<u>Cross-Memorial Memory in the Digital Age: A Computational Analysis of Genocide Commemoration</u>

Final Project: Social Media Data Analysis

submitted by:

Dajana Hennig

Lecturer:

Prof. Dr. David Garcia

21.09.2025

1. Introduction

1.1 Motivation

Does the commemoration of one genocide trigger collective attention to other genocides? This study investigates cross-memorial memory patterns by analysing how the remembrance days of genocides influence Wikipedia pageviews for other genocide-related articles. While existing research has examined how individual historical events are remembered over time, I want to explore whether scheduled commemorative events create patterns of historical reflection across related events and topics in digital media. Right now, the digital age provides exceptional opportunities to measure these patterns at scale, allowing to move beyond small-scale surveys and interviews to observe information-seeking behaviour across entire populations.

Understanding these dynamics addresses fundamental questions about collective memory in digital environments. If commemorative practices generate broader historical consciousness, they have implications for how societies maintain awareness of multiple historical events, not only honouring the past, but actively shaping public awareness. Conversely, if memorial days create narrow, event-specific attention patterns, this suggests more fragmented approaches to historical memory. By studying commemorations, I want to further investigate, whether attention to one event is triggered by thematical or geographical linkage, which could further reveal patterns of collective memory.

In the following, I aim to answer, whether the memorial days of genocides lead to additional attention for other historical genocides during that time. If this effect can be found, the goal is to further investigate the role of similarity factors between two events (time separation, location, number of deaths).

1.2 Background

Collective memory research originated with Maurice Halbwachs' foundational argument that individual memories operate within socially constructed group contexts that unify communities across time and space (Halbwachs, 1992). Contemporary scholarship emphasizes media technologies' role in shaping collective memories, noting that "culture and individual memory are constantly produced through, and mediated by, the technologies of memory" (Sturken, 2008).

Traditional methodologies for studying collective memory rely on content analysis, surveys, interviews, and discourse analysis (Kligler-Vilenchik et al, 2014). These approaches, while valuable for understanding memory construction processes, face limitations in scale and temporal scope, often requiring substantial resources to analyse memory patterns across extended periods.

García-Gavilanes et al. (2017) pioneered computational approaches to collective memory research using Wikipedia pageview data, introducing the "view flow" concept to measure how current events trigger attention to past events. Their aircraft accident analysis demonstrated that new crashes generate substantial attention flows to historical crashes, with secondary attention often exceeding primary attention to current events.

This computational approach to collective memory research connects directly to broader themes in digital sociology and computational social science covered in recent coursework. The methodological framework exemplifies how large-scale digital behavioural data can be used to test theoretical propositions about social phenomena that were previously difficult to observe systematically.

However, the García-Gavilanes framework was developed for spontaneous news events like disasters. Memorial days represent fundamentally different commemorative events: they are scheduled, recurring, and explicitly designed to trigger historical reflection. This distinction raises questions about whether memory-triggering mechanisms operate differently for planned commemorative events compared to sudden news developments.

The study adapts the García-Gavilanes methodology to examine cross-memorial memory patterns, specifically testing whether genocide remembrance days create measurable attention flows to other genocide-related content. This adaptation addresses gaps in existing literature by examining how scheduled commemorative practices influence broader historical attention patterns in digital environments.

2. Data

Data was collected primarily from Wikipedia pageview dumps for daily traffic statistics, supplemented by web scraping of Wikipedia articles for metadata on memorial days and historical genocide events. Article content from Wikipedia's 'List of genocides' provided structured information on locations, time periods, and casualty estimates for each historical event.

The choice of Wikipedia as a data source follows established methodological precedent in computational social science research. García-Gavilanes et al. (2017) demonstrated high correlation between Wikipedia pageviews and Google search volumes, validating Wikipedia traffic as a reliable proxy for general internet attention patterns. Wikipedia's comprehensive coverage, standardized article structure, and real-time reflection of collective information-seeking behaviour make it particularly suitable for studying memorial-triggered attention patterns. While the platform's user demographics skew toward educated populations in developed countries, this bias remains consistent across the dataset, allowing for valid relative comparisons between different commemorative events. The educational focus of Wikipedia aligns well with historical memory research, as users actively seeking information about genocides represent genuine engagement with historical content rather than passive media consumption.

