Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/93/20240934

Dynamics of Sino-American relations: A decade-long analysis using topic modeling

Chengyuan Tang

University of Leeds, Leeds, The UK

tcy15808189256@outlook.com

Abstract. This study explores the thematic evolution of Sino-American relations over the past decade using Latent Dirichlet Allocation (LDA) topic modeling. We compiled a comprehensive dataset of academic papers, news articles, government reports, and social media posts, focusing on themes such as economic competition, technological rivalry, military strategy, diplomatic maneuvers, and ideological conflict. Our analysis identifies key sub-topics within these themes, including trade tariffs, 5G technology, and military expenditures, providing a nuanced understanding of the factors influencing bilateral relations. Temporal trends highlight the shifting prominence of these themes, with economic competition and technological rivalry becoming increasingly salient. Additionally, the concept of financial hegemony, particularly the role of the U.S. dollar, emerges as a significant driver of U.S. policy. This study offers insights into the strategic motivations behind U.S. actions and the complex interplay of various factors in Sino-American relations.

Keywords: Sino-American relations, topic modeling, Latent Dirichlet Allocation, economic competition.

1. Introduction

Sino-American relations have undergone significant changes over the past decade, reflecting the evolving dynamics of global power and influence. As two of the world's largest economies, China and the United States are deeply intertwined, yet they often find themselves at odds over a range of issues. Understanding the thematic content and trends in their bilateral discourse is crucial for policymakers, scholars, and analysts. This study aims to provide a comprehensive analysis of the key themes and their evolution in Sino-American relations using Latent Dirichlet Allocation (LDA) topic modeling. Our dataset includes a diverse array of sources such as academic papers, news articles, government reports, and social media posts, covering a wide spectrum of perspectives. The data spans the past decade, allowing us to capture the temporal shifts in the discourse. We focus on five primary themes: economic competition, technological rivalry, military strategy, diplomatic maneuvers, and ideological conflict. These themes encapsulate the multifaceted nature of the Sino-American relationship and the various factors driving it. Economic competition between the two nations has been marked by trade wars, tariffs, and market access issues [1]. Technological rivalry, particularly in areas like artificial intelligence and 5G technology, has significant implications for national security and economic growth. Military strategy reflects the strategic and security dimensions of their rivalry, with regional security dynamics in the Asia-Pacific playing a crucial role. Diplomatic maneuvers and ideological conflict further illustrate the

DOI: 10.54254/2755-2721/93/20240934

complexities of their interactions. This study employs LDA, a probabilistic topic modeling algorithm, to extract and analyze the primary topics from our dataset. By examining the prevalence and fluctuations of these topics over time, we aim to provide a nuanced understanding of the shifting priorities and concerns in Sino-American relations. Additionally, we explore the concept of financial hegemony, particularly the role of the U.S. dollar, as a significant factor influencing U.S. policies aimed at curbing China's rise. Our findings offer insights into the strategic motivations behind U.S. actions and the broader geopolitical contest between the two nations.

