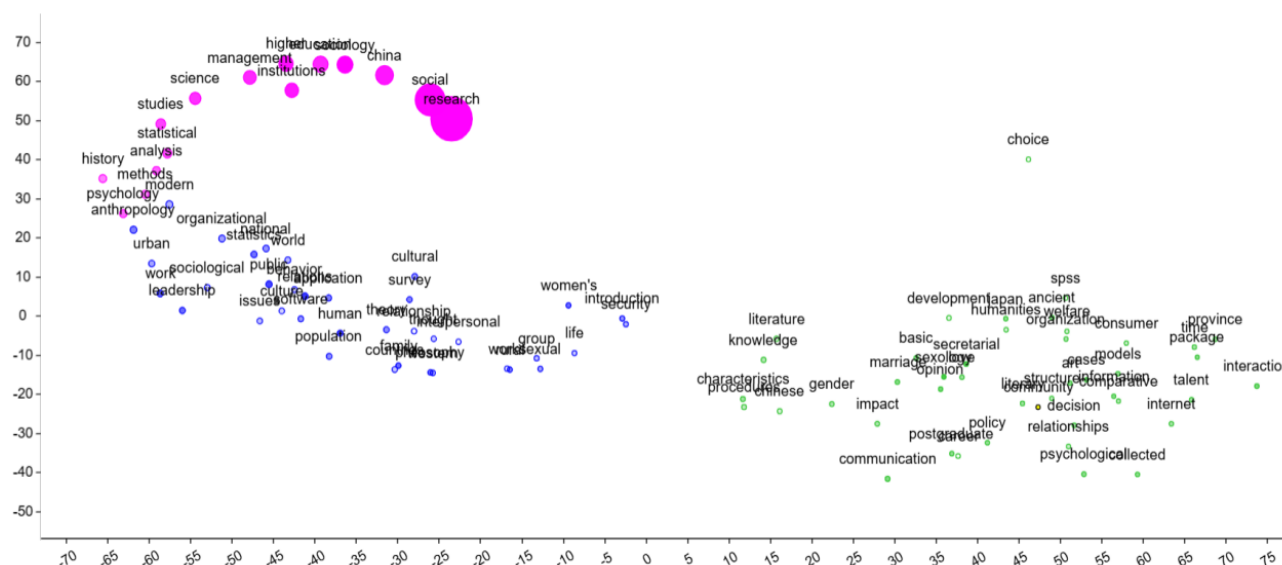


Figure 2. Top 100 subject words (2016-2018)

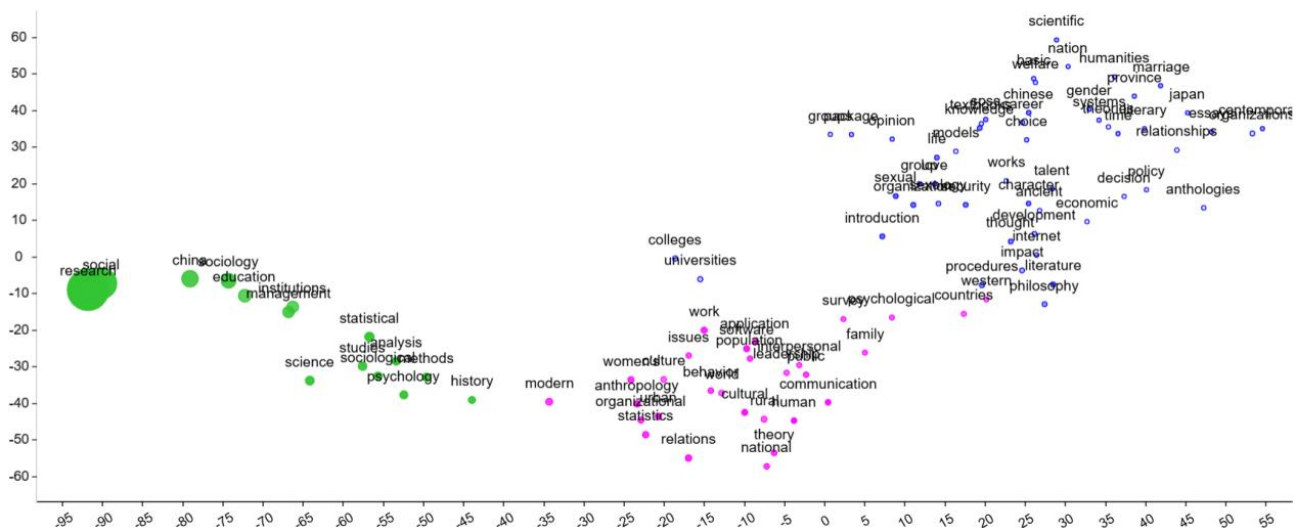


Figure 3. Top 100 subject words (2019-2021)

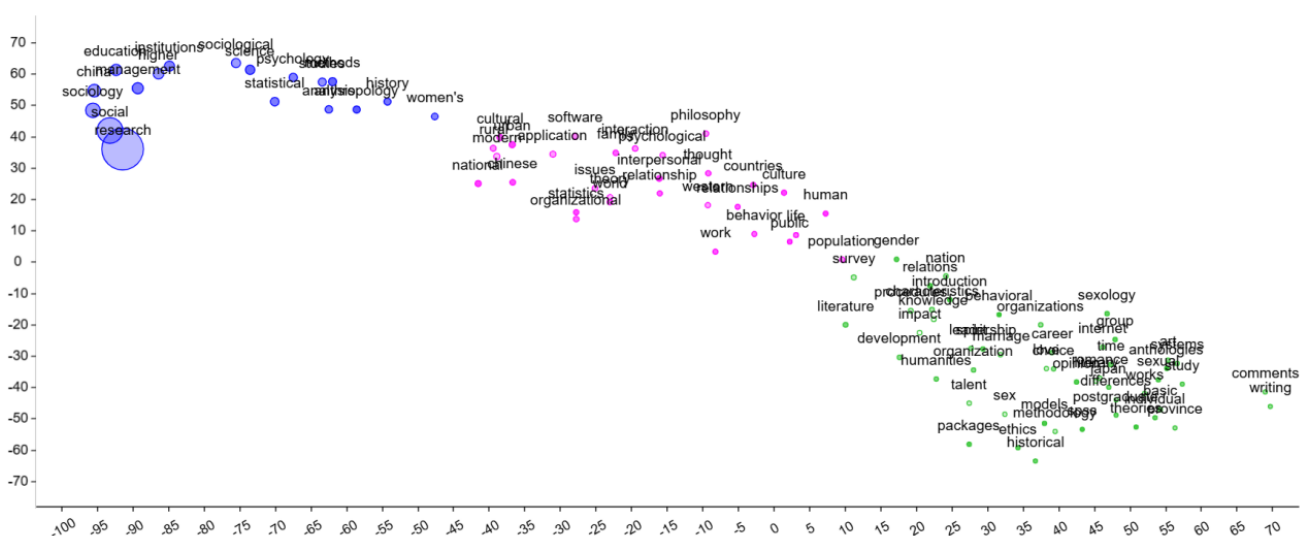


Figure 4. Top 100 subject words (2022-2024)

The top 100 subject high-frequency words across the three time periods (2016–2018, 2019–2021, and 2022–2024) reveals a strong concentration of cross-period core themes in sociology books. This high degree of overlap indicates that readers' attention to core sociological themes has remained stable over nearly a decade, primarily covering research methods and paradigms, social structure and institutional analysis, education and social development, interdisciplinary connections, cultural and behavioral studies, social groups and identity, and global and regional contexts.

Among the high-frequency words across the three time periods of 2016–2018, 2019–2021, and 2022–2024, there are 31 words that appear in only one period. Single-period unique words reveal a shift from "practical specialization" (2016–2018) and "educational text focus" (2019–2021) to "technology integration, social issues, and opinion dynamics" (2022–2024). This trend reflects the adaptability of sociology to social changes and the expanding scope of reader interest in response to contemporary challenges:

11 words appearing only in 2016-2018 (11%): These words are unique to the 2016–2018 period and did not appear in the subsequent two periods. They mainly consist of specialized terminology in specific fields and low-frequency words, covering areas such as secretarial work, leisure, and consulting, such as "secretarial" (82

occurrences), "leisure" (39 occurrences), and "consulting" (42 occurrences). They also include general domain words like "information" (47 occurrences) and "cases" (47 occurrences).

5 words appearing only in 2019-2021 (5%): The unique words in this period focus on educational institutions and text types, with core education-related vocabulary such as "universities" (110 occurrences) and "colleges" (110 occurrences). There are also text-related words like "essays" (38 occurrences), character-related words like "character" (61 occurrences), and change-related words like "changes" (28 occurrences). Overall, they reflect the period's phased focus on educational scenarios and text analysis.

15 words appearing only in 2022-2024 (15%): The unique words in this period are the most numerous and cover a wider range of fields, including emerging areas such as technology ("technology", 19 occurrences), social issues ("aging", 19 occurrences; "feminist", 18 occurrences), and individual-related terms ("individual", 23 occurrences; "youth", 18 occurrences). They also include newly emerging focuses like "network" (18 occurrences), "ethics" (16 occurrences), and opinions ("criticism", 19 occurrences; "comments", 22 occurrences). Overall, they reflect the addition and extension of recent research themes.

4.3 Theme Cluster

Figures 5, 6, and 7 identify the thematic word clusters and the top 50 high-frequency words. The cyan - colored "Bin" in the figures represents clustering. In the context of sociological book topic analysis, these bins group together terms that have strong thematic connections. In different time periods (2016 - 2018, 2019 - 2021, 2022 - 2024), the cyan bins aggregate sociological concepts and

topics that share similar research focuses or semantic relationships, helping to visualize the clustering structure and evolution of sociological research themes over time.

