nature human behaviour

SUPPLEMENTARY INFORMATION
https://doi.org/10.1038/s41562-018-0474-5

In the format provided by the authors and unedited.

The universal decay of collective memory and attention

Cristian Candia[1,2,3]*, C. Jara-Figueroa[1], Carlos Rodriguez-Sickert[3], Albert-László Barabási[2] and César A. Hidalgo[1]*

[1]Collective Learning Group, The MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Network Science Institute, Northeastern University, Boston, MA, USA. [3]Centro de Investigación en Complejidad Social (CICS), Facultad de Gobierno, Universidad del Desarrollo, Santiago, Chile. *e-mail: ccandiav@mit.edu; hidalgo@mit.edu

The Universal Decay of Human Collective Memory

Cristian Candia[a,b,c,*], C. Jara-Figueroa[a], Carlos Rodriguez-Sickert[c], Albert-László Barabási[b], and César A. Hidalgo[a,*]

[a]Collective Learning Group, The MIT Media Lab, Massachusetts Institute of Technology

[b]Network Science Institute, Northeastern University, Boston, MA 02115, USA

[c]Centro de Investigación en Complejidad Social (CICS), Facultad de Gobierno, Universidad del Desarrollo, Santiago, Chile.

[*]Corresponding Authors: hidalgo@mit.edu, ccandiav@mit.edu

Supplementary Material

In this study, we connect two different bodies of literature: Collective memory and knowledge diffusion. In the newest remarkable empirical studies of collective memory, Roediger and DeSoto[1] describes how society forgets US presidents. The found a U-shaped behavior, or three-step process, this is an example of what psychologists call the serial position effect, which is the tendency for people to remember most prominently the first (primacy effect) and last (recency effect) items of a list[2]. Roediger and DeSoto[1] report this effect with how people remember presidents and they also found that presidents who held office during a subject's life were recalled at significantly higher rates. On the other hand, knowledge diffusion has focused on citations curves as a proxy of attention of ideas. It has been shown that in the temporal dimension of the problem there is no consensus[3–15].

Here we propose two mechanisms inspired by collective memory studies which, after a mathematical operationalization described on section SM 2.1, they can explain the temporal dimension of knowledge diffusion, also showing the emergence of universal behavior. We observe a recency effect and then a longer and slower decay, in the words of Neruda, fast and intense love and a longer and slower forgetting, which is universal.

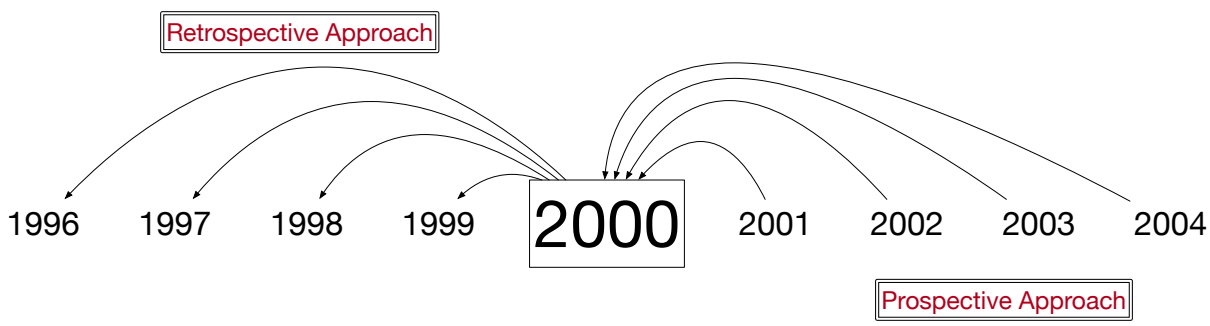
Besides the contribution about the cultural mechanisms that explain how society forgets cultural content over time and the universality of the emerging behavior, the utility of this approach relies on its potential to study collective memory on systems where is difficult to get time series data. Using this approach, controlling by preferential attachment and considering just a "snapshot" of a system, is enough to describe and understand knowledge diffusion of cultural goods and ideas, in the same way, how was doing in this study using songs, movies, and biographies.

Supplementary Methods

Used Approach

Figure 1 shows the two different approach to study citations patterns described in the literature. In words of Bouabid 2011[16], "The first considers papers cited by a publication during a particular year and then analyze the distribution of their ages retrospectively. This approach is called 'synchronous distribution' (Nakamoto 1988), 'citations from'

approach (Redner 2004) and 'retrospective citation' approach (Burrell 2002; Glanzel 2004). The second approach consists of analyzing the distribution of citations gained over time by a paper (or papers) published in a given year. This approach is called 'diachronous distribution' (Nakamoto 1988), 'citations to' approach (Redner 2004) and 'prospective citation' approach (Burrell 2002; Glanzel 2004). Nakamoto suggested that the synchronous and diachronous distributions follow similar curves and are symmetric. The two approaches were compared by Stinson and Lancaster (1987) in measuring the obsolescence". Glanzel 2004[17] has empirically shown that the best approach to study these systems is the prospective approach. Also, Yin and Wang[18] have proved the mathematical equivalence of both. So, this decision doesn't have any impact on the results.



Supplementary Figure 1: Retrospective and Prospective approach used to study citations patters.

Inflation Factor

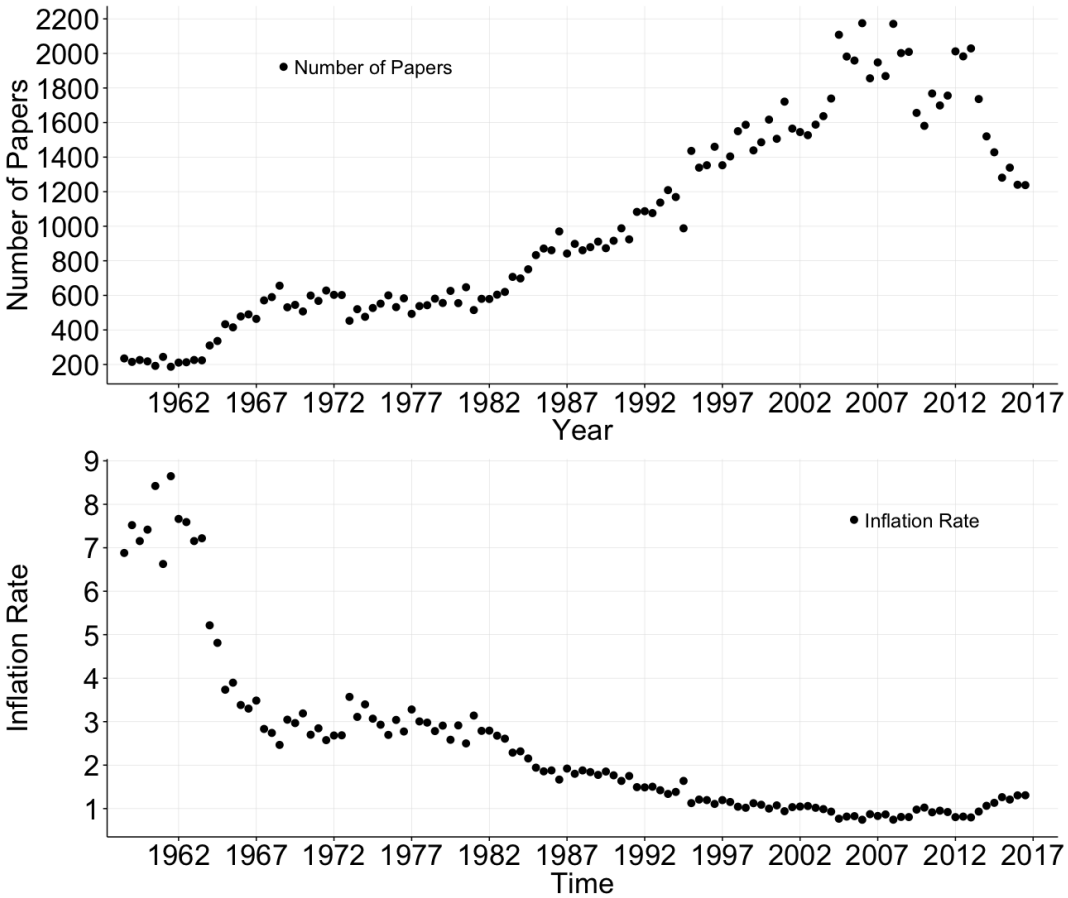
These time series were constructed using a time window of six months. Using a different time window has been shown not to change the results[14,15]. For patents and papers we discount the citations by an inflation rate[14] by fixing a base year and re-scaling all the citations obtained to this base year–for more information see SM 0.2.

External factors affect the number of publications each year, this factors could be, for instance, a policy decision of a journal, more resources, among others. These kinds of factors impact the citation rate, covering the temporal patterns of forgetting, because the probability that a paper obtain a cite in a future year change if, for instance, there are more resources to publish more papers[14,15]. To separate the effects of inflation and knowledge obsolescence, it means, to unveil the temporal pattern of collective forgetting, we multiply all the citations obtained by the inflation rate described in equation 1 as proposed in[15].

$$I_f = \frac{N_J(T_0, T + \Delta T)}{N_J(T, T + \Delta T)}, \tag{1}$$

where $N_J(T, T + \Delta T)$ is the number of paper published by the citing journal J at the interval $(T, T + \Delta T)$, and $N_J(T_0, T + \Delta T)$ is the number of paper published by the citing journal J at the interval when the paper was published. For instance, let's consider an article published in PRL in the first semester of 2000 (base year). If it gets six citations the second semester of 2010, but the number of papers published by PRL in that time is twice the amount of paper published in PRL in the base year, the six citations adjusted by the inflation factor become in three. It means an inflation factor of 0.5 adjusts the citations. The inflation factor allows us to control by external shocks and the exponential growth of science[14,19].

Figure 2 shows the number of papers published by PRL every six months and the inflation rate using 2000 as a base year. Thus, we deflated each citation obtained by a paper by its corresponding inflation rate given the journal that is citing and the time when the new paper is citing.



Supplementary Figure 2: A) Number of papers published by PRL each six months. B) Inflation rate for PRL each six months with $T_0 = 2000$. We observe that before 2000 the inflation rate is bigger than 1, this means that there were less papers published before 2000. After 2000, the citation rate is smaller than 1, this means that there were more papers published after 2000, with exception of after 2012, when the inflation rate increases. We also observe in A) this fact.