Memorial Day articles were identified through Wikipedia's 'Genocide remembrance days' category, with selection criteria requiring clear linkages to specific historical genocides. This filtering approach serves three methodological purposes. Firstly, it ensures conceptual coherence by maintaining focus on historical atrocity commemoration rather than mixing diverse commemorative event types (national holidays, independence celebrations, etc.) that might dilute cross-memorial effects. Furthermore, it enables the creation of meaningful source-target pairs where both the commemorative event and historical genocide can be clearly identified and analysed.

After filtering for articles with sufficient metadata and clear historical referents, the sample includes seven memorial days. The memorial days, their event date and the paired historical event can be found in Table 1. I collected information about the historical genocides corresponding to the memorial days by scraping Wikipedia’s official list of genocides using beautifulsoup in python, which provides information about the location, the period and estimated victims for each genocide.

Article	Event_Date	Event
Armenian Genocide Remembrance Day	April 24	Armenian genocide
Circassian Day of Mourning	May 21	Circassian genocide
Genocide Remembrance Day (Bangladesh)	March 25	Bangladesh genocide
Holodomor Memorial Day	fourth Saturday in November	Holodomor
International Day of Reflection on the 1994 Rw...	April 7	Rwandan genocide
National Day of Remembrance (Cambodia)	May 20	Cambodian genocide
International Holocaust Remembrance Day	January 27	Holocaust

Table 1: Remembrance Days paired with historical event and event date

Daily pageview statistics were extracted from Wikipedia dumps spanning 2015-2024, representing the most recent eight-year period for which complete pageview data was available at the time of analysis. This temporal range was selected to capture contemporary commemorative patterns while ensuring sufficient historical depth for baseline calculations. The data normalization followed García-Gavilanes et al.'s procedures, adjusting for global English Wikipedia traffic fluctuations to enable cross-temporal comparisons.

Memorial Day Selection
Filter Wikipedia "Genocide remembrance days" category

7 memorial days
Require clear linkage to specific historical genocides

Historical Genocide Matching
Pair memorial days with corresponding genocide articles

7 memorial days × 7 genocide articles = 49 pairs
Create source-target combinations

Self-Pair Exclusion
Remove memorial days paired with their own genocide

49 → 42 source-target pairs

Final Analytical Dataset
Cross-memorial view flow calculations

378 source-target-year observations
Each observation represents one memorial day's effect on one genocide article in one year

Table 2: Specified sample sizes and detailed data filtering

I constructed source-target pairs by pairing each Memorial Day (source) with all other genocide articles (targets), creating event windows spanning three days before and after each commemoration date (7-day windows total). To prevent contamination effects, I removed all source-target pairs where memorial days occurred within 10 days of each other, ensuring that observed attention flows reflect genuine cross-memorial memory rather than temporal proximity effects. The 10-day exclusion window was chosen to exceed the 7-day analysis window, preventing overlap between source event periods and ensuring clean separation between commemorative effects. See Table 2 for the detailed data filtering steps and specified sample size.

3. Methods

3.1 View Flow Calculation

I adapted García-Gavilanes et al.'s "view flow" methodology to measure cross-memorial attention patterns. View flow quantifies excess attention to target genocide articles during source memorial event windows, calculated as the sum of daily pageviews minus baseline levels over 7-day commemoration periods.

For each target genocide article in each analysis year, baseline attention levels were calculated as the mean daily pageviews during all non-event-window periods. This approach excludes days falling within any memorial event window to prevent contamination from other commemorative activities that might artificially inflate baseline estimates.

For each day within memorial event windows, daily excess views were computed as the difference between observed normalized pageviews and the baseline level, capturing the attention above or below normal levels during commemorative periods:

$$daily\_excess = normalized\_views - baseline\_views.$$

View flow, the primary dependent variable, was calculated by summing daily excess views over the complete 7-day event window for each source-target-year combination:

$$view\_flow = \Sigma(daily\_excess)$$

over the event window. This aggregation follows García-Gavilanes et al.'s approach, representing the total additional attention (positive or negative) that target articles receive during source memorial periods.

