scientific reports

OPEN

Check for updates

A two-phase model of collective memory decay with a dynamical switching point

Naoki Igarashi¹, Yukihiko Okada^{2,3}, Hiroki Sayama^{4,5} & Yukie Sano²

Public memories of significant events shared within societies and groups have been conceptualized and studied as collective memory since the 1920s. Thanks to the recent advancement in digitization of public-domain knowledge and online user behaviors, collective memory has now become a subject of rigorous quantitative investigation using large-scale empirical data. Earlier studies, however, typically considered only one dynamical process applied to data obtained in just one specific event category. Here we propose a two-phase mathematical model of collective memory decay that combines exponential and power-law phases, which represent fast (linear) and slow (nonlinear) decay dynamics, respectively. We applied the proposed model to the Wikipedia page view data for articles on significant events in five categories: earthquakes, deaths of notable persons, aviation accidents, mass murder incidents, and terrorist attacks. Results showed that the proposed two-phase model compared favorably with other existing models of collective memory decay in most of the event categories. The estimated model parameters were found to be similar across all the event categories. The proposed model also allowed for detection of a dynamical switching point when the dominant decay dynamics exhibit a phase shift from exponential to power-law. Such decay phase shifts typically occurred about 10 to 11 days after the peak in all of the five event categories.

Memories of past significant events, such as natural disasters and wars, that are shared by members of social groups like countries and families are called "collective memory". This concept was proposed in 1925 by a sociologist Halbawchs¹. Later, Assman classified collective memory into communicative memory and cultural memory depending on how the memory is passed down to future generations². Communicative memory is maintained by everyday communications such as conversations with close people. By contrast, cultural memory is maintained by cultural formation (texts, rites, monuments) and institutional communication (recitation, practice, observance)^{2,3}. Although these were initially treated as sociological concepts, recently, it has begun to attract attention as a target for empirical research. How much people remember World War II is investigated by country⁴ and age⁵ through self-reported surveys. Roediger et al. examined how much people forgotten the U.S. presidents through interviews with students⁶ and found two different functions that characterize forgetting.

Thanks to the recent advancement in digitization of public-domain knowledge and online user behaviors, there has been a growing effort to study collective memory quantitatively using large-scale empirical data. For example, Michel et al.⁷ investigated collective memory using the word frequencies in digitized books. Au Yeung et al.⁸ measured the extent to which collective memories were retained in different countries using large data sets of news articles. Page views and edit histories of Wikipedia articles about significant events, such as natural and man-made disasters, aviation accidents, and terrorist attacks, have been frequently used as indicators of collective memory⁹⁻¹². Some studies used Wikipedia not only to measure the level of collective memory but also to understand the collective nature of people in more general sense, such as revealing the relationship between page views and turnout in elections¹³, building a model of people's browsing behavior considering external factors¹⁴, and predicting the popularity of movies from page views ¹⁵. Singer et al.¹⁶ showed that mass media and current events (30% and 13% of respondents, respectively) dominated the motivation for people to access Wikipedia pages. It was found that Wikipedia page view activity strongly correlates with Google search activity^{17,18}. These earlier studies warrant the use of Wikipedia page views as a quantitative metric of the general user behavior on the Internet.

¹Graduate School of Science and Technology, University of Tsukuba, Ibaraki 305-8573, Japan. ²Faculty of Engineering, Information and Systems, University of Tsukuba, Ibaraki 305-8573, Japan. ³Center for Artificial Intelligence Research, University of Tsukuba, Ibaraki 305-8577, Japan. ⁴Department of Systems Science and Industrial Engineering, State University of New York, Binghamton University, Binghamton 13902-6000, USA. ⁵Faculty of Commerce, Waseda University, Tokyo 169-8505, Japan. [⊠]email: sano@sk.tsukuba.ac.jp

Category	URL of summary article	# of Events
earthquake	List_of_earthquakes_2011-2020	74
deaths	Deaths_in_[January-December]_[2015-2020]	22,319
aviation	Category:Aviation_Accidents_and_incidents_in_[2015-2020]	107
murder	Category:Mass_murder_in_[2015-2020]	121
terror	List_of_terrorist_incidents_in_[2015-2020]	403

 Table 1.
 Summary of Wikipedia page data. URL following: https://en.wikipedia.org/wiki/.