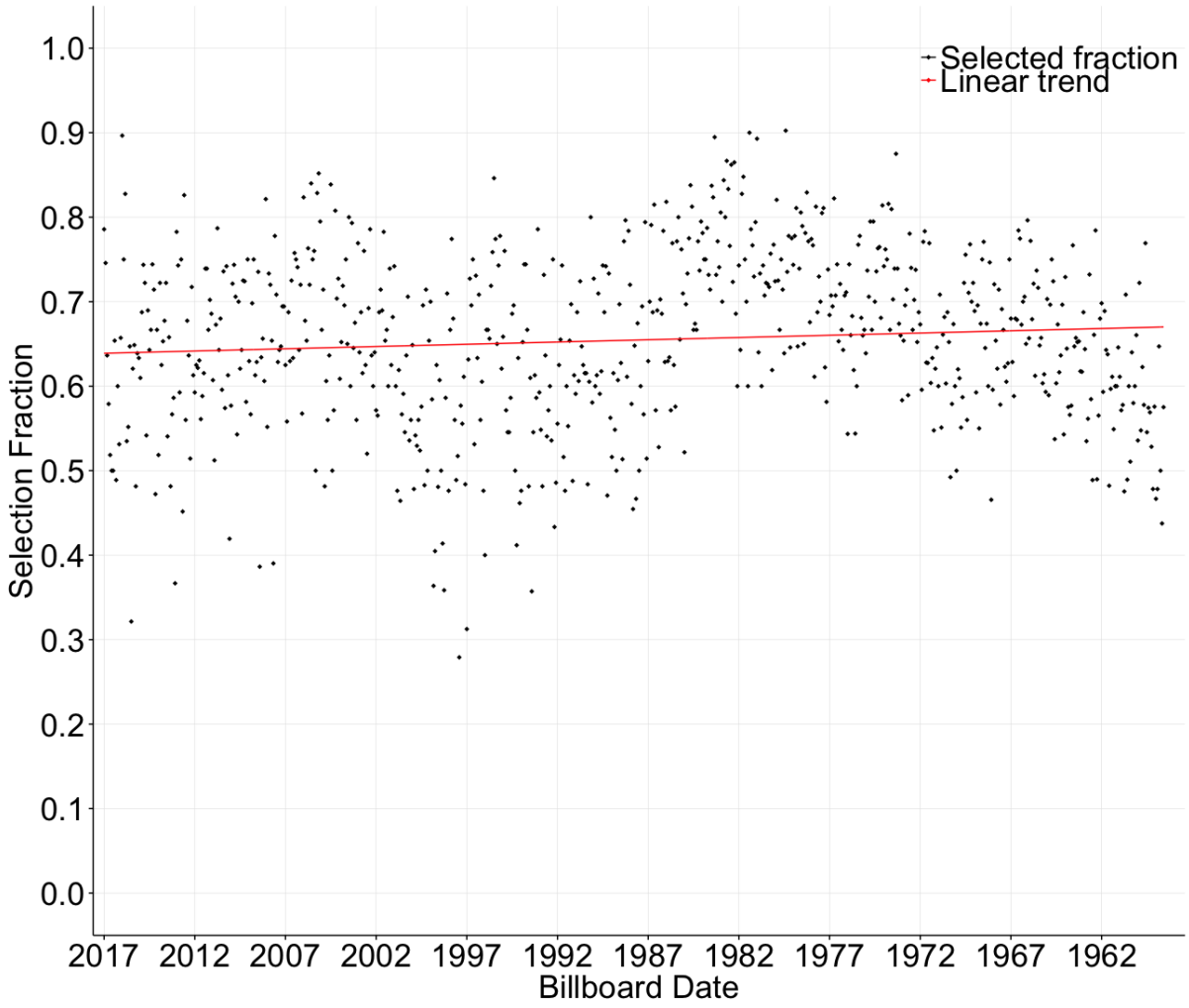
Inclusion Criterion of Data

The inclusion criterion of each cultural piece depends only on a proxy of its accomplishment, and is, in principle, independent of present-day popularity. For example, we cannot include biographies based on their number of language editions (as it has been done in the past[20]) because our sample would be biased against old, unpopular, biographies. Instead, here the inclusion criteria of each type of cultural piece depend only on a proxy of its initial accomplishment: Billboard hot 100 ranking for songs, more than 1.000 votes in IMDB for movies, and athletic performance ranking for biographies, making it independent of their present-day popularity.

Then, to extend this study, you must select content whose selection criteria does not be correlated with the popularity measure.

Bias Check of Music Data

Since not all Billboard songs are present in Spotify and last.FM, we explore the nearly 20 thousand songs that were present in both. We check for selection bias when we match data from Billboard Magazine and Spotify. Given the linear trend in Fig. 3, we can see that there is not any selection bias in our sub-sample.



Supplementary Figure 3: We can see that there is not a significant bias in the matching of songs between Billboard (complete data) and Spotify (selected data). The same plot is for Last.fm data, because we use the same songs filtered by songs availables on Spotify.

Supplementary Model

There is an alternative way to solve the model as an uncoupled differential equation system, which give us the same mathematical result:

$$
\begin{aligned}
S_{t+1} &= n_t(1-p) + \eta_t(1-q) && (2)\\
&= n_t + \eta_t - n_t p - q\eta_t && (3)\\
&= n_t + \eta_t - (n_t + \eta_t)p + p\eta_t - q\eta_t && (4)\\
&= n_t + \eta_t - (n_t + \eta_t)p + \eta_t(p-q) && (5)\\
&= S_t - S_t p + \eta_t(p-q) && (6)\\
S_t &= n_t + \eta_t && (7)\\
\eta_{t+1} &= (1-q)\eta_t && (8)
\end{aligned}
$$

We can rewrite these recurrence relations 5 and 8 as differential equations:

$$
\begin{aligned}
\frac{dS}{dt} &= -S_t p + \eta_t(p-q) && (9)\\
\frac{d\eta}{dt} &= -q\eta(t) && (10)\\
S(t=0) &= 1 && (11)\\
\eta(t=0) &= \eta_0. && (12)
\end{aligned}
$$

Solving first equation 10 and using the initial condition giving by equation 12, we have

$$
\begin{aligned}
\frac{d\eta}{dt} &= -q\eta(t) && (13)\\
\int \frac{d\eta}{\eta(t)} &= \int -q dt && (14)\\
log(\eta(t)) &= -qt + C && (15)\\
\eta(t) &= \eta_0 e^{-qt} && (16)
\end{aligned}
$$

Now, replacing 16 in 9, solving the differential equation transforming it into an exact equation.

$$
\begin{aligned}
\frac{dS}{dt} &= -S_t p + \eta_0 e^{-qt}(p-q) && (17)\\
-\eta_0(p-q)e^{-qt} + pS + \frac{dS}{dt} &= 0 && (18)
\end{aligned}
$$

Let $M(t,S) = pS - \eta_0(p-q)e^{-qt}$ and $N(t,S) = 1$, we note that this is not an exact equation, because:

$$\frac{\partial M(t,S)}{\partial S} = p \neq 0 = \frac{\partial N(t,S)}{\partial t} \tag{19}$$

Therefore, we need to find an integrating factor $\mu(t)$ such that:

$$\big(\mu(t)M(t,S)\big) + \big(\mu(t)\frac{dS}{dt}N(t,S)\big) = 0 \tag{20}$$

is exact, this is:

$$\frac{\partial}{\partial S}\big(\mu(t)M(t,S)\big) = \frac{\partial}{\partial t}\big(\mu(t)N(t,S)\big) \tag{21}$$

$$p\mu(t) = \frac{d\mu(t)}{dt} \tag{22}$$

$$\frac{\frac{d\mu(t)}{dt}}{\mu(t)} = p \tag{23}$$

$$\int \frac{\frac{d\mu(t)}{dt}}{\mu(t)} dt = \int p dt \tag{24}$$

$$log(\mu(t)) = pt \tag{25}$$

$$\mu(t) = e^{pt} \tag{26}$$

Now, replacing 26 in 20 and using $M(t,S)$ and $N(t,S)$ we have:

$$e^{pt}\big(pS - \eta_0(p-q)e^{-qt}\big) + e^{pt}\big(\frac{dS}{dt}\big) = 0 \tag{27}$$

Let $M_2(t,S) = e^{pt}\big(pS - \eta_0(p-q)e^{-qt}\big)$ and $N_2(t,S) = e^{pt}$, we note that this is an exact equation, because:

$$\frac{\partial M_2(t,S)}{\partial S} = pe^{pt} = \frac{\partial N_2(t,S)}{\partial t} \tag{28}$$

Now, let $f(t,S)$ such that:

$$\frac{\partial f(t,S)}{\partial t} = M_2(t,S) \tag{29}$$

$$\frac{\partial f(t,S)}{\partial S} = N_2(t,S) \tag{30}$$

Then, the solution will be:

$$f(t,S) = C_1 \tag{31}$$

Where C_1 is an arbitrary constant. In order to find $f(t,S)$:

$$\frac{\partial f(t,S)}{\partial t} = e^{pt}\big(pS - \eta_0(p-q)e^{-qt}\big) \tag{32}$$

$$\int \frac{\partial f(t,S)}{\partial t} dt = \int e^{pt}\big(pS - \eta_0(p-q)e^{-qt}\big) dt \tag{33}$$

$$f(t,S) = -\eta_0 e^{(p-q)t} + Se^{pt} + g(S) \tag{34}$$

Where $g(S)$ is an arbitrary function of S. In order to fin $g(S)$:

$$\frac{\partial f(t,S)}{\partial S} = \frac{\partial}{\partial S}\big(- \eta_0 e^{(p-q)t} + Se^{pt} + g(S)\big) \tag{35}$$

$$\frac{\partial f(t,S)}{\partial S} = e^{pt} + \frac{dg(S)}{dS} \tag{36}$$

Using the definition of $N_2(t,S)$ and equations 30 and 36 we have:

$$\frac{\partial f(t,S)}{\partial S} = e^{pt} \tag{37}$$

$$e^{pt} = e^{pt} + \frac{dg(S)}{dS} \tag{38}$$

$$\frac{dg(S)}{dS} = 0 \tag{39}$$

$$\int \frac{dg(S)}{dS} dS = \int 0 dS \tag{40}$$

$$g(S) = 0 \tag{41}$$

Replacing equation 41 in equation 34 and using equation 31 we have:

$$f(t,S) = -\eta_0 e^{(p-q)t} + Se^{pt} + 0 \tag{42}$$

$$-\eta_0 e^{(p-q)t} + Se^{pt} = C_1 \tag{43}$$

$$S(t) = e^{-pt}\big(\eta_0 e^{(p-q)t} + C_1\big) \tag{44}$$

Finally, using the initial condition 11:

$$S(t) = C_1 e^{-pt} + \eta_0 e^{-qt} \tag{45}$$

$$S(t=0) = 1 = C_1 + \eta_0 \tag{46}$$

$$C_1 = 1 - \eta_0 \tag{47}$$

$$S(t) = (1-\eta_0)e^{-pt} + \eta_0 e^{-qt} \tag{48}$$

$$S(t) = e^{-pt} + \eta_0(e^{-qt} - e^{-pt}) \tag{49}$$

Transition Time

There are others alternative ways to estimate the critical time. Using Eq. 49, the critical time, t_c, can also be found by noting that the second part of the model $eta_0(e^{-qt} - e^{-pt})$ has a maximum when the relevant process changes significantly from the second to the first at t_{c1}:

$$\frac{dS}{dt} = \eta_0(-qe^{-qt} + pe^{-pt}) \tag{50}$$

$$\eta_0(-qe^{-qt_{c1}} + pe^{-pt_{c1}}) = 0 \tag{51}$$

$$qe^{-qt_{c1}} = pe^{-pt_{c1}} \tag{52}$$

$$log(q) - qt_{c1} = log(p) - pt_{c1} \tag{53}$$

$$pt - qt_{c1} = log(p) - log(q) \tag{54}$$

$$t_{c1} = \frac{log(p/q)}{p-q} \tag{55}$$

And also, the derivative of the second process has a minimum when the main process start to deviate from the second process at t_{c2}:

$$\frac{d^2S}{dt^2} = \eta_0(q^2e^{-qt} - p^2e^{-pt}) \tag{56}$$

$$\eta_0(q^2e^{-qt_{c2}} - p^2e^{-pt_{c2}}) = 0 \tag{57}$$

$$p^2e^{-pt_{c2}} = q^2e^{-qt_{c2}} \tag{58}$$

$$2log(q) - qt_{c2} = 2log(p) - pt_{c2} \tag{59}$$

$$pt - qt_{c2} = 2(log(p) - log(q)) \tag{60}$$

$$t_{c2} = 2\frac{log(p/q)}{p-q} \tag{61}$$

Another way to find the time noting that the second part of the equation 49, $\eta_0(e^{-qt} - e^{-pt}$, has a maximum

when the relevant process changes significantly from the second process to the sum of the processes at t_{c1}:

$$\frac{dS}{dt} = \eta_0(-qe^{-qt} + pe^{-pt}) = 0$$
$$t_{c1} = \frac{log(p/q)}{p-q} \tag{62}$$