The final analytical dataset consisted of source-target-year observations, where each row represents the view flow from one Memorial Day to one genocide article in a specific year, enabling regression analysis of cross-memorial attention patterns.

3.2 Similarity Variables

To test whether similarity factors influenced *view flow* between commemorative events and genocide articles, I applied a non-parametric group comparison strategy, where the dependent variable is the *view flow* value, and the independent variables (similarity factors) are:

- *Temporal similarity*: Calculated as the absolute difference in start years between the source genocide (commemorated by the Memorial Day) and the target genocide.

- *Geographic similarity*: Determined by mapping genocide locations to continental regions and creating a binary variable indicating whether the source and target genocides occurred on the same continent (1) or different continents (0).

- *Scale similarity*: Victim counts were categorized into five ordinal bins based on highest estimated killings: (1) <100,000, (2) 100,000-499,999, (3) 500,000-999,999, (4) 1,000,000-4,999,999, and (5) ≥5,000,000. Scale similarity was coded as a binary variable where genocides in the same victim count category received a value of 1, and those in different categories received 0.

3.3 Statistical test

The analysis employed a two-stage approach to examine cross-memorial attention patterns. First, descriptive comparisons were conducted using the Mann-Whitney U test to assess whether similarity factors influenced view flow distributions. Since view flow values were highly skewed and violated normality assumptions, this non-parametric test was appropriate for comparing independent groups without distributional assumptions. For each similarity factor (temporal, geographic, and scale), the sample was divided into two groups and tested for significant differences in view flow distributions.

Second, a multiple regression model was fitted to quantify the combined explanatory power of all similarity variables. The regression model incorporated temporal similarity, scale similarity, geographic similarity, and baseline attention levels as predictors of view flow. Model performance was evaluated using R-squared values to assess the proportion of variance explained by the similarity factors. Bootstrap confidence intervals (10,000 samples) were calculated to provide robust estimates of model performance and parameter uncertainty.

4. Results

4.1 Descriptive Results

Figure 1 shows the distribution of cross-memorial view flows across all 378 source–target pairs. The distribution is highly cantered around zero, with the majority of flows clustering tightly and only a few extreme positive or negative cases. Nearly half of the flows are positive (46.6%), indicating that in these cases, the source Memorial Day was associated with increased attention to another genocide-related article. However, the absolute values remain small: the mean view flow is 2.07×10^{-6}, and the standard deviation is about an order of magnitude larger (2.15×10^{-5}), suggesting that cross-memorial effects exist but are generally weak and variable.

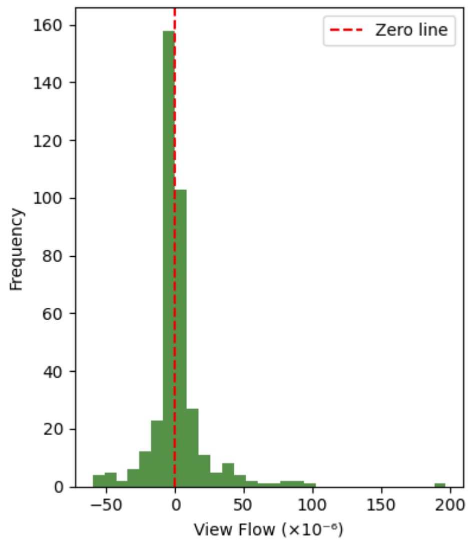


Figure 1: Distribution of Cross-Memorial view flow

Figure 2 ranks individual memorial days by their average cross-memorial effect. Some commemorations are associated with more consistent positive flows: notably the Genocide Remembrance Day in Bangladesh, the Cambodian National Day of Remembrance, and the Armenian Genocide Remembrance Day all show mean positive effects on other genocide articles. Other memorial days, such as Holocaust Remembrance Day and Holodomor Memorial Day, show slightly negative average flows. This heterogeneity suggests that the strength of cross-memorial attention differs substantially by commemoration, with some days more strongly triggering associative remembering than others. The evolution of the mean view-flow over time can be found in the appendix (A1).