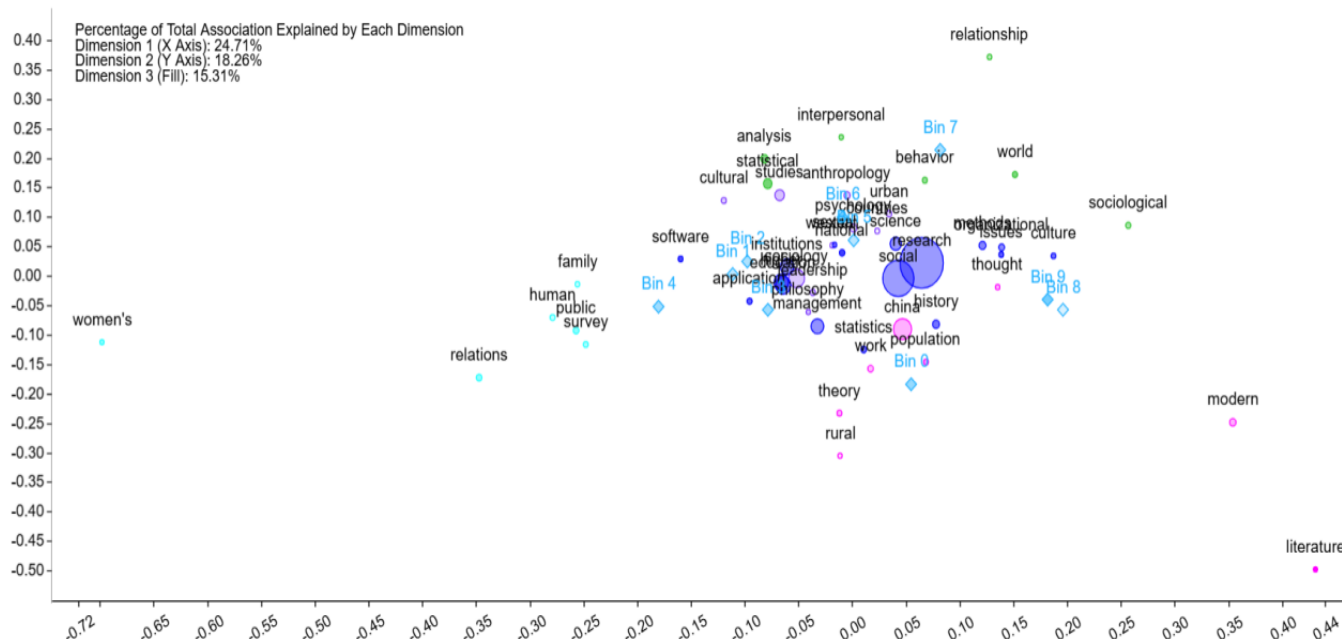


Figure 5. Theme clusters (2016-2018)

2016 – 2018 core themes: Topics cluster around interpersonal relationships, family, cultural anthropology, and social structure (evidenced by terms like "interpersonal", "family", "cultural anthropology", "social structure"). There's also a focus on gender

studies (with "women's" present).

Scope: Themes are relatively traditional, covering foundational sociological areas like interpersonal dynamics and cultural exploration within a societal context.

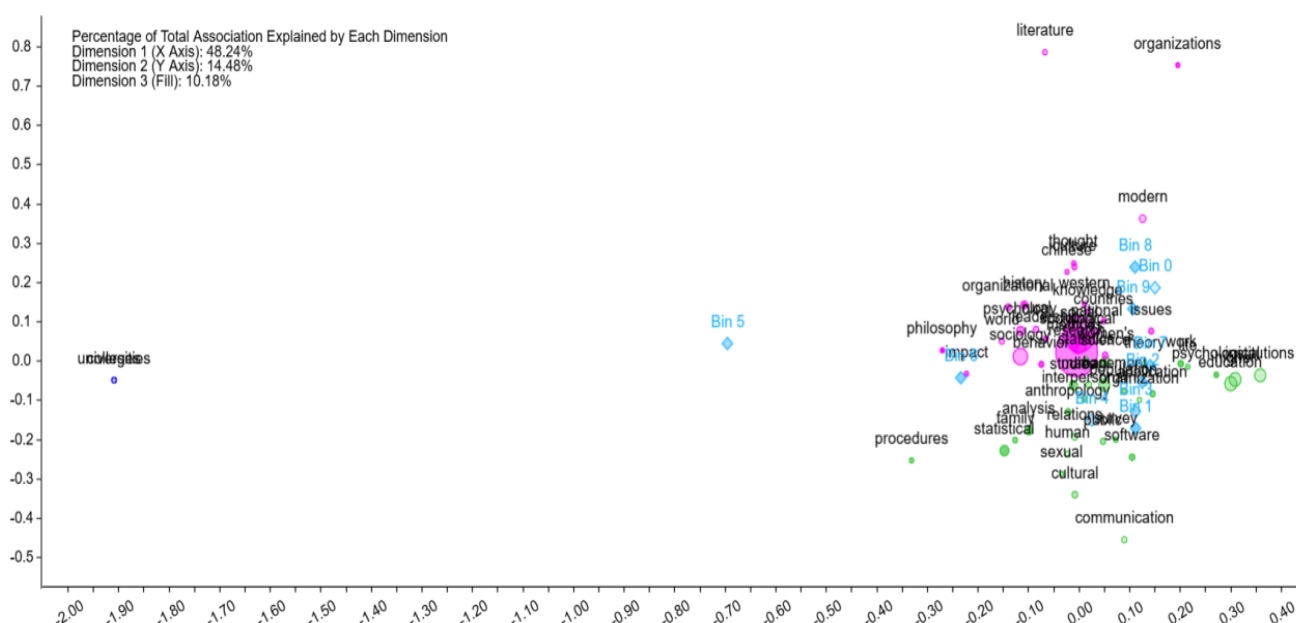


Figure 6. Theme clusters (2019-2021)

2019 – 2021 shift in focus: The cluster centers shift toward modern social issues (e.g., "modern", "urbanization", "organizations") and practical sociology applications (terms like "procedures", "communication", "public policy").

New emphases: Emergence of topics related to urban development,

organizational behavior, and sociological methods in real - world settings, reflecting a move toward applied sociology responding to contemporary societal changes.

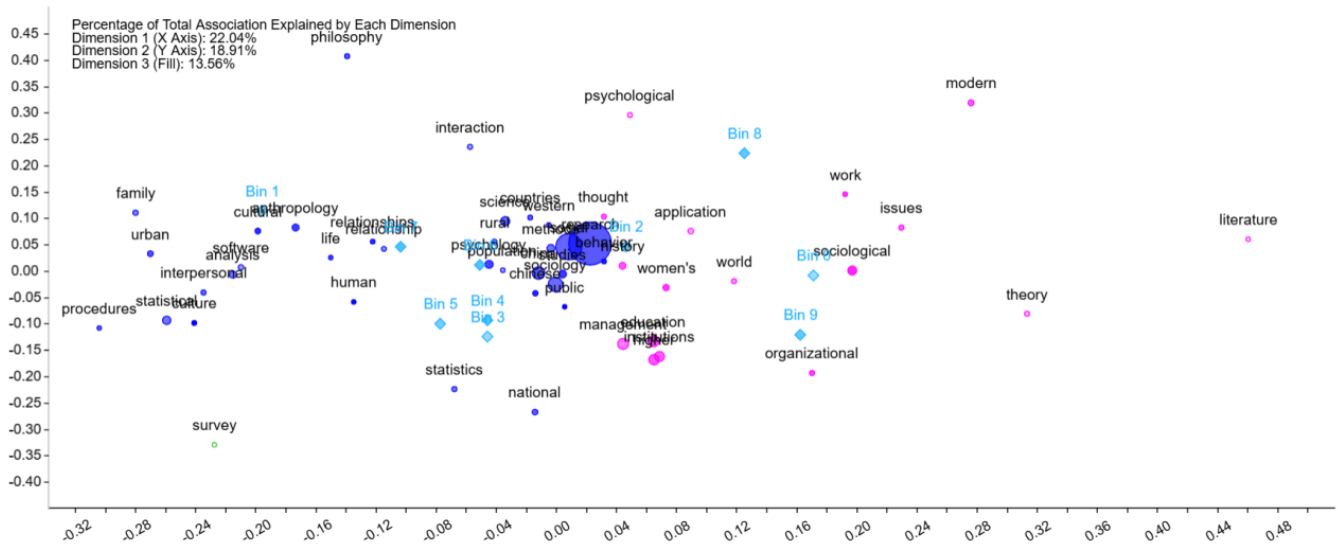


Figure 7. Theme clusters (2022-2024)

2022 – 2024 theoretical & interdisciplinary blend: Clusters now integrate sociological theory ("theory"), interdisciplinary intersections (e.g., "psychological", "philosophy", "statistics"), and contemporary global issues ("national", "interaction", "ecological").

Evolution trajectory: The field has evolved to embrace theoretical depth, cross - discipline collaboration, and response to large - scale societal challenges (like ecological concerns, national - level dynamics), showing a more holistic and forward - looking

sociological inquiry.

4.4 Linear Regression

The accuracy of linear model 62.1%. This represents how well the model predicts "Total Borrowings", though in linear regression. A 62.1% accuracy suggests the model explains a moderate portion of the variance in total borrowings.

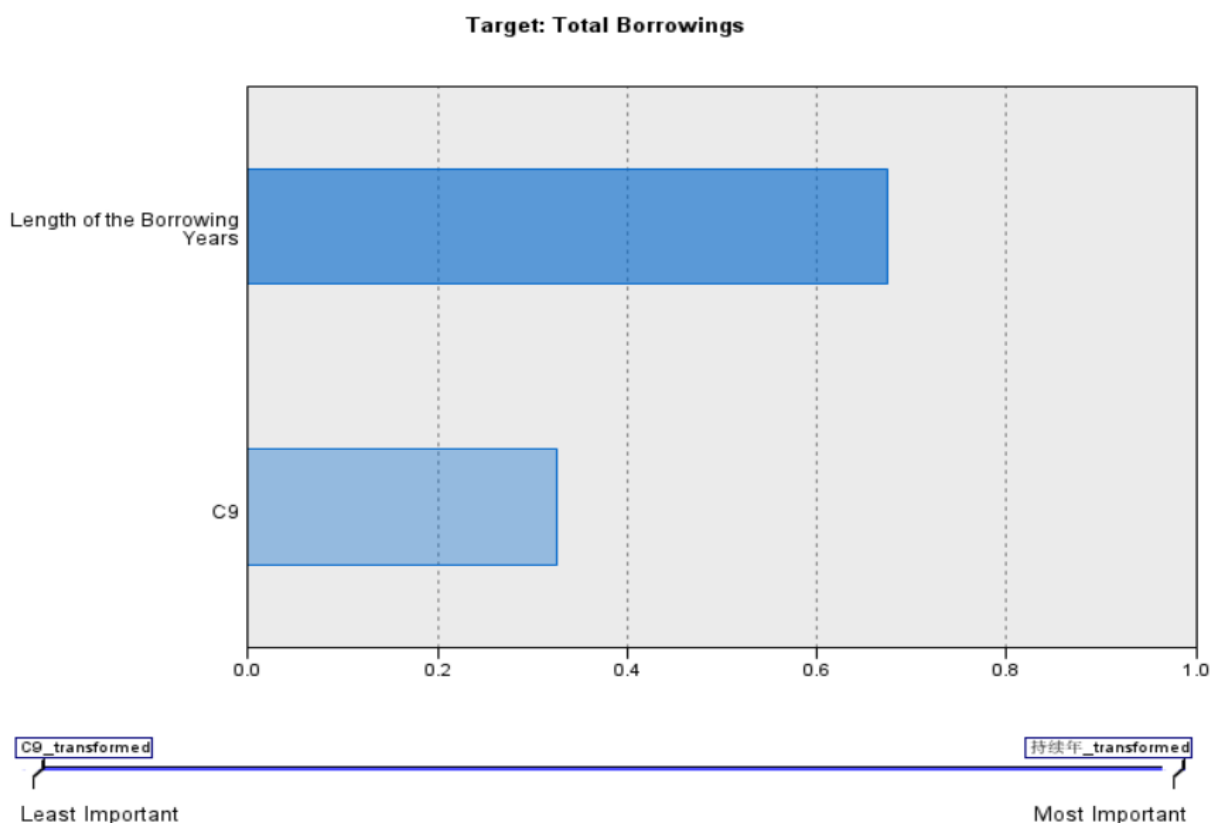


Figure 8. Predictor importance (2016-2024)

Figure 8 assesses the relative importance of predictors for "Total Borrowings". "Length of the Borrowing Years" has a much longer bar, indicating it is the most important predictor. It has a strong influence on predicting total borrowings. "C9" has a shorter bar,

meaning it is less important than "Length of the Borrowing Years" but still contributes to the model. The horizontal axis ranges from "Least Important" to "Most Important", clearly showing the hierarchy of predictor influence.