Mathematical models were proposed to describe collective memory decay and validated with various empirical data, including Wikipedia page views. Candia et al. revealed the universal nature of decay patterns on a yearly time scale³. They showed that the decay of collective memory can be modeled by a biexponential function $C_1 e^{-\alpha t} + C_2 e^{-\beta t}$ using the number of citations of papers and patents as well as online attention to songs, movies, and biographies of Wikipedia views on a yearly scale.

Collective memory decay have also been investigated daily time scale. Kim et al.¹⁹ proposed a stretched exponential function $e^{-(t/\alpha)^{y/t^0}}$ to describe daily page views of online academic articles. They successfully depict the dynamics that decay quickly in the beginning and slowly in the latter using the stretched exponential function. West et al.²⁰ studied the daily collective user behavior of Twitter and news sites on the news about the deaths of celebrities between 2009 and 2014. They showed that the mention frequency can be modeled by a shifted power-law function $C_1t^{-\alpha} + C_2$ with their exponents are $\alpha = 1.34$ and $\alpha = 1.54$ for news and Twitter, respectively. García-Gavilanes et al.²¹ analyzed daily Wikipedia page view dynamics on articles of aviation accidents and found that the collective memory decays exponentially after it reaches a maximal value. They proposed a segmented model that assumes separated regimes behind the dynamics.

The earlier studies on daily collective memory decay dynamics typically considered only one dynamical process applied to data obtained in just one specific event category. Whereas a universal model ³ was proposed for annual collective memory decay using the data for multiple event categories, such year-by-year dynamics are only relevant at a slow, historical time scale, which would not be applicable to day-to-day dynamics. There is hence a need for a universal model of collective memory decay for a faster, daily time scale.

Here we propose a two-phase decay model for collective memory of various types of significant events and evaluate its validity using Wikipedia page view data. We compare the proposed decay model to several other existing decay models developed using data from Wikipedia²², blogs²³, Twitter^{24,25}, YouTube²⁶, news sites^{27,28}, book sales ²⁹, and the number of articles read¹⁹. These earlier studies modeled collective memory decay in various mathematical forms, such as power-law, exponential, and stretched exponential, to which the proposed model is compared for performance evaluation.

Data and methodology Data.

In this study we analyzed collective memory decay using English Wikipedia page view data. We selected the following five categories of significant events for analysis: earthquakes, deaths of notable persons, aviation accidents, mass murder incidents, and terrorist attacks. These events were also used in previous collective memory studies^{10,11,20,21,30}. For the events in these categories, the date and location of the event are precise, which allows for the collection of unambiguous time series data. We obtained the Wikipedia pages of events listed in the summary article of each category in the English version of Wikipedia. The target period of event occurrence is from July 1st, 2015, to June 30th, 2020. Table 1 shows a summary of the dataset we obtained from Wikipedia.

Figure 1 shows two examples of Wikipedia page view decay from the event occurrence date (one for the 2016 earthquakes in Kumamoto, Japan, and the other for the death of Alan Rickman). The two examples commonly show that the Wikipedia page views peaked around the date of the event and gradually decayed over time. In addition, the peak height of the page views (i.e., how much attention an event receives) and the decay rate (i.e., how quickly people forget it) varied greatly from event to event.

For each of the collected Wikipedia pages, we obtained daily page view counts since the event occurrence date for 300 days from the infobox in each event Wikipedia page by using Wikimedia REST API (https://wikimedia. org/api/restv1/). The length of the data collection period was set to 300 days, shorter than one year, in order to avoid a possible "anniversary" page view increase toward the end of the 365-day cycle (such increase was seen in Fig. 1 right). If the page view peak was less than 1000 or occurred 5 or more days after the event date, the data was excluded from the analysis since we considered such events did not trigger significant collective memory responses. With these criteria, we acquired valid page view data for 34 earthquakes, 8684 deaths of notable persons, 43 aviation accidents, 37 mass murder incidents, and 123 terrorist attacks.

