Does President Xi Make like Mao? A Broader Historical Perspective Using TAD

Xiangming Zeng

1. Introduction

President Xi's centralizing behavior, the removal of term limits, and his apparent imitation of Mao's rhetoric have sparked widespread concern among observers: Is China heading toward a future marked by a return to the ideological fervor of the Mao era? As one article in Time put it, "Xi is using some of Mao's strategies to unite the masses and burnish his personal rule, injecting Marxist and Maoist ideology back into Chinese life." (Beech, March 31, 2016) However, to date, no study has employed a text-as-data approach to empirically examine this important question. This paper represents an initial attempt to do so.

Speeches serve as crucial data for understanding the agendas of political figures. While political speeches have emerged as critical data in political science, TAD-based studies of speeches of Chinese leaders remain scarce. Lim et al. (2025) conducted a preliminary exploration of Xi's agenda using *The Database of Xi Jinping's Important Speech Series*, identifying 25 topics and illustrating the temporal trends of their proportion. However, since this study utilizes a variety of overlapping sources, duplicate documents may introduce biases in the estimated topic proportions. Therefore, this paper aims to apply **supervised machine learning** to identify unique speeches and then classify them using **structural topic model** to better understand the shift in Xi's agenda over time.

Furthermore, although the study by Lim et al. (2025) explores changes in Xi's agenda, these insights are solely derived from the temporal trends in estimated topic proportions,

lacking a nuanced discussion of how these changes are reflected in semantic shifts. Therefore, this paper aims to apply **word embedding** to map semantic shifts in key terms, such as "reform," over time. In addition, this also facilitates my ability to address the research question of whether President Xi is imitating Mao.

2. Data and Methods

Similar to Lim et al. (2025), this study also used web scraping to collect all speeches from *The Database of Xi Jinping's Important Speech Series*, resulting in a total of 847 articles. However, I found that these articles are not all pure speech transcripts; a substantial portion consists of news reports on speeches. If all documents are used directly as the corpus, the repetition of speech content within news reports may introduce estimation bias in key measures, such as topic proportions. Therefore, it is necessary to classify articles and retain only speech transcripts.

By comparing the two types of documents, I found that a small set of distinctive keywords could be used for classification. For example, words such as "Xi Jinping," "pointed out," and "emphasized" appeared only in news reports, while terms like "dear" and "thank you" appeared exclusively in speech transcripts. After identifying around 30 such keywords, I used their term frequencies as features to train a random forest model. I manually labeled 200 documents and applied an 80-20 train-test split. The testing accuracy reached 100%.

I then applied the model to the remaining 647 unlabeled documents to predict their categories and manually validated the predictions by reading the first few sentences of each document. During the validation process, I removed 18 clearly duplicated files—

though not all duplicates—which led me to consider duplicate removal as the next step. Manual validation showed that the model achieved an accuracy of 97.3%, further demonstrating its strong performance.

For data cleaning, the first challenge I encountered is that documents contain various "noise" elements. These include titles, the speaker's name appearing before the speech content, and source information, such as the publishing newspaper and the editor's name, appended at the end. By leveraging the specific patterns in which this "noise" appeared, I was able to remove it and retain only the main body of the speeches.

The second challenge is, unlike English, Chinese text does not naturally include spaces between words. Therefore, it must first be segmented into tokens with spaces before applying corpus(). For this task, I used pkuseg, a Chinese word segmentation tool developed by Peking University. Since it is currently only available in Python, I exported the cleaned data during the final preprocessing step and then re-imported it after performing word segmentation in Python.

Next, I identified duplicate speeches by measuring the maximum cosine similarity of each document with all other documents. Files with a maximum cosine similarity greater than 0.96 were flagged as duplicates, and I removed the one with the later timestamp. Before the first round of deletions, over 250 documents had a maximum cosine similarity above 0.98, indicating a severe duplication problem. After removal, this number dropped to 18, indicating that the issue was significantly mitigated. Repeating the same process again resulted in no document having a maximum cosine similarity greater than 0.96, effectively resolving the problem of duplicate speeches.

3. Analysis

With a clean corpus in hand, the formal analysis could begin. First, I aimed to obtain an overview of President Xi's agenda using LDA. After applying FindTopicsNumber() and evaluating four different metrics, I ultimately decided to use 19 topics.