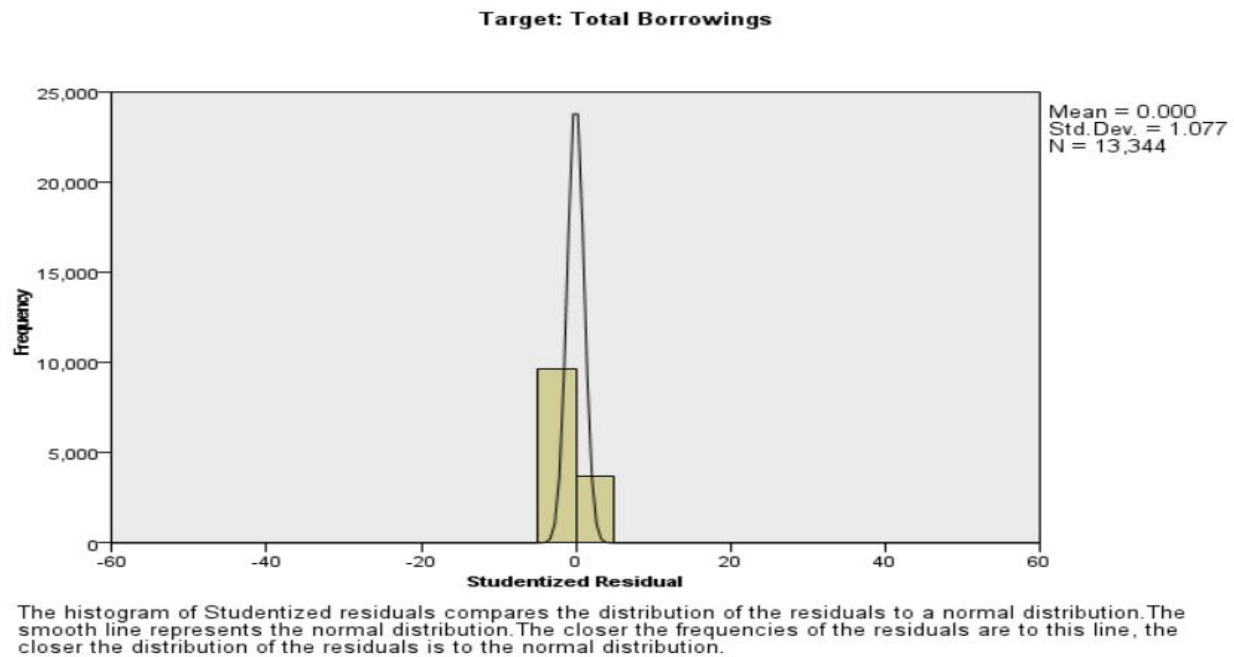


Figure 9. Residuals (2016-2024)

The histogram in Figure 9 shows the distribution of studentized residuals. Studentized residuals adjust for differences in residual variance across observations.

Mean and Standard Deviation: The mean of studentized residuals is 0.000, and the standard deviation is 1.077. For a well - behaved linear regression model, residuals should ideally have a mean of 0 (which they do here) and a standard deviation close to 1.

Distribution vs. Normal Distribution: The smooth line represents the normal distribution. The histogram of residuals does not closely match this smooth normal curve. The frequencies of residuals deviate from the normal distribution line, indicating that the residuals do not perfectly follow a normal distribution. While the

$n = 13344$

Bootstrap = 5000

mean is correct, the shape suggests there might be some non - normal patterns in the errors (e.g., skewness or kurtosis different from a normal distribution), which could impact the validity of some regression assumptions (though moderate deviations might be acceptable in large samples, as here $N=13,344$).

4.5 Mediating Effect

Based on the output results of PROCESS Model 4, researcher comprehensively analyze the mediation effect of "Length of Borrowing Years (LBY, X) $\rightarrow$ Mediator Variable C9 $\rightarrow$ Total Borrowings (TB, Y)" from three core dimensions: model fit, significance of path coefficients, and mediation effect test (see Figure 10).

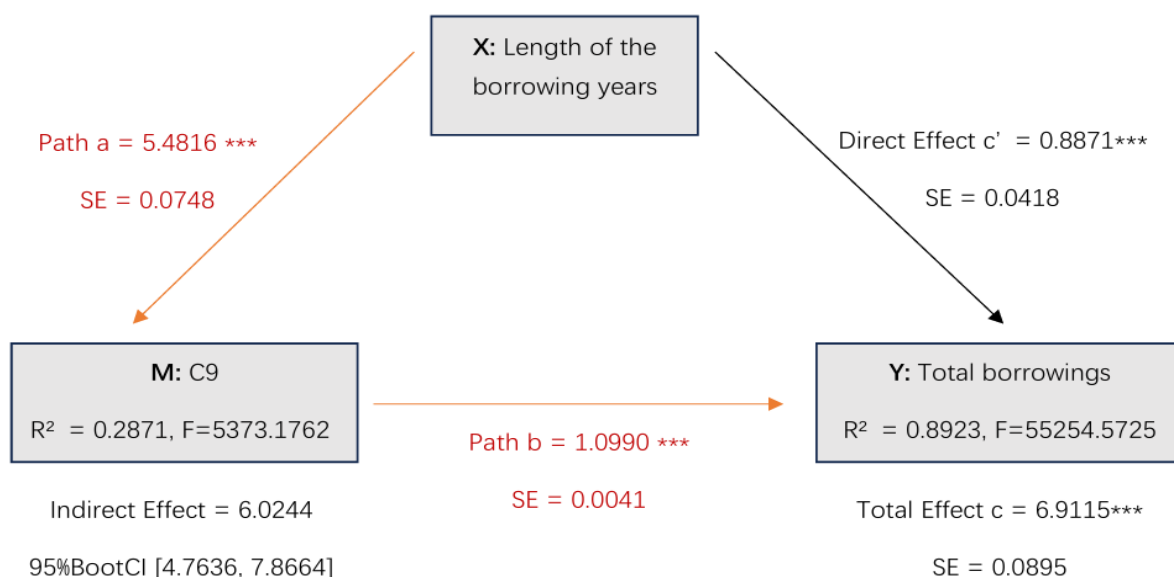


Figure 10. Mediation Role of C9 in the Relationship Between Length of Borrowing Years and Total Borrowings (2016-2024)

4.5.1 Basic Information and Fit of the Model

Sample size: 13,344. The extremely large sample size results in high statistical power and strong result stability.

Bootstrap settings: 5,000 repeated samplings (a commonly used industry standard) are adopted, with a 95% confidence interval (CI), and the results are highly reliable.

Regression model of mediator variable C9 ($X \rightarrow M$): $R^2=0.2871$, indicating that the length of borrowing years (LBY) can explain 28.71% of the variation in C9; the F-value is 5,373.1762 ($p < 0.001$), and the model is highly significant overall.

Regression model of total borrowings TB ($X + M \rightarrow Y$): $R^2=0.8923$, indicating that the length of borrowing years (LBY) and

C9 together explain 89.23% of the variation in total borrowings; the F-value is 55,254.5725 ($p < 0.001$), and the model fit is extremely high.

Total effect model ($X \rightarrow Y$): $R^2=0.3091$, indicating that only the length of borrowing years can explain 30.91% of the variation in total borrowings. After adding C9, R^2 increases significantly, suggesting that the mediation effect of C9 may be significant.

4.5.2 Analysis of Core Path Coefficients (Stepwise Test)

The premise of the mediation effect is that the paths " $X \rightarrow M$ ", " $M \rightarrow Y$ ", and " $X \rightarrow Y$ (total effect)" are all significant (or at least the first two steps are significant, combined with Bootstrap results). Table 1 presents the coefficients.

Table 1. Coefficient (2016-2024)

Path	Coefficient	SE	t-value	p-value	95% CI	significance
1. $X \rightarrow M$ (LBY $\rightarrow$ C9)	5.4816	0.0748	73.3020	< 0.001	[5.3350, 5.6282]	Highly significant
2. $M \rightarrow Y$ (C9 $\rightarrow$ TB)	1.0990	0.0041	268.7475	< 0.001	[1.0910, 1.1070]	Highly significant
3. $X \rightarrow Y$ (direct effect)	0.8871	0.0418	21.2044	< 0.001	[0.8051, 0.9691]	Highly significant
4. $X \rightarrow Y$ (total effect)	6.9115	0.0895	77.2623	< 0.001	[6.7362, 7.0868]	Highly significant

$X \rightarrow M$: For each 1-unit increase in the length of borrowing years (LBY), the mediator variable C9 increases significantly by 5.4816 units ($p < 0.001$), indicating that the longer the borrowing years, the higher the level of C9.

$M \rightarrow Y$: After controlling for the length of borrowing years, for each 1-unit increase in C9, the total borrowings (TB) increase significantly by 1.0990 units ($p < 0.001$), indicating that the higher the level of C9, the higher the total borrowings.

Direct effect: After controlling for C9, for each 1-unit increase in the length of borrowing years, the total borrowings still increase significantly by 0.8871 units ($p < 0.001$), indicating that there is a "direct impact" of the length of borrowing years on total borrowings.

Total effect: When C9 is not controlled, for each 1-unit increase in the length of borrowing years, the total borrowings increase significantly by 6.9115 units ($p < 0.001$). The total effect = direct effect + indirect effect (verified later).

4.5.3 Bootstrap Mediation Effect Test (Core Conclusion)

The stepwise test can only preliminarily determine the significance of the path. The Bootstrap method directly estimates the confidence interval of the indirect effect through the sampling distribution, which is the "gold standard" for current mediation effect tests. The judgment rule is: if the 95% Bootstrap CI does not contain 0, the mediation effect is significant.

The value of the indirect effect = "path coefficient of $X \rightarrow M$ " $\times$ "path coefficient of $M \rightarrow Y$ " = $5.4816 \times 1.0990 \approx 6.0244$ (consistent with

the output results), verifying that "total effect = direct effect + indirect effect": 6.9115 (total effect) = 0.8871 (direct effect) + 6.0244 (indirect effect).