The second critical time can be found noting that the derivative of the second process has a minimum when this starts to deviate from the main process slightly from the second process at t_{c2}:

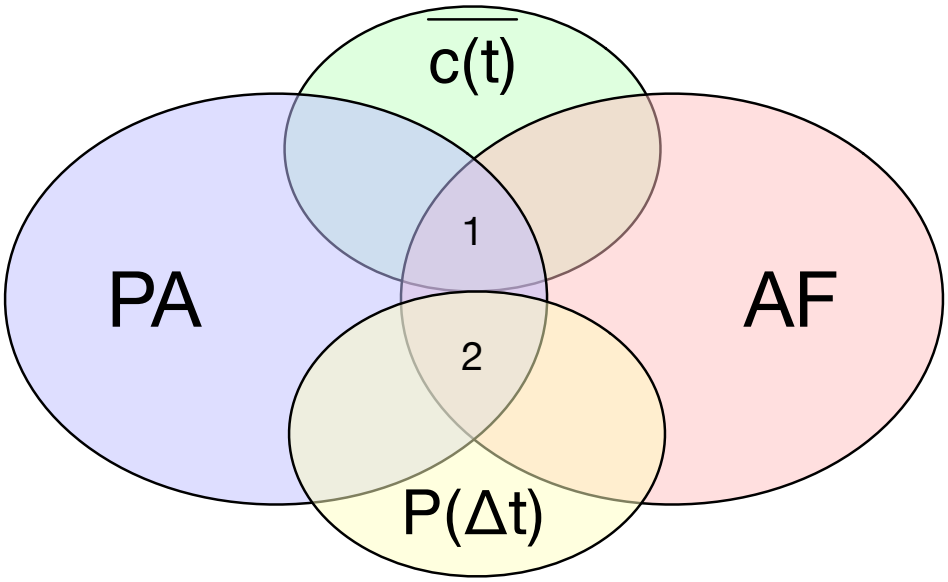
$$\frac{d^2S}{dt^2} = \eta_0(q^2e^{-qt} - p^2e^{-pt}) = 0$$
$$t_{c2} = 2\frac{log(p/q)}{p-q} \tag{63}$$

Supplementary Note 1

Literature Intersection

Here we talk about the connection of knowledge diffusion and science of science. We can divide the literature on knowledge diffusion that uses the citations as an indicator of knowledge flows in two different macro-dimensions, these are geographical and temporal macro-dimensions[21]. The geographical macro-dimension is about how knowledge flow from one place to another. It has been shown that the knowledge flows quicker to proximate areas[22], besides after controlling by geographical proximity, knowledge also flows faster from regions that have embedded related knowledge[23]. For instance, regions in China, which have related industries (e.g., socks and t-shirts), can learn from each other because they share similar knowledge[24], this implies a rising in the probability of knowledge flux from one place to another, and it's proportional to the overlapping of previous sharing knowledge. However, here we focus on the second macro-dimension, this is the temporal macro-dimension of knowledge diffusion.

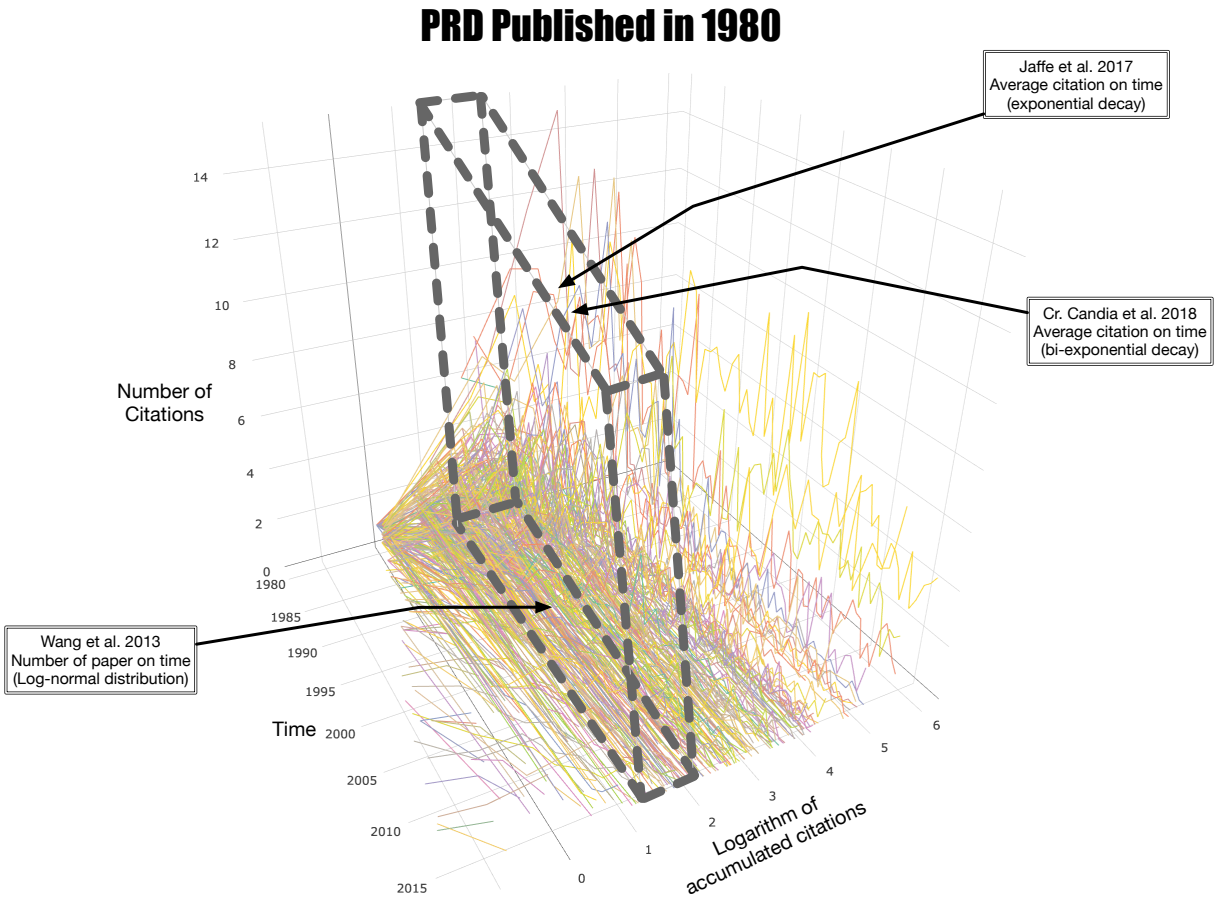
Literature at the temporal macro-dimension can be classified considering two different mechanisms: 1) cumulative advantage mechanism or preferential attachment (PA). 2) Knowledge obsolescence or aging function (AF). Several models describe the citation pattern using one or both mechanisms combined. There is almost a plenty consensus on the existence of PA and AF mechanisms, however, is still not clear how to operationalize and to model the AF mechanisms. Our contribution is focusing on how to measure and separate the AF mechanism empirically and how to model it. We are also able to explain the whole decay curve, in contrast with new literature which is using a similar framework[14,15], and we also claim that following our method it's possible to unveil a hidden and universal temporal pattern across many different domains, such as patents, scientific papers, songs, movies and even cultural icons (people).



Supplementary Figure 4: Literature characterization in function of mechanisms of preferential attachment (PA) and aging (AF) mechanisms, and approach for AF: average number of citations on time $\overline{\Delta c(t)}$ and probability distribution of citation time $P(\Delta t)$. Our contribution is located on 1, this means we recognize both mechanisms, PA and AF, in human collective forgetting, and we use as a empirical proxy for AF the average citations on time.

As Fig 4 shows, we can divide the literature on AF mechanisms, regardless of PA mechanism, considering two

approaches: 1) Calculating the citation rate of papers on time, $\overline{\Delta c(t)}$[14,15,25,26]. We note that our contribution is located at this intersection. 2) Estimating the probability of time since the cultural piece was released to get a new citation, $P(\Delta t)$[6,27–30]. We note that the first one, (1), considers more information than the second one (2), this is because in the first one the analyze is 'binary,' this means we only consider if a piece of content is cited (or not) in a given time, while in the second one the analyze is 'weighted,' because we consider if the piece of content is cited or not, but also it considers the number of citations that a piece of content gets on a given time.



Supplementary Figure 5: We observe how our model works, and how it is related with the two different bodies of literature, characterized by Higham, Jaffe et al.[14] and Wang et al.[6].

Let's consider the example plotted on figure 5. We observe all papers published in 1980 by Physical Review D. The three dimensions that completely describe the system are: 1) Age of the paper (Time), Number of Citation at each time (analogous to a particle's position), and Number of accumulated citation at each time (analogous to a particle's momentum). As we sentenced before, to explain, understand, and make predictions in the system, there are two significant ways on how to study the AF mechanism and both follow the same analyze strategy: fix the number of accumulated citations (this is equivalent to control by preferential attachment). The dashed box on the figure represents this strategy. In other words, all calculations and estimation are doing in a box that is moving across the whole axis which represents the number of accumulated citations.

For example, Higham, Jaffe et al.[14] calculate the average number of citations received on time inside of the box.

This means they calculate for a given number of accumulated citations and for a given time the number of citations received. This implies that they are using the three dimensions of the system.

Wang et al.[6], collapses one of the three relevant dimensions, this is the Number of Citations at each time, and then they estimate that the probability distribution of the time to get a new citation, for each level of the number of accumulated citations, corresponds to a log-normal distribution. On the figure, this is equivalent to work just with the number of papers that received an extra citation inside of the box, what is equivalent to say they work with the projection of the box to the plane described by Accumulated citations and Time, which implies that the Number of Citations dimension is not relevant to the analysis.