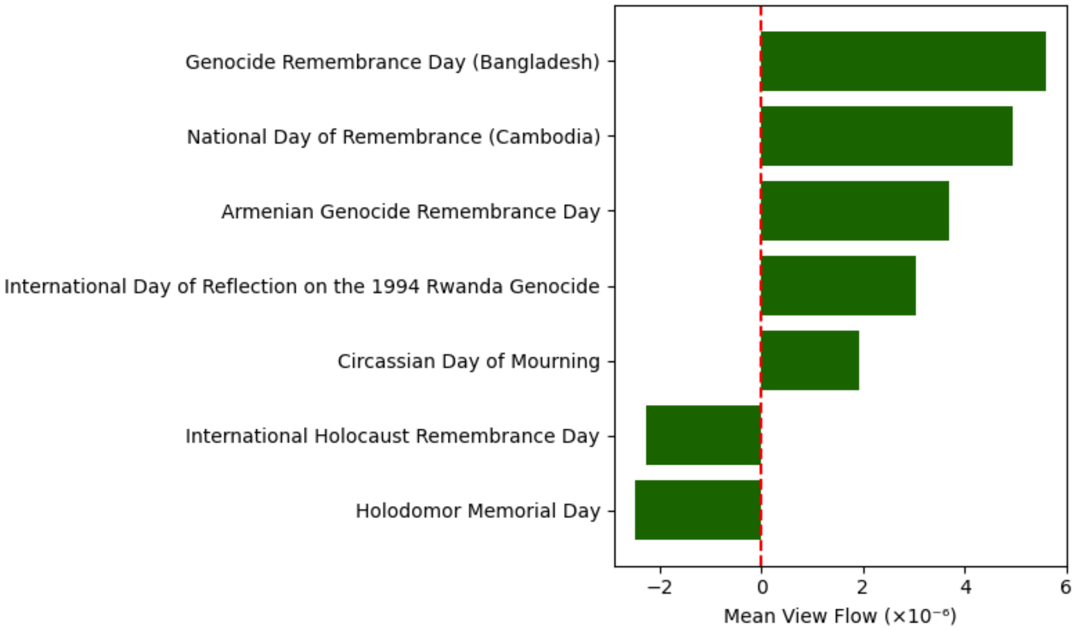


Figure 2: Memorial Days by Cross-Memorial Effect

4.2 Similarity Factor Analysis

I examined three hypothesized factors that might predict stronger cross-memorial memory effects: temporal proximity between genocides, victim scale similarity, and geographic proximity of affected regions. Figure 3 shows the view-flow by similarity factors.

Temporal Similarity emerged as the only statistically significant predictor of cross-memorial attention patterns ($p = 0.025$). Counter-intuitively, memorial days commemorating older genocides generated significantly stronger cross-memorial effects than those commemorating more recent genocides (3.68×10^{-6} vs. 2.08×10^{-7} mean view flow). This pattern suggests that established historical events with settled narratives may be more effective at triggering broader historical reflection than recent genocides that remain politically sensitive or contested.

Scale Similarity showed no significant effect on cross-memorial flows ($p = 0.209$). Memorial days commemorating genocides with similar victim scales did not generate stronger attention to other genocides than those commemorating events with dissimilar scales (-1.94×10^{-7} vs. 4.12×10^{-6} mean view flow). This finding diverges from our hypothesis that genocides of comparable magnitude would trigger stronger associative memory effects.

Geographic Similarity similarly showed no significant relationship with cross-memorial attention patterns ($p = 0.263$). Commemorations of genocides occurring on the same continent did not generate stronger cross-memorial effects than those occurring on different continents (1.25×10^{-6} vs. 2.20×10^{-6} mean view flow). This suggests that cross-memorial genocide memory may transcend geographic boundaries more readily than other types of historical memory.