LBY $\rightarrow$ C9: Bootstrap CI = [4.8619, 6.3174] (does not contain 0), the path is significant;

C9 $\rightarrow$ TB: Bootstrap CI = [0.9694, 1.2630] (does not contain 0), the path is significant;

LBY $\rightarrow$ TB (direct effect): Bootstrap CI = [0.2991, 1.3955] (does not contain 0), the path is significant.

The Bootstrap CIs of all paths do not contain 0, indicating that even considering sampling errors, the significance of each path is still robust, and the mediation effect is reliable.

4.6 CRT (Classification and Regression Tree) Analysis for Renewal Behavior

Figure 11 shows the 2016–2018 period decision tree model, with 29 nodes (15 terminal nodes) and a depth of 5.

Risk Assessment: The resubstitution risk estimate is 0.153 with a standard error of 0.004; the cross - validation risk estimate is 0.157 with a standard error of 0.004, and the risk is at a moderate level.

Classification Results: For the cases that are actually "Non - renewal (Category 0)", there are 5,600 cases predicted as Category 0 and 178 cases predicted as Category 1, with a correct rate as high as 96.9%. When the cases are actually "Renewal (Category 1)", there are 900 cases predicted as Category 0 and 355 cases predicted as Category 1, with a correct rate of only 28.3%. The overall correct rate is 84.7%.

It can be seen that the model has an excellent prediction effect on the "Non - renewal" category, but a weak prediction ability on the

"Renewal" category.

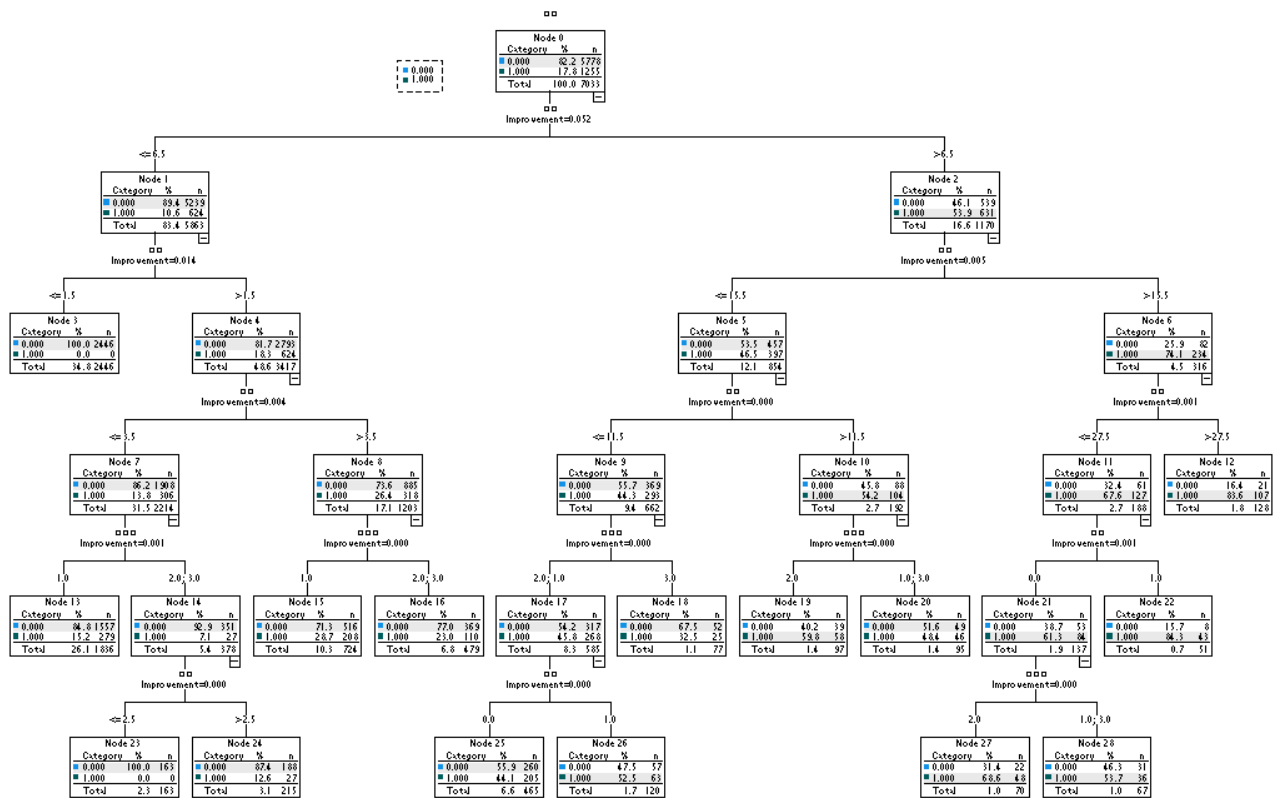


Figure 11. Decision tree model (2016–2018)

Figure 12 presents the decision tree model for the 2019–2021 Period. There are 19 nodes, 10 terminal nodes, and the depth is 4, so the model structure is simpler than before.

Risk Assessment: The resubstitution risk decreases to 0.135 with a standard error of 0.005; the cross - validation risk is 0.138 with a standard error of 0.005, and the risk is lower than that in 2016 - 2018.

Classification Results: When the actual category is 0, there are 3,811

cases predicted as Category 0 and 136 cases predicted as Category 1, with a correct rate of 96.6%. When the actual category is 1, there are 491 cases predicted as Category 0 and 221 cases predicted as Category 1, with a correct rate of 31.0%. The overall correct rate is 86.5%. The prediction for the "Non - renewal" category is still very accurate, and the prediction for the "Renewal" category has been improved to a certain extent, so the overall performance has been improved.

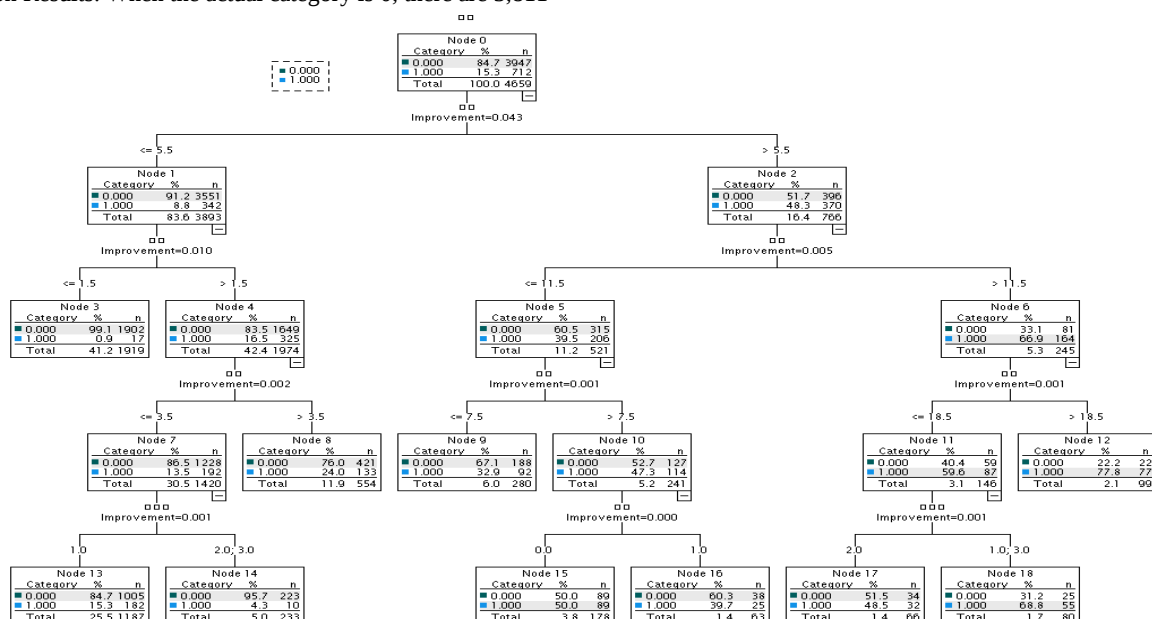


Figure 12. Decision tree model (2019–2021)

Figure 13 presents the decision tree model for the 2022–2024 Period. There are 19 nodes, 10 terminal nodes, and the depth is 4.

Risk Assessment: The resubstitution risk increases to 0.172 with a standard error of 0.007; the cross - validation risk is 0.177 with a standard error of 0.007, and the risk is significantly higher than that in the previous two periods.

Classification Results: When the actual category is 0, there are 2,532 cases predicted as Category 0 and 93 cases predicted as Category 1,

with a correct rate of 96.5%. When the actual category is 1, there are 474 cases predicted as Category 0 and 196 cases predicted as Category 1, with a correct rate of 29.3%. The overall correct rate is 82.8%. The prediction for the "Non - renewal" category remains good, but the prediction for the "Renewal" category becomes worse again. Moreover, the number of independent variables included in the model is reduced, which may affect the prediction ability, and the overall performance declines.