As we show on the figure, we are closer to Higham, Jaffe et al. 2017[14] work. But, we proposed a significant contribution here, the time pattern of knowledge obsolescence follows a bi-exponential decay, which can explain the complete curve of the time-dependence. Also, our model is motivated by two broadly theoretically studied memory mechanisms: Communicative memory and cultural memory. We must note that the whole body of literature that has focused on calculate the average number of citations controlling by accumulated citations have not been able to explain the decay curves at very short-term nor very long-term. Also, we explain why focusing on the distribution of papers which received a new citation on time is not a comprehensive approach to understand the system.

Supplementary Note 2

Comparison with Citation Models

We also claim here that the approach based on estimating the probability distribution of the number of papers that are cited on time, $P(\Delta t)$[6] is not comprehensive. Therefore, it's less informative on the complete dynamic of the system, because it doesn't consider information related to the 'momentum' of the system, this is the number of citation accrued by time. For example, Wang et al. 2013[6] focus on the intersection 2 showed on Fig 4, it means they consider the projection of the box in Fig. 5 in the plane formed by time and accumulated citations. They estimate the probability distribution of citation time by counting the number of paper which received citations in each time for a fixed level of accumulated citation. It implies that the information about the number of citations is implicit in the model. To understand what the authors are assuming in this simplification, let's use their model to obtain an explicit form of the citation rate on time for a fixed number of accumulated citation. We start from equation 3 in Wang et al. 2013[6]:

$$c_i^t = m\left[e^{\lambda_i \Phi\left(\frac{ln(t)-\mu_i}{\sigma_i}\right)} - 1\right] \tag{64}$$

Where c_i^t is the number of accumulated citations, m is the average number of references in a paper, λ_i captures the relative importance of the paper i relative to other papers (fitness), μ_i is the immediacy,σ_i is the longevity, and $\Phi(x) = (2\pi)^{-1/2}\int_{-\infty}^{x} e^{\frac{-y^2}{2}}dy$. So now, by construction, the number of citations on time is equal to the derivative of c_i^t on time.

$$s(t) = \frac{\partial c_i^t}{\partial t} = me^{\lambda_i \Phi\left(\frac{ln(t)-\mu_i}{\sigma_i}\right)}\lambda_i \frac{d\Phi\left(\frac{ln(t)-\mu_i}{\sigma_i}\right)}{dt} \tag{65}$$

We note that $me^{\lambda_i \Phi\left(\frac{ln(t)-\mu_i}{\sigma_i}\right)} = (c_i^t + m)$, and deriving the term using the definition of $\Phi(x)$ and the fundamental theorem of calculus we have:

$$s(t) = \frac{\lambda_i}{\sqrt{2\pi}\sigma_i}(c_i^t + m)\frac{1}{t}e^{\left(\frac{-(ln(t)-\mu_i)^2}{2\sigma_i^2}\right)} \tag{66}$$

Given that both strategies (ours and Wang et al. 2013[6]) focus on separate the data using a fixed level of accumulated citation, this in order to separate the effect of PA from AF, we must consider a fixed level of $c_i^t = c_i^t|_C$, thus applying a logarithm function in both sides of Eq. 66, and ordering terms, we have:

$$ln(s(t)) = \beta_0 + \beta_1 ln(t) - \beta_2 (ln(t))^2 \tag{67}$$

Where, $\beta_0 = ln\left(\frac{\lambda_i}{\sqrt{2\pi}\sigma_i}(c_i^t + m)\right) - \mu_i^2$, $\beta_1 = \frac{\mu_i}{\sigma_i^2} - 1$, and $\beta_2 = \frac{1}{2\sigma_i^2}$. We note that Eq. 67 correspond to a parabola in a log-log scale, or equivalently to a log-normal decay. Now, we can apply an exponential function, and we have:

$$s(t) = \beta_0^{'} t^{\beta_1} e^{-\beta_2 (ln^2(t))} \tag{68}$$

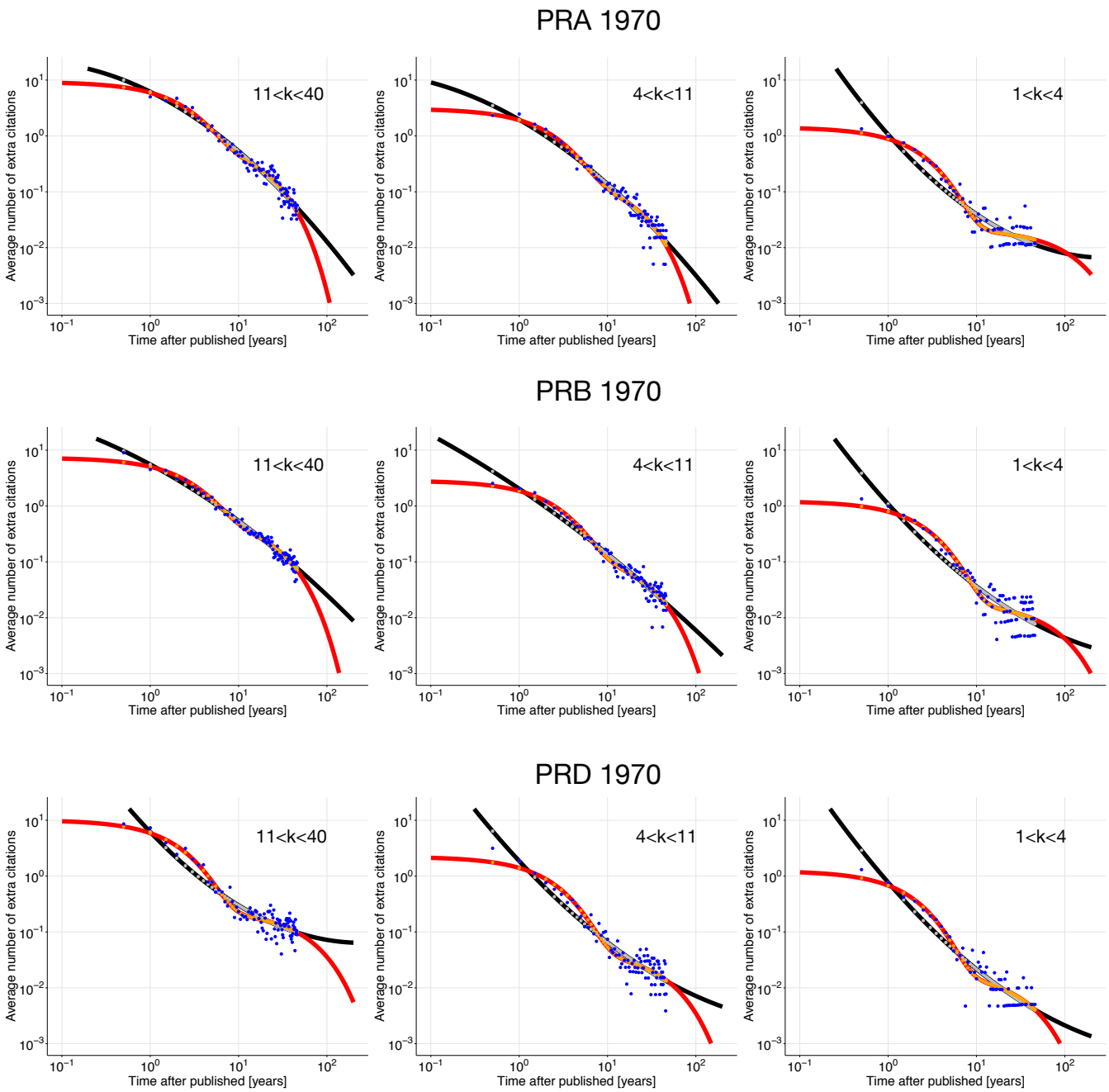
where $\beta_0^{'} = e^{\beta_0}$. As we said before, the implicit form that the authors consider in their analysis corresponds to a log-normal decay or equivalently to a power-law grow multiply by an exponential decay.

Now, let's make the empirical exercise of fitting the curve. Here we proceed on both, fitting the parameters with no bounds and fitting the parameters with the bounds given by the theoretical derivation of Eq. 67. On Fig. 6 we observe the fitting without bounds, it means we are going to allow to the algorithm to fit the best parameters regardless of the theoretical restrictions. We fit Eq. 67, obtained from Wang et al. 2013[6], and we observe that the curve fits quite well to the data, with a few exceptions on the very beginning and at the long term, particularly we observe data (blue dots) decays nearer to our model (red line) than Wang's model (red line). Therefore, we can see in Fig. 7 that both models do an excellent job explaining the behavior, maybe with some minor troubles for Wang's model with the highest values. However, when we consider the theoretical restrictions for the parameters, they are $\beta_1 > 0 and \beta_2 > 0$, we can see two theoretical issues. First, we observe several curves concave up ($-\beta_2 - \frac{1}{2\sigma_i^2} > 0$), this is implausible for a decay function, just because at the limit, $t \to \infty$, the number of citations goes to infinite, and also because by definition $\sigma_i > 0$. Second, we observe for all curves in Fig. 6, the parabola vertex is located at the negative part or $ln(t)$, this implies that $\beta_1/4\beta_2 < 0$, which is impossible because always $\mu_i > \sigma_i^2$. Therefore, we fix to zero the lower limit for β_1 and β_2. We observe in Fig. 8 that Eq. 67 with its theoretical restrictions fails completely to fit the very beginning of the temporal decay (with a fixed level of accumulated citations) assigning an underestimated level of attention to cultural goods. It also has some troubles with the number of citations in the very long term, but this could be probably because of data resolution. However, we observe on Fig. 8 that blue dots (data) decay closer to red lines (our temporal model). Fig. 9 show how our model explains much better the highest values of citations (beginning of the decay curve).

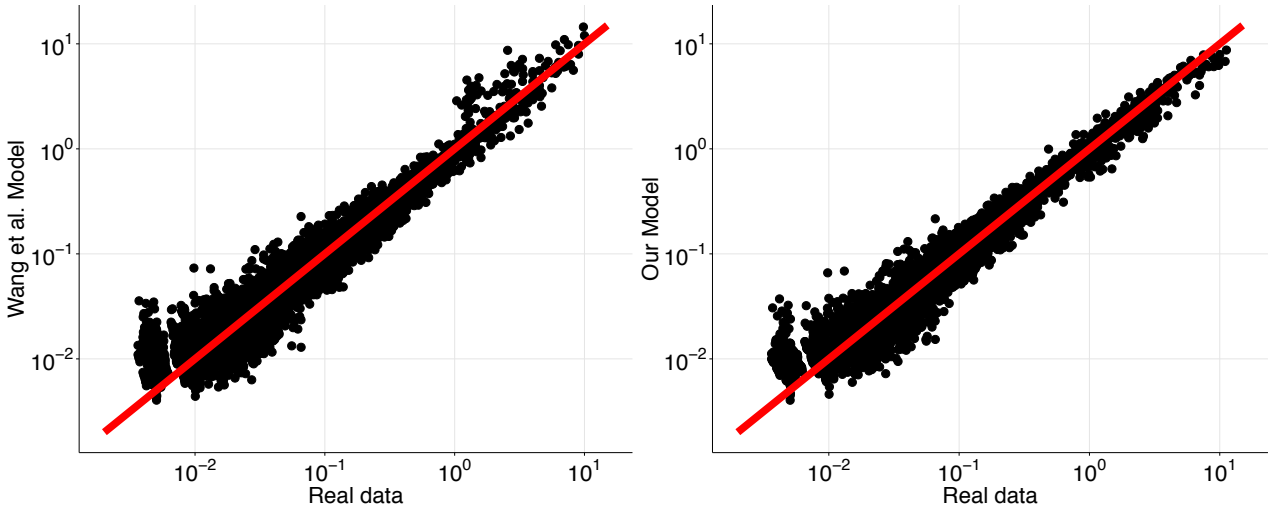
Finally, we can say the approach used by Wang et al. 2013 do a good job fitting the decay curve (Fig. 7), but without considering theoretical restrictions for parameters. However, when we include the restrictions, we observe that it fails at the beginning. It is because the model it's less comprehensive, therefore less accurate on the explanatory

power of the phenomenon, than our model, which is based on collective memory mechanisms. The main reason for this is they are collapsing the information contained at the "Number of Citations on Time" dimension (Fig 5). Given this lack of information, they are assuming a functional form to the temporal decay that presents some theoretical restrictions that are just realizable when the expression is explicit. When the expression is implicit like in Wang et al., is easy to estimate the level of attention accurately but it is just an artifact of the fitting algorithms as we observe in Fig. 6.

