Supplementary Information for

The Time Dimension of Science: Connecting the Past to the Future

Yian Yin and Dashun Wang

Contents

S1 Fitting procedure for $P^{\leftarrow}(t)$ and $P^{\rightarrow}(t)$

To get the best fit for observed data, likelihood ratio test is used to compare the goodness of various distributions. While what we have is only the total number for each year, it is unwise to regard them as discrete and get a maximum likelihood estimation directly. Actually, the citation numbers within a year is a sum (integration) over the year, thus can be treated as binned data with length 1 [1]. For example, when we collect data for a specific time, the number of citations to the last time unit corresponds to the bin (0.5,1.5). The idea of using binned data to describe the continuity of time can be traced back to [2].

Another characteristic of our data is that it could not exactly follow a power law distribution (or exponential, lognormal, etc.), even if they should. One important reason for this lies in that we only get a finite time to study the citation dynamics, and citation beyond this period would never occur. Therefore, a cutoff would emerge in the age distribution, and we should fit the data with bounded distributions instead.

We simply restrict the distribution onto $[0.5, t + 0.5]$, where t is the length of observation time. Therefore, the appropriate normalization constant C should satisfy

$$C_{\text{pow}} \int_{x_{\min}}^{x_{\max}} f(x)dx = 1 \tag{S1}$$

For power law distribution, this leads to

$$C_{\text{pow}} = (\alpha - 1)/(x_{\min}^{1-\alpha} - x_{\max}^{1-\alpha}) \tag{S2}$$

Analogously, for power law with exponential cutoff, lognormal, exponential and Weibull, we obtain

$$C_{\text{pwe}} = \lambda^{1-\alpha}/(\Gamma(1-\alpha, \lambda x_{\min}) - \Gamma(1-\alpha, \lambda x_{\max})) \tag{S3}$$

$$C_{\text{LN}} = \left(\Phi(\frac{x_{\max} - \mu}{\sigma}) - \Phi(\frac{x_{\min} - \mu}{\sigma}) \right)^{-1} \tag{S4}$$

$$C_{\exp} = \lambda / (e^{-\lambda x_{\min}} - e^{-\lambda x_{\max}}) \tag{S5}$$

$$C_{\mathrm{wei}} = \lambda / (e^{-\beta \lambda x_{\min}^{\beta}} - e^{-\lambda x_{\max}^{\beta}}) \tag{S6}$$

S2 Temporal changes on $P^{\leftarrow}(t)$

The development of information and computer technology have conflicting influences on the age structure of citations. Indeed, online access to digitized collections makes literature more easily available, enhancing the use of older papers, which leads to Hypothesis 1 (*H1*): the average age of citations increases as we cite more older articles [3, 4]. On the other hand, searching online quickly puts researchers in touch with prevailing opinion through hyperlinks, narrowing the range of findings and ideas to build upon [5]. Therefore, the average age of citations should decrease, as we rely more on recent literature and avoid older articles (*H2*). To better understand temporal changes of $P^{\leftarrow}(t)$, we repeat two recent studies using our data.

We find the retrospective distributions become broader over time, indicating a increasing fraction of old papers are now being cited in references. We first use the approaches by Verstak *et al.* [6] to measure the fraction of citations to 'older papers', i.e. papers older than a given age, over time. Over the course of last two decades, the fraction of citations to papers that are at least 10 years old has increased from 30.55% in 1990 to 37.23% in 2013. These results indicate the aging structure of citations is not static but dynamic. Older papers are not forgotten, but gathering more attention over time. We repeated this analysis for references published at least 15 or 20 years ago. We find fraction of citations to papers that are at least 15 (20) years old has grown from 16.46% (9.24%) to 20.80% (12.16%) in the same period, indicating our results are independent of our definitions of 'old papers'. These results are consistent with those obtained by using Google Scholar data [6], as well as other data sources [7, 8]. Similar to earlier studies using Google Scholar dataset [6], we also find an acceleration in the temporal shift in our Web of Science corpus. Going from 1990 to 2001, the fraction of citations that are no less than 10 years old changed from 30.55% to 32.74%, corresponding to a 7.15% increase. Yet for the next 11 years, the fraction of old papers increased from 32.87% in 2002 to 37.23% in 2013, corresponding to a 14.28% change. Therefore, the change in the second half nearly doubles that in the first half, documenting an evident acceleration in the temporal shift in the Web of Science dataset. The discrepancy may be rooted in the explosion of the Web in the second period creating more web-based scholarly content.

Next, we compare our results with a recent study by Pan *et al.* [9], which reports a more complicated mechanism for aging. We test the fraction of citations older than 6 years and less than 50 years old, finding 41.04% increase from 1965 to 2012. We go on to examine whether this significant increase is caused by the decline in the percentage of extremely new and old citations. For citations within 3 years, we observe a significant decrease in percentage, from 42.69% in 1965 to 26.52% in 2012. However, we do not observe any decrease of fraction for extremely old citations. For example, papers older than 50 years age take up 0.56% in 1965 and 0.8% in 2012. Indeed, although these extremely old papers may receive less attention , the number of extremely old papers has a steady increase due to our definition. For example, the extremely old papers for 1965 are defined as papers published between 1900 and 1915, while extremely old papers for 2012 are defined as papers published in 1900 and 1962. Therefore, the temporal decay of individuals is counteracted by the growth of extremely old papers.