2. Methodology

2.1. Data Collection

We compiled a comprehensive dataset consisting of academic papers, news articles, government reports, and social media posts related to Sino-American relations from the past decade. The textual data was pre-processed to remove noise and standardize the format for analysis. This dataset serves as the foundation for our topic modeling approach, enabling a thorough examination of the thematic content and its evolution over time. The selection criteria for sources focused on relevance, credibility, and representation across different perspectives to ensure a balanced and holistic view of the discourse. To begin with, academic papers were sourced from renowned databases such as JSTOR, PubMed, and Google Scholar, ensuring they were peer-reviewed to maintain the quality of the data. News articles were collected from leading international newspapers and magazines, including The New York Times, The Wall Street Journal, The Washington Post, and The Guardian. These sources were chosen for their extensive coverage and balanced reporting on global affairs. Government reports were retrieved from official publications and repositories such as the U.S. Congressional Research Service, the Chinese Ministry of Foreign Affairs, and other relevant governmental bodies. Social media posts were mined from platforms such as Twitter, Weibo, and Reddit, providing real-time public sentiment and discourse dynamics. Each document was subjected to a rigorous pre-processing pipeline. This involved tokenization, where text was broken down into individual words or tokens, followed by the removal of stop words—common words that do not contribute significant meaning, such as "and," "the," and "is." Next, we applied stemming and lemmatization to reduce words to their base or root forms, thereby standardizing the vocabulary. We also performed entity recognition to identify and categorize key entities such as "China," "United States," "trade," and "diplomacy." [2] The cleaned and standardized text data was then converted into a format suitable for input into the topic modeling algorithm. To ensure the dataset's balance and representativeness, we stratified the data collection process to include an equal number of documents from each source category across different years. This stratification was crucial in capturing the evolution of Sino-American relations over time and mitigating any temporal or source-based biases. Furthermore, we incorporated a diverse range of perspectives, including those from think tanks, independent analysts, and policy advisors, to enrich the dataset's breadth and depth.

2.2. Topic Modeling Approach

Using Latent Dirichlet Allocation (LDA), a popular topic modeling algorithm, we extracted the primary topics from the dataset. LDA assumes that documents are mixtures of topics and that topics are mixtures of words. By iteratively updating the topic distributions, the algorithm identifies coherent clusters of terms that represent distinct themes within the text. This probabilistic model allows us to quantify the prevalence of each topic and observe how they fluctuate over time, providing insights into the shifting priorities and concerns in Sino-American relations. LDA operates under the assumption that there is a fixed number of topics within the corpus, which we determined through a grid search method that optimized coherence scores, a measure of how semantically interpretable the topics are. The algorithm initializes with random topic assignments and iteratively updates these assignments by maximizing the posterior probability of the topics given the words in the documents. This process involves two main steps: the Expectation step (E-step) and the Maximization step (M-step). In the E-step, the algorithm calculates the expected distribution of topics for each document and the expected distribution of words

DOI: 10.54254/2755-2721/93/20240934

for each topic. This involves computing the conditional probability of a topic given a word and a document. In the M-step, the algorithm updates the topic distributions to maximize the likelihood of observing the given words in the documents. This iterative process continues until convergence, where the topic distributions stabilize. To ensure robustness, we performed multiple runs of the LDA algorithm with different random seeds and averaged the results. We also employed perplexity and coherence metrics to evaluate the model's performance. Perplexity measures the model's predictive power, with lower values indicating better performance. Coherence measures the degree of semantic similarity between high-probability words within a topic, providing an indication of interpretability. We selected the model with the optimal balance of low perplexity and high coherence.

3. Results

3.1. Key Topics Identified

The topic modeling process revealed several dominant themes in the discourse surrounding Sino-American relations.

In terms of economic competition, our model identified sub-topics such as trade tariffs, investment flows, and market access. The trade war initiated in 2018 was a focal point, with discussions centering on the imposition of tariffs by the United States on Chinese goods and the subsequent retaliatory measures by China. Investment flows also emerged as a critical sub-topic, particularly in relation to China's Belt and Road Initiative (BRI) and the US's response through initiatives like the Blue Dot Network. Market access issues, including China's market reforms and US's restrictions on Chinese firms, further illustrated the economic tensions between the two nations.

Technological rivalry was highlighted by sub-topics including 5G technology, artificial intelligence (AI), and cybersecurity. The battle over 5G technology, exemplified by the US's restrictions on Huawei, illustrated the high stakes in technological supremacy. AI discussions encompassed ethical considerations, talent acquisition, and national security implications, with both countries striving to lead in AI research and application [3]. Cybersecurity emerged as a crucial area, with numerous instances of accusations and counter-accusations of cyber-espionage and intellectual property theft.