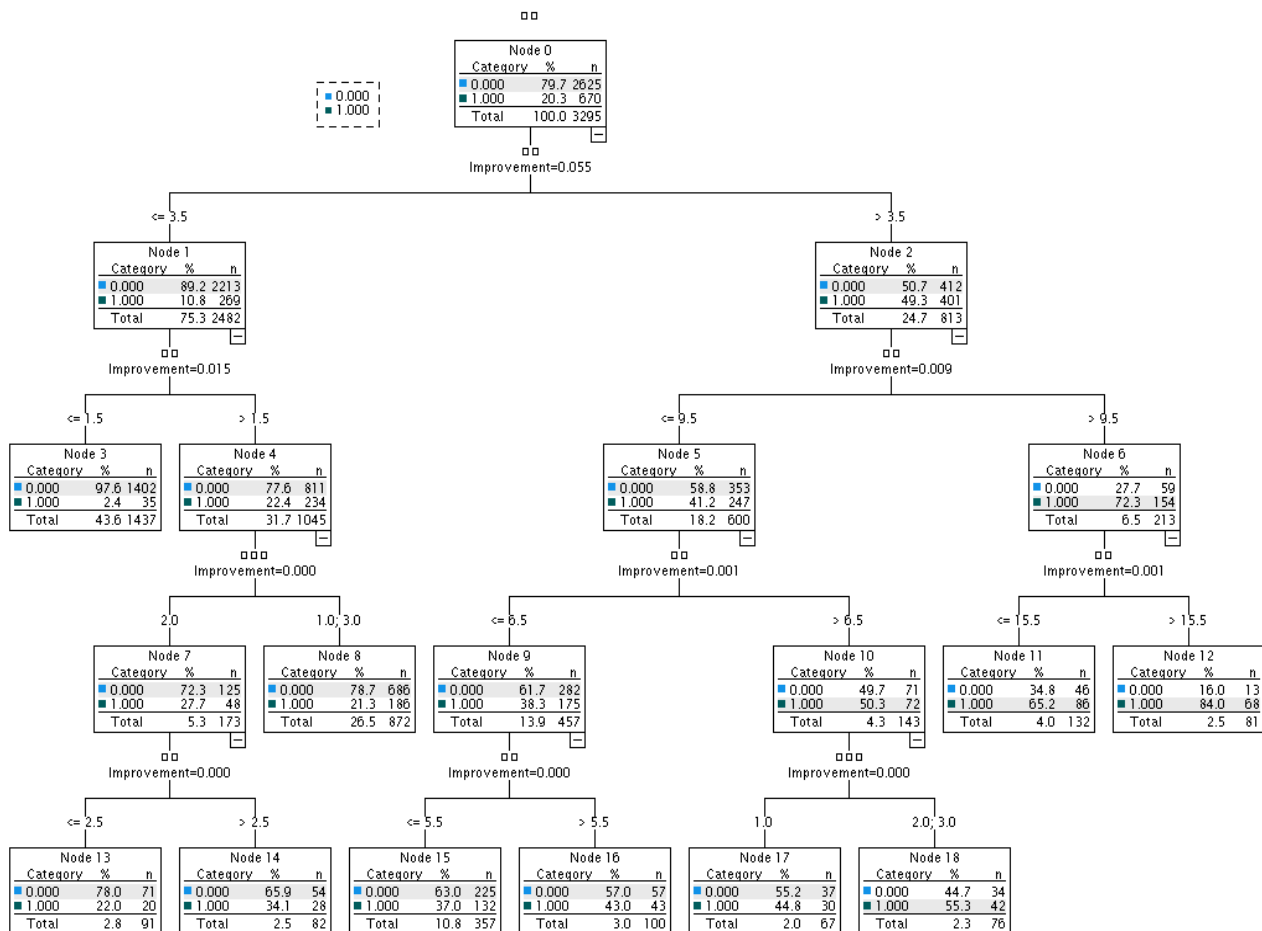


Figure 13. Decision tree model (2022–2024)

5. Discussion

5.1 Key Characteristics of Sociology's Popular Topics

Sociology's popular topics take core disciplines as the foundation, which construct the discipline's theoretical framework, shape research paradigms, and provide methodological tools, ensuring the stability and professionalism of its development. The core disciplines (e.g., institutional sociology, rural studies) define the discipline's research scope and logical system, standardize the thinking for exploring social phenomena, and offer technical support for other specialized topics. Their popularity reflects sociology's emphasis on its own theoretical system and research paradigms, a trend echoed in empirical studies of sociological reading behavior. For example, Trotsuk (2023) found that readers interested in Russian social realities are specifically drawn to sociological themes like social polarization and migration myths—topics rooted in core sociological analyses of social structure and inequality—proving that core disciplinary content is a primary driver of sociological reading engagement, aligning with the foundational role of core disciplines in shaping the field.

Interdisciplinary integration is a major trend in sociology's popular topics, linking sociology with literature, art, philosophy, psychology, and more to expand research scope and enhance comprehensiveness. "Literature, Art & Philosophy" explores the social expression of human thoughts and values through cross-field integration, while "Psychology & Social Security" studies well-being and social stability by combining psychology and social security with sociology. This integration is not only theoretical but also manifests in reading practices: Sirisena et al. (2024)'s "Reading for Wellbeing" initiative shows that reading for pleasure—often blending sociological themes (e.g., social connection) with psychological well-being—links content interest to positive social experiences and improved self-efficacy. Similarly, Shen et al. (2024) found that Chinese older adults' leisure reading of sociological books on aging and community (integrating sociological analysis of aging societies with psychological insights into mental health) reduces depression, demonstrating how interdisciplinary sociological content bridges theory and practical reading outcomes.

A practical application orientation permeates many of sociology's

popular topics, focusing on solving real-world issues like education, regional development, and social equity to highlight the discipline's pragmatism. "Higher & Vocational Education" reforms education systems to connect education with social needs, "National & Regional Studies" supports international cooperation through regional analysis, and "Sexual & Gender Studies" promotes gender equality to drive social progress. This orientation is strongly reflected in how content interest mediates reading behavior for practical problems: Gonzalez et al. (2024) implemented a content-based shared book reading (SBR) intervention for preschool dual language learners (DLLs), targeting science and social studies; aligning materials with curricular content (addressing educational equity) enhanced content interest, ultimately improving DLLs' vocabulary skills. Lee et al. (2023) also noted that Head Start families engaging with sociological books on food justice (a practical public health topic) showed sustained reading behavior when content matched their interests, as content interest mediated intervention integration and reading practices.

Research methods and statistical tools form an independent popular topic in sociology, reflecting the discipline's focus on scientific rigor and serving as a methodological guarantee for other topics. Covering survey statistics, sampling control, SPSS application, and questionnaire design, these tools ensure the logic and accuracy of social analysis. This focus is indirectly supported by sociological reading studies using rigorous frameworks to examine content interest: Li et al. (2025) used mediation analysis in an online extensive reading (ER) environment, finding that ER engagement mediates the effect of planning behavior and discussion participation on English performance; though focused on language learning, the study's methodological rigor (a hallmark of sociological research) allows inferences about content interest as a latent driver of ER engagement. Zheng (2024)'s digital library recommendation model—using an attention mechanism to model readers' interests with 97.24% accuracy in science reading—also demonstrates how statistical tools (core to sociological methods) quantify content interest, linking methodological rigor to reading behavior research.

Sociology's popular topics cover "macro-meso-micro" multi-level issues, forming a diversified, three-dimensional research system that meets social cognition needs at different scales and reflects the discipline's breadth. Macro-level topics (e.g., "National & Regional Studies") focus on societal operation, meso-level topics (e.g., "Higher & Vocational Education") center on group-system interactions, and micro-level topics (e.g., "Psychology & Social Security") focus on individual behavior. This structure is mirrored in sociological reading research, where content interest operates as a mediator across scales: At the macro level, Nenonen et al. (2025) framed reading as cultural participation (a macro-sociological concept) and found content interest mediates long-term reading participation and consistent behavior, with implications for population-level outcomes like reduced mortality. At the meso level, Wang and Deng (2024) studied community book systems, showing social value (a meso community factor) promotes sustained sociological reading via content interest-driven emotional value. At the micro level, Westerveld et al. (2024) found that content interest mediates parent-child literacy interactions (micro interpersonal factors) and reading behavior for autistic children with high support needs—collectively proving content interest's unifying role across the macro-meso-micro spectrum.

5.2 Key Insights Extracted from Sociology's Top 100 High-Frequency Thematic Words

The extremely high overlap rate of high-frequency words across

periods and the stability of core themes over the decade reflect the continuity of readers' attention to sociological content. Among the top 100 high-frequency thematic words in 2016–2018, 2019–2021, and 2022–2024, 86 are cross-period co-occurring words (over 80%), covering core sociological scopes like research methods, social structure, and institutional analysis. 27 high-frequency words with a total frequency of ≥ 1000 are basic iconic terms in sociology, and their cross-period stability confirms that the discipline's theoretical framework, research paradigms, and core focus have remained consistent. This continuity is directly mirrored in sociological reading behavior: Hiidenmaa et al. (2024) found that "reader equality" (no differences by education or gender) among Finnish readers stems from shared content interest in these stable core themes, not demographics—proving that sustained attention to cross-period core topics drives persistent sociological reading, aligning with the stability of high-frequency thematic words.

Cross-period co-occurring words show obvious frequency stratification, reflecting the primary-secondary difference within core themes, proving that period-unique words directly shape the direction of sociological reading materials and interventions. The evolutionary trend of period-unique words confirms the adaptability and expandability of sociological themes, with core stability and dynamic adjustment coexisting. While cross-period core words (e.g., social structure, research methods) maintain disciplinary foundation, period-unique words reflect active adaptation to social changes—for example, 2022–2024's focus on technology responds to digital transformation, and 2019–2021's educational focus addresses societal demands for education-society integration. This coexistence is reflected in long-term reading persistence: Sun and Xi (2025) found that childhood access to sociological books (focused on stable core themes) fosters enduring content interest. Together, this shows that stable core themes ensure consistent readership, while dynamic unique themes attract new readers, embodying the discipline's adaptability and expandability.

5.3 Phased characteristics of sociological thematic content, evolutionary direction of research focus, and internal logic of disciplinary development.

Sociological thematic content exhibits distinct period-specific differentiation across 2016–2018, 2019–2021, and 2022–2024, with each phase having a clear core focus. The 2016–2018 phase centers on traditional fundamental themes, encompassing core foundational domains like social micro-interactions, macro-structures, and cultural basic exploration, while excluding complex interdisciplinary or practical application topics—emphasizing the discipline's theoretical roots. This focus is reflected in reading behavior: Lerasulimskiy et al. (2023)'s study of late 19th–early 20th-century Yaroslavl parish libraries (a precursor to modern foundational theme focus) found that sustained sociological reading was driven by content interest in core social/cultural topics, mirroring 2016–2018's emphasis on traditional foundations.

The 2019–2021 phase shifts to modern practical issues and application orientation, moving beyond basic theoretical discussions to prioritize responses to immediate social changes. This transition is evident in reading interventions: Pollard-Durodola et al. (2024) implemented a content-focused preschool vocabulary intervention (targeting practical science and social studies topics) during this period, and found that content interest mediated the sustainability of the intervention—teachers continued using the approach only when materials aligned with practical educational needs, reflecting the phase's application-driven focus. Additionally, Kadriu-Citaku and Malaj-Abdyli (2024)'s focus on multicultural sociological materials

in Kosovo curricula (addressing real-world cultural cohesion issues) further embodies this phase's shift to practicality.

The 2022–2024 phase features three-dimensional integration: strengthening disciplinary theoretical depth, highlighting interdisciplinary intersections (via high-frequency terms linked to psychology, philosophy, and statistics), and focusing on global/macro-social challenges (via terms related to national dynamics and ecology). This integration is reflected in diverse reading practices: Feng et al. (2025)'s study of diabetic/hypertensive patients (2022–2024) found that sociological books on health equity (integrating sociology with healthcare, an interdisciplinary focus) reduced anxiety, with content interest mediating sustained reading—aligning with the phase's interdisciplinary emphasis. Abimbola (2025)'s focus on climate change reading (a global macro-challenge) also fits this phase, as guided reading on environmental sustainability (a sociological theme) enhanced literacy via content interest, demonstrating the integration of macro-issues and theoretical depth.

Sociology's research focus follows a distinct evolutionary trajectory—from traditional foundation to practical reality, and further to three-dimensional integration of theory, interdisciplinarity, and macro-issues—driven by social changes and reflecting the discipline's adaptability. This evolution is not random but represents active response to stage-specific social needs: 2016–2018's foundational focus built disciplinary consensus, 2019–2021's practical shift addressed urgent societal problems, and 2022–2024's integration responded to complex global challenges. This trajectory is mirrored in long-term reading outcomes: Lewis et al. (2023) found that European adults who accessed sociological books in childhood (2016–2018's foundational phase) developed early content interest, which later mediated reduced cognitive impairment risk—showing how early foundational focus laid the groundwork for later engagement with practical and integrated themes.