S3 Rescaling and collapse of $P_N^{\leftarrow}(t)$ and $P_r^{\rightarrow}(t)$

In this section, we give a detailed illustration on rescaling $P_N^{\leftarrow}(t|t_0)$ for different t_0. For convenience, we assume that the observation time of the system is exactly $[1, T]$. As we previously measured, $P_N^{\leftarrow}(t|t_0)$ can be well fitted by lognormal distribution, i.e.

$$P_N^{\leftarrow}(t|t_0) \propto \frac{1}{\sqrt{2\pi}\sigma(t_0 - t)} e^{-\frac{(\ln(t_0-t)-\mu_0)^2}{2\sigma_0^2}} \tag{S7}$$

Thus we have

$$\frac{P_N^{\leftarrow}(t|t_0)}{\sum_{t=0}^{t_0-2} P_N^{\leftarrow}(t|t_0)} = \frac{\frac{1}{\sqrt{2\pi}\sigma(t_0-t)} e^{-\frac{(\ln(t_0-t)-\mu_0)^2}{2\sigma_0^2}}}{\int_0^{t_0-2} \frac{1}{\sqrt{2\pi}\sigma(t_0-t)} e^{-\frac{(\ln(t_0-t)-\mu_0)^2}{2\sigma_0^2}} dt} \tag{S8}$$

which can be reformulated into

$$\frac{t_0-t}{[(t_0-t)e^{-\mu}]^{1/\sigma}} \frac{P_N^{\leftarrow}(t|t_0)}{\sum_{t=0}^{t_0-2} P_N^{\leftarrow}(t|t_0)} \int_0^{t_0-2} \frac{1}{\sqrt{2\pi}\sigma(t_0-t)} e^{-\frac{(\ln(t_0-t)-\mu_0)^2}{2\sigma_0^2}} dt = \frac{1}{\sqrt{2\pi}\sigma[(t_0-t)e^{-\mu}]^{1/\sigma}} e^{-\frac{(\ln(t_0-t)-\mu_0)^2}{2\sigma_0^2}} \tag{S9}$$

Indeed, if we do the variable substitution as

$$\begin{cases} X = \sigma[(t_0-t)e^{-\mu}]^{1/\sigma} \\ Z = A(t_0-t)^{1-1/\sigma} e^{\mu/\sigma} P_N^{\leftarrow}(t|t_0) \end{cases} \tag{S10}$$

where

$$A = \frac{\int_0^{t_0-2} \frac{1}{\sqrt{2\pi}\sigma(t_0-t)} e^{-\frac{(\ln(t_0-t)-\mu_0)^2}{2\sigma_0^2}} dt}{\sum_{t=0}^{t_0-2} P_N^{\leftarrow}(t|t_0)} \tag{S11}$$

Hence, we obtain

$$Z = \frac{1}{\sqrt{2\pi}X} e^{-(\ln X)^2/2} \tag{S12}$$

In other words, (X, Z) always lies in the probability density function curve of a standard lognormal distribution.

Since the derivation only depends on the fact that $P_N^{\leftarrow}(t)$ is lognormal, similar collapse would happen for $P_r^{\rightarrow}(t)$ as well, i.e. if we define

$$\begin{cases} X' = \sigma[(t-t_0)e^{-\mu}]^{1/\sigma} \\ Z' = A'(t-t_0)^{1-1/\sigma} e^{\mu/\sigma} P_r^{\rightarrow}(t|t_0) \end{cases} \tag{S13}$$

where

$$A' = \frac{\int_{t_0+2}^{T} \frac{1}{\sqrt{2\pi}\sigma(t-t_0)} e^{-\frac{(\ln(t-t_0)-\mu_0)^2}{2\sigma_0^2}} dt}{\sum_{t=t_0+2}^{T} P_r^{\rightarrow}(t|t_0)} \tag{S14}$$

We find (X', Z') collapses to the same curve as well.

S4 Validation on model systems

In this section, we calculate $P^{\leftarrow}(t)$ and $P^{\rightarrow}(t)$ for Price's model, Bianconi-Barabasi model and Wang-Song-Barabasi model.

S4.1 Price's model

Price's model [10] predicts

$$P(t_1, t_2) \sim \frac{c_0}{c_0 + m} t_1^{-c_0/(c_0+m)} t_2^{-m/(c_0+m)} \Theta(t_1 - t_2) \tag{S15}$$

Integrating over t_2, we can obtain $P_{out}(t_1)$, defined as $\frac{M(t_1)}{\int_0^T M(t)dt}$, from $P(t_1, t_2)$: $P_{out}(t_1) = \int P(t_1, t_2) dt_2$. Evaluating this integral and after normalization we have

$$\langle P_{out}(t_1) \rangle \sim N(t_1) = \text{constant} \tag{S16}$$

$P_{out}(t_1)$ captures the probability of citations originated from time t_1. In Price's model, new nodes carrying new citations arrive at a constant rate from a mean field perspective. Hence the fact that $P_{out}(t_1)$ is independent with time agrees with the model prediction.