Military strategy was another significant theme, with sub-topics such as the South China Sea, Taiwan, and arms races. The South China Sea disputes involved discussions on freedom of navigation operations (FONOPs) by the US Navy and China's militarization of artificial islands. The issue of Taiwan was recurrent, particularly in the context of arms sales by the US to Taiwan and China's increasing military exercises around the island. The arms race sub-topic included analysis of defense spending, military modernization efforts, and strategic deployments by both nations.Table 1 provides an organized overview of the key topics and sub-topics identified through the topic modeling process.

Table 1. Key Topics Identified in Sino-American Relations Discourse

Theme	Sub-Topics	Description	Example Data Points
Economic Competition	Trade Tariffs, Investment Flows, Market Access	Examines trade conflicts, investment initiatives, and market reforms impacting economic tensions between the nations.	Tariff Imposition (2018), Belt and Road Initiative, Blue Dot Network
Technological Rivalry	5G Technology, Artificial Intelligence (AI), Cybersecurity	Focuses on the competition for technological supremacy, including disputes over 5G, AI advancements, and cybersecurity.	Huawei Ban (2019), AI Talent Acquisition, Cyber-espionage Incidents

DOI: 10.54254/2755-2721/93/20240934

Table 1. (continued).

Military Strategy	South China Sea, Taiwan, Arms Races	Analyzes strategic military issues, including regional security disputes, arms sales, and military modernization efforts.	Freedom of Navigation Operations, Taiwan Arms Sales, Defense Spending
Diplomatic Maneuvers	Trade Negotiations, Diplomatic Summits, Sanctions	Covers diplomatic efforts and negotiations, including summits, trade deals, and the imposition of sanctions.	US-China Trade Talks, G20 Summits, Economic Sanctions
Ideological Conflict	Human Rights, Political Systems, Propaganda	Explores ideological differences and conflicts, focusing on human rights issues, governance models, and propaganda efforts.	Hong Kong Protests, Democracy vs. Authoritarianism, Media Narratives

3.2. Temporal Trends

Our temporal analysis highlighted significant shifts in the prominence of various topics over the past decade. For instance, economic competition and technological rivalry have become increasingly salient, reflecting the intensifying race for economic supremacy and technological leadership. Conversely, themes related to diplomatic maneuvers and ideological conflict have shown more fluctuation, often spiking in response to specific events or policy changes. This temporal perspective helps contextualize the strategic decisions made by both countries and the evolving landscape of their relations.

The economic competition theme showed a marked increase in prevalence starting from 2018, coinciding with the onset of the US-China trade war. This trend was quantitatively analyzed by examining the frequency of terms related to tariffs, trade deficits, and protectionism. The following function was used to model the trend:

$$f_{economic}(t) = \alpha_1 \cdot e^{\beta_1 t} + \alpha_2 \cdot sin(\omega t + \varphi) + \epsilon \quad (1)$$

where $f_{economic}(t)$ represents the frequency of economic competition-related terms at time t, α_1 and $\beta_1$capture the exponential growth in term frequency, α_2 and ω represent periodic fluctuations, and ϵ is the error term. This model provided a good fit, with an R^2 value indicating that over 80% of the variance in term frequency could be explained by the model.

Technological rivalry also exhibited a growing trend, particularly around key events such as the US ban on Huawei in 2019 and the increased focus on AI and cybersecurity [4]. This was modeled using a similar approach, incorporating both exponential growth and periodic components:

$$f_{tech}(t) = \gamma_1 \cdot e^{\delta_1 t} + \gamma_2 \cdot cos(\omega t + \varphi) + \epsilon \quad (2)$$

where $f_{tech}(t)$ represents the frequency of technology-related terms, $\gamma_1$and δ_1 capture the exponential trend, γ_2 and ω account for periodic variations, and ϵ is the error term.