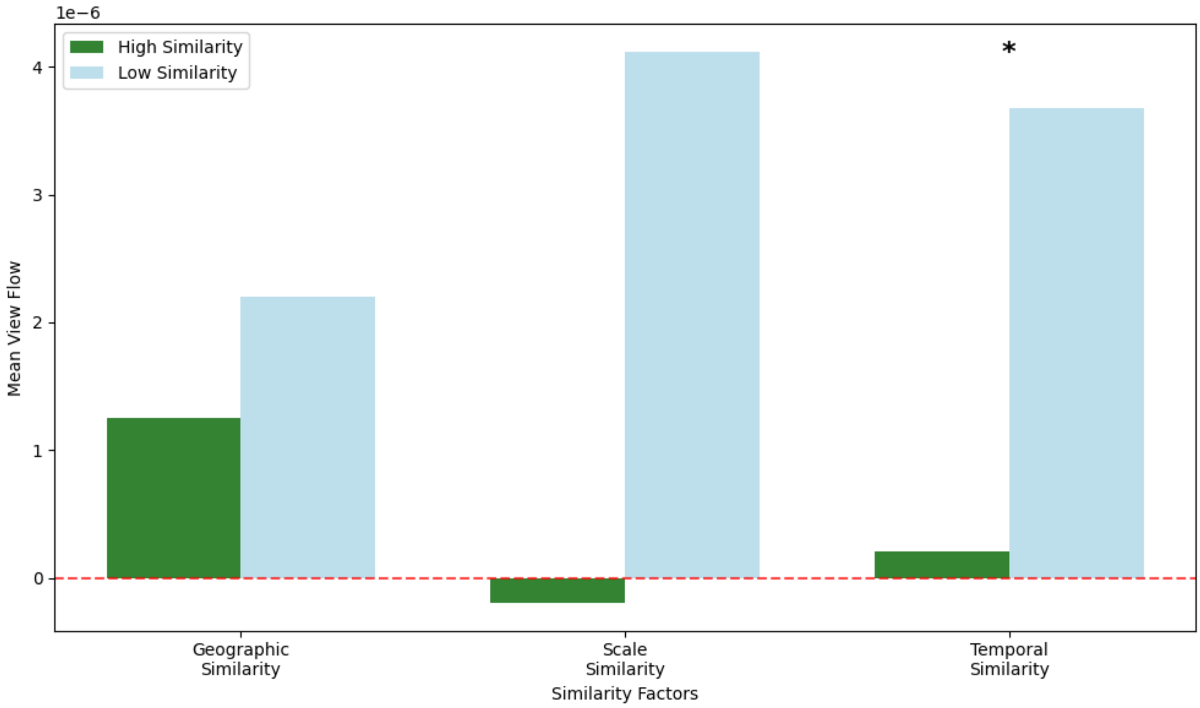


Figure 3: Cross-Memorial view-flow by similarity factors

4.3 Statistical Model Performance

A regression model incorporating temporal similarity, scale similarity, and baseline attention levels demonstrated limited explanatory power for cross-memorial view flows ($R^2 = 0.007$, 95% CI: [0.000, 0.020]). The model explained less than 1% of the variance in cross-memorial attention patterns, with regression coefficients approaching zero for all predictors: baseline attention (-3.10×10^{-15}), scale similarity (-5.72×10^{-19}), and temporal similarity (-5.32×10^{-8}).

This minimal explanatory power represents a contrast to the original García-Gavilanes aircraft crash study, which achieved R^2 values of approximately 0.35 using structurally similar variables. The bootstrap confidence interval for model performance includes zero, indicating substantial uncertainty about whether the observed predictive capacity exceeds random chance.

The absence of meaningful predictive relationships suggests that cross-memorial genocide memory operates through mechanisms not captured by structural similarity measures. Unlike aircraft disasters, where geographic proximity, temporal separation, and event magnitude predicted memory cascades, genocide commemoration appears to trigger cross-memorial attention through more complex pathways that may involve cultural narratives, political contexts, or historical frameworks not readily quantifiable through Wikipedia metadata.

5. Conclusion

This study adapted the García-Gavilanes et al. (2017) computational memory framework to examine whether genocide memorial days trigger cross-memorial attention to other historical genocides on Wikipedia. The analysis of 378 cross-memorial observations across seven genocide commemorations revealed limited but theoretically significant patterns that diverge from the original study's findings about disaster-related memory cascades.