To calculate $P_{in}(t_2)$, defined as $\frac{L(t_2)}{\int_0^T L(t)dt}$, we integrate $P(t_1, t_2)$ over t_1, obtaining $P_{in}(t_2) = \int P(t_1, t_2) dt_1$:

$$P_{in}(t_2) \sim M(t_2) \sim \left(\frac{T}{t_2}\right)^{\frac{m}{c_0+m}} - 1 \tag{S17}$$

where T measures the cut off time of the system. $P_{in}(t_2)$ captures the probability of a citation received by other papers that are published at time t_2. As (S17) indicates, this probability is a function of t_2 as well as T. When t_2 increases, the node spends less time in the system hence its probability to receive citations decreases. As t_2 approaches T, the node was added most recently to the system, and its probability to be cited by the next new node is close to zero.

Combining (S15) with (S16) and (S17) allows us to derive the citation age distributions for both retrospective and prospective approaches. Recall the retrospective approach measures citation age by looking back in time. We calculate $P^{\leftarrow}(t_2|t_1) = P(t_1,t_2)/P_{out}(t_1)$. By defining $\Delta t = t_1 - t_2$, we obtain

$$P^{\leftarrow}(t_2|t_1) = \frac{c_0}{c_0+m} t_1^{-\frac{c_0}{c_0+m}} t_2^{-\frac{m}{c_0+m}} \sim \frac{1}{t_1}\left(1-\frac{\Delta t}{t_1}\right)^{-\frac{m}{c_0+m}} \tag{S18}$$

where $\Delta t \in [0, t_1]$. (S18) also predicts that the retrospective age distributions collapse into a universal function for different t_1 as

$$P^{\leftarrow}(t_2|t_1) = \frac{c_0}{t_1(c_0+m)} x^{-m/(c_0+m)} \tag{S19}$$

after rescaling $x \equiv t_2/t_1$. When taking the prospective approach, we need to calculate $P^{\rightarrow}(t_1|t_2) = P(t_1,t_2)/P_{in}(t_2)$ to measure the citation age distribution. We obtain

$$P^{\rightarrow}(t_1|t_2) = \frac{m t_1^{-c_0/(c_0+m)}}{(c_0+m)(T^{m/(c_0+m)} - t_2^{m/(c_0+m)})} \sim \frac{1}{T^{m/(c_0+m)} - t_2^{m/(c_0+m)}} (t_2+\Delta t)^{-c_0/(c_0+m)} \tag{S20}$$

where $\Delta t \in [0, T - t_2]$. This equation could also be rescaled into a universal form as

$$P^{\rightarrow}(t_1|t_2) = \frac{m t_2^{-c_0/(c_0+m)}}{(c_0+m)(T^{m/(c_0+m)} - t_2^{m/(c_0+m)})} x^{-c_0/(c_0+m)} \tag{S21}$$

If we use (S56) and (S18) to derive $P^{\rightarrow}(t)$, note that $\langle M(t)\rangle$ is a constant in Price's model, we have

$$P^{\rightarrow}(t_1|t_2) = \frac{t_1^{-c_0/(c_0+m)} t_2^{-m/(c_0+m)}}{\int_{t_2}^{T} t_1^{-c_0/(c_0+m)} t_2^{-m/(c_0+m)} dt_1} = \frac{m t_1^{-c_0/(c_0+m)}}{(c_0+m)(T^{m/(c_0+m)} - t_2^{m/(c_0+m)})} \tag{S22}$$

On the contrary, given the form of $P^{\rightarrow}$, from (S59) we obtain

$$L(t_2) = c_0 \left(\frac{T}{t_2}\right)^{\frac{m}{c_0+m}} - c_0 \tag{S23}$$

Therefore

$$P^{\leftarrow}(t_2|t_1) = c_0\left(\left(\frac{T}{t_2}\right)^{\frac{m}{c_0+m}} - 1\right)\frac{m}{c_0+m}\frac{t_1^{-c_0/(c_0+m)}}{T^{m/(c_0+m)} - t_2^{m/(c_0+m)}}/c = \frac{c_0}{c_0+m} t_1^{-\frac{c_0}{c_0+m}} t_2^{-\frac{m}{c_0+m}} \tag{S24}$$

which is exactly the same as (S20).