Among the 19 topics, five are closely related to ideology. However, LDA alone cannot capture trends in the proportions of these topics over time. Therefore, I further employed Structural Topic Modeling (STM) to compare topic proportions between President Xi's first term and his subsequent terms.
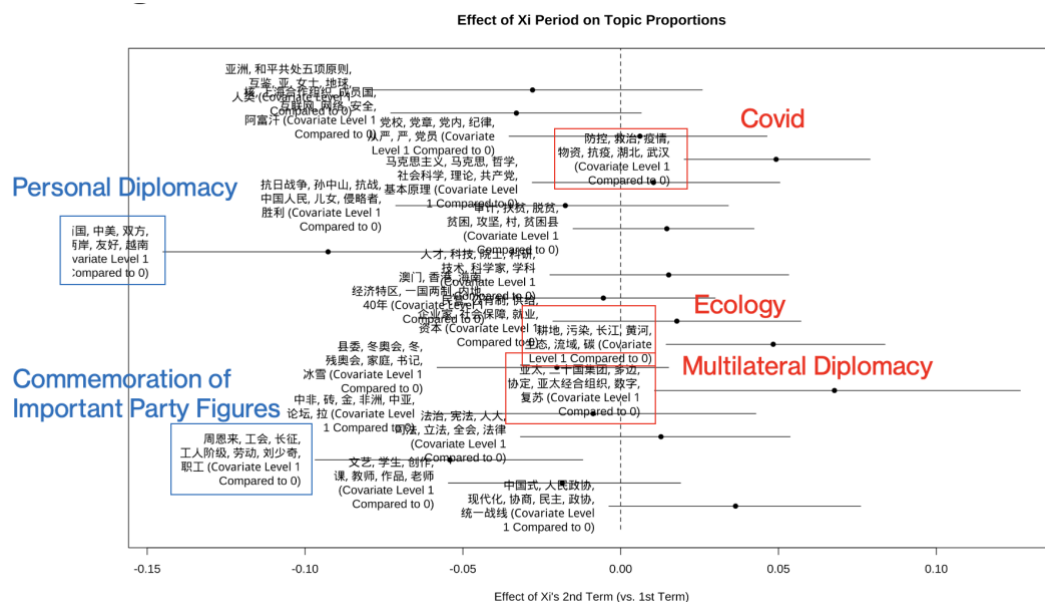

Figure 1. Outcomes of STM

As shown in Figure 1., three topics that saw a significant increase in proportion during Xi's second term were "COVID-19," "Ecology," and "Multilateral diplomacy." Surprisingly, ideology-related topics did not exhibit any notable increase in proportion. Therefore, it is essential to adopt a broader historical perspective by comparing Xi's rhetoric with that of his predecessors, especially Mao.

Word embedding enables the comparisons by capturing subtle semantic shifts. To train

word embeddings for other leaders, I introduced a new corpus consisting of officially published selected works of CCP leaders. Specifically, there are five volumes for Mao (1949-1976), and three volumes each for Deng Xiaoping (1978-1989), Jiang Zemin (1989-2002), and Hu Jintao (2002-2012).

After getting all the embeddings, I wrote a function allowing me to type into any two words to get the cosine similarity between them for Xi and his predecessors. After having this function, I could measure the semantic shift of some interesting terms.

4. Results

4.1. Topic 1: Role of Law VS. Party Discipline

The first interesting question is whether President Xi is undermining the legal reforms of his predecessors. The results are shown in Figure 2.
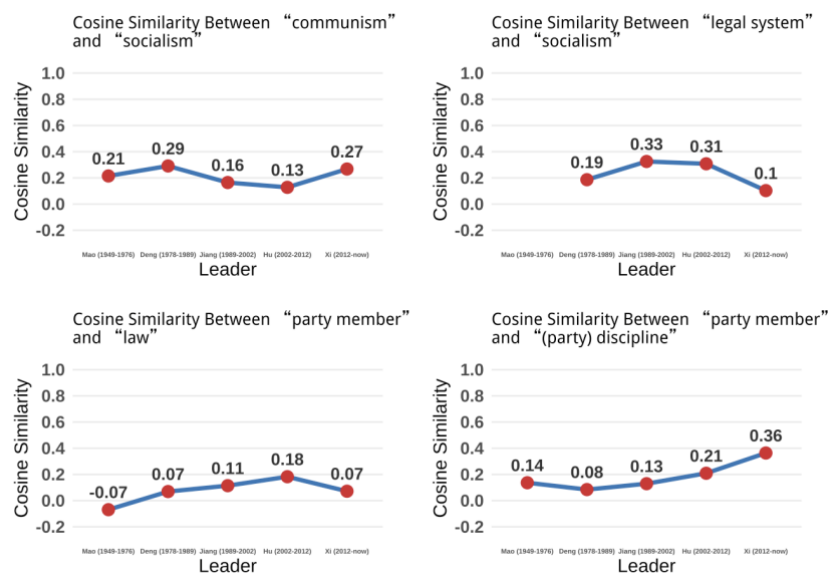

Figure 2. Role of Law VS. Party Discipline

One of the first observable trends is that during the Reform and Opening-Up period, the semantic distance between "socialism" and "communism" grew increasingly large, indicating a de-emphasis on ideological concerns by leaders in favor of pragmatic

economic development. However, in President Xi's speeches, the two concepts have become closer in meaning, suggesting a possible return to the ideological fervor of the Mao era. To explore the nuance further, I also measured the changing semantic distance between "socialism" and "legal system," which appears to confirm the concern noted above — that the rule of law is being de-emphasized under President Xi.

