Temporal Topic Modeling of Netflix Descriptions Using TF-IDF and NMF to Map the Evolution of Digital Storytelling Themes

Tri Wahyuningsih[1,*], Andhika Rafi Hananto[2]

[1] Department of Computer Science, Faculty of Information Technology, Universitas Kristen Satya Wacana, Salatiga, Indonesia

[2] Master Program in Computer Science, Department of Computer Science and Electronics, Faculty of Mathematics and Natural Science, Universitas Gadjah Mada,Yogyakarta, Indonesia

ABSTRACT

This study investigates the thematic evolution of storytelling within Netflix's global catalog between 1990 and 2021 through a computational text analysis framework. Using TF-IDF vectorization and Non-Negative Matrix Factorization (NMF), the research identifies latent narrative patterns from plot descriptions and traces their temporal trajectories to map the digital zeitgeist of streaming media. The dataset, derived from Netflix's public metadata, was preprocessed to remove noise and standardized through tokenization and lemmatization. Vocabulary pruning produced a corpus of 2,516 unique terms, which served as the input for topic modeling. After evaluating model coherence and interpretability, a ten-topic configuration was selected as optimal. The extracted themes encompassed crime and action, adventure and fantasy, documentary and biography, romance and youth, family life, and comedy. Temporal aggregation and visualization revealed a clear diversification of thematic focus over three decades: early periods were dominated by crime and domestic dramas, while later years showed the rise of documentary and fantasy narratives alongside enduring romantic motifs. The findings highlight how Netflix's content strategy mirrors broader cultural shifts, balancing realism, escapism, and emotional intimacy. Although limited by its focus on short textual summaries, the study demonstrates the capability of unsupervised topic modeling to capture evolving storytelling trends and provides an empirical basis for understanding how global streaming narratives reflect and shape cultural consciousness in the digital age.

Keywords Netflix; Topic Modeling; TF-IDF; Narrative Evolution; Digital Zeitgeist

Submitted 6 October 2025
Accepted 6 November 2025
Published 1 December 2025

*Corresponding author
Tri Wahyuningsih,
982022001@student.uksw.edu

Additional Information and Declarations can be found on page 296

Introduction

The emergence of streaming platforms has fundamentally transformed global cultural production, positioning services like Netflix at the forefront. As a dominant global content hub, Netflix not only influences storytelling patterns but also shapes viewer preferences and genres through its extensive library of original productions [1], [2]. The platform engages in a thoughtful curation of local cultures, packaging content in ways intended to resonate with diverse global audiences while simultaneously addressing authenticity concerns [3], [4]. Accompanying this growth is a substantial accumulation of metadata and descriptions, which serve as a rich textual corpus reflecting socio-cultural trends. This metadata not only aids in content discovery but also provides insights into consumption behaviors and the evolving dynamics of cultural narratives online [5]. Netflix's strategic embedding of diverse cultural stories showcases the platform's role as a pivotal player in the ongoing globalization of media production [6].

How to cite this article: T.Wahyuningsih and A.R.Hananto, "Temporal Topic Modeling of Netflix Descriptions Using TF-IDF and NMF to Map the Evolution of Digital Storytelling Themes," *J. Digit. Soc.*, vol. 1, no. 4, pp. 287-298, 2025.

In a digital context, the term zeitgeist refers to the collective themes, attitudes, and sentiments prevalent during a particular era as manifested through media content [7]. As streaming platforms like Netflix reshape narrative forms, computational methods can be employed to quantify and visualize cultural evolution within these texts, revealing how storytelling adapts to audience preferences and societal changes over time. These methods include text analysis and data visualization techniques that allow researchers to track thematic trends and shifts in narrative structures [7]. The capacity to analyze large-scale textual data has opened new pathways for exploring cultural evolution with a degree of precision that traditional qualitative methods could not easily achieve.

This study aims to map the thematic evolution of storytelling across time, illustrating how media reflects and shapes cultural values. By examining the interplay between viewer engagement and content production, particularly within the streaming paradigm, it becomes possible to identify how Netflix's catalog mirrors broader societal dynamics. Such a temporal perspective not only captures changes in genre popularity but also uncovers shifts in the emotional and ideological tones of storytelling. Through this lens, digital narratives can be interpreted as evolving reflections of contemporary cultural consciousness [8].