Model. In this study, we propose a unique two-phase mathematical model of collective memory decay that combines exponential and power-law phases. Our model does not assume a regime shift in the decay of collective memory but rather a change within the population that forms collective memory. First, we define the normalized daily page views t days after the peak t_c for each event as $S(t) = S^{raw}(t)/S^{raw}(0)$, where $S^{raw}(t)$ is the raw number of daily page views t days after the peak t_c (and therefore $S^{raw}(0)$ is the number of page views at the peak t_c). Next, we assume that there are two types of users: the first type is "temporary interest users", whose page views decay rapidly as an exponential function of time with no interaction, and the second type is "long inter-



Figure 1. Examples of daily Wikipedia page view decay for 400 days in log-log scale. The panels show page views decay after an event occurrence in 2016: Kumamoto earthquake (left) and the death of Alan Rickman (right). In Alan Rickman's case, there is a spontaneous increase of page views around 365 days.



Figure 2. Proposed two-phase model of collective memory decay. The model consists of exponential decay, which corresponds to temporal interest users without interactions, and power-law decay, which corresponds to long-term interest users with interactions.

est users", whose page views decay following a power-law function of time which implies non-trivial interactions among those users. We made this assumption based on Ebbinghaus' forgetting curve that the independent individual memory decays exponentially³¹. Combining these two types of users determines the total number of page views in our model (Fig. 2). This model can capture the shift from "fast decay" to "slow decay". The model formula is mathematically expressed as follows:

$$S(t) = C_1 e^{-\beta t} + C_2 t^{-\alpha} \tag{1}$$

 C_1 and C_2 are constant parameters representing the amplitudes of the two decay dynamics. β is the decay rate of the initial exponential decay, and α is the decay rate of the mid- to long-term power-law decay. This proposed model is different from the models of the previous research, and the idea that the basic properties of the user can be divided into two distinct groups is also unique to our research.

To evaluate the validity of this model, we quantitatively compared the accuracy of the proposed model with that of other models in the previous research, including bi-exponential¹⁹, stretched exponential³, and shifted power-law²⁰ by measuring the coefficient of determination R^2 and the Akaike Information Criterion (*AIC*).

Results

Model fitting. We performed model fitting for each normalized time-series data of page views S(t) with the following four nonlinear models that do not assume a regime shift: bi-exponential $C_1e^{-\alpha t} + C_2e^{-\beta t}$, stretched exponential $e^{-(t/\alpha)^{y^{t-1}}}$, shifted power-law $C_1t^{-\alpha} + C_2^{20}$, and the proposed model $C_1e^{-\beta t} + C_2t^{-\alpha}$. In model fitting, we added a constant value ϵ to each individual time series, where ϵ was the minimum nonzero value across all individual time series. Then, we took base-10 logarithms of the empirical data and conducted parameter fitting of each model formula to the data in a log-log space using a nonlinear least-squares method, following the method by West et al. ²⁰. Figure 3 shows examples of model fitting. Compared to the purely exponential (blue, dashed) and purely power-law (red, dash-dotted) decays, our proposed model (yellow, solid) can capture both the initial exponential decay and the mid- to long-term power-law decay simultaneously.

We compared the median of R^2 and AIC of each model formula for each event category to compare the model performance. Tables 2 and 3 show the results. The proposed model showed the best performance for earthquakes,



Figure 3. Fitting examples shown in Fig. 1 of the proposed two-phase model for 300 days. Blue dashed line and red dash-dotted line correspond to the simple exponential decay and power-law decay, respectively. Yellow solid line shows the proposed two-phase model which combined the exponential and power-law functions.

Model	earthquake	death	aviation	murder	terror
Bi-exponential ³	0.765	0.723	0.822	0.800	0.753
Stretched-exponential ¹⁹	0.758	0.730	0.796	0.799	0.730
Shifted power-law ²⁰	0.768	0.737	0.813	0.801	0.740
Proposed model	0.786	0.733	0.846	0.798	0.757

Table 2. Median of R^2 for four decay models.