S4.2 Bianconi-Barabasi model

Taking into consideration fitness of a paper, η, we calculate $P(t_1, t_2)$ for Bianconi-Barabasi model [11]

$$P(t_1, t_2) \sim \langle \beta(\eta) t_1^{\beta(\eta)-1} t_2^{-\beta(\eta)} \rangle \Theta(t_1 - t_2) \tag{S25}$$

where $\beta(\eta) = \eta/C$ and $C = \int \rho(\eta) \frac{\eta}{1-\beta(\eta)} d\eta$. In addition, by definition we have

$$P_{out}(t_1) = \text{const} \tag{S26}$$

$$P_{in}(t_2) \sim \langle \left(\frac{T}{t_2}\right)^{\beta(\eta)} \rangle - 1, \tag{S27}$$

allowing us to obtain the retrospective and prospective distributions:

$$P^{\leftarrow}(t_2|t_1) \sim \langle \beta(\eta) t_1^{\beta(\eta)-1} t_2^{-\beta(\eta)} \rangle \tag{S28}$$

$$P^{\rightarrow}(t_1|t_2) \sim \frac{\langle \beta(\eta) t_1^{\beta(\eta)-1} t_2^{-\beta(\eta)} \rangle}{\langle (T/t_2)^{\beta(\eta)} \rangle - 1} \tag{S29}$$

Using (S56), we obtain the prospective distribution:

$$P^{\rightarrow}(t_1|t_2) \sim \frac{\langle \beta(\eta) t_1^{\beta(\eta)-1} t_2^{-\beta(\eta)} \rangle}{\langle \beta(\eta)/(1-\beta(\eta)) \rangle} \Big/ \int_{t_2}^{T} \frac{\langle \beta(\eta) t_1^{\beta(\eta)-1} t_2^{-\beta(\eta)} \rangle}{\langle \beta(\eta)/(1-\beta(\eta)) \rangle} dt_1 \sim \frac{\langle \beta(\eta) t_1^{\beta(\eta)-1} t_2^{-\beta(\eta)} \rangle}{\langle (T/t_2)^{\beta(\eta)} \rangle - 1} \tag{S30}$$

While we start from prospective distribution, by solving (S59) we get

$$L(t_2) \sim \langle \left(\frac{T}{t_2}\right)^{\beta(\eta)} \rangle - 1 \tag{S31}$$

Plugging (S31) into (S57) we have

$$P^{\leftarrow}(t_2|t_1) \sim (\langle (T/t_2)^{\beta(\eta)}\rangle - 1)\frac{\langle \beta(\eta) t_1^{\beta(\eta)-1} t_2^{-\beta(\eta)}\rangle}{\langle (T/t_2)^{\beta(\eta)}\rangle - 1} \sim \langle \beta(\eta) t_1^{\beta(\eta)-1} t_2^{-\beta(\eta)}\rangle \tag{S32}$$

which is equivalent with (S28) after normalization. Assuming a truncated exponential fitness distribution,

$$\rho(\eta) = \begin{cases} \frac{e^{-\eta/\eta_0}}{\eta_0(1-e^{-\eta_{\max}/\eta_0})} & \eta \in [0, \eta_{\max}] \\ 0 & \text{otherwise} \end{cases} \tag{S33}$$

we have

$$P^{\leftarrow}(t_2|t_1) = \frac{\alpha_1^{-2}[1-(\alpha_1\eta_{\max}+1)e^{-\alpha_1\eta_{\max}}]}{C\eta_0 t_1(1-e^{-\eta_{\max}/\eta_0})} \tag{S34}$$

$$P^{\rightarrow}(t_1|t_2) = \frac{\alpha_1^{-2}[1-(\alpha_1\eta_{\max}+1)e^{-\alpha_1\eta_{\max}}]}{C t_1[\alpha_2^{-1}(1-e^{-\alpha_2\eta_{\max}}) - \eta_0(1-e^{-\eta_{\max}/\eta_0})]} \tag{S35}$$

where $\alpha_1 = \frac{1}{\eta_0} - \frac{1}{C}\ln(t_1/t_2)$ and $\alpha_2 = \frac{1}{\eta_0} - \frac{1}{C}\ln(T/t_2)$.

Taking $\eta_{\max} \to \infty$ we obtain

$$P^{\leftarrow}(t_2|t_1) = \frac{1}{C\eta_0 t_1 \alpha_1^2} = \frac{1}{C\eta_0 t_1}\left(\frac{1}{\eta_0} + \frac{1}{C}\ln x\right)^{-2} \tag{S36}$$

$$P^{\rightarrow}(t_1|t_2) = \frac{\alpha_2}{C t_1 \alpha_1^2 (1-\alpha_2\eta_0)} = \frac{\alpha_2 x}{C t_2 (1-\alpha_2\eta_0)}\left(\frac{1}{\eta_0} + \frac{1}{C}\ln x\right)^{-2} \tag{S37}$$

S4.3 Wang-Song-Barabasi model

For Wang-Song-Barabasi model [12], we have

$$\Pi_i(\Delta t_i) \sim \eta_i c_i^t P(\Delta t_i) \tag{S38}$$

yielding the rate equation:

$$\frac{dc_i^t}{dN(t)} = m\frac{c_i\eta_i P_t(\beta^{-1}\ln(N/i))}{\sum_{i=0}^{N} c_i\eta_i P_t(\beta^{-1}\ln(N/i))} \tag{S39}$$