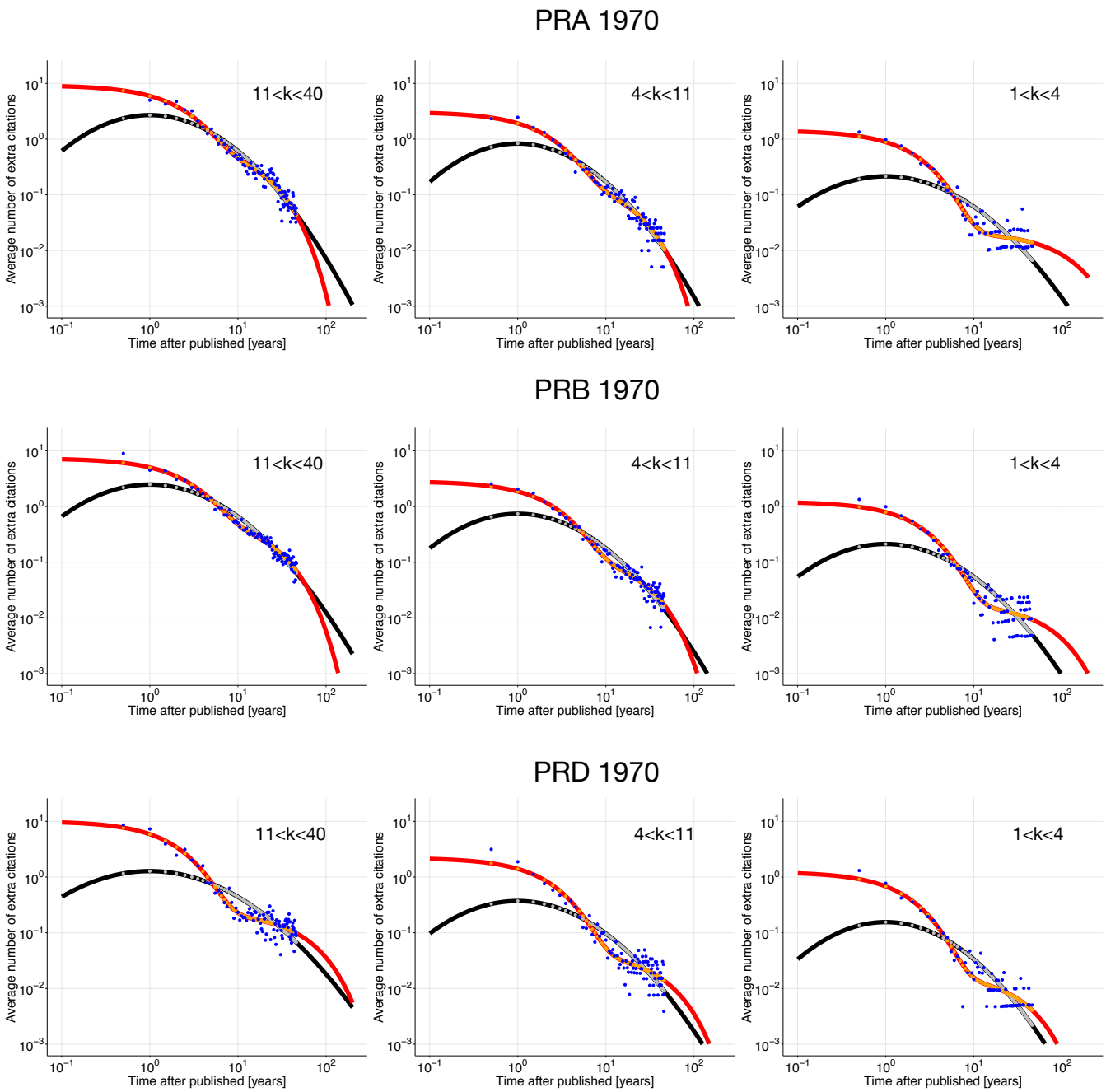
We are able to explain the whole curve accurately due to our model consider communicative memory as a part of its formulation. Another interesting consequence of using our approach is that we are unveiling a universal pattern on the temporal dimension of knowledge diffusion, a pattern that we called the universal temporal pattern of human collective forgetting, which exists in many different cultural domains, such as, songs, movies, patents, papers, and even people. It can be wholly explained using the model presented in this work (Fig 9, which comes from theoretical social anthropological studies on collective memory, taking in account two mechanisms that co-exist in all time: Communicative memory and cultural memory.



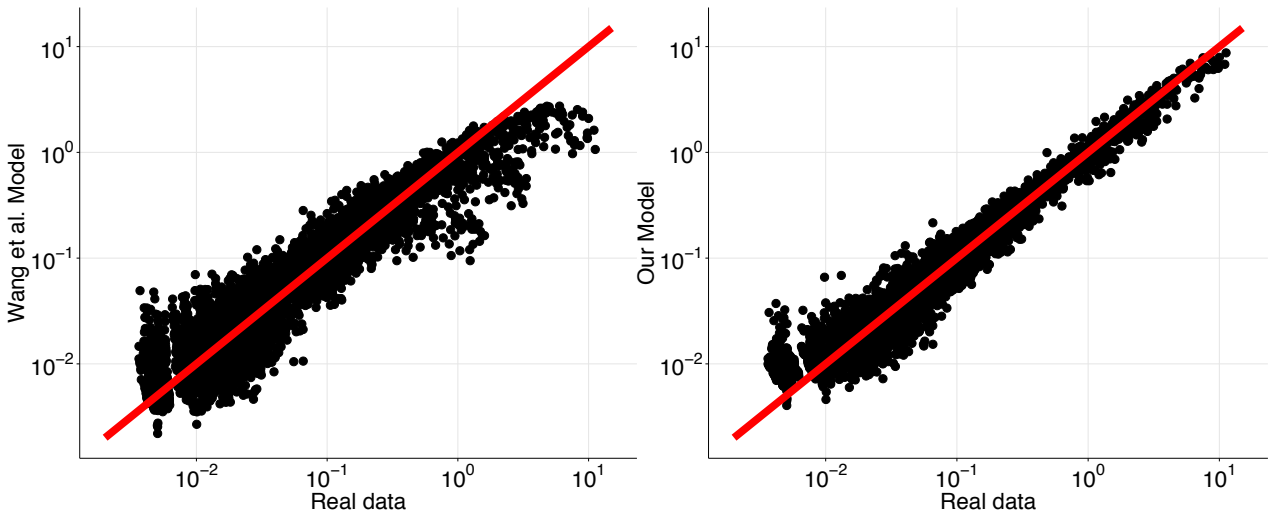
Supplementary Figure 6: Figure shows all papers published by PRA, PRB, and PRD in 1970 for different ranges of accumulated citations ($11 < k < 40$, $4 < k < 11$, $1 < k < 4$). Black lines represent the equation 67 and red lines represent our model for temporal decay explicited at the equation **??**. Blue dots are the data.



Supplementary Figure 7: Figure shows the comparison between real data (x-axis) and data explained by the models for all published by PRA, PRB, PRC, PRD, PRL in 1970, 1971, 1972 and 1973.



Supplementary Figure 8: Figure shows all papers published by PRA, PRB, and PRD in 1970 for different ranges of accumulated citations ($11 < k < 40$, $4 < k < 11$, $1 < k < 4$). Black lines represent the equation 67 and red lines represent our model for temporal decay explicited at the equation **??**. Blue dots are the data. Here we fix the signs of the parameters according to theoretical results.



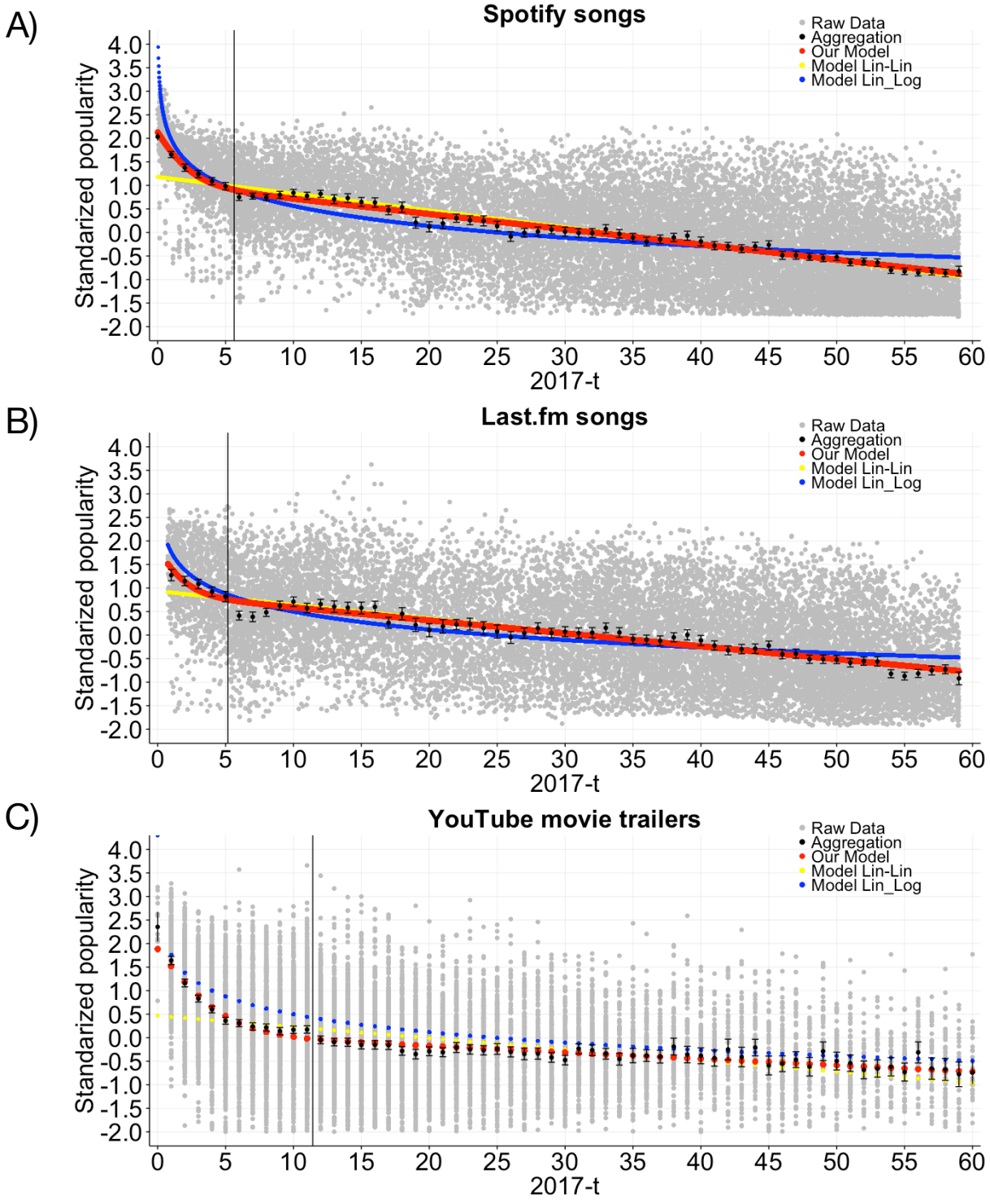
Supplementary Figure 9: Figure shows the comparison between real data (x-axis) and data explained by the models for all published by PRA, PRB, PRC, PRD, PRL in 1970, 1971, 1972 and 1973. Here we fix the signs of the parameters according to theoretical results.