Sociological theme clustering embodies the discipline's developmental logic: expanding from simplified to diversified research scopes, and upgrading from static social description to dynamic participation. The shift from single disciplinary fields to interdisciplinary collaboration (e.g., sociology + healthcare in 2022–2024) reflects boundary expansion, while the move from passive phenomenon description to active problem-solving (e.g., 2019–2021's educational interventions) represents functional upgrading. This logic is evident in community reading systems: book systems evolved from providing basic sociological texts (2016–2018) to practical materials (2019–2021) and finally to interdisciplinary, macro-themed books (2022–2024), with content interest mediating sustained use—confirming that the discipline's developmental logic directly shapes reading resource provision and engagement. Ultimately, this evolution positions sociology as a pragmatic discipline aligned with real-world needs, balancing theoretical rigor with practical impact.

5.4 Comparison of Differences Between Linear Regression Model and Mediation Effect Model

5.4.1 Differences in Model Positioning and Research Objectives

Both models take "Total Borrowings (TB)" as the dependent variable. The linear regression model is an association descriptive model. Its core objective is only to quantify the strength of linear association between each predictor variable and total borrowings, answering questions such as "which variables affect total borrowings" and "to what extent do they affect it", without involving the influence mechanism between variables. For example, it can

only determine that Length of Borrowing Years (LBY) and C9 both affect total borrowings but cannot explain whether LBY affects total borrowings directly or indirectly through other variables.

The mediation effect model is a mechanism-explanatory model. Its core objective focuses on "how LBY affects total borrowings". By introducing the mediating variable C9, it clarifies the transmission path of "LBY→C9→Total Borrowings", decomposes the direct effect (LBY affects total borrowings without going through C9) and indirect effect (LBY affects total borrowings indirectly through C9), and finally answers "what is the specific mechanism by which LBY affects total borrowings", rather than staying at the superficial judgment of "whether an impact exists".

5.4.2 Differences in Variable Composition and Relationship Assumptions

The variable composition of the linear regression model is broader, including multiple predictor variables. However, it only analyzes the direct association between each predictor variable and total borrowings and does not define the causal order between variables. Its relationship assumptions are also relatively simple, only requiring a linear relationship between predictor variables and total borrowings, without involving assumptions about "whether there is intermediate transmission between variables". For example, it does not preset that "LBY affects C9 first and then affects total borrowings"; instead, it treats C9 as a predictor variable parallel to LBY.

The variable composition of the mediation effect model has strict definitions, which must include three types of variables: independent variable (LBY, X), mediating variable (C9, M), and dependent variable (Total Borrowings, Y). Moreover, it clearly sets the causal order as "X→M→Y". Its relationship assumptions are more complex: in addition to satisfying the linear relationship, it also needs to assume that LBY can significantly affect C9 and C9 can significantly affect total borrowings. At the same time, it defaults to no reverse causality (e.g., C9 does not affect LBY in turn). Through a clear causal chain, it focuses on transmission mechanism rather than parallel association.

5.4.3 Differences in Model Fitting Effect and Explanatory Power

The fitting effect of the linear regression model is at a moderate level. According to the results, the model prediction accuracy is 62.1%. Additionally, there is a problem of non-normal residuals (the residual histogram does not match the normal curve, and only the mean of 0.000 and standard deviation of 1.077 meet some assumptions). Combining with the total effect model $R^2=0.3091$ containing only LBY in the mediation effect model, it can be inferred that even if the linear regression model includes C9, its overall explanatory power is still far lower than that of the mediation effect model.

The fitting effect of the mediation effect model is at an extremely high level, and the goodness of fit is verified from multiple dimensions. Firstly, the regression model of "LBY→C9" has an R^2 of 0.2871 (LBY can explain 28.71% of the variation in C9), with an F-value of 5,373.1762 ($p<0.001$), indicating that the model is highly significant overall. Secondly, the regression model of "LBY + C9→Total Borrowings" has an R^2 of 0.8923 (the two variables together explain 89.23% of the variation in total borrowings), with an F-value of 55,254.5725 ($p<0.001$), and its explanatory power is far higher than that of the linear regression model. Finally, although residuals are not reported separately, the Bootstrap sampling (5,000 repetitions) is used to avoid the impact of non-normal residuals on

the conclusions, making the model more robust.

5.4.4 Differences in Core Conclusions and Effect Quantification

The conclusions of the linear regression model only stay at predictor importance ranking and superficial association. It can only determine that LBY is the most important predictor variable (with the longest bar), and C9 has lower importance (with a shorter bar), but cannot quantify the indirect impact of LBY on total borrowings through C9. Its effect indicators are also relatively simple: it can only obtain the total association coefficient of each predictor variable on total borrowings (e.g., the coefficient of LBY on total borrowings may be close to 6.9115), but cannot decompose the proportion of direct impact and indirect impact, resulting in a lack of depth in the conclusions.

The conclusions of the mediation effect model realize effect decomposition and mechanism clarification. Through the calculation of path coefficients, it quantifies three types of key effects: the indirect effect (LBY→C9→Total Borrowings) is 5.4816 (the coefficient of LBY on C9) × 1.0990 (the coefficient of C9 on total borrowings) ≈ 6.0244, accounting for 87.16% of the total effect (6.9115) and serving as the core impact path; the direct effect (LBY→Total Borrowings) is 0.8871, accounting for 12.84% of the total effect, with a weak but significant impact; at the same time, it verifies that "total effect = direct effect + indirect effect" ($6.9115 = 0.8871 + 6.0244$). Finally, it clearly confirms that "more than 87% of the impact of LBY on total borrowings is achieved through C9", rather than the linear regression model's conclusion that LBY directly dominates the impact.

5.4.5 Differences in Testing Logic and Methods

The testing logic of the linear regression model is relatively simple. It only uses the t-test to determine the significance of the coefficients of each predictor variable (e.g., determining whether the coefficients of LBY and C9 are significantly non-zero) and uses the residual histogram to test whether residuals conform to basic assumptions such as normal distribution. This type of test can only verify "whether an association exists between variables" but cannot verify the causal path between variables, which easily confuses association with causality. For example, it cannot prove whether LBY is the cause and C9 is the effect or C9 is the cause and LBY is the effect.

The testing logic of the mediation effect model is more hierarchical and rigorous. The first step is to verify the prerequisites of the mediation effect through the stepwise test: first, test that the coefficients of "LBY→C9", "C9→Total Borrowings", and "LBY→Total Borrowings (total effect)" are all significant (in the results, the p-values of the three are all <0.001, meeting the prerequisites). The second step is to directly estimate the 95% confidence interval of the indirect effect through the "Bootstrap test" (5,000 repeated samplings); if the interval does not contain 0, the mediation effect is significant. In the results, the Bootstrap confidence intervals of "LBY→C9", "C9→Total Borrowings", and "LBY→Total Borrowings (direct effect)" all do not contain 0. This not only avoids the defect of low power in the stepwise test but also strengthens the causal explanatory power of the conclusions through the verification of a clear causal path, rather than staying at the association level.

5.4.6 Differences in Practical Application Value

The practical value of the linear regression model is relatively limited. It can only be used for predicting total borrowings or identifying key predictor variables. For example, based on the conclusion that "LBY is the most important predictor variable", it

can be concluded that "to affect total borrowings, priority should be given to LBY", but it cannot guide "how to specifically affect total borrowings through LBY" or answer "whether C9 can be regulated to indirectly affect total borrowings", lacking actionable intervention directions.

The practical value of the mediation effect model is significantly improved. It can not only predict but also provide targeted intervention plans. Based on the mechanism that LBY affects total borrowings through C9, two core intervention ideas can be derived: first, if total borrowings are to be increased through LBY, the core focus is to increase the level of C9 by extending LBY (for each 1-unit increase in LBY, C9 increases by 5.4816 units, which in turn increases total borrowings by 6.0244 units); second, if total borrowings are to be regulated more efficiently, C9 can be directly intervened (for each 1-unit increase in C9, total borrowings increase by 1.0990 units) without relying solely on LBY. Such conclusions upgrade from identifying influencing factors to providing solutions, and have more specific guiding significance for policy formulation or management decision-making.

5.5 CRT Model Characteristics, Prediction Performance Differences, and Risk Variation Trends of Renewal Behavior

5.5.1 Model Construction and Characteristics of Variables Influencing Renewal Behavior

The prediction models for renewal behavior in 2016–2018 and 2019–2021 both incorporated three independent variables: "Total", "Gender", and "Persistence", indicating these factors were key to shaping renewal behavior during these periods. By 2022–2024, the model retained only "Total" and "Persistence" while excluding "Gender", suggesting gender's impact on renewal behavior weakened and it no longer functioned as an effective predictive variable. This shift aligns with broader trends in sociological reading research, where demographic factors (like gender) are increasingly overshadowed by content-related and environmental drivers of engagement. In the context of reading environments that foster renewal behavior (e.g., sustained engagement with sociological materials), variables tied to content relevance (e.g., matching materials to reader interests) and persistence (e.g., long-term access to resources) prove more influential than gender, mirroring the CRT model's variable adjustment.

The 2016–2018 model had the most complex structure (29 nodes, 15 terminal nodes, depth of 5), while the 2019–2021 and 2022–2024 models simplified to 19 nodes, 10 terminal nodes, and a depth of 4. This simplification reflects a growing concentration in the mechanism influencing renewal behavior: later periods required fewer variables and structural layers to predict engagement, as core drivers (e.g., content interest-mediated persistence) became more distinct. This trend is supported by studies on reading persistence, such as O'Brien et al. (2024)'s research on bilingual home literacy environments (HLE), where "child interest" emerged as a key, streamlined factor mediating HLE resources and sustained reading—reducing the need for complex demographic or contextual variables to explain long-term engagement (a parallel to renewal behavior). Additionally, Yu et al. (2024) found that authoritative parents' provision of interest-aligned sociological books directly mediates sustained reading, further demonstrating how focused, content-driven factors simplify the prediction of persistent engagement (and by extension, renewal behavior).

Across all three periods, models used the CRT growing method and cross-validation with unified parameters (maximum tree depth = 5, minimum parent node cases = 100, minimum child node cases = 50),

ensuring comparability and eliminating parameter interference. This methodological consistency is critical for linking model findings to real-world reading dynamics, where stable, replicable factors (e.g., access to interest-matched content) consistently drive renewal behavior. For example, Webber et al. (2025)'s study of school librarians—who curate sociologically relevant books reflecting students' lives—shows that consistent curation practices (parallel to the model's unified parameters) lead to predictable student engagement, as content interest reliably mediates curation efforts and sustained interaction with materials.