Model	earthquake	death	aviation	murder	terror
Bi-exponential ³	-277.0	-229.1	-423.1	-269.7	-244.5
Stretched-exponential ¹⁹	-284.4	-244.5	-408.7	-279.3	-227.7
Shifted power-law ²⁰	-294.0	-250.1	-412.5	-281.3	-240.6
Proposed model	-311.1	-245.4	-436.9	-280.9	-262.1

 Table 3. Median of AIC for four decay models.

aviation accidents, and terrorist attacks. For deaths of notable persons and mass murder incidents, the shifted power-law model 20 performed slightly better, but the differences between its R^2 and *AIC* values and those of our model were small. In fact, when we determined each sample individually, our model performed better in more than half of the cases in all categories. 82%, 59%, 59%, 55%, and 58% for earthquake, notable death, aviation, mass murder, and terror incidents, respectively. Interestingly, the shifted power-law model proposed to describe obituaries also performed well in our data targeting negative events.

Decay parameters. The initial fast exponential decay is characterized by β and the late slow power-law decay by α . Figures 4 and 5 show probability density distributions of the parameter values of β and α . Note that 26 (0.3%) outliers ($\beta > 2$) are not shown in the distribution of death of notable persons (N = 8, 684). These distributions show a clear unimodal distribution with a distinct characteristic value for each parameter whose medians are shown in Table 4. These results suggest that there is a common pattern of collective memory decay, first in the fast exponential decay immediately after the event with the exponent β around 0.4, followed by the slow power-law decay with the exponent α around 0.3.

An interesting observation is that the value of α may be loosely related to the lasting societal impact of the events. Earthquakes tend to cause a massive damage to society and the characteristic value of α for this category was large (0.48), implying that there was meaningful long-term collective memory decay going on for a long period of time. Meanwhile, deaths of notable persons would have minimal impact on society and its characteristic value of α was small (0.22), implying that the long-term behavior was closer to a flat line ($\alpha = 0$) and more likely dominated by constant random page views. Events in other categories would have societal impacts at intermediate levels, which may be reflected on their intermediate characteristic α values as well. This observation remains largely speculative and would need further systematic investigation.

Switching point of collective memory decay dynamics. Our proposed model allows for detection of the "switching point" of collective memory decay where the dominant component in the model formula $S(t) = C_1 e^{-\beta t} + C_2 t^{-\alpha}$ switches from exponential to power-law. Such a switching point t^* is defined as the first time point at which $C_2 t^{-\alpha} > C_1 e^{-\beta t}$ in the fitted model (Fig. 6).



Figure 4. Probability density distributions of parameter values of α obtained using the proposed model in five categories. Most of the samples fall within a certain range ($0 < \alpha < 1$) and show similar values ($\alpha \sim 0.3$) independent of the category.

Figure 7 shows the probability density distributions of the switching points detected for five categories. The median values for all categories were quite similar (earthquake: $t^* = 10$; deaths of notable persons: $t^* = 11$; aviation accidents: $t^* = 10$; mass murder incidents: $t^* = 11$; and terrorist attacks: $t^* = 11$), indicating a common pattern of the shift of collective memory decay dynamics at about the same timing (around 10 to 11 days after the peak), regardless of the event category.

Discussions

In this study, we collected daily English Wikipedia page view counts for five event categories and modeled their decay processes using a new two-phase model that combined initial exponential decay and mid- to long-term power-law decay in a single mathematical formula. To the limit of our knowledge, this study was the first attempt to develop a universal model of collective memory decay applicable to multiple event categories at daily time scales. We found that our proposed model showed consistently high accuracy across multiple event categories, and closely matching the best performance in the previously proposed decay models.

Our model also allowed for the detection of a "switching point" in collective memory decay at which the dominant decay dynamics switches from exponential to power-law. We found that the decay phase switches about 10 to 11 days after the peak, irrespective of the event category. This number is similar to what was reported in García-Gavilanes et al.²¹ that the first break point of the segmentation was 3-10 days for both English and Spanish





	earthquake	death	aviation	murder	terror
β	0.39	0.42	0.45	0.41	0.43
α	0.48	0.22	0.31	0.30	0.33

Table 4. Median of β and α for five categories for the proposed model.