Moreover, given China's party-state system, where Party members hold the core of political power, the question of how to constrain their behavior is crucial. Examining the channels through which constraints of power are emphasized can offer valuable insights into shifts in the role of law. I found that the term "Party member" has grown increasingly distant from "law," while becoming closer to "discipline." This further demonstrates a tendency to constrain Party members through non-transparent channels.

4.2. Topic 2: Party Spirit Education and Loyalty

In addition to emphasizing Party discipline, another means by which President Xi strengthens control over party members is through ideological education, which manifests in two key ways. First, Xi places great emphasis on the need for party members to possess firm belief and ideals. Second, President Xi has reinforced the emphasis on political loyalty. As shown in Figure 3., the semantic distance between terms such as "Party member," "cadre," "comrade," and "Party" and the term "loyalty" has significantly narrowed in Xi's speeches.
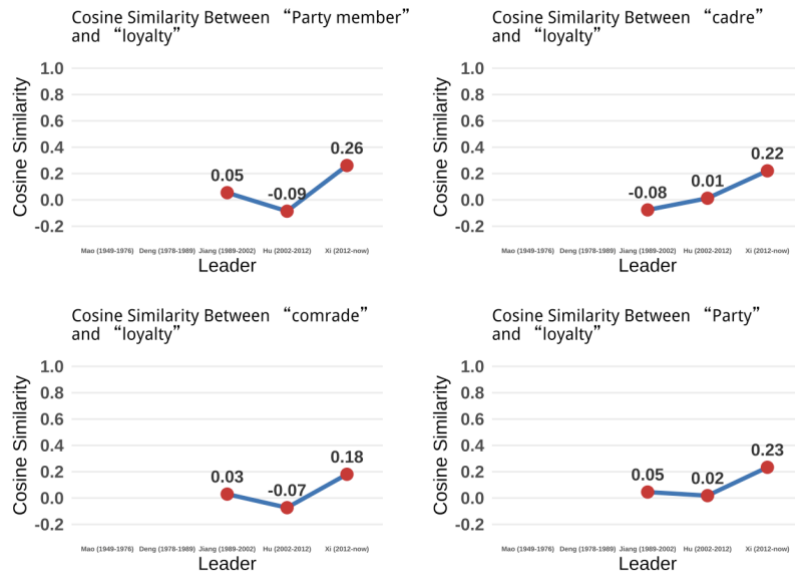
Figure 3. Party Spirit Education and Loyalty

4.3. Topic 3: Concentration of Power

In addition, President Xi has been consolidating power, which is specifically reflected in the decreasing semantic distance between the term "Party Central Committee" (党中央) and concepts related to political authority. The proximity of "Party Central Committee" to words such as "deploy," "lead," "decision-making," and "implement" illustrates this trend. The results are shown in Figure 4.
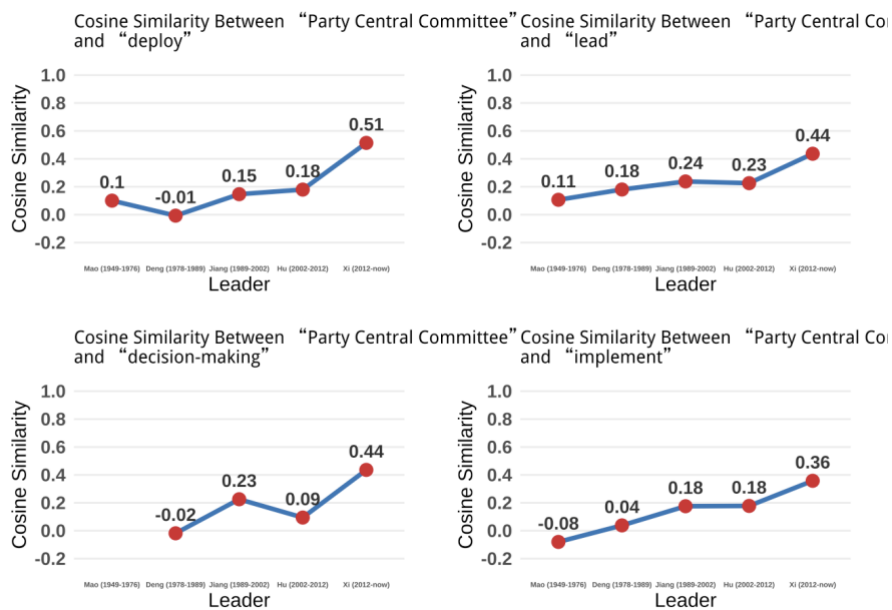
Figure 4. Concentration of Power