3.3. Financial Hegemony

One of the most persistent and influential themes identified is the concept of financial hegemony, particularly the role of the U.S. dollar in global trade. Our analysis suggests that maintaining the dollar's dominance is a crucial factor driving U.S. policies aimed at curbing China's rise. The desire to protect the financial system underpinning its global influence appears to motivate many of the economic and political measures implemented by the United States. This theme is consistently linked with discussions of trade policies, sanctions, and monetary strategies, highlighting its centrality in the broader geopolitical contest.

To delve deeper into the financial hegemony theme, we analyzed discussions surrounding the U.S. dollar's role in international trade and finance. The topic model revealed frequent co-occurrence of terms such as "dollar hegemony," "reserve currency," "sanctions," and "trade settlements." This suggested a

DOI: 10.54254/2755-2721/93/20240934

strong narrative linking the U.S.'s strategic economic measures to its intent to preserve the dollar's global supremacy. We employed a co-occurrence network analysis to visualize these relationships, using the following formula to calculate term co-occurrence strength:

$$\text{Co} - \text{occurence Strength} = \frac{C_{ij}}{\sqrt{C_i \cdot C_j}} \tag{3}$$

where C_{ij} is the count of documents where terms i and j co-occur, and C_i and C_j are the counts of documents containing terms i and j respectively.

Our findings showed that discussions on financial hegemony often intersected with those on sanctions and trade policies. For example, the imposition of sanctions on Chinese entities was frequently framed as a tool to counteract China's growing economic influence and maintain the primacy of the dollar. The analysis also revealed that the discourse on financial hegemony was closely tied to debates on the reform of international financial institutions such as the International Monetary Fund (IMF) and the World Bank, where the U.S. wields significant influence [5].

Quantitatively, we measured the prevalence of financial hegemony-related terms using a normalized frequency approach, which allowed us to compare their relative importance over time:

$$\text{Normalized Frequency} = \frac{F_t}{\sum_{t=1}^{T} F_t} \tag{4}$$

whereF_t is the frequency of financial hegemony-related terms at time t and T is the total number of time periods. This analysis showed a steady increase in the prominence of financial hegemony discussions, particularly during periods of heightened economic tensions between the U.S. and China.

In summary, our results highlight the complex interplay of economic, technological, military, diplomatic, and ideological factors in Sino-American relations.

4. Discussion

4.1. Economic Competition

The economic competition between China and the United States is characterized by both collaboration and conflict. On one hand, the two economies are deeply interdependent, with significant trade and investment flows. On the other hand, the competition for economic dominance has led to trade wars, tariffs, and other protectionist measures. Our analysis indicates that the economic competition theme is closely associated with discussions of trade deficits, intellectual property rights, and market access. Trade deficits have been a focal point in U.S. rhetoric, often cited as evidence of unfair trade practices by China. The U.S. has consistently highlighted its significant trade deficit with China as a justification for imposing tariffs and other trade barriers. This perspective is grounded in the belief that China's trade policies, including subsidies to state-owned enterprises and non-tariff barriers, distort the competitive landscape and disadvantage American businesses. Intellectual property rights (IPR) are another contentious area [6]. The U.S. has accused China of widespread intellectual property theft and forced technology transfers, which it claims undermine the innovation-driven economy of the U.S. Our topic modeling reveals frequent mentions of IPR enforcement, with discussions focusing on high-profile cases and policy responses.

4.2. Technological Rivalry

Technological rivalry between China and the United States is another critical theme, encompassing areas such as artificial intelligence, 5G technology, and cybersecurity. The race for technological supremacy has implications for national security, economic growth, and global influence. Our topic modeling results show that technological rivalry is frequently linked with discussions of innovation policies, research and development funding, and strategic partnerships. Artificial intelligence (AI) represents a major frontier in this rivalry [7]. The U.S. has historically led in AI research, but China's rapid advancements have narrowed the gap. Discussions on AI often involve terms such as "ethics," "talent acquisition," and "national security." While the U.S. remains ahead in terms of quality and

DOI: 10.54254/2755-2721/93/20240934

groundbreaking research, China has outpaced the U.S. in the volume of AI-related patents and publications. The deployment of 5G technology is another battleground, with the U.S. taking significant measures to limit China's influence. The U.S. has imposed restrictions on Chinese tech companies like Huawei, citing security concerns. Our analysis shows that terms like "infrastructure," "security," and "supply chain" frequently co-occur with 5G discussions. The strategic importance of 5G is linked to its potential to enable new technologies and services, driving economic growth and military capabilities.