Denoting $\Delta t_i = t - t_i$ and $c_i = m(f(\eta_i, \delta t_i) - 1)$, we obtain

$$\frac{df(\eta_i, \Delta t_i)}{d\Delta t_i} = \beta\frac{\eta_i f(\eta_i, \Delta t_i)P_t(\Delta t_i)}{A} \tag{S40}$$

where the normalization constant

$$\begin{aligned} A &\equiv \lim_{N\to\infty}(N(t))^{-1}\langle\sum_{i=1}^{N(t)}\eta_i P_t(\beta^{-1}\ln(N(t)/i))f(\eta_i, \beta^{-1}\ln(N(t)/i))\rangle \\ &= \lim_{N\to\infty}\langle\int_1^{N(t)}\eta_i P_t(\beta^{-1}\ln(N(t)/i))f(\eta_i, \beta^{-1}\ln(N(t)/i))d(i/N(t))\rangle \\ &= \beta\int d\eta\rho(\eta)\int_0^{\infty}\eta P_t(t')f(\eta, t')e^{-\beta t'}dt' \end{aligned} \tag{S41}$$

The solution for (S39) is

$$f(\eta_i, \Delta t_i) = e^{\frac{\beta}{A}\eta_i\int_{t_i}^{t}P_t(t'-t_i)dt'} - 1 \tag{S42}$$

Therefore, we get

$$c_i^t = m\left(e^{\frac{\beta}{A}\eta_i\int_{t_i}^{t}P_t(t'-t_i)dt'} - 1\right) \tag{S43}$$

If we substitute $\frac{\beta}{A}\eta_i$ with λ_i and assume $P(\lambda_i) \sim e^{-\lambda_i/\lambda_0}$, we have

$$\langle c_i^t\rangle = \frac{m\lambda_0\Phi(\frac{\ln(t-t_i)-\mu_i}{\sigma_i})}{1-\lambda_0\Phi(\frac{\ln(t-t_i)-\mu_i}{\sigma_i})} \tag{S44}$$

and the probability for citation from t_1 to t_2 can be formulated as

$$P(t_1, t_2) \sim \frac{m\lambda_0 e^{\beta t_2-(\ln(t_1-t_2)-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma(t_1-t_2)[1-\lambda_0\Phi(\frac{\ln(t_1-t_2)-\mu}{\sigma})]^2} \tag{S45}$$

Taking the derivative of (S43) with respect to t, we get the form of prospective distribution:

$$P^{\rightarrow}(t_1|t_2) = \frac{m\lambda_0 e^{-(\ln(t_1-t_2)-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma(t_1-t_2)[1-\lambda_0\Phi(\frac{\ln(t_1-t_2)-\mu}{\sigma})]^2} / \frac{m\lambda_0\Phi(\frac{\ln(T-t_2)-\mu}{\sigma})}{1-\lambda_0\Phi(\frac{\ln(T-t_2)-\mu}{\sigma})}$$
$$= \frac{e^{-(\ln(t_1-t_2)-\mu)^2/2\sigma^2}[1-\lambda_0\Phi(\frac{\ln(T-t_2)-\mu}{\sigma})]}{\sqrt{2\pi}\sigma(t_1-t_2)\Phi(\frac{\ln(T-t_2)-\mu}{\sigma})[1-\lambda_0\Phi(\frac{\ln(t_1-t_2)-\mu}{\sigma})]^2} \tag{S46}$$

Plugging (S42) into (S39), the retrospective distribution can be written as:

$$P^{\leftarrow}(t_2|t_1) = \frac{m\lambda_0 e^{\beta t_2-(\ln(t_1-t_2)-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma(t_1-t_2)[1-\lambda_0\Phi(\frac{\ln(t_1-t_2)-\mu}{\sigma})]^2} / me^{\beta t_1}$$
$$= \frac{\lambda_0 e^{-\beta(t_1-t_2)-(\ln(t_1-t_2)-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma(t_1-t_2)[1-\lambda_0\Phi(\frac{\ln(t_1-t_2)-\mu}{\sigma})]^2} \tag{S47}$$

Note that $\Phi(x)$ converges to 1 for $x \to \infty$, for $t_1 - t_2 \to \infty$ we have

$$P^{\leftarrow}(t_2|t_1) \to \frac{\lambda_0 e^{-\beta(t_1-t_2)-(\ln(t_1-t_2)-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma(t_1-t_2)(1-\lambda_0)^2} \tag{S48}$$

and

$$P^{\rightarrow}(t_1|t_2) \to \frac{1-\lambda_0\Phi(\frac{\ln(T-t_2)-\mu}{\sigma})}{\lambda_0\Phi(\frac{\ln(T-t_2)-\mu}{\sigma})(1-\lambda_0)^2} \frac{e^{-(\ln(t_1-t_2)-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma(t_1-t_2)} \tag{S49}$$

Since $T \geq t_1$, this can be further simplified as

$$P^{\rightarrow}(t_1|t_2) \to \frac{e^{-(\ln(t_1-t_2)-\mu)^2/2\sigma^2}}{\lambda_0(1-\lambda_0)\sqrt{2\pi}\sigma(t_1-t_2)} \tag{S50}$$