Wikipedia page views of aviation accidents. This is a unique, non-trivial finding because it indicates a universal property of our society's "collective attention span" which shows immediate attention period of the news.

There are some limitations in our study. Firstly, we only validated our model by the English Wikipedia page views of events that related to a negative impact on society. Therefore one still needs to be careful in considering the generality of the obtained results by using other data for events with a positive impact on society, such as a scientist winning a Nobel Prize and an actor winning an Academy Award. We expect that positive events' decay patterns will be similar to the negative ones, because our previous study showed that the word frequency of names related to obituaries and Nobel Prizes exhibited a similar decay pattern in Japanese blog data²³. Similarly, other than Wikipedia page views, Twitter mentions and number citations are also considered for future tasks. The assumptions of the model should also be noted. Here we focused on aggregated behavior of the user population, and we did not consider each individual user's behavioral changes.

Therefore, future directions of research include consideration of more detailed information about specific events and modeling their influences on collective memory decay, such as more detailed event types, the popularity of the event, and the size of societal impacts the event created. Such systematic analysis will help understand the nature of collective memory in greater depth, possibly revealing the quantitative relationship between the event's impact and the value of α as indicated above. Also, we found the spontaneous increases in collective memory decay around 365 days which could be attributed to year-to-year recall. We recognize that investigating such spontaneous increases is another interesting future direction.



Figure 7. Probability density distributions of switching point t^* detected for five categories. Most of the samples fall within $0 < t^* < 40$ days, which is relatively short compared to the maximum length of the time series (300 days). The median values of t^* are 10 to 11 days for all categories.

.....

Data availability

The datasets used and analyzed during the current study are available through the Wikimedia REST API.

Received: 17 September 2022; Accepted: 6 December 2022 Published online: 12 December 2022

References

- 1. Halbwachs, M. On Collective Memory (University of Chicago Press, Chicago, 1992).
- 2. Assmann, J. & Czaplicka, J. Collective memory and cultural identity. New German Critique 125-133 (1995).
 - Candia, C., Jara-Figueroa, C., Rodriguez-Sickert, C., Barabási, A.-L. & Hidalgo, C. A. The universal decay of collective memory and attention. *Nat. Hum. Behav.* 3, 82–91 (2019).