Prior studies examining the evolution of narrative themes on streaming platforms like Netflix have often focused on viewership data and ratings, rather than performing textual analyses of the content itself [9]. Moreover, they typically rely on manual genre classifications, which can be subjective and lack the granularity necessary for a comprehensive understanding [10]. Many analyses are conducted as one-time snapshots, missing the opportunity to explore how narratives shift over time [11]. While these approaches reveal audience engagement patterns, they do not adequately capture the underlying thematic evolution embedded within the media text.

Although recent research has begun to employ computational approaches for cultural analysis, few studies have systematically utilized unsupervised topic modeling to examine the dynamic evolution of Netflix's narrative themes [12]. Topic modeling enables the automatic discovery of latent thematic structures within large corpora, allowing for data-driven interpretations of how stories change across decades. This gap suggests an underexplored potential for understanding media as a living dataset that evolves with its audience and cultural environment. Addressing this limitation can enrich the understanding of how global streaming narratives both respond to and influence shifting cultural norms.

The current research builds upon this opportunity by employing TF-IDF vectorization and Non-Negative Matrix Factorization (NMF) to analyze Netflix title descriptions across multiple release years. These methods collectively enable the extraction of coherent thematic topics and the visualization of their temporal trajectories. By focusing on textual metadata rather than viewership statistics, this study provides a fresh perspective on the cultural evolution embedded in narrative content itself. The combination of text mining and temporal trend analysis thus offers a more nuanced lens to observe how streaming platforms mediate the relationship between global culture and digital storytelling.

Accordingly, this study sets out three core objectives: (1) to extract latent narrative topics from Netflix title descriptions using NMF; (2) to trace how these topics evolve across release years; and (3) to interpret the socio-cultural implications of these temporal shifts in storytelling. These objectives align with the broader goal of quantifying cultural evolution in digital media while maintaining interpretive depth. The resulting thematic maps are expected to highlight not only shifts in popular genres but also transitions in the emotional and moral vocabulary that define the streaming era's narrative ecosystem.

Ultimately, by integrating computational text analysis with cultural interpretation, this research contributes to both the computational social science and digital humanities domains. It demonstrates how algorithmic tools can illuminate macro-level cultural transformations, offering a replicable framework for future studies of digital storytelling. Through its focus on Netflix as a microcosm of the global entertainment industry, the study underscores how data-driven analysis can deepen our understanding of cultural production, audience adaptation, and the broader contours of the digital zeitgeist.

Literature Review

Evolution of Storytelling in the Streaming Era

Studies on content diversity and genre expansion in streaming platforms have highlighted significant transformations in how narratives are produced, classified, and consumed. The rise of streaming services has disrupted traditional broadcasting models, allowing for unprecedented experimentation with genre boundaries and storytelling techniques. Research [13] observes that streaming media's algorithmic infrastructure has redefined genre classification, facilitating personalized content delivery while simultaneously diversifying available narratives. This evolution has enabled the coexistence of global blockbusters and localized storytelling, reflecting the platform's responsiveness to heterogeneous audience demands.

Building on this perspective, Lobato's analysis, as interpreted by [14], underscores the strategic use of viewer data in shaping narrative structures on Netflix. By leveraging algorithmic recommendation systems, Netflix fosters a participatory form of curation in which user preferences actively inform production and distribution decisions. This approach transcends conventional genre labels by creating dynamic, data-driven narrative clusters that blur the boundaries between fiction categories. As a result, audiences experience a more fluid and customized form of storytelling that reflects both global and local cultural dynamics. Collectively, these studies illustrate how advancements in technology and data analytics significantly influence the evolution of storytelling within the streaming ecosystem.

Text Mining and Topic Modeling in Media Studies

Topic modeling has emerged as a pivotal computational technique for uncovering latent themes within large-scale textual datasets, enabling researchers to analyze complex narrative corpora in systematic and reproducible ways [15]. Through statistical inference, topic modeling identifies co-occurring word patterns that represent hidden thematic structures within a corpus. Two prominent approaches dominate this field: Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF). LDA, a probabilistic generative model, is particularly effective for large and heterogeneous text collections, offering interpretable topic distributions that support qualitative interpretation. However, as [16] note, its performance tends to decline when applied to short, descriptive texts—such as film synopses—where context is limited and word co-occurrence patterns are sparse.