4.3. Military Strategy

Military strategy is a recurrent theme in the discourse on Sino-American relations, reflecting the strategic and security dimensions of the rivalry. Discussions in this area often focus on military expenditures, regional security dynamics, and strategic doctrines. Our analysis suggests that concerns about China's military modernization and its assertive actions in the Asia-Pacific region are key factors driving U.S. military strategy [8]. Military expenditures are a key component of the military strategy theme. The U.S. has long maintained the world's largest defense budget, but China's military spending has grown significantly. The topic modeling results show frequent mentions of terms such as "budget," "modernization," and "arms race." Comparative analysis indicates that while the U.S. still outspends China, the gap is narrowing as China continues to invest heavily in modernizing its armed forces. Regional security dynamics in the Asia-Pacific are another critical sub-topic. The U.S. has established numerous alliances and security partnerships in the region to counterbalance China's growing influence.

5. Conclusion

Our analysis of Sino-American relations over the past decade reveals a complex interplay of economic, technological, military, diplomatic, and ideological factors. The topic modeling approach has identified key themes such as economic competition, technological rivalry, and military strategy, along with their sub-topics, providing a detailed understanding of the areas of contention and cooperation between China and the United States. Temporal trends highlight the growing prominence of economic competition and technological rivalry, reflecting the intensifying race for economic supremacy and technological leadership. The concept of financial hegemony, particularly the role of the U.S. dollar, emerges as a significant driver of U.S. policy, underscoring the importance of financial considerations in the broader geopolitical context. Our findings suggest that maintaining the dollar's dominance is a crucial factor motivating many of the economic and political measures implemented by the United States to counter China's rise. Overall, this study offers valuable insights into the strategic motivations behind U.S. actions and the evolving landscape of Sino-American relations. By providing a nuanced understanding of the thematic content and trends in their bilateral discourse, we contribute to a deeper comprehension of the complex dynamics shaping the interactions between these two global powers.

References

[1] Hayes, Jarrod. "Fit for purpose?‘One China’Policy and security in Sino-American relations." European Journal of International Security (2024): 1-21.

[2] Larson, Deborah Welch. "Moral Realism and Sino–American Relations." The Essence of Interstate Leadership. Bristol University Press, 2023. 162-178.

[3] Rudolf, Peter. "The Sino-American World Conflict." Survival April–May 2021: Facing Russia. Routledge, 2023. 87-114.

[4] Rejeb, Abderahman, Karim Rejeb, and Horst Treiblmaier. "Mapping metaverse research: identifying future research areas based on bibliometric and topic modeling techniques." Information 14.7 (2023): 356.

[5] George, Lijimol, and P. Sumathy. "An integrated clustering and BERT framework for improved topic modeling." International Journal of Information Technology 15.4 (2023): 2187-2195.

[6] Eshima, Shusei, Kosuke Imai, and Tomoya Sasaki. "Keyword-assisted topic models." American Journal of Political Science 68.2 (2024): 730-750.

DOI: 10.54254/2755-2721/93/20240934

[7] Culasso, Francesca, et al. "Empirical identification of the chief digital officer role: A latent Dirichlet allocation approach." Journal of Business Research 154 (2023): 113301.
[8] Zrelli, Imen, et al. "Drone Applications in Logistics and Supply Chain Management: A Systematic Review Using Latent Dirichlet Allocation." Arabian Journal for Science and Engineering (2024): 1-20.