Supplementary Note 3

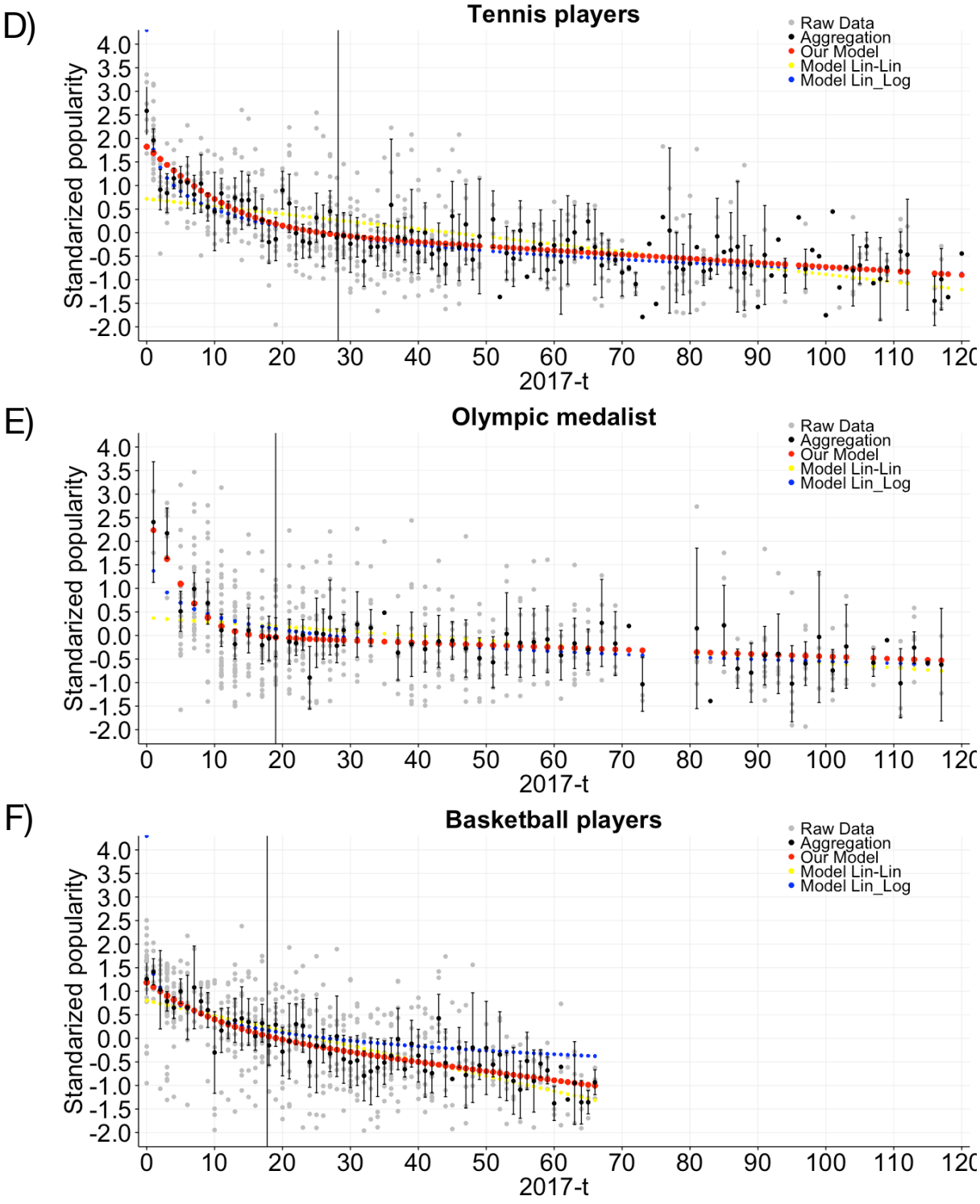
Comparison with Forgetting Models

At the individual level, forgetting has been modeled using exponential, logarithmic, power-law, and hyperbolic[31,32] functions, whereas at the collective level, scholars have used different data sources to study the dynamics of collective forgetting, including the fraction of the population that remembers a person or event[1,33] and the present online popularity of a piece of content[20,34–37]. Zaromb et al. (2014), interviewed adults about three major wars in U.S. history and found that younger adults experienced more consensus than older adults when recalling war events[33]. Roediger and DeSoto[1] asked hundreds of Americans to name their current and past presidents and found that three stages characterize forgetting: an initial fast rate of forgetting for recent presidents (a "recency effect"), followed by a slower decay for past presidents, and finally, a high recall rate for the first few presidents, which they call a "primacy effect"[2]. Wang et al. used the citation patterns of academic publications[6], to show that the decay of citations can be modeled using an aging function, and a normalization across each paper. Higham et al.[14] separated the preferential attachment and the aging effects to predict citation rates in patent data.

Figure 10 and 11 shows different plausible models from forgetting literature for each data set. We our model describe more accurately the date in both at the beginning and at the end.



Supplementary Figure 10: Comparing models



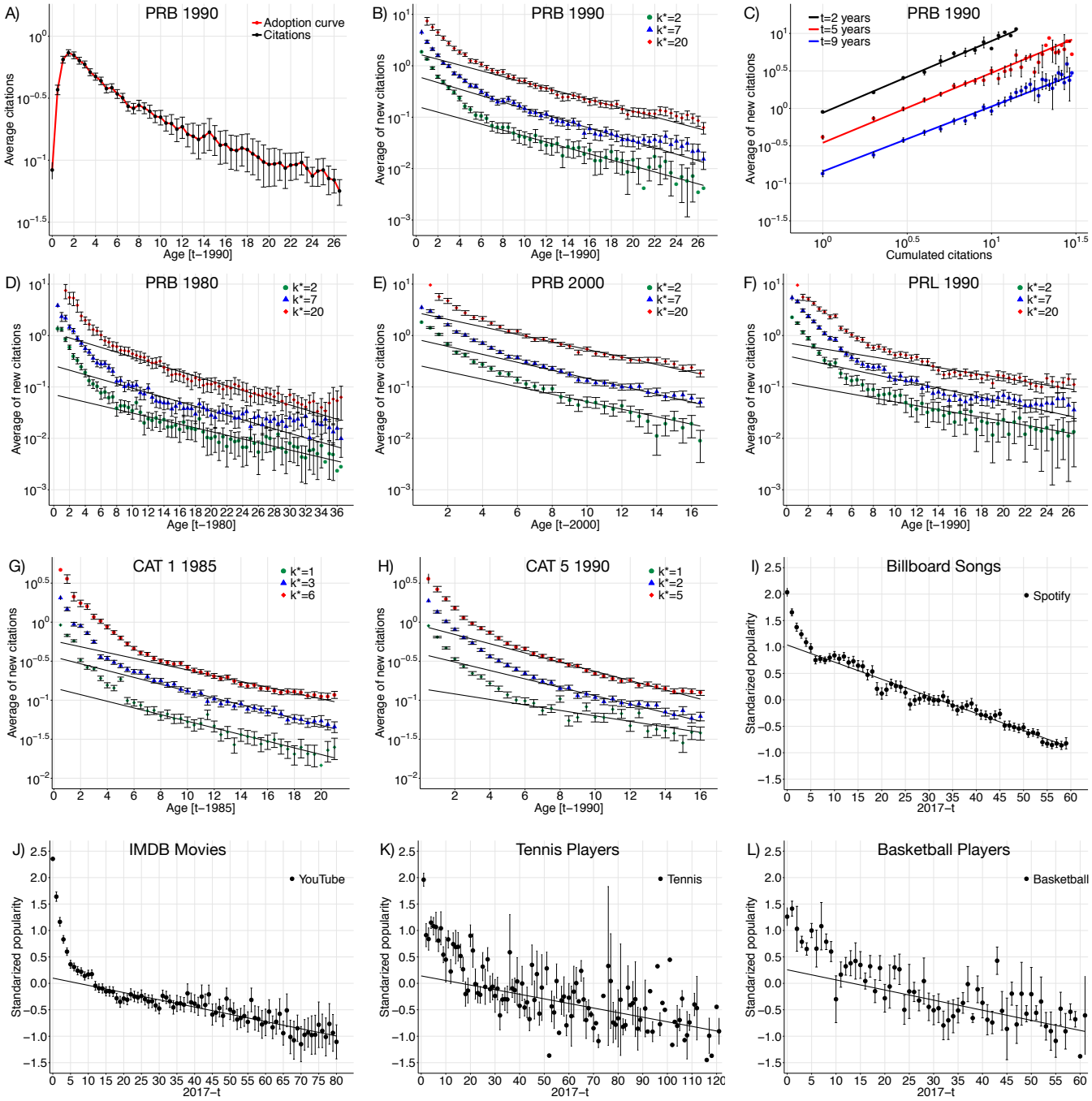
Supplementary Figure 11: Comparing models