In contrast, NMF provides an additive and deterministic approach, producing stable topic decompositions that are especially suitable for descriptive or domain-specific corpora [17]. NMF's capacity to represent each document as a combination of coherent topic components allows for more interpretable and semantically meaningful results in media studies applications. This stability makes it particularly advantageous when analyzing streaming metadata, where text length is limited but linguistic variation is high. Overall, both methods offer distinct advantages: LDA excels in exploring broader thematic structures across vast datasets, while NMF offers sharper interpretability and consistency for compact, content-rich text sources. The choice between them ultimately depends on the research goals and the textual nature of the dataset under analysis.

Temporal Topic Modeling Approaches and Methodological Considerations

Longitudinal topic analysis plays a vital role in revealing how cultural and thematic patterns evolve over time by systematically examining narrative shifts within text corpora. Recent frameworks proposed by [18] extend topic modeling into dynamic or time-segmented analyses, allowing researchers to trace how dominant topics emerge, fade, or transform across specific temporal intervals. These methods provide valuable insights into how societal change, audience behavior, and production priorities influence the content landscape. While temporal topic modeling has been successfully applied to domains such as film, literature, and online discourse, its application to streaming platforms remains comparatively underexplored. This scarcity of streaming-specific studies highlights the potential to uncover new dimensions of narrative evolution within the rapidly growing digital entertainment ecosystem.

Methodologically, Term Frequency–Inverse Document Frequency (TF-IDF) remains a robust baseline for feature extraction in text mining, effectively weighting terms based on their relative importance within a corpus [19]. When paired with NMF, TF-IDF enables the decomposition of textual data into meaningful thematic structures that capture the distinctiveness of narrative elements across years. The interpretability and additive structure of NMF make it especially appropriate for analyzing short, descriptive metadata typical of Netflix titles [20]. Furthermore, topic coherence serves as an essential evaluation metric to determine the optimal number of topics (k), ensuring that each extracted theme maintains internal consistency and semantic clarity [21]. Together, these methodological strategies form a reliable foundation for conducting temporal topic modeling, enabling a systematic mapping of thematic evolution across streaming narratives.

Method

This study adopts a computational text analysis framework to investigate the temporal evolution of storytelling themes in Netflix's global catalog. By combining textual metadata with unsupervised machine learning, the research aims to quantify how narrative themes shift across years and to visualize these trends as indicators of the digital zeitgeist. The analytical process follows a sequential structure encompassing data preparation, text preprocessing, feature extraction through TF-IDF, topic modeling with Non-Negative Matrix Factorization (NMF), thematic interpretation, and temporal trend visualization.

Data Selection and Preparation

The dataset utilized in this study is the publicly available netflix_titles.csv, which contains metadata describing thousands of films and television series available on the platform. Two attributes were selected as the core analytical variables: description, serving as the textual corpus, and release_year, representing the temporal dimension of each title. To ensure data quality and temporal consistency, entries containing missing or anomalous values were removed. Additionally, titles released prior to 1990 were excluded to focus on the modern streaming era, where digital media production and global distribution accelerated significantly. This filtering ensured sufficient annual data density and allowed for a longitudinal analysis that aligns with the growth of Netflix as a cultural and narrative producer.

Text Preprocessing

Text preprocessing was conducted to transform the raw descriptions into a standardized and analyzable corpus. Each text was first converted to lowercase to eliminate case sensitivity inconsistencies. Punctuation marks, numbers, and non-alphabetic characters were stripped from the corpus to reduce noise and prevent distortion in term frequency counts. Stop words—common words such as “the,” “is,” and “at”—were removed using a standard English stop word list to enhance semantic precision. A domain-specific stop-word list was also developed to exclude recurring but uninformative terms such as “movie,” “series,” “season,” and “Netflix.” After cleaning, the remaining words were tokenized into individual terms, and lemmatization was applied to reduce each token to its base or dictionary form (e.g., studies → study). These steps ensured that conceptually related words were grouped together, enhancing the coherence of later topic modeling.

Feature Engineering through TF-IDF Vectorization

To prepare the cleaned corpus for machine learning analysis, textual data were converted into numerical form using the Term Frequency–Inverse Document Frequency (TF-IDF) method. This approach assigns weights to terms based on their frequency in a document relative to their prevalence across the corpus, giving prominence to distinctive words while minimizing the influence of overly common ones. The process generated a Document–Term Matrix (DTM), where each row represents a document (Netflix title description) and each column corresponds to a unique word in the vocabulary. To optimize computational efficiency and improve signal clarity, vocabulary pruning was applied by setting a minimum document frequency threshold to exclude rare words and a maximum document frequency limit to remove overly frequent ones. The resulting matrix provided a robust, high-dimensional representation of textual data suitable for unsupervised topic modeling.