- 4. Roediger, H. L. et al. Competing national memories of world war II. Proc. Natl. Acad. Sci. USA 116, 16678-16686 (2019).
- Zaromb, F., Butler, A. C., Agarwal, P. K. & Roediger, H. L. III. Collective memories of three wars in united states history in younger and older adults. *Memory Cognit.* 42, 383–399 (2014).
 - 6. Roediger, H. L. & DeSoto, K. A. Forgetting the presidents. Science 346, 1106-1109 (2014).
- 7. Michel, J.-B. et al. Quantitative analysis of culture using millions of digitized books. Science 331, 176–182 (2011).
- Au Yeung, C.-M. & Jatowt, A. Studying how the past is remembered: towards computational history through large scale text mining. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management, 1231–1240 (2011).
- 9. Ferron, M. & Massa, P. Beyond the encyclopedia: collective memories in Wikipedia. Memory Stud. 7, 22-45 (2014).
- Kanhabua, N., Nguyen, T. N. & Niederée, C. What triggers human remembering of events? A large-scale analysis of catalysts for collective memory in Wikipedia. In *IEEE/ACM Joint Conference on Digital Libraries*, 341–350 (IEEE, 2014).
- 11. García-Gavilanes, R., Mollgaard, A., Tsvetkova, M. & Yasseri, T. The memory remains: understanding collective memory in the digital age. *Sci. Adv.* **3**, e1602368 (2017).
- Ferron, M. & Massa, P. Collective memory building in Wikipedia: the case of north african uprisings. In Proceedings of the 7th International Symposium on Wikis and Open Collaboration, 114–123 (2011).
- 13. Yasseri, T. & Bright, J. Wikipedia traffic data and electoral prediction: towards theoretically informed models. *EPJ Data Sci.* 5, 1–15 (2016).
- 14. Ratkiewicz, J., Fortunato, S., Flammini, A., Menczer, F. & Vespignani, A. Characterizing and modeling the dynamics of online popularity. *Phys. Rev. Lett.* **105**, 158701 (2010).
- 15. Mestyán, M., Yasseri, T. & Kertész, J. Early prediction of movie box office success based on Wikipedia activity big data. *PLoS ONE* 8, e71226 (2013).
- Singer, P. et al. Why we read Wikipedia. In Proceedings of the 26th International Conference on World Wide Web, 1591–1600 (2017).
 Ratkiewicz, J., Flammini, A. & Menczer, F. Traffic in social media I: Paths through information networks. In 2010 IEEE Second International Conference on Social Computing, 452–458, https://doi.org/10.1109/SocialCom.2010.72 (2010).
- Yoshida, M., Arase, Y., Tsunoda, T. & Yamamoto, M. Wikipedia page view reflects web search trend. In *Proceedings of the ACM Web Science Conference*, WebSci '15, https://doi.org/10.1145/2786451.2786495 (Association for Computing Machinery, New York, NY, USA, 2015).
- 19. Kim, Y. & Weon, B. M. Stretched exponential dynamics in online article views. Front. Phys. 8, 614 (2021).
- 20. West, R., Leskovec, J. & Potts, C. Postmortem memory of public figures in news and social media. *Proceedings of the National Academy of Sciences of the United States of America* **118** (2021).
- 21. García-Gavilanes, R., Tsvetkova, M. & Yasseri, T. Dynamics and biases of online attention: the case of aircraft crashes. R. Soc. Open Sci. 3, 160460 (2016).
- Kobayashi, R., Gildersleve, P., Uno, T. & Lambiotte, R. Modeling collective anticipation and response on Wikipedia. arXiv preprint arXiv:2105.10900 (2021).
- Sano, Y., Yamada, K., Watanabe, H., Takayasu, H. & Takayasu, M. Empirical analysis of collective human behavior for extraordinary events in the blogosphere. *Phys. Rev. E* 87, 012805 (2013).
- Asur, S., Huberman, B. A., Szabo, G. & Wang, C. Trends in social media: persistence and decay. In Fifth International AAAI Conference on Weblogs and Social Media (2011).
- Lorenz-Spreen, P., Mønsted, B. M., Hövel, P. & Lehmann, S. Accelerating dynamics of collective attention. Nat. Commun. 10, 1–9 (2019).
- Crane, R. & Sornette, D. Robust dynamic classes revealed by measuring the response function of a social system. Proc. Natl. Acad. Sci. USA 105, 15649–15653 (2008).
- 27. Wu, F. & Huberman, B. A. Novelty and collective attention. Proc. Natl. Acad. Sci. USA 104, 17599-17601 (2007).
- 28. Dezsö, Z. et al. Dynamics of information access on the web. Phys. Rev. E 73, 066132 (2006).
- Sornette, D., Deschâtres, F., Gilbert, T. & Ageon, Y. Endogenous versus exogenous shocks in complex networks: an empirical test using book sale rankings. *Phys. Rev. Lett.* 93, 228701 (2004).
- 30. Watanabe, G. *Empirical analysis of trend duration period in online media (in Japanese)*. Master's thesis, Graduate School of Systems and Information Engineering, University of Tsukuba, Japan (2020).
- Ebbinghaus, H. Memory: A Contribution to Experimental Psychology (Teachers College, Columbia University, New York City, 1885/1913).

Acknowledgements

We would like to thank Dr. Hiroyasu Ando and Dr. Kiyoshi Kanazawa for the fruitful discussions. This work was supported by JSPS Grants-in-Aid for Scientific Research Grant Numbers 17K12783, 20K19928, 22H00895.

Author contributions

N.I., Y.S., and Y.O. designed the research. H.S. contributed to advised the research design and analysis. N.I. preprocessed and analyzed the data. Y.S., N.I. and H.S. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022