S5 Derivation of rescaling formulas

$P(t_1, t_2)$ can be measured by

$$P(t_1, t_2) = \frac{P^{\leftarrow}(t_2|t_1)M(t_1)}{\int_0^T \int_0^{t_1} P^{\leftarrow}(t_2|t_1)M(t_1)dt_2 dt_1} \tag{S51}$$

To capture the time constraint $t_1 > t_2$, we introduce a step function $\Theta(x)$

$$\Theta(x) = \begin{cases} 1 & x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{S52}$$

The number of citations pointing toward papers published at time t_2, defined as $L(t_2)$, can be formulated as

$$L(t_2) = \int_0^T P^{\leftarrow}(t_2|t_1)M(t_1)\Theta(t_1 - t_2)dt_1 = \int_{t_2}^T P^{\leftarrow}(t_2|t_1)M(t_1)dt_1 \tag{S53}$$

Among them, citations originated from t_1 takes up a fraction of $P^{\rightarrow}(t_1|t_2)$, thus the number of citations from t_1 to t_2, equals to

$$P^{\leftarrow}(t_2|t_1)M(t_1) = P^{\rightarrow}(t_1|t_2)L(t_2) \tag{S54}$$

Combining (S53) and (S54), we have

$$P^{\leftarrow}(t_2|t_1)M(t_1) = P^{\rightarrow}(t_1|t_2)\int_{t_2}^T P^{\leftarrow}(t_2|t_1)M(t_1)dt_1 \tag{S55}$$

Therefore, we obtain our rescaling formula for $P^{\rightarrow}(t)$:

$$P^{\rightarrow}(t_1|t_2) = \frac{P^{\leftarrow}(t_2|t_1)M(t_1)}{\int_{t_2}^T P^{\leftarrow}(t_2|t_1)M(t_1)dt_1} \tag{S56}$$

Using (S56), one can derive $P^{\rightarrow}$ when given $P^{\leftarrow}$ and M.

To derive $P^{\rightarrow}(t)$ we first transform (S56) into

$$P^{\rightarrow}(t_1|t_2) = \frac{P^{\leftarrow}(t_2|t_1)M(t_1)}{L(t_2)} \tag{S57}$$

Once we get $L(t_2)$, $P^{\leftarrow}$ can be given by (S57). Note that

$$\int_0^T P^{\leftarrow}(t_2|t_1)dt_2 = 1 \tag{S58}$$

Plugging (S57) into (S58), we get

$$\int_0^T P^{\rightarrow}(t_1|t_2)L(t_2)dt_2 = M(t_1) \tag{S59}$$

Next we solve the discrete form of (S59):

$$\begin{cases} N(0)m(0) = P^{\rightarrow}(0|0)L(0) \\ N(1)m(1) = P^{\rightarrow}(1|0)L(0) + P^{\rightarrow}(1|1)L(1) \\ N(2)m(2) = P^{\rightarrow}(2|0)L(0) + P^{\rightarrow}(2|1)L(1) + P^{\rightarrow}(2|2)L(2) \\ \dots \\ N(T)m(T) = P^{\rightarrow}(T|0)L(0) + P^{\rightarrow}(T|1)L(1) + P^{\rightarrow}(T|2)L(2) + \cdots + P^{\rightarrow}(T|T)L(t) \end{cases} \tag{S60}$$

Using the notations defined in main text, (S60) can be written in the form of matrix equation:

$$\boldsymbol{P}^{\rightarrow}\boldsymbol{L} = \boldsymbol{M} \tag{S61}$$

Solution of (S61) can be formulated as

$$\boldsymbol{L} = (\boldsymbol{P}^{\rightarrow})^{-1}\boldsymbol{M} \tag{S62}$$

Therefore, the rescaling formula for $P^{\leftarrow}(t)$ is

$$P^{\leftarrow}(t_2|t_1) = \frac{P^{\rightarrow}(t_1|t_2)[(\boldsymbol{P}^{\rightarrow})^{-1}\boldsymbol{M}](t_2)}{\boldsymbol{M}(t_1)} \tag{S63}$$

S6 Testing rescaling formulas on WoS dataset

S6.1 Predicting $P^{\rightarrow}(t)$ from $P^{\leftarrow}(t)$

Here we present our approach for estimating $M(t)$ and $P^{\leftarrow}(1990|t)$ for $t \in [1990, 2010]$. For $M(t)$, we fit it with an exponential function, yielding

$$\hat{M}(t) \sim e^{0.058t} \tag{S64}$$

Next we estimate retrospective distributions for years after 1990. As we discussed in main text, lognormal distribution with exponential cutoff is the best fit for $P^{\leftarrow}$. Next, we approximate the three parameters for this distribution from history data, i.e.

$$P^{\leftarrow}(t_2|t_1) \sim \frac{1}{\sqrt{2\pi}\sigma(t_1 - t_2)} \exp\left[-\frac{(\ln(t_1 - t_2) - \mu)^2}{2\sigma^2} - \beta(t_1 - t_2)\right] \tag{S65}$$

where β captures the growth of number of papers, and we take $\beta = 0.04$. For the other two parameters, we find σ is very steady over different t_1, while μ increases slowly with t_1, which is consistent with our observations in Sec S2. Thus, we take $\mu = 0.0072t_1 + 1.5715$ and $\sigma = 1.2$, obtaining

$$\hat{P}^{\leftarrow}(t_2|t_1) \sim \frac{1}{t_1 - t_2} \exp\left[-\frac{(\ln(t_1 - t_2) - 0.0072t_1 - 1.5715)^2}{2.88} - 0.058(t_1 - t_2)\right] \tag{S66}$$

With $\hat{M}$ and $\hat{P}^{\leftarrow}(t)$, we are now able to predict $P^{\rightarrow}(1990|t)$ through different models (main text).