Unsupervised Topic Modeling with Non-Negative Matrix Factorization (NMF)

The extracted TF-IDF matrix was analyzed using NMF, an unsupervised learning algorithm widely used in text mining for its interpretability and stability. NMF decomposes the document–term matrix into two non-negative matrices: one representing document-to-topic relationships and another representing topic-to-word distributions. This additive structure enables clear semantic interpretation, as each topic can be viewed as a distinct component that contributes to the meaning of each document. To determine the optimal number of topics (k), the model was trained iteratively with varying k values ranging from 5 to 50. For each candidate model, topic coherence scores were computed to evaluate semantic consistency, while manual inspection of the top keywords for each topic was conducted to assess interpretability and thematic distinction. The value of k yielding the highest coherence and least thematic overlap was selected for the final model, which was subsequently trained on the complete dataset to ensure stability.

Topic Labeling and Interpretation

After the final NMF model was trained, the top 15–20 words with the highest weights within each topic were examined to identify the dominant narrative elements. These clusters of words were then manually labeled with intuitive and human-readable titles such as True Crime and Investigation, Teen Romance, or Dystopian Futures. This interpretive process translated machine-derived statistical groupings into culturally meaningful categories that represent recurring storytelling patterns across Netflix content. Each document was assigned topic weights that quantified its degree of association with each thematic cluster. This mapping of documents to topic distributions provided a foundation for longitudinal aggregation, allowing the subsequent examination of thematic prevalence across different release years.

Temporal Trend Analysis

To capture the evolution of storytelling over time, topic-weight data were merged with the release_year variable. Titles were grouped by year, and for each topic, the mean topic weight was calculated to represent its relative prominence in that particular year. The resulting temporal dataset was used to generate stacked area and line plots visualizing the fluctuation of thematic prevalence across time. These visualizations illustrate the rise, dominance, and decline of specific narrative motifs, revealing, for instance, how dystopian and science fiction themes intensified during the mid-2010s or how family-oriented and inspirational narratives regained prominence in the early 2020s. Such representations provide both macro-level and fine-grained insights into how streaming-era storytelling responds to changing audience sensibilities and cultural contexts.

Result and Discussion

Model Performance and Topic Optimization

After the preprocessing and feature extraction stages, the final TF-IDF vectorization produced a vocabulary of 2,516 unique terms, representing the most meaningful and discriminative tokens from the Netflix title descriptions. The Non-Negative Matrix Factorization (NMF) model was then trained iteratively with topic numbers ranging from k = 5 to k = 15 to identify the configuration that yielded the most coherent and distinct themes. Models with fewer than eight topics showed substantial thematic overlap, while those with more than twelve produced fragmented clusters with weak interpretability. Through a combination of quantitative and qualitative evaluation, the ten-topic configuration (k = 10) demonstrated the highest topic coherence and clearest conceptual separation. This optimal balance ensured that the resulting model captured the diversity of Netflix narratives while maintaining internal semantic consistency. The ten-topic model was therefore selected for final analysis, forming the foundation for thematic labeling and temporal mapping.

Identified Topics and Thematic Overview

The ten-topic model generated a rich and interpretable structure that reflects the narrative diversity within Netflix's global catalog. Each topic represented a distinct thematic domain emerging from the data-driven analysis of the title descriptions. The first topic consisted predominantly of words such as man, young, cop, drug, revenge, and mission, representing crime and action-oriented stories focused on heroism, law enforcement, and personal redemption. The second topic captured adventure and science fiction narratives characterized by team, battle, power, evil, and force, aligning with fantasy-driven and heroic arcs. The third topic represented documentary and biographical storytelling, highlighting words such as life, film, history, personal, and story, pointing to factual and reflective content. A fourth cluster centered on teenage and romantic dramas, including terms like love, student, relationship, and high school, reflecting themes of youth and emotion.