5.5.2 Phased Differences in Renewal Behavior Prediction Performance

The correct prediction rate for "non-renewal" behavior remained above 96% across periods (96.9% in 2016–2018, 96.6% in 2019–2021, 96.5% in 2022–2024), with minimal fluctuation. This stability suggests "non-renewal" is driven by consistent, easily identifiable barriers—often tied to poor alignment between reader needs and reading resources. In sociological reading research, such barriers are well-documented: Arroyo-Vazquez et al. (2023) found that only 21.0% of recommended sociological monographs for Spanish social sciences degrees are digital, and poor accessibility of interest-aligned content breaks the mediation of content interest between persistence factors and reading behavior—directly leading to non-renewal. Similarly, Grams (2025) noted that poor accessibility of academic ebooks (e.g., lacking text search or annotation features) weakens content interest, reducing long-term engagement and increasing non-renewal risk. These studies confirm that non-renewal's stability in the CRT model reflects real-world, persistent barriers to accessible, interest-matched reading.

By contrast, "renewal" (Category 1) prediction performance was weak and followed a "rise-then-fall" trend: 28.3% (2016–2018), 31.0% (2019–2021), and 29.3% (2022–2024). This volatility indicates renewal behavior depends on dynamic, context-dependent factors—many of which are tied to content interest and environmental support, which shift over time. For example, in 2019–2021 (a period of improved renewal prediction), sociological reading interventions increasingly focused on practical, interest-driven content. By 2022–2024, renewal prediction declined, possibly due to the rise of diverse, fragmented reading contexts (e.g., digital vs. physical materials) that introduce new variables (e.g., technology familiarity) not captured by the model. Rodriguez-Gallego et al. (2024)'s study of Booktuber communities—where digital engagement mediates sociological reading via content interest—highlights how emerging contexts add complexity to renewal behavior, making it harder for fixed variables to predict.

The overall correct rate mirrored "renewal" prediction trends: 84.7% (2016–2018), 86.5% (2019–2021), 82.8% (2022–2024), confirming that renewal behavior is the core bottleneck for model effectiveness. This aligns with sociological reading research, where sustained engagement (renewal) is the most variable outcome—dependent on factors like content relevance, resource access, and environmental support. For instance, Dinotola and Testa (2025) found that public libraries' transmedia shelf design enhances serendipitous access to sociological content, and content interest mediates exposure to reading behavior—but this effect varies by library layout and reader demographics, leading to inconsistent renewal. Similarly, Bates et al. (2023) noted that sociological theme interest mediates instructional design and active reading for health science students, but this mediation weakens if themes do not align with students' evolving career interests—further explaining why renewal (and thus overall model performance) fluctuates.

5.5.3 Risk Changes in Renewal Behavior Prediction and Potential Causes

Model risk followed a "fall-then-rise" trend, inversely correlated with overall prediction performance: moderate risk in 2016–2018 (resubstitution = 0.153, cross-validation = 0.157), lowest in 2019–2021 (0.135, 0.138), and highest in 2022–2024 (0.172, 0.177). This inverse relationship confirms that risk primarily stems from unstable renewal prediction—a pattern supported by real-world reading dynamics, where renewal depends on mutable factors like content interest and resource availability. For example, in 2019–2021 (low risk), libraries and schools expanded access to interest-aligned sociological materials. By 2022–2024 (high risk), factors like limited digital access to sociological books and shifting reader preferences (e.g., toward short-form digital content) introduced unpredictability into renewal behavior, increasing model risk.

The 2022–2024 risk increase may be linked to the exclusion of "Gender" and reduced variable dimensions, which weakened the model's explanatory power. However, renewal behavior in this period likely depends on unincorporated variables tied to content and context—factors more influential than gender. Additionally, the rise of interdisciplinary sociological reading (e.g., linking sociology to healthcare or environmental studies; Dirgar et al., 2024; Abimbola, 2025) introduces new drivers of renewal that the model's simplified variables fail to account for—further increasing risk.

Conclusion

This study systematically analyzed the thematic characteristics of sociological books, the dynamics of high-frequency subject words across time, and the factors influencing borrowing and renewal behaviors, providing comprehensive insights into the reader attention trends and behavioral mechanisms in sociological book consumption.

First, regarding thematic content, the top 10 topics in sociological books reflect a balanced structure of foundational disciplines, interdisciplinary integration, and practical orientation. Core sociology disciplines (e.g., institutional sociology, rural studies) form the theoretical backbone, while interdisciplinary themes (e.g., Literature, Art & Philosophy; Psychology & Social Security) expand the research scope by linking sociology to humanities and social welfare. Practical themes such as Higher & Vocational Education and Academic Exchange further highlight the discipline's responsiveness to real-world needs, bridging academic inquiry with social development.

Second, the analysis of top 100 high-frequency subject words across three periods (2016–2018, 2019–2021, 2022–2024) reveals two key trends: temporal stability of core themes and phased evolution of emerging topics. Eighty-six cross-period words (86%)—including "research," "social," "sociology," and "education"—confirm that readers' attention to foundational sociological domains (e.g., research methods, social structure, education) has remained consistent over nearly a decade. In contrast, period-unique words reflect adaptive shifts: from "practical specialization" (e.g., "secretarial," "consulting") in 2016–2018 and "educational text focus" (e.g., "universities," "essays") in 2019–2021 to "technology integration and contemporary social issues" (e.g., "technology," "aging," "feminist") in 2022–2024. This evolution underscores sociology's ability to align with societal changes and expanding reader interests.

Third, theme cluster analysis further illustrates the progressive expansion of sociological inquiry. The 2016–2018 clusters centered on traditional foundational topics (e.g., interpersonal relationships,

cultural anthropology); 2019–2021 shifted to applied modern issues (e.g., urbanization, public policy); and 2022–2024 integrated theoretical depth, interdisciplinary collaboration (e.g., psychology, statistics), and global challenges (e.g., ecological concerns). This trajectory reflects a shift from static description of social phenomena to dynamic, holistic engagement with complex contemporary issues.

Fourth, linear regression and mediation effect analyses clarified the drivers of “Total Borrowings (TB).” The linear model explained 62.1% of TB variance, with “Length of Borrowing Years (LBY)” identified as the most influential predictor. Mediation analysis (PROCESS Model 4, $N=13,344$) confirmed that “Mediator Variable C9” plays a significant partial mediating role in the LBY→TB relationship: LBY explains 28.71% of C9 variance, and LBY + C9 together explain 89.23% of TB variance (far exceeding the 30.91% explained by LBY alone). The indirect effect (6.0244) and direct effect (0.8871) were both highly significant ($p<0.001$), with Bootstrap 95% CIs excluding 0, verifying the robustness of the mediation mechanism.

Finally, CRT (Classification and Regression Tree) analysis of renewal behavior revealed phased differences in prediction performance. Across all periods, the model exhibited excellent accuracy in predicting “Non-renewal” (96.5%–96.9%) but weak performance in predicting “Renewal” (28.3%–31.0%). The 2019–2021 model (19 nodes, 4 depth) showed the lowest risk (resubstitution risk=0.135) and highest overall correct rate (86.5%), while the 2022–2024 model (excluding “Gender” as a predictor) saw increased risk (0.172) and decreased accuracy (82.8%). This suggests that renewal behavior is shaped by complex, time-varying factors, and excluding potential predictors may undermine model effectiveness.

Despite its contributions, this study has several limitations that warrant attention in future research. This study focused on high-frequency subject words and broad themes but did not explore the depth of thematic content (e.g., how specific sub-themes of “gender studies” evolve) or reader demographics (e.g., whether student vs. professional readers differ in thematic preferences). The mediation analysis confirmed C9’s role but did not define its theoretical or operational meaning (e.g., whether C9 represents “reading frequency,” “content relevance,” or “reader satisfaction”). Clarifying C9’s conceptual definition and validating it with additional datasets would strengthen the generalizability of the mediation mechanism. The 2022–2024 CRT model excluded “Gender” and still failed to predict “Renewal” effectively, indicating unmeasured variables (e.g., digital vs. physical book format, seasonal borrowing trends, or reader engagement with specific themes) may influence renewal. The study analyzed aggregate data without comparing thematic or behavioral differences across regions (e.g., urban vs. rural libraries) or countries. Cross-regional studies could reveal how cultural, economic, or educational contexts shape sociological book consumption. While the study compared three time periods, it did not adopt a continuous longitudinal design to capture real-time shifts in reader behavior (e.g., how events like the COVID-19 pandemic impacted borrowing patterns). Future research should address these limitations by integrating multi-method (qualitative + quantitative) designs, expanding predictor variables, and incorporating contextual and longitudinal data. This will further advance our understanding of sociological book consumption and provide evidence-based guidance for library collection development and reader service optimization.

Funding Statement

This research was funded by the 2025 Teaching Reform Research Project of Jiangsu University Library Working Committee: Research on Demand Identification and Resource Adaptation Based on Readers' Reading Behaviors, China (2025JTYB39).

References

1. Abimbola, M. O. (2025). Climate change literacy for young readers through guided reading: A case study of Soaring Crew Reading Club. *Journal of Poetry Therapy*. <https://doi.org/10.1080/08893675.2025.2486259>
2. Agafonova, A. B. (2023). Urban everyday life and electrification in context of fuel crises in first quarter of 20th century. *Nauchnyi Dialog*, 12(8), 348–371. <https://doi.org/10.24224/2227-1295-2023-12-8-348-371>
3. Aleksic, G., & Durus, N. M. (2025). Home literacy environment of multilingual preschool children. *Journal of Educational Psychology*, 117(5), 717–736. <https://doi.org/10.1037/edu0000938>
4. Ananthakrishnan, U. M., Basavaraj, N., Karmegam, S. R., Sen, A., & Smith, M. D. (2025). Book Bans in American Libraries: Impact of Politics on Inclusive Content Consumption. *Marketing Science*, 44(4), 933–953. <https://doi.org/10.1287/mksc.2024.0716>
5. Arroyo-Vazquez, N., Alvite-Diez, M. L., Rodriguez-Bravo, B., & Gonzalez-Taladiz, L. de C. (2023). E-books in Spanish university libraries: An analysis of online social sciences degrees. *Journal of Academic Librarianship*, 49(6), Article 102809. <https://doi.org/10.1016/j.acalib.2023.102809>
6. Bates, S., Fisher, K., Watson, J., Fair, K., Orwig, T., & Walker, D. (2023). Structural competence and equity-minded interprofessional education: A common reading approach to learning. *Education Sciences*, 13(3), Article 287. <https://doi.org/10.3390/educsci13030287>
7. Busetta, G., Campolo, M. G., & Panarello, D. (2024). Changes in daily life habits during COVID-19 and their transitory and permanent effects on Italian university students' anxiety level. *Applied Research in Quality of Life*, 19(4), 1711–1733. <https://doi.org/10.1007/s11482-024-10304-1>
8. Chilufya, E. M., Arvola, M., Severinsson, S., Bylund, A. M., Stenliden, L., Mortazavi, A., & Ziemke, T. (2025). The BookBot project: Conceptual design of a social robot facilitating reading motivation. In A. Folstad, S. Papadopoulos, T. Araujo, E. L. Law, E. Luger, S. Hobert, & P. B. Brandtzaeg (Eds.), *Chatbots and Human-Centered AI, Conversations 2024* (Vol. 15545, pp. 132–149). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-031-88045-2_9
9. Dinotola, S., & Testa, R. (2025). Transmedia library shelf experience: Innovative research approaches. *Journal of Librarianship and Information Science*. <https://doi.org/10.1177/09610006241310908>
10. Dirgar, E., Berse, S., Bor, N. A., & Tosun, B. (2024). The effect of concept mapping in nursing education on critical thinking motivation: A path analysis. *Nurse Education*