S6.2 Validating functional forms of aging functions

To further validate our measurements on functional forms of $P^{\leftarrow}$ and $P^{\rightarrow}$, we derive them through rescaling formula (S57,S63).

Consider a steady system where papers from different years follow the same pattern of aging.

Given

$$P^{\leftarrow}(t_1-\Delta t|t_1)=\frac{e^{-(\ln(\Delta t)-\mu)^2/(2\sigma^2)-\beta\Delta t}/\Delta t}{\int_0^{t_1}e^{-(\ln(t')-\mu)^2/(2\sigma^2)-\beta t'}/t'dt'}\sim\frac{1}{\sqrt{2\pi}\sigma\Delta t}e^{-(\ln(\Delta t)-\mu)^2/(2\sigma^2)-\beta\Delta t} \tag{S67}$$

From (S57) we have

$$P^{\rightarrow}(t_2+\Delta t|t_2)\sim\frac{m(t_1)}{\int_0^{t_1}e^{-(\ln(t')-\mu)^2/(2\sigma^2)-\beta t'}/t'dt'}\frac{e^{-(\ln(\Delta t)-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma\Delta t} \tag{S68}$$

Thus for $t_2\to\infty$ we obtain

$$P^{\rightarrow}(t_2+\Delta t|t_2)\sim\frac{m(t_1)}{\int_0^{\infty}e^{-(\ln(t')-\mu)^2/(2\sigma^2)-\beta t'}/t'dt'}\frac{e^{-(\ln(\Delta t)-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma\Delta t}\sim\frac{m(t_1)}{\sqrt{2\pi}\sigma\Delta t}e^{-(\ln(\Delta t)-\mu)^2/(2\sigma^2)} \tag{S69}$$

S7 Calculating average citation growth for WSB model

From Wang-Song-Barabasi model, we have

$$c_i^t\sim e^{\lambda\int_0^{t-t_i}P(s)ds}-1 \tag{S70}$$

In time $(t,t+\Delta t)$, there are $\Delta t(dc_i^t/dt)$ new citations towards the paper, each with age t, so we have

$$a_i(t)=\frac{\int_0^{t-t_i}sdc_i^s}{c_i^t} \tag{S71}$$

At a specific time t, the number of citations by then is

$$c_i^t=m\left(e^{\lambda\Phi(\frac{\ln(t-t_i)-\mu}{\sigma})}-1\right) \tag{S72}$$

and the average age is

$$a_i(t) = \frac{\lambda \int_0^{t-t_i} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} \left(e^{\lambda\Phi(\frac{\ln(x)-\mu}{\sigma})}\right) dx}{e^{\lambda\Phi(\frac{\ln(t-t_i)-\mu}{\sigma})} - 1} \tag{S73}$$

To simplify our calculation, we choose $m = 30$, the same as [12] did, and we will see that the value of m does not influence our result.

First, we use three groups of parameters from empirical data, i.e.(3.0,8.8,1.2),(1.9,7.5,0.9) and (6.7,9.2,1.0). Tests over these data indicate that c_i^t and $a_i(t)$ has a near linear relationship in the log-log panel when c_i^t is not very close to c_i^∞ (main text). Thus, the average age growth can be approximated as power law.

References

[1] Yogesh Virkar, Aaron Clauset, et al. Power-law distributions in binned empirical data. *The Annals of Applied Statistics*, 8(1):89–119, 2014.

[2] VM Motylev. The main problems of studying literature aging. *Scientometrics*, 15(1-2):97–109, 1989.

[3] Andrew Odlyzko. The rapid evolution of scholarly communication. *Learned Publishing*, 15(1):7–19, 2002.

[4] Vincent Larivière, Éric Archambault, and Yves Gingras. Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004). *Journal of the American Society for Information Science and technology*, 59(2):288–296, 2008.

[5] James A Evans. Electronic publication and the narrowing of science and scholarship. *science*, 321(5887):395–399, 2008.

[6] Alex Verstak, Anurag Acharya, Helder Suzuki, Sean Henderson, Mikhail Iakhiaev, Cliff Chiung Yu Lin, and Namit Shetty. On the shoulders of giants: The growing impact of older articles. *arXiv preprint arXiv:1411.0275*, 2014.

[7] Dag IK Sjøberg. Confronting the myth of rapid obsolescence in computing research. *Communications of the ACM*, 53(9):62–67, 2010.