Supplementary Note 4

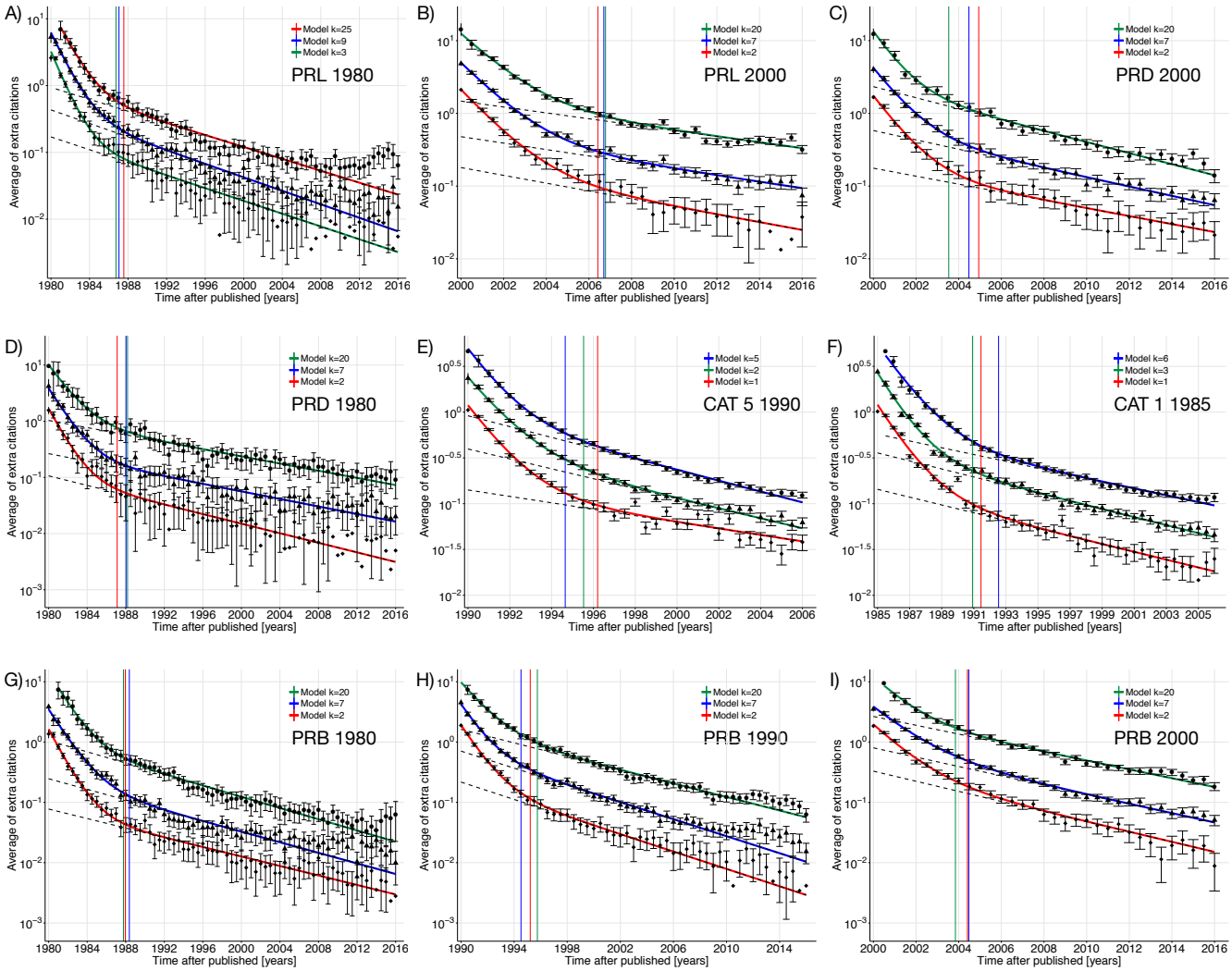
Citation Curve Decomposition

We illustrate this decomposition using data from papers and patents. Figure 12 C and D show, respectively, the decay curve for all papers published in PRL in 2000 and for all patents granted in 1990 for mechanical inventions, and also, decompose these curves by considering groups of papers and patents that have received the same total number of citations. Here we can see that the aggregate adoption curves (the one considering papers with all levels of preferential attachment) rise and fall, whereas the curves that consider only papers and patents with the same level of preferential attachment have an initial fast decay followed by a slower decay.

Fig. 12 shows that the attention received by songs (A and B), scientific papers (C) and patents (D) decays following a two-step process characterized by a fast initial short decay followed by a slower longer decay. Songs experience five years of "love" on average, and a relatively long time of "forgetting." This two-step process appears to be universal, in the sense that the functional form is the same for songs, measured using data from two different online platforms, scientific publications, and patents.



Supplementary Figure 12: Universal patterns in the decay of human collective memory. Attention decay for songs using data from Last.fm (A) and Spotify (B), for PRL papers published in 2000 (C), and for patents in Mechanical (CAT 5) category granted on 1990 (D). Each dot in A and B represents a song, x-axis corresponds to the date when a song reach the billboard ranking, and y-axis corresponds to the current popularity measure. Black dots represent a month aggregation and the red broken line represents an ad-hoc segmented regression[38] to guide the eye and to show the presence of a two-step process. Spotify popularity index (B) goes from 0 to 100, and it's a logarithmic function of play counts on a temporal window. For C and D, the x-axis is the time after publication (C) and time after granted (D), y-axis represents the number of extra citations in each time. k denotes the accomplishment level, defined as the middle point of the logarithm of accumulated citations. The dashed line corresponds to an exponential fit, which indicates that presence of another process.



Supplementary Figure 13: Forgetting cultural goods. Average number of extra citations for A) All papers published by Physical Review Letters in 1980, B) All papers published by Physical Review Letters in 2000, C) All papers published by Physical Review D in 2000, D) All papers published by Physical Review D in 1980, E) All Mechanical patents granted in 1990, and F) All Chemical patents granted in 1985. G) All papers published by Physical Review B in 1980. H) All papers published by Physical Review B in 1990. I) All papers published by Physical Review B in 2000. X-axis represents the time starting when the paper (patent) was published (granted), k represents the logarithmic middle point of the accrued citation level. Dashed lines represent an exponential fit to the cultural memory process, and vertical lines represent the critical time t_c.

Supplementary Tables

Here we show the table associated with Figure 3 in the main paper. We note that for papers and patents, we use the aggregated data to make the fit. We provide AICc for exponential, log-normal, and bi-exponential models over the same data, and we observe that AICc's are better for the bi-exponential model, which means that our model left less information without explaining. AICs penalize by the number of parameters and the number of data points of each model.

Supplementary Table 1: Regression Results Model: Patents

	Dependent variable: Level of attention (Citations)					
	Cat 5 1990 (3)	Cat 5 1990 (2)	Cat 5 1990 (1)	Cat 1 1985 (3)	Cat 1 1985 (2)	Cat 1 1985 (1)
p	0.6728***	0.5464***	0.5992***	0.5683***	0.6452***	0.5386***
	(0.0812)	(0.0221)	(0.0243)	(0.0520)	(0.0317)	(0.0232)
q	0.0953***	0.0977***	0.1161***	0.0944***	0.0978***	0.0764***
	(0.0135)	(0.0069)	(0.0052)	(0.0084)	(0.0036)	(0.0043)
r	0.0930***	0.0641***	0.0895***	0.0642***	0.0799***	0.0471***
	(0.0208)	(0.0069)	(0.0079)	(0.0112)	(0.0063)	(0.0038)
N	1.1646***	2.3123***	4.6492***	1.1067***	2.7126***	5.4459***
	(0.1287)	(0.0864)	(0.1646)	(0.1055)	(0.1366)	(0.2817)
t_c	5.8327	7.3694	6.0539	7.2489	6.2617	8.4825
	(0.2956)	(0.1948)	(0.1834)	(0.2507)	(0.1719)	(0.1647)
Obs	30	30	30	40	40	39
pseudo-R^2	0.9758	0.9973	0.9978	0.983	0.9954	0.9959
AICc Bi-exponential	-20.954	-87.4829	-94.3849	-34.175	-93.5709	-100.2472
AICc Log-Normal	-14.326	-69.924	-103.2909	-31.2382	-83.117	-66.4649
AICc Exponential	-7.7364	-36.0969	-45.8428	-27.8928	-62.4075	-51.8984

Note: *p<0.1; **p<0.05; ***p<0.01
AICc penalizes by the number of parameters and data points of each model

Supplementary Table 2: Regression Results Model: Cultural Goods

	Dependent variable: Current level of attention (Play counts and page views)					
	Spotify	Last.fm	Youtube	Tennis Players	Olympic Medalists	Basketball Players
p	0.4685***	0.4843***	0.3843***	0.1345***	0.3303***	0.1001***
	(0.0516)	(0.1353)	(0.0202)	(0.0216)	(0.0844)	(0.0313)
q	0.0324***	0.0275***	0.0137***	0.0087***	0.0049***	0.0191***
	(0.0004)	(0.0005)	(0.0006)	(0.0013)	(0.0016)	(0.0051)
r	0.2197***	0.2752***	0.0746***	0.0287***	0.0285***	0.0528*
	(0.0164)	(0.0285)	(0.0028)	(0.0052)	(0.0069)	(0.0294)
N	8.4335***	6.2993***	6.5895***	6.1946***	12.9281**	3.2750***
	(0.5019)	(1.0467)	(0.3418)	(0.7615)	(5.4875)	(0.3727)
t_c	5.7076	5.2252	11.4898	28.5498	19.0278	18.7355
	(0.222)	(0.3551)	(0.1369)	(0.2728)	(0.2955)	(0.5538)
df	18303	15271	14629	620	522	588
pseudo-R^2	0.3966	0.2525	0.2342	0.451	0.1853	0.3595
RSE	0.7768	0.8646	0.8752	0.7427	0.9052	0.8024

Note: *p<0.1; **p<0.05; ***p<0.01

Supplementary Table 3: Regression Results Model: Papers

	Dependent variable: Level of attention (Citations)										
	PRB 1980 (3)	PRB 1980 (2)	PRB 1980 (1)	PRD 1980 (3)	PRD 1980 (2)	PRD 1980 (1)	PRL 1980 (2)	PRL 1980 (1)	PRL 1990 (3)	PRL 1990 (2)	PRL 1990 (1)
p	0.4875***	0.5037***	0.3757***	0.5518***	0.5581***	0.4497***	0.4417***	0.3273***	0.5666***	0.4626***	0.4321***
	(0.0467)	(0.0261)	(0.0210)	(0.0575)	(0.0517)	(0.0355)	(0.0269)	(0.0201)	(0.0442)	(0.0219)	(0.0227)
q	0.0726***	0.0630***	0.0458***	0.0593***	0.0593***	0.0589***	0.0612***	0.0207**	0.0919***	0.0449***	0.0563***
	(0.0066)	(0.0038)	(0.0055)	(0.0070)	(0.0054)	(0.0038)	(0.0053)	(0.0082)	(0.0079)	(0.0080)	(0.0080)
r	0.0157***	0.0130***	0.0090***	0.0129***	0.0168***	0.0286***	0.0112***	0.0053***	0.0227***	0.0100***	0.0143***
	(0.0031)	(0.0014)	(0.0014)	(0.0027)	(0.0028)	(0.0035)	(0.0017)	(0.0012)	(0.0038)	(0.0016)	(0.0023)
N	1.1549***	3.0476***	5.7689***	1.9250***	5.8589***	11.2699***	5.2186***	8.3888***	2.6288***	5.8895***	12.6049***
	(0.2038)	(0.3073)	(0.6160)	(0.4035)	(1.0166)	(1.2887)	(0.6090)	(1.0011)	(0.3464)	(0.5075)	(1.0629)
t_c	12.096	12.3985	16.8917	11.671	10.9777	11.2302	14.1216	21.9045	9.8512	14.2342	13.6944
	(0.2259)	(0.1622)	(0.1918)	(0.2259)	(0.2079)	(0.1977)	(0.1882)	(0.2604)	(0.216)	(0.2023)	(0.2082)
Obs	70	70	69	70	70	70	70	69	50	50	50
pseudo-R^2	0.9586	0.9851	0.9804	0.9392	0.956	0.9757	0.9788	0.9685	0.9807	0.9879	0.9881
AICc Bi-exponential	42.8709	-37.5985	-24.5303	63.8473	29.6509	-32.7756	-6.0015	6.3698	-12.1371	-40.3657	-45.2021
AICc Log-Normal	55.8737	23.819	24.308	87.4583	52.3836	-19.8774	20.5884	49.3654	2.9487	18.7356	-9.69
AICc Exponential	63.3675	28.4005	51.0032	79.6036	46.8568	0.0375	47.3897	63.2738	10.4387	18.8642	7.2593

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

AICc penalizes by the number of parameters and data points of each model

Supplementary Discussion

We found a two-step process in the human forgetting, characterized by the relevance of a piece of information in communicative memory and cultural memory. This process is described according to the literature, and we found a typical decay function: smooth, monotonic, at first, but then leveling[31,39]. This work suggests empirical evidence to support the forgetting as annulment and forgetting as planned obsolescence[40]. Finally, our findings confirm that the dynamic of human forgetting is characterized by a narrow set of mathematical functions and not only contribute to our understanding on human behavior, but also offer a reliable way to quantify the impact of a new piece of information in the society, and this may have potential industrial and policy implications.