- Today, 143, 106386. <https://doi.org/10.1016/j.nedt.2024.106386>
11. Feng, Z., Liu, H., Zhang, W., Liu, Y., Xiong, N., Chen, W., Yang, J., Wu, X., & Dai, W. (2025). Associations of lifestyle behaviors with anxiety and depression in Type 2 Diabetes Mellitus patients with hypertension: A cross-sectional study. *Psychiatric Quarterly*. <https://doi.org/10.1007/s11126-025-10136-8>
 12. Garcia-Gonzalez, M., Velazquez-Burgos, R., Veliz, S., Oyarzun, J. D., & Saona, I. (2023). Discursive fluctuations of the school library program in Chile: A critical policy analysis. *Libri-International Journal of Libraries and Information Studies*, 73(3), 199–208. <https://doi.org/10.1515/libri-2023-0005>
 13. Gonzalez, J. E., Kim, H., Anderson, J., & Pollard-Durodola, S. (2024). The effects of a science and social studies content rich shared reading intervention on the vocabulary learning of preschool dual language learners. *Early Childhood Research Quarterly*, 66, 34–47. <https://doi.org/10.1016/j.ecresq.2023.08.011>
 14. Grams, K. (2025). An increase in academic ebook preferences: A decade comparison of ebook use versus non-use. *Evidence Based Library and Information Practice*, 20(2), 92–95. <https://doi.org/10.18438/ebliip30732>
 15. Hiebert, E. H. (2023). Thinking through research and the science of reading. *Phi Delta Kappan*, 105(2), 37–41. <https://doi.org/10.1177/00317217231205940>
 16. Hiidenmaa, P., Lindh, I., Linko, M., Suomalainen, R., & Tossavainen, T. (2024). Reading culture as shared ethos: A study of Finnish self-identified readers. *Poetics*, 105, Article 101912. <https://doi.org/10.1016/j.poetic.2024.101912>
 17. Jones, B., & Lynch, J. (2023). Teachers' and Black students' views on the incorporation of African American children's literature in an after-school book club: Collaborative and culturally based learning. *Literacy*, 57(3), 209–220. <https://doi.org/10.1111/lit.12326>
 18. Kadriu-Citaku, S., & Malaj-Abdyli, F. (2024). Literature in primary schools in Kosovo, its artistic and educational character. *Eurasian Journal of Educational Research*, (110), 281–296. <https://doi.org/10.14689/ejer.2024.110.17>
 19. Lee, J., Coffie, B., Davis, K. E., Warren, C., & Zhang, T. (2023). Promising practices to promote physical activity and healthy eating among Head Start caregivers and their children. *American Journal of Lifestyle Medicine*. <https://doi.org/10.1177/15598276231221512>
 20. Lerasimskiy, Y. Y., Ierusalimskaya, S. Y., & Haritonov, G. P. (2023). Parish libraries of the Yaroslavl diocese in the second half of the 19th–early 20th centuries. *Vestnik Tomskogo Gosudarstvennogo Universiteta Istoriya-Tomsk State University Journal of History*, (85), 39–46. <https://doi.org/10.17223/19988613/85/5>
 21. Lewis, N. A., Yoneda, T., Melis, R. J. F., Mroczek, D. K., Hofer, S. M., & Muniz-Terrera, G. (2023). Availability of cognitive resources in early life predicts transitions between cognitive states in middle and older adults from Europe. *Innovation in Aging*, 7(9), igad124. <https://doi.org/10.1093/geroni/igad124>
 22. Li, H., Majumdar, R., & Ogata, H. (2025). Self-directed extensive reading with social support: Effect on reading and learning performance of high and low English proficiency students. *Research and Practice in Technology Enhanced Learning*, 20, Article 25. <https://doi.org/10.58459/rptel.2025.20025>
 23. Mohammed, S., Afaya, A., & Abukari, A. S. (2023). Reading, singing, and storytelling: The impact of caregiver-child interaction and child access to books and preschool on early childhood development in Ghana. *Scientific Reports*, 13(1), Article 13751. <https://doi.org/10.1038/s41598-023-38439-5>
 24. Nausa, R., Zivkovic, J., & Sichko, L. (2024). Needs analysis and design of a master's level academic reading course in English. *Profile-Issues in Teachers Professional Development*, 26(2), 89–106. <https://doi.org/10.15446/profile.v26n2.110905>
 25. Nenonen, T., Kouvonen, A., Harkanen, T., & Martelin, T. (2025). Cultural activities and all-cause mortality among Finnish adults: A 19-year follow-up. *BMC Public Health*, 25(1), 1906. <https://doi.org/10.1186/s12889-025-23032-4>
 26. O'Brien, B. A., Arshad, A., & Ng, S. C. (2024). Bilingual families and the home literacy environment: An examination of English and ethnic language activities and outcomes. *Journal of Research in Reading*, 47(4), 580–600. <https://doi.org/10.1111/1467-9817.12465>
 27. Otani, S., Sato, K., & Kondo, N. (2025). Public libraries and functional disability: A cohort study of Japanese older adults. *SSM-Population Health*, 29, 101762. <https://doi.org/10.1016/j.ssmph.2025.101762>
 28. Pollard-Durodola, S. D., Gonzalez, J. E., Saenz, L., & Davis, H. S. (2024). The sustainability of content enriched shared book reading vocabulary practices and preschool emergent bilinguals. *International Journal of Bilingual Education and Bilingualism*, 27(3), 315–331. <https://doi.org/10.1080/13670050.2023.2174372>
 29. Rodriguez-Gallego, M. R., De-Cecilia-Rodriguez, C., Domene-Martos, S., & Ordóñez-Sierra, R. (2024). Booktuber as a literary practice in secondary education: A systematic review. *Ocnos-Journal of Reading Research*, 23(1), Article e22549099. <https://doi.org/10.18239/ocnos.2024.23.1.405>
 30. Shen, X., Wang, J., Chen, J., Zhang, H., Shen, S., & Zhao, X. (2024). Relationship between participation in leisure activities and the maintenance of successful aging in older Chinese adults: A 4-year longitudinal study. *BMC Geriatrics*, 24(1), Article 989. <https://doi.org/10.1186/s12877-024-05574-1>
 31. Shigemasa, K., Kono, M., Ikemoto, S., & Akabayashi, H. (2024). Causal effect of parental reading on later development of children: Demonstrating a Bayesian approach. *British Journal of Developmental Psychology*, 42(3), 293–304. <https://doi.org/10.1111/bjdp.12482>
 32. Sianturi, M., & Hurit, A. A. (2024). “I want to read this book again!” Decolonizing children's literature to support

- indigenous children in reading and mathematics learning. *Journal of Intercultural Studies*, 45(2), 338–362. <https://doi.org/10.1080/07256868.2023.2247345>
33. Sirisena, M., Lhussier, M., Kaner, E., Wearn, A., Gray, J., James, R., & Redgate, S. (2024). “The book's a conversation starter”: A realist exploration of the salutogenic potential of reading for pleasure. *Medical Humanities*, 50(3), 504–512. <https://doi.org/10.1136/medhum-2023-012880>
 34. Srivastava, P., & Srivastava, M. (2023). Prevalence and correlates of loneliness in the later life—insights from Longitudinal Ageing Study in India (LASI) Wave-1. *Indian Journal of Psychiatry*, 65(9), 914–921. <https://doi.org/10.4103/indianjpsychiatry.indianjpsychiatry.594.22>
 35. Sun, H., & Xi, Y. (2025). Reading roots of resilience: The role of childhood reading environments in closing the educational gap in later-life cognitive functioning. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*, 80(8), Article gbaf128. <https://doi.org/10.1093/geronb/gbaf128>
 36. Tani, Y., Kawahara, T., Sugihara, G., Machida, M., Amagasa, S., Murayama, H., Inoue, S., Fujiwara, T., & Shobugawa, Y. (2024). Childhood book availability helps to preserve cognitive function in older adults with low education: Results from the NEIGE study. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*, 79(6), Article gbae052. <https://doi.org/10.1093/geronb/gbae052>
 37. Tran, T. T., Snasel, V., & Nguyen, L. T. (2023). Combining social relations and interaction data in recommender system with graph convolution collaborative filtering. *IEEE Access*, 11, 139759–139770. <https://doi.org/10.1109/ACCESS.2023.3340209>
 38. Trotsuk, I. (2023). A book on economics for sociological reading. *Journal of Economic Sociology-Ekonomicheskaya Sotsiologiya*, 24(1), 133–148. <https://doi.org/10.17323/1726-3247-2023-1-133-148>
 39. Wang, Z., & Deng, R. (2024). Study on user intentions for a sustainable community idle book recycling and reading service system from the perspective of value co-creation: A design proposal and validation based on Chinese community residents. *Environment Development and Sustainability*. <https://doi.org/10.1007/s10668-024-05781-1>
 40. Webber, C., Bohan, K., Collins, H., Ewens, J., Wilkinson, K., & McGeown, S. (2025). Beyond books: High school librarians as champions of pupil inclusivity, autonomy, and reader development. *Journal of Librarianship and Information Science*, 57(3), 674–686. <https://doi.org/10.1177/09610006241230101>
 41. Westerveld, M. F., Malone, S. A., Clendon, S., Bowen, R., Hayley, G., & Paynter, J. (2024). The home literacy environment of school-age autistic children with high support needs. *Journal of Applied Research in Intellectual Disabilities*, 37(5), e13284. <https://doi.org/10.1111/jar.13284>
 42. Wu, R., Liu, F., Situ, X., & Huang, W. (2023). “Reading a book, by a gain in your wit”—How to cultivate high school students' sustainable reading: A multi-factor analysis. *Heliyon*, 9(12), Article e23333. <https://doi.org/10.1016/j.heliyon.2023.e23333>
 43. Yang, H.-T., Liao, C.-Y., Syu, C.-M., & Chan, T.-W. (2023). Exploring the impact on student reading preferences in gamified reading portfolio. In 31st International Conference on Computers in Education (ICCE 2023), Vol. I (pp. 704–706). Singapore: IEEE. <https://doi.org/10.1109/ICCE58068.2023.00111>
 44. Yu, L., Huang, J., Liu, P. D., Yeung, S. S.-S., Lin, D., Cheung, H., & Tong, X. (2024). How parenting styles affect the development of language skills and reading comprehension in primary school students. *British Journal of Educational Psychology*, 94(4), 1245–1270. <https://doi.org/10.1111/bjep.12718>
 45. Zheng, W. (2024). Reading recommendation technology in digital libraries based on readers' social relationships and readers' interests. *International Journal of Advanced Computer Science and Applications*, 15(7), 689–697. <https://doi.org/10.17223/15617793/505/2>