The fifth topic corresponded to urban life and self-discovery, defined by city, job, move, and start, symbolizing aspiration and adaptation. The sixth cluster reflected friendship and exploration, emphasizing adventure, friend, childhood, and journey, which captured themes of companionship and discovery. The seventh topic focused on humor and performance, with strong associations to comedy, standup, stage, and live, mirroring Netflix's investment in stand-up specials and variety content. The eighth topic revolved around family narratives, showing frequent terms such as father, mother, son, and home, which underscore familial bonds and domestic struggles. The ninth cluster dealt with love and emotional journeys, marked by dream, journey, heart, and form, depicting inspirational or transformative experiences. Finally, the tenth topic captured mystery and true-crime narratives with terms like murder, detective, killer, and truth, illustrating an enduring fascination with suspense and investigation. Collectively, these ten thematic domains highlight the breadth of Netflix's storytelling landscape, encompassing both escapist and intimate forms of narrative expression.

Temporal Evolution of Storytelling Themes

To trace the progression of narrative themes over time, the topic-weight matrix produced by the NMF model was merged with the release_year variable, and the mean weight of each topic was computed for every year between 1990 and 2021. The results were visualized in Figure 1. The visualization revealed several distinct temporal phases in Netflix's thematic evolution. In the early years (1990–2005), the platform's catalog was dominated by crime, family, and drama-oriented content, reflecting its reliance on pre-existing licensed productions from Hollywood and traditional television. During this period, genres emphasizing heroism, justice, and family bonds maintained the highest proportion of thematic focus, while lighter or experimental narratives appeared less frequently.

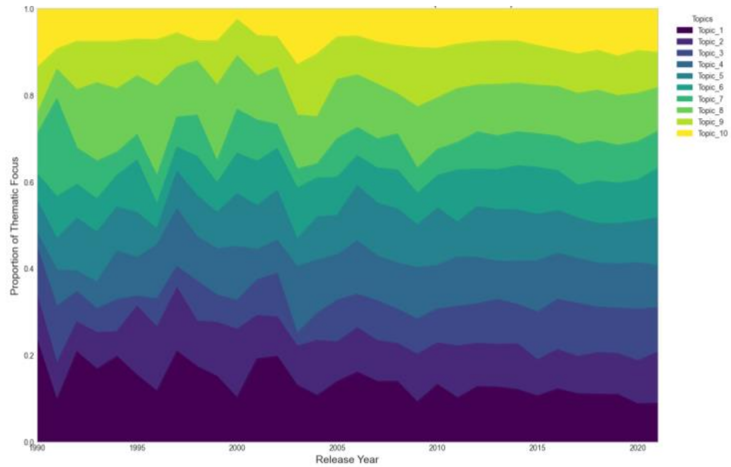

Figure 1 Thematic Evolution of Netflix Content (1990-2021)

From 2005 to 2015, the thematic composition began to diversify markedly. The rise of documentary and biographical storytelling was evident through an increased proportion of content related to real-life narratives, historical retrospectives, and personal journeys. This trend coincided with Netflix's transition from a distribution service to a producer of original content. Similarly, comedy and stand-up performance topics gained visibility during this phase, marking the platform's growing role in popularizing live entertainment and comedian-centered specials. At the same time, romantic and coming-of-age dramas maintained steady representation, suggesting that emotional storytelling remained a core audience draw despite changes in production strategy.

In the post-2015 streaming era, new patterns emerged as Netflix expanded globally and embraced more diverse storytelling formats. Adventure and fantasy-driven topics experienced a significant rise, likely reflecting the success of high-budget original series and visually rich speculative fiction. Simultaneously, family-centric and romantic journey topics sustained a stable presence, indicating Netflix's consistent investment in universally relatable themes. The overall landscape from 2015 onward thus demonstrated a balanced coexistence between large-scale, visually ambitious genres and smaller, emotionally intimate narratives. The visualization's layered structure clearly illustrates how thematic diversity expanded in parallel with Netflix's transformation into a global content producer.

Thematic Balance and Cultural Implications

The longitudinal analysis underscores Netflix's adaptive storytelling strategy, which mirrors broader shifts in audience preference and cultural sensibility. The sustained prominence of crime and action narratives across all periods reflects the enduring audience appetite for suspense, moral conflict, and resolution. The concurrent rise of documentary and personal storytelling in the late 2000s aligns with a global turn toward authenticity, where viewers increasingly value real-life experiences and emotionally grounded narratives. Similarly, the persistent presence of romantic and family-related themes highlights the continued importance of interpersonal relationships as universal storytelling anchors.