[8] Alberto Martín-Martín, Enrique Orduña-Malea, Juan Manuel Ayllón, and Emilio Delgado López-Cózar. Reviving the past: the growth of citations to old documents. *arXiv preprint arXiv:1501.02084*, 2015.

[9] Raj K Pan, Alexander M Petersen, Fabio Pammolli, and Santo Fortunato. The memory of science: Inflation, myopia, and the knowledge network. *arXiv preprint arXiv:1607.05606*, 2016.

[10] Derek J de Solla Price. Networks of scientific papers. *Science*, 149(3683):510–515, 1965.

[11] Ginestra Bianconi and A-L Barabási. Competition and multiscaling in evolving networks. *EPL (Europhysics Letters)*, 54(4):436, 2001.

[12] Dashun Wang, Chaoming Song, and Albert-László Barabási. Quantifying long-term scientific impact. *Science*, 342(6154):127–132, 2013.

Figure S1: $P^{\leftarrow}(t_2|t_1)$ for t_1 in $[1984, 2013]$, black circles are empirical data, red lines are results from our model.

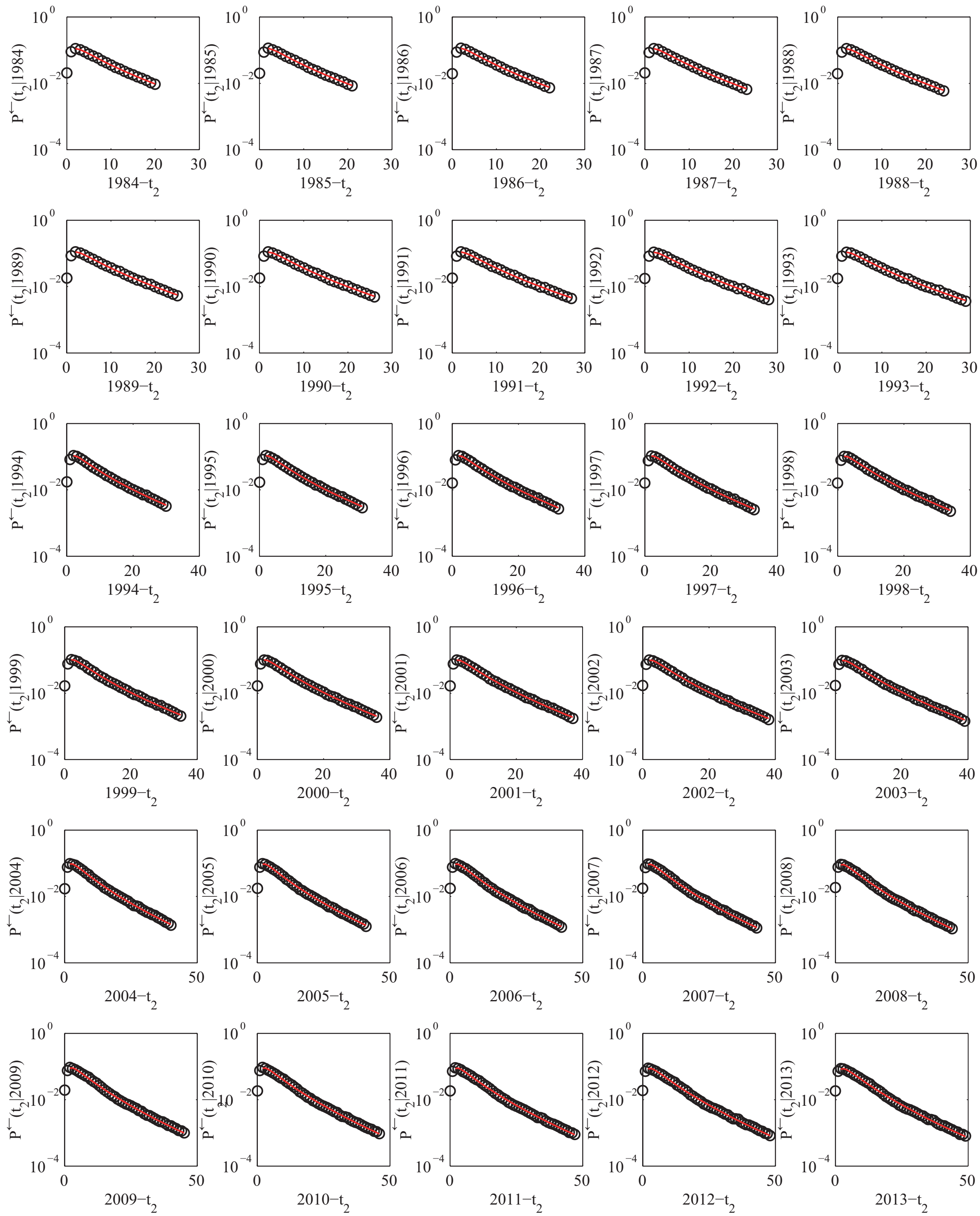

Figure S2: $P^{\rightarrow}(t_1|t_2)$ for t_2 in $[1948, 1977]$, black circles are empirical data, red lines are results from our model.