Cross-memorial memory effects between genocide commemorations exist but operate at substantially smaller magnitudes than those observed for spontaneous news events. Only 46.6% of Memorial Day-genocide pairs showed positive attention flows, with effects averaging 2.07×10^{-6} normalized pageviews. This suggests that scheduled commemorative events generate more constrained attention patterns than breaking news about disasters.

The temporal similarity analysis yielded the study's most significant finding: memorial days for historically established genocides trigger stronger cross-memorial effects than those for recent events ($p = 0.025$). This counter-intuitive pattern may reflect that older genocides with settled historical narratives are less politically contested and thus more effective at prompting broader historical reflection. Recent genocides, despite their temporal proximity, may remain too sensitive or contested to generate associative memory effects.

Notably, geographic proximity and victim scale similarity showed no significant predictive power for cross-memorial attention, contrasting sharply with the original study's findings. This suggests that genocide memory operates through different mechanisms than disaster memory, potentially transcending geographic boundaries and magnitude categories that structure other types of collective remembering.

The study demonstrates both the potential and limitations of adapting computational memory frameworks across different domains. While the García-Gavilanes approach successfully identified cross-memorial patterns, the extremely low explanatory power ($R^2 = 0.007$) indicates that structural similarity measures inadequately capture the mechanisms driving genocide memory. This finding has important implications for computational social science: methodological frameworks may not transfer seamlessly across event types, requiring domain-specific adaptations and variable selection.

The research also reveals methodological challenges in studying scheduled versus spontaneous commemorative events. Memorial days, being predictable and recurring, may generate different attention dynamics than breaking news events, necessitating different analytical approaches for understanding planned versus reactive collective memory processes.

The findings suggest that cross-memorial genocide memory operates through cultural, political, and narrative pathways not easily quantified through Wikipedia metadata. Unlike disaster memory, which responds to structural similarities like geographic proximity and event magnitude, genocide commemoration appears to involve more complex associative processes that may depend on historical narratives, educational curricula, or cultural frameworks linking different atrocities.

The study's reliance on Wikipedia as the primary data source introduces several methodological limitations. Wikipedia's user base skews toward educated, technologically literate populations in developed countries, potentially misrepresenting broader patterns of historical memory and commemoration. The platform's editorial processes may systematically bias genocide representation through Western historical perspectives, affecting both article content and the salience of different commemorative events.

Furthermore, Wikipedia pageviews capture only active information seeking, while missing other important dimensions of collective memory such as passive exposure through media, educational contexts, or social discussion. The decision to focus exclusively on English Wikipedia limits generalizability to non-English speaking populations, who may have different commemorative calendars, historical narratives, and cross-memorial association patterns.

The study's scope was necessarily limited to seven genocide commemorations, constrained by the availability of clear Wikipedia categorization and sufficient pageview data. This small sample size may have limited statistical power to detect subtle cross-memorial effects that exist but fall below measurement thresholds. The extremely low R^2 value (0.007) could reflect either the absence of meaningful relationships or insufficient sample size to detect weak but real effects.

Future research could expand beyond Wikipedia to examine cross-memorial patterns across multiple platforms, incorporate textual analysis to understand the narrative mechanisms linking different genocides, or focus on educational contexts where cross-memorial learning may be more systematically structured. Additionally, qualitative research could explore how individual users navigate between different genocide-related content during memorial periods.

Methodological improvements might include multilingual analysis, incorporation of social media data, or collaboration with memorial organizations to understand institutional commemorative strategies. Cross-cultural studies comparing commemoration patterns across different linguistic and cultural contexts would address the current study's Western bias.

This research contributes to understanding how digital technologies mediate collective memory processes, particularly for historical atrocities. While cross-memorial genocide memory appears weaker and more complex than disaster memory, its existence demonstrates that commemorative practices do generate measurable, if subtle, patterns of broader historical attention. This has implications for how memorial organizations, educators, and policymakers might design commemorative practices to promote comprehensive historical awareness rather than event-specific remembering.