The interplay between realism and escapism—between factual authenticity and imaginative spectacle—emerges as a defining feature of Netflix's narrative evolution. As digital audiences navigate complex global realities, Netflix's catalog appears to mediate this tension by offering both introspective reflection through documentary formats and emotional catharsis through fictional universes. This duality captures the essence of the digital zeitgeist: a cultural environment characterized by simultaneous desires for information, identification, and immersion. The thematic balance observed across decades therefore reflects not only changing production strategies but also shifting patterns in global cultural consumption.

Limitations

Although the present study provides a comprehensive examination of thematic evolution within Netflix's catalog, several methodological limitations must be acknowledged. First, the analysis relies exclusively on plot descriptions rather than full scripts or subtitles, which means the extracted topics reflect marketing-oriented summaries rather than the complete narrative content. This limitation could result in partial representations of complex storylines or character development. Second, the dataset is platform-specific and limited to Netflix, excluding content from competing services such as Amazon Prime Video or Disney+, thereby constraining the generalizability of the results across the broader streaming ecosystem. Third, topic modeling techniques like NMF are model-dependent; the topics generated are not absolute categories but probabilistic abstractions that require interpretive labeling. Finally, temporal aggregation was performed at the yearly level, which may smooth over more granular shifts in thematic emphasis, particularly for years with smaller sample sizes. Despite these constraints, the approach remains a robust and interpretable framework for tracing cultural and narrative trends in streaming media.

Future Research Directions

Future research can expand on these findings by incorporating multi-platform comparisons to capture cross-ecosystem narrative dynamics in the global streaming industry. Analyzing script-level or subtitle-based corpora would allow for deeper semantic modeling and a more accurate reflection of character dialogue, emotional tone, and narrative complexity. Integrating sentiment analysis and emotion recognition models could further reveal how narrative tone evolves alongside thematic content. Additionally, employing dynamic topic models or transformer-based embeddings (such as BERT or BERTopic) would provide a more nuanced view of temporal topic shifts and contextual dependencies. Future studies may also investigate the relationship between audience engagement metrics and thematic prevalence to understand how content consumption patterns reinforce or challenge existing cultural narratives. By combining advanced linguistic modeling with cross-platform data, future research can extend this study's contribution toward a more holistic understanding of the digital zeitgeist.

Conclusion

This study examined the thematic evolution of Netflix content between 1990 and 2021 using TF-IDF vectorization and Non-Negative Matrix Factorization (NMF). The findings revealed ten dominant thematic clusters encompassing genres such as crime, adventure, documentary, romance, family, and comedy, each reflecting distinct storytelling modes within the streaming ecosystem. The temporal analysis showed that while early Netflix content was dominated by crime and family-oriented narratives, later years witnessed increasing diversity with the emergence of documentary, fantasy, and stand-up comedy genres. The coexistence of factual and fictional storytelling suggests that Netflix balances realism and escapism to align with evolving audience preferences. By mapping these longitudinal shifts, this study demonstrates the potential of computational text analysis to uncover cultural and narrative transformations within large-scale media datasets. The results highlight Netflix's dual role as both a mirror of global cultural trends and a mediator that shapes them through algorithmic curation and production strategy. Ultimately, this research underscores how data-driven approaches can bridge computational methods and cultural theory to reveal how the digital zeitgeist unfolds through the stories people watch.

Declarations

Author Contributions

Conceptualization: T.W.; Methodology: A.R.H.; Software: T.W.; Validation: A.R.H.; Formal Analysis: T.W.; Investigation: A.R.H.; Resources: T.W.; Data Curation: A.R.H.; Writing Original Draft Preparation: T.W.; Writing Review and Editing: T.W.; Visualization: A.R.H.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D. R. Idiz, J. Noordegraaf, and R. Vliegenthart, "Culture as Window Dressing? A Threefold Methodological Framework for Researching the Locality of Netflix Series," *Crit. Stud. Telev. Int. J. Telev. Stud.*, 2024, doi: 10.1177/17496020241235579.

[2] P. Katerynych, V. Goian, and O. Goian, "Exploring the Evolution of Storytelling in the Streaming Era: A Study of Narrative Trends in Netflix Original Content," *Ct*, 2023, doi: 10.34135/communicationtoday.2023.vol.14.no.2.3.

[3] B. M. Han, D. Y. Jin, and K. Yoon, "The Lure of Cultural Authenticity: Netflix

and Speculative Koreanness in the Global Media Market," *Telev. New Media*, 2024, doi: 10.1177/15274764241303735.