These results are according with the literature, that suggests a smooth, monotonic, decreasing rapidly at first, but then leveling decay function[31,39]. On the other hand, this is evidence about the recency effect observed in others empirical works like Roediger and DeSoto[1] or Kelley et al.[32].

According to Connerton[40] we can distinguish seven different types of forgetting: i) repressive erasure, ii) prescriptive forgetting, iii) forgetting that is constitutive in the formation of a new identity, iv) structural amnesia, v) forgetting as annulment, vi) forgetting as planned obsolescence, and vii) forgetting as humiliated silence. However, in this work, we see evidence on tow of Connerton's types of forgetting, v) forgetting as annulment and vi) forgetting as planned obsolescence.

Forgetting as annulment, refers basically to the excess of information and our inability to filter information promotes the forgetting[40]. It is dangerous because we are very likely to forget valuable information if we don't have any mechanism to remember information from historical memory[41,42]. Given this, a system with more information or more dynamic should be forgotten first than the other with less available information. In this work, we find that songs are forgotten faster than movies and movies are forgotten more quickly than people. And we know, for example, in 2016 the number of songs which reached the Billboard ranking was 600 approximately– note that those are just the top best songs of the year– is very similar to the number of all movies that Hollywood produced in 2016, approximately 700.

Forgetting as planned obsolescence refers to those ideas, events, productions, or people that will be forgotten as part of market cycles and social pressures. Cultural productions, for example, are expected to be forgotten fast at first, given the social incentives for consumers of this market to forget quickly, and after this massive process of discarding of information, a transition to another process characterized by a slower forgetting rate, which is described by our model.

In words of Connerton[40] *Paradigm shifts modulate the discarding of information in academic productions. This need to discard is felt most acutely, of course, in the natural sciences. As long ago as 1963 it was calculated that 75 percent of all citations in the area of physics were taken from writings that were less than ten years old. Every scientist needs to learn how to forget in this way if his or her research activity is not to be crippled by chronic over information at the very outset. Indeed, Kuhn's concept of the scientific paradigm is an idea about forgetting. Kuhn sees the development of science as one in which every shift in scientific evolution unburdens scientific memory, where every collapse of a paradigm is always an act of forgetting of great importance for the economy of scientific effort. The paradigm that has been surpassed is one that can be forgotten. Even if the historical disciplines are not subject to*

such a drastic process of inbuilt obsolescence, they also have been marked by a paradigm shift and its corresponding cultural forgetting. This same process is what we found on Scientific data in the APS corpus, the transition time between communicative memory (when papers and patents received the bulk of citations) to cultural memory is around five years in average, slightly varying with the subfield.

Supplementary References

1. Roediger, H. & DeSoto, K. Forgetting the presidents. *Science* **346,** 1106–1109 (2014).
2. Ebbinghaus, H., Bussenius, C. E. & Ruger, H. A. *Memory; a contribution to experimental psychology* (Teachers College, Columbia University, 1913).
3. Fortunato, S. *et al.* Science of science. *Science* **359** (2018).
4. Radicchi, F., Fortunato, S. & Castellano, C. Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences* **105,** 17268–17272 (2008).
5. Kuhn, T., Perc, M. c. v. & Helbing, D. Inheritance Patterns in Citation Networks Reveal Scientific Memes. *Phys. Rev. X* **4,** 041036 (4 Nov. 2014).
6. Wang, D., Song, C. & Barabási, A.-L. Quantifying long-term scientific impact. *Science (New York, N.Y.)* **342,** 127–32 (2013).
7. Sinatra, R., Wang, D., Deville, P., Song, C. & Barabasi, A.-L. Quantifying the evolution of individual scientific impact. *Science* **354** (Nov. 2016).
8. King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J. & West, J. D. Men Set Their Own Cites High: Gender and Self-citation across Fields and over Time. *Socius* **3** (2017).
9. Chu, J. S. G. & Evans, J. A. Too Many Papers? Slowed Canonical Progress in Large Fields of Science. *Preprint at osf.io/preprints/socarxiv/jk63c* (Mar. 2018).
10. Yook, S.-H., Radicchi, F. & Meyer-Ortmanns, H. Self-similar scale-free networks and disassortativity. *Phys. Rev. E* **72,** 045105 (4 Oct. 2005).
11. Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. Atypical Combinations and Scientific Impact. *Science* **342,** 468–472 (2013).
12. Mukherjee, S., Uzzi, B., Jones, B. F. & Stringer, M. in *Knowledge and Networks* (eds Glückler, J., Lazega, E. & Hammer, I.) 243–267 (Springer International Publishing, Cham, 2017).
13. Mukherjee, S., Romero, D. M., Jones, B. & Uzzi, B. The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot. *Science Advances* **3** (2017).
14. Higham, K. W., Governale, M., Jaffe, A. B. & Zülicke, U. Fame and obsolescence: Disentangling growth and aging dynamics of patent citations. *Physical Review E* **95** (2017).
15. Higham, K., Governale, M., Jaffe, A. & Zülicke, U. Unraveling the dynamics of growth, aging and inflation for citations to scientific articles from specific research fields. *Journal of Informetrics* **11,** 1190 –1200 (2017).

16. Bouabid, H. Revisiting citation aging: a model for citation distribution and life-cycle prediction. *Scientometrics* **88,** 199 (Mar. 2011).

17. Glänzel, W. Towards a model for diachronous and synchronous citation analyses. *Scientometrics* **60,** 511–522 (Aug. 2004).

18. Yin, Y. & Wang, D. The time dimension of science: Connecting the past to the future. *Journal of Informetrics* **11,** 608 –621 (2017).

19. Hall, B. H., Jaffe, A. B. & Trajtenberg, M. *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools* NBER Working Papers at www.nber.org/papers/w8498 8498 (National Bureau of Economic Research, Inc, Oct. 2001).

20. Yu, A. Z., Ronen, S., Hu, K., Lu, T. & Hidalgo, C. A. Pantheon 1.0, a manually verified dataset of globally famous biographies. *Scientific Data* **3,** 150075 (Jan. 2016).

21. Jaffe, A. B. & de Rassenfosse, G. Patent citation data in social science research: Overview and best practices. *Journal of the Association for Information Science and Technology* **68,** 1360–1374 (June 2017).

22. Jaffe, A. B., Trajtenberg, M. & Henderson, R. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics* **108,** 577–598 (Aug. 1993).

23. Hidalgo, C. A., Klinger, B., Barabasi, A.-L. & Hausmann, R. The Product Space Conditions the Development of Nations. *Science* **317,** 482–487 (July 2007).

24. Gao, J., Jun, B., Pentland, A. S., Zhou, T. & Hidalgo, C. A. Collective Learning in China's Regional Economic Development. *arXiv* (Mar. 2017).

25. Simkin, M. V. & Roychowdhury, V. P. A mathematical theory of citing. *Journal of the American Society for Information Science and Technology* **58,** 1661–1673 (Sept. 2007).

26. Medo, M., Cimini, G. & Gualdi, S. Temporal Effects in the Growth of Networks. *Physical Review Letters* **107,** 238701 (Dec. 2011).

27. Valverde, S., Solé, R. V., Bedau, M. A. & Packard, N. Topology and evolution of technology innovation networks. *Physical Review E* **76,** 056118 (Nov. 2007).

28. Burrell, Q. L. Modelling citation age data: Simple graphical methods from reliability theory. *Scientometrics* **55,** 273–285 (2002).

29. Gupta, B. M. Analysis of distribution of the age of citations in theoretical population genetics. *Scientometrics* **40,** 139–162 (Sept. 1997).

30. Egghe, L. & Ravichandra rao, I. Citation age data and the obsolescence function: Fits and explanations. *Information Processing & Management* **28,** 201–217 (Jan. 1992).

31. Rubin, D. C. & Wenzel, A. E. One hundred years of forgetting: A quantitative description of retention. *Psychological Review* **103,** 734–760 (1996).

32. Kelley, M. R., Neath, I. & Surprenant, A. M. Serial position functions in general knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **41** (6 2015).

33. Zaromb, F., Butler, A. C., Agarwal, P. K. & Roediger III, H. L. Collective memories of three wars in United States history in younger and older adults. *Memory & cognition* **42,** 383–399 (2014).

34. Murray, C. *Human accomplishment: The pursuit of excellence in the arts and sciences, 800 BC to 1950* (Harper Collins, 2003).

35. Michel, J.-B. *et al.* Quantitative analysis of culture using millions of digitized books. *Science (New York, N.Y.)* **331,** 176–82 (2011).

36. Jara-Figueroa, C., Yu, A. Z. & Hidalgo, C. A. Estimating technological breaks in the size and composition of human collective memory from biographical data. *Preprint at https://arxiv.org/abs/1512.05020* (2015).

37. García-Gavilanes, R., Mollgaard, A., Tsvetkova, M. & Yasseri, T. The memory remains: Understanding collective memory in the digital age. *Science Advances* **3** (2017).

38. Muggeo, V. M. R. Estimating regression models with unknown break-points. *Statistics in Medicine* **22,** 3055–3071 (Oct. 2003).

39. Wixted, J. T. & Ebbesen, E. B. Genuine power curves in forgetting: a quantitative analysis of individual subject forgetting functions. *Memory & Cognition* **25,** 731–739 (1997).

40. Connerton, P. Seven types of forgetting. *Memory Studies* **1,** 59–71 (2008).

41. Halbwachs, M. *La mémoire collective* (Albin Michel, 1997).

42. Nora, P. Between Memory and History: Les Liex de Mémoire. *Representations* **26,** 7–25 (1989).