RESEARCH ARTICLE

Citations driven by social connections? A multi-layer representation of coauthorship networks

Christian Zingg, Vahan Nanumyan, and Frank Schweitzer

Chair of Systems Design, ETH Zurich, Zurich, Switzerland

Keywords: citation network, citation rate, collaboration network, collective attention, network centrality, two-layer network

ABSTRACT

To what extent is the citation rate of new papers influenced by the past social relations of their authors? To answer this question, we present a data-driven analysis of nine different physics journals. Our analysis is based on a two-layer network representation constructed from two large-scale data sets, INSPIREHEP and APS. The social layer contains authors as nodes and coauthorship relations as links. This allows us to quantify the social relations of each author, prior to the publication of a new paper. The publication layer contains papers as nodes and citations between papers as links. This layer allows us to quantify scientific attention as measured by the change of the citation rate over time. We particularly study how this change correlates with the social relations of their authors, prior to publication. We find that on average the maximum value of the citation rate is reached sooner for authors who have either published more papers or who have had more coauthors in previous papers. We also find that for these authors the decay in the citation rate is faster, meaning that their papers are forgotten sooner.

1. INTRODUCTION

The availability of large-scale data sets about journals and scientific publications therein, their authors, institutions, cited references, and citations obtained in other papers has boosted scientometric research in recent years. They allow us to address new research questions that go beyond the calculation of mere bibliographic indicators. These particularly concern the role of *social influences* on the success of papers, for example coauthorship relations (Sarigol, Pfitzner, et al., 2014) or the relations between authors and handling editors (Sarigol, Garcia, et al., 2017). Such investigations have contributed to a new scientific discipline, *the science of success* (Jadidi, Karimi, et al., 2018; Sinatra & Lambiotte, 2018).

But such data also allow us to redo traditional scientometric analyses on a much larger scale. In Parolo, Pan, et al. (2015), the dynamics of the citation *rate* (i.e., the change in the number of citations during a fixed time interval) is analyzed. The authors find that the change of the average citation rate follows two characteristic phases: first a growth phase and then a decay phase. Interestingly, the duration of the first and the speed of the second phase have changed over the years. This allows us to draw conclusions about how the *collective attention* of scientists towards a given paper has evolved between early and recent times.

In general, the dynamics of citations are extensively studied in the bibliometric literature. For example, the relation between the current number of citations and the citation rate was studied

an open access journal

Check for updates

Citation: Zingg, C., Nanumyan, V., & Schweitzer, F. (2020). Citations driven by social connections? A multi-layer representation of coauthorship networks. *Quantitative Science Studies*, *1*(4), 1493–1509. https://doi.org/10.1162/qss_a_00092

DOI: https://doi.org/10.1162/qss_a_00092

Supporting Information: https://doi.org/10.1162/qss_a_00092

Received: 18 October 2019
Accepted: 27 September 2020

Corresponding Author: Christian Zingg
czingg@ethz.ch

Handling Editor: Ludo Waltman

The MIT Press

in Jeong, Néda, and Barabási, (2003). Citations were found to occur in bursts, with large bursts within a few years after publication (Eom & Fortunato, 2011). Concerning the scientific field of a paper, citations from papers in the same field tend to be obtained earlier than citations from papers in other fields (Rinia, Van Leeuwen, et al., 2001). Citation rates have also been used to classify papers (Avramescu, 1979; Li & Ye, 2014). Such classes often identify papers that receive citations earlier or later than the majority of papers (Ciotti, Bonaventura, et al., 2016; Colavizza & Franceschet, 2016; Costas, van Leeuwen, & van Raan, 2010). Papers in the second class (i.e., which receive their citations only a long time after publication) are often called *sleeping beauties* or *delayed* (Burrell, 2005; van Raan, 2004). Their citation rate and how it differs from other papers was studied extensively in Lachance and Larivière (2014). This class has also been thoroughly studied outside paper classification settings. It was found that "sleeping beauties" are extremely rare, and only 0.04% of papers published in 1988 were identified as such (van Raan, 2004). They were also found to occur especially often in multidisciplinary data sets (Ke, Ferrara, et al., 2015).