[4] J. F. Khalil, "The Netflix Paradox in the Middle East: Diversity, Inclusivity, and Authenticity?," *Telev. New Media*, 2024, doi: 10.1177/15274764241285563.

[5] D. Warch, P. Stellbauer, and P. Neis, "Advanced Techniques for Geospatial Referencing in Online Media Repositories," *Future Internet*, 2024, doi: 10.3390/fi16030087.

[6] A. Afilipoaie, C. Iordache, and T. Raats, "The 'Netflix Original' and What It Means for the Production of European Television Content," *Crit. Stud. Telev. Int. J. Telev. Stud.*, 2021, doi: 10.1177/17496020211023318.

[7] M. Puohiniemi and M. Verkasalo, "Zeitgeist Effects, Fragmentation of Media Use, and Value Consensus," *J. Soc. Polit. Psychol.*, 2020, doi: 10.5964/jspp.v8i1.1117.

[8] D. Hassler-Forest, "'Two Birds With One Stone': Transmedia Serialisation in ≪em>Twin Peaks</Em>;," *MC J.*, 2018, doi: 10.5204/mcj.1364.

[9] K. Watanabe and A. Baturo, "Seeded Sequential LDA: A Semi-Supervised Algorithm for Topic-Specific Analysis of Sentences," *Soc. Sci. Comput. Rev.*, 2023, doi: 10.1177/08944393231178605.

[10] R. Ventura, V. Blanco-Fernández, and J. J. Sánchez Soriano, "Queer Temporalities in TV Series Veneno: Breaking Down the Spiral of Silence Through Trans Visibility," *Miguel Hernández Commun. J.*, 2024, doi: 10.21134/mhjournal.v15i.2091.

[11] P. S. Vishnuprasad, G. Nogara, F. M. Cardoso, S. Cresci, S. Giordano, and L. Luceri, "Tracking Fringe and Coordinated Activity on Twitter Leading Up to the US Capitol Attack," *Proc. Int. Aaai Conf. Web Soc. Media*, 2024, doi: 10.1609/icwsm.v18i1.31409.

[12] R. Albalawi, T. Yeap, and M. Benyoucef, "Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis," *Front. Artif. Intell.*, 2020, doi: 10.3389/frai.2020.00042.

[13] S. Sahoo, "The Impact of Streaming Services on Youth Television Viewing Habits and Media Literacy," *Shodhkosh J. Vis. Perform. Arts*, 2024, doi: 10.29121/shodhkosh.v5.i1.2024.946.

[14] D. Susilo and H. Harliantara, "The Digital Promotion of Japanese and Korean Movie in OTT Platform by Netflix," *Indones. J. Bus. Anal.*, 2023, doi: 10.55927/ijba.v3i5.6418.

[15] M. I. ul haq, Q. Li, J. Hou, and A. Iftekhar, "Detecting the Research Structure and Topic Trends of Social Media Using Static and Dynamic Probabilistic Topic Models," *Aslib J. Inf. Manag.*, 2022, doi: 10.1108/ajim-02-2022-0091.

[16] G. Ruan and L. Xia, "Knowledge Discovery of News Text Based on Artificial Intelligence," *Acm Trans. Asian Low-Resour. Lang. Inf. Process.*, 2020, doi: 10.1145/3418062.

[17] S. J. Blair, Y. Bi, and M. Mulvenna, "Aggregated Topic Models for Increasing Social Media Topic Coherence," *Appl. Intell.*, 2019, doi: 10.1007/s10489-019-01438-z.

[18] T. Ramamoorthy, V. Kulothungan, and B. Mappillairaju, "Topic Modeling and Social Network Analysis Approach to Explore Diabetes Discourse on Twitter in India," *Front. Artif. Intell.*, 2024, doi: 10.3389/frai.2024.1329185.

[19] R. Egger and J. Yu, "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts," *Front. Sociol.*, 2022,

doi: 10.3389/fsoc.2022.886498.

[20] J. Qiang, Z. Qian, Y. Li, Y. Yuan, and X. Wu, “Short Text Topic Modeling Techniques, Applications, and Performance: A Survey,” *Ieee Trans. Knowl. Data Eng.*, 2022, doi: 10.1109/tkde.2020.2992485.

[21] P. Pavithra and Savitha, “Topic Modeling for Evolving Textual Data Using LDA, HDP, NMF, BERTOPIC, and DTM With a Focus on Research Papers,” *Joti*, 2024, doi: 10.37802/joti.v5i2.618.