The study ultimately reveals that collective memory operates through domain-specific mechanisms that resist simple generalization across event types. While computational approaches offer valuable tools for studying memory at scale, they must be carefully adapted to the unique characteristics of different historical phenomena and commemorative practices.

Replication Data

The code used for the data collection can be found in the Github Repository: https://github.com/DajanaHennig/Social-Media-Data-Analysis-Final-Project

References

N. Kligler-Vilenchik, Y. Tsfati, O. Meyers, Setting the collective memory agenda: Examining mainstream media influence on individuals' perceptions of the past. Mem. Stud. 7, 484–499 (2014).

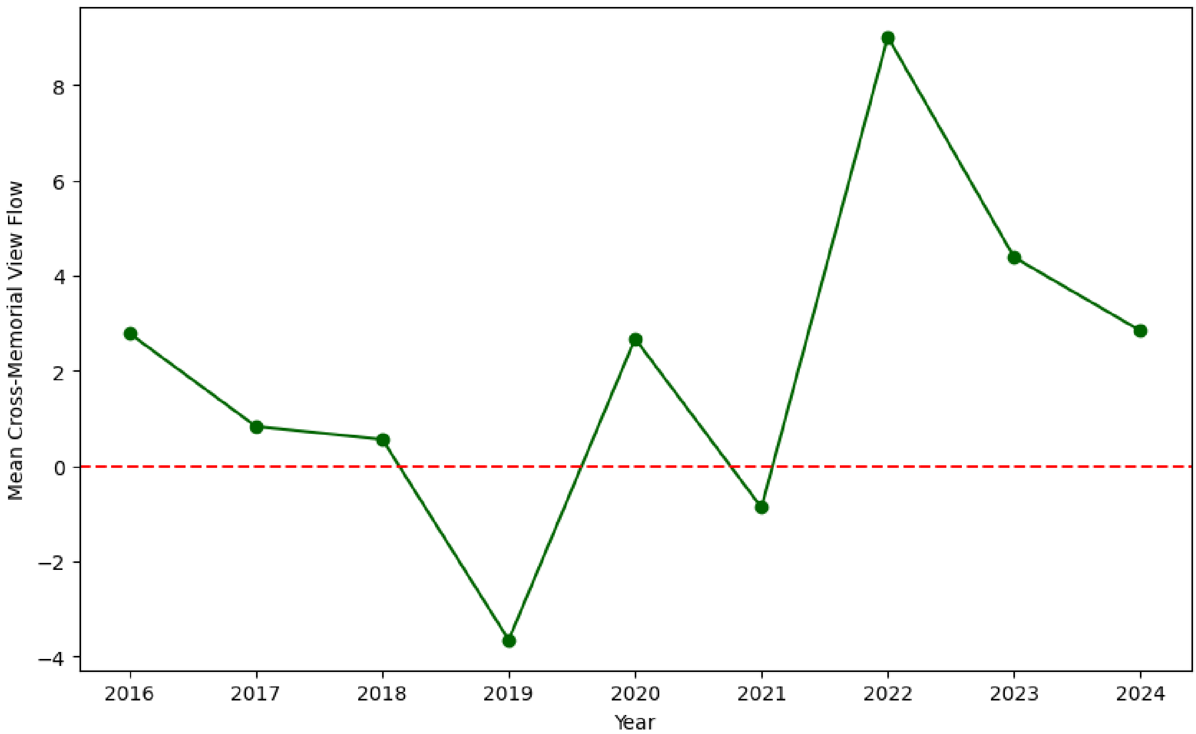
M. Halbwachs, L. A. Coser, On Collective Memory (University of Chicago Press, 1992).

M. Sturken, Memory, consumerism and media: Reflections on the emergence of the field. Mem. Stud. 1,73–78 (2008).

R. García-Gavilanes, A. Mollgaard, M. Tsvetkova, T. Yasseri, The memory remains: Understanding collective memory in the digital age. Sci. Adv. 3, e1602368 (2017)

Appendix

A: Additional Visualizations:



A1: Cross-Memorial Effects over time