4.4. Topic 4: The Rhetoric of Nationalism

Although President Xi has shown a return to Mao's centralization, there is a significant difference between the two: Xi is markedly more inclined to use nationalist language. As shown in Figure 5., in Xi's speeches, the semantic distances between "Chinese nation" and terms like "revival" and "great" have significantly decreased. Moreover, this emphasis on nationalism also reflects a dilution of traditional ideological language, by tying abstract concepts like socialism to the Chinese nation and stressing "socialism with Chinese characteristics." This is specifically evidenced by the narrowing semantic distances between "Chinese nation" and "socialism," as well as between "socialism" and "characteristics."
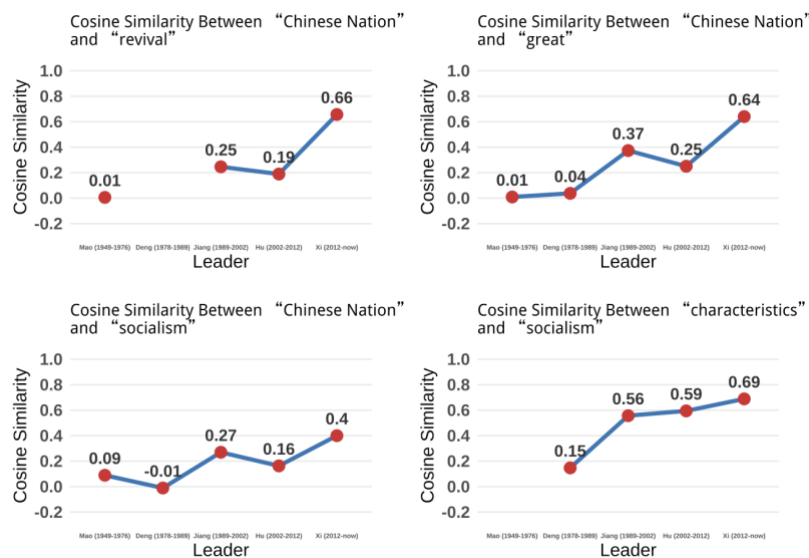

Figure 5. The Rhetoric of Nationalism

4.5. Topic 5: The Direction of Reform

Another common criticism of President Xi is that market-oriented reforms have stalled. As shown in Figure 6., the distance between "reform" and "system" has widened under

President Xi, suggesting that critics are, to some extent, correct in arguing that he has rejected systemic reform. Moreover, the distances between "reform" and terms like "market economy" and "enterprise" have also grown, further supporting the critics' perspective. However, this does not mean that President Xi has pursued no reforms at all. Compared to his predecessors, Xi's emphasizes modernizing the country's governance capacity and governance system. This is reflected in the decreasing semantic distance between "governance" and "reform."
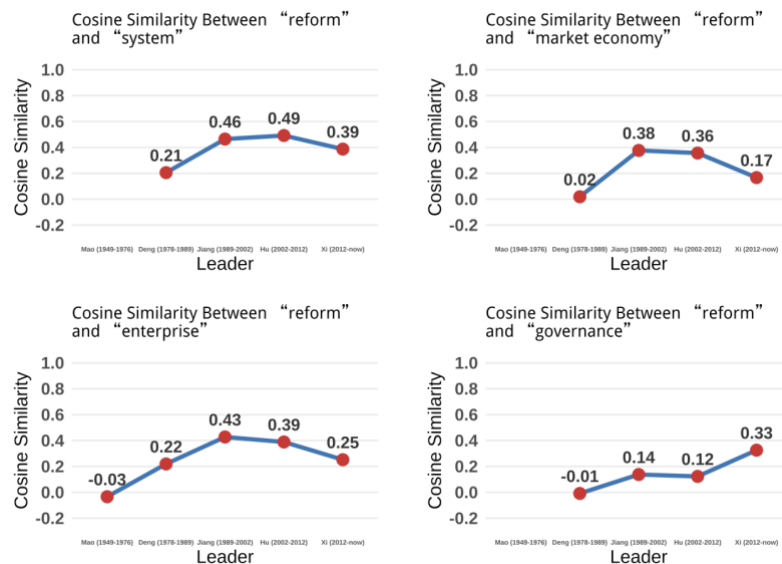

Figure 6. The Direction of Reform

4.6.Topic 6: Attitude to "Struggle"

One sign that has raised concern among observers is that President Xi appears to be imitating Mao's language by placing renewed emphasis on "struggle." As shown in Figure 7., the semantic distances between "struggle" and terms like "great" and "development" have narrowed. However, in my robustness checks, I found that Xi is merely "borrowing" Mao's language. The distance between "struggle" and "class" has been steadily increasing. Meanwhile, the distance between "struggle" and "Chinese

nation" has significantly narrowed in Xi's speeches. The conclusion is that although Xi has borrowed the Marxist-ideological term "struggle" from Mao, its meaning has been replaced by a nationalist narrative
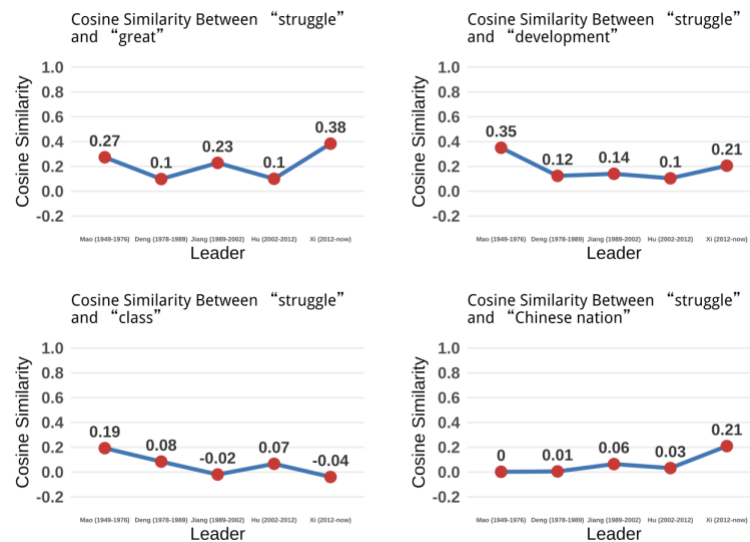

Figure 7. Attitude to "Struggle"

5. Discussion

Is President Xi imitating Mao? The answer is both yes and no. While both leaders share a deep mistrust of the rule of law, Xi relies more heavily on institutionalized tools to build his personal rule. For example, he emphasizes intra-Party discipline to constrain party members and uses institutionalized party spirit education to strengthen ideological purity and loyalty within the Party. In addition, Xi consolidates power by reinforcing the authority of the Party Central Committee, whereas Mao derived personal authority from his charisma as a revolutionary leader.

Moreover, although Xi adopts Maoist language—such as the term "struggle"—this borrowing serves nationalist purposes and departs from the original meaning rooted in orthodox Marxism. This shift is largely a consequence of the ideological vacuum left

by the collapse of the Soviet Union, after which the CCP increasingly turned to nationalism to maintain regime legitimacy and national cohesion. This trend has intensified under Xi, particularly as China rises as a global superpower and confronts strategic competition with the United States. In conclusion, it would be an oversimplification to say that Xi is merely imitating Mao. In reality, Xi is borrowing Mao's language while constructing his own version of a more institutionalized and nationalist form of neo-authoritarianism.

This study builds on the work of Lim et al. (2025) by not only examining topic shifts in President Xi's speeches across different terms, but also incorporating word embeddings to compare semantic shifts in key concepts between Xi and his predecessors. More importantly, this study seeks to answer whether Xi is imitating Mao by analyzing six key themes in contemporary Chinese political discourse. Given that Xi has abolished term limits, understanding his agenda and political preferences offers valuable insights into China's political and policy trajectory in the years to come.

This study primarily focuses on comparing Xi with Mao; however, it does not delve deeply into comparisons between Xi and his predecessors from the reform era. Future research could explore these distinctions, which would be valuable for understanding the future trajectory of China's economic and political landscape. Additionally, this study relies on a single corpus, namely, speeches and selected works of the leaders. Future work could incorporate other sources, such as newspapers, to enable cross-validation.

Reference

Hannah Beech. (March 31, 2016). China's Chairman Builds a Cult of Personality. Time.

https://time.com/4277504/chinas-chairman/

Lim, J., Ito, A., & Zhang, H. (2025). Uncovering Xi Jinping's Policy Agenda: Text As

Data Approach. *Developing Economies*, *63*(1), 9–46.

https://doi.org/10.1111/deve.12418