Recent progress in the study of scientometric systems has very much relied on representing them as networks. A first example is *citation networks*, representing papers as nodes and citations as their (directed) links. Such networks can be seen as a knowledge map of science (Leydesdorff, Carley, & Rafols, 2013). They can be also used to predict scientific success (Mazloumian, 2012). A second example is *coauthorship networks*, representing scientists as nodes and their coauthorships as links. While sociological studies (Cetina, 2009) just report that communication between coauthors can be very intricate, formal models of how such collaborations form on the structural level have also been developed (Guimera, 2005; Tomasello, Vaccario, & Schweitzer, 2017). To study collaboration patterns in a university faculty (Claudel, Massaro, et al., 2017), such coauthorship networks have been combined with a network encoding the physical distance between the faculty members. It was also analyzed how communities detected on a coauthorship network overlap with different research topics (Battiston, Iacovacci, et al., 2016).

These investigations have the drawback that they study citation networks and coauthorship networks separately from each other. As already emphasized (Clauset, Larremore, & Sinatra, 2017; Schweitzer, 2014), this becomes a problem if one wants to study social influence on citation dynamics. For example, based on a data set of *Physical Review*, it was shown that scientists cite former coauthors more often (Martin, Ball, et al., 2013). Therefore, a better approach is to combine both the citation and the coauthorship network in a *multilayer network*. Links between the citation and the coauthorship layer express the authorship of papers. Using such a representation, a method to detect citation cartels was proposed (Fister, Fister, & Perc, 2016). Further, the rate of citations dependence on the authors' total number of citations was studied (Petersen, Fortunato, et al., 2014). However, it has not yet been investigated how the *position* of authors in the coauthorship network influences *when* their papers are cited. In this paper we study exactly this question.

Our analysis extends recent studies that focus on the success of papers as measured by their total number of citations. In Sarigol et al. (2014), this success was related to the position of the authors in a coauthorship network. It was shown that authors of successful papers are considerably more central (as quantified by *various* centrality measures) in the coauthorship network. We extend this by an analysis of the dynamics of the citation rate over time (i.e., *when* their papers are cited). To parametrize the citation dynamics, we resort to the phases identified in Parolo et al. (2015). We extend this work by relating these phases to the social relations of the authors.

Our paper is structured as follows. In section 2.1 we explain how citation dynamics can be measured by means of *citation histories*, which represent the collective attention given to a paper. In section 2.2 we describe the data sets used for our analysis. In section 3.1 we introduce the multilayer

network to combine social information about authors with citation data. We then turn to our research question and study in sections 3.2 and 3.3 how the social relations of authors in the coauthorship network influence the collective attention. Lastly, in section 4 we conclude our findings.

2. METHODS AND DATA

2.1. Dynamics of Citation Rates

2.1.1. Measuring attention

Citations are often used as a measure of the *success* of a paper, accumulated over time. They have the advantage that they are objective in the sense that they are protocolled in the reference lists of citing papers. But the sheer number of citations does not utilize the *temporal information* (i.e., how many of these citations arrive at a given time). This is captured in the citation *rate*, which better estimates the *attention* a paper receives in a given time (interval). Individual attention (i.e., *who* cites a given paper at a given time), is not of interest for our study. We focus on *collective* attention (i.e., the aggregate over all authors who cite this paper during a given time interval). Obviously, the citation rate is only a proxy for this collective attention. One could additionally consider other attention measures like the `altmetric` score. But such information is only available for very recent publications and further is strongly biased against the use of social media. Therefore, we decide to restrict our study to using only the citation rate as a proxy for collective attention. Most papers are still cited because they have caught in some way the attention of the authors of the citing papers. Furthermore, citation counts were found to be a good approximation of scientific impact as perceived by scientists from the same field as the paper (Radicchi, Weissman, & Bollen, 2017).

2.1.2. Citation histories

We measure the collective attention of a paper by the number of citations it receives over a particular time interval (i.e., its citation rate). More precisely, for paper i published at time δ_i, the citation rate at $t = \delta - \delta_i$ time units after publication is

$$c_i(t) = \frac{k_i^{\text{in}}(\delta + \Delta t) - k_i^{\text{in}}(\delta)}{\Delta t} \tag{1}$$

where $k_i^{\text{in}}(\delta)$ denotes the total number of "incoming" citations the paper has received at time δ. The dynamics of the citation rate $c_i(t)$ is also called the *citation history* of paper i (Parolo et al., 2015). To compare citation histories across papers we further normalize them by their respective maximum value $c_i^{\max} = \max_t\{c_i(t)\}$:

$$\tilde{c}_i(t) = c_i(t)/c_i^{\max}. \tag{2}$$

2.1.3. Two phases in citation histories

Parolo et al. (2015) find two characteristic phases in the dynamics of normalized citation histories $\tilde{c}_i(t)$ of a paper i. In the first phase, which lasts for 2–7 years, it grows and eventually reaches a peak at a time t_i^{peak}. After the peak there is the second phase, in which the citation rate decays over time. For the majority of papers this decay was found to be well described by an exponential function:

$$\tilde{c}_i(t) \propto \exp(-t/\tau_i), \tag{3}$$

The parameter τ_i is called the "lifetime," and it determines the speed of the decay. The *larger* τ_i is, the *faster* is the decay. Figure 1 illustrates the two phases of $\tilde{c}_i(t)$.

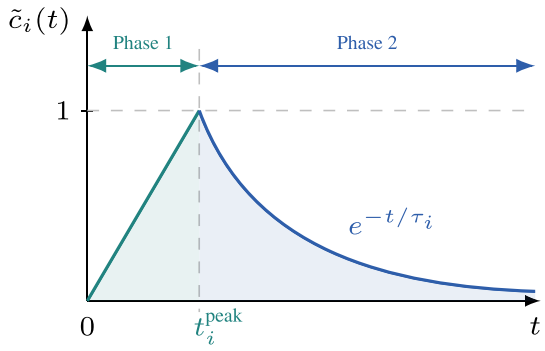

Figure 1. Illustration of the two characteristic phases in the normalized citation histories $\tilde{c}_i(t)$ of most papers.

2.2. Bibliographic Databases

As we argued in section 2, citations are particularly suitable to quantify the collective attention by scientists from the same field as a given paper. Therefore, in our analysis we study different journals separately, because each describes a topic-related community of authors and their papers. To obtain the data for our study we resort to large bibliographic databases which index papers across journals. They collect information such as a paper's title, the list of authors, the date of publication, and also the list of references that a paper cites. We extracted this set of information for nine journals from two such databases in the same way as in Nanumyan, Gote, and Schweitzer (2020) and as explained below.

2.2.1. APS database

This indexes papers published in journals by the American Physical Society (APS). Access to the database can be requested for research purposes at https://journals.aps.org/datasets. We extracted the journals *Physical Review* (PR), *Physical Review A* (PRA), *Physical Review C* (PRC), *Physical Review E* (PRE), and *Reviews of Modern Physics* (RMP) to cover a wide range of physics sub-fields.

The APS database has the known issue of name disambiguation, because it indexes authors by their name and not by a unique identifier. This means that different authors with the same name are indexed as one author. Such a "multiauthor" then owns all papers and coauthorships that were actually accumulated by multiple authors. In contrast, one author whose name can be spelled in different ways may be indexed as different authors in the database. The consequence for our study is that such undisambiguated authors bias measures involving (co)authorships. This problem has already been discussed in the scientific literature, and a disambiguation algorithm specifically for authors in the APS database was proposed (Sinatra, Wang, et al., 2016). We applied this algorithm to the APS database to lower the bias from undisambiguated authors.

2.2.2. INSPIREHEP database

The second database, called INSPIREHEP, indexes papers relevant for the field of high-energy physics. This database can be downloaded at http://inspirehep.net/dumps/inspire-dump.html. In this database authors are disambiguated, because each author is indexed by a unique identifier. We extracted the journals *Journal of High Energy Physics* (JHEP), *Physics Letters* (Phys. Lett.), *Nuclear Physics* (Nuc. Phys.), and high energy physics literature in *Physical Review* journals (PR-HEP) from this database. These were the four largest journals in terms of number of citations from papers in the same journal (i.e., the citations which we will use to compute citation rates in the later sections).

Table 1. Overview of the extracted journals from the APS database and the INSPIREHEP database (IH). $|V^p|$ is the number of papers, $|V^a|$ is the number of authors, $|E^{pc}|$ is the number of citations between papers, and $|E^a|$ is the number of authorships

Database	Journal	$\|V^p\|$	$\|V^a\|$	$\|E^{pc}\|$	$\|E^a\|$
APS	PR	46728	24307	253312	87386
	PRA	69147	41428	416639	144806
	PRC	36039	22672	253948	108844
	PRE	49118	36382	182701	95796
	RMP	3006	3788	5282	5044
IH	JHEP	15739	7994	191990	39056
	PR-HEP	44829	33908	213625	115237
	Phys. Lett.	22786	18078	56332	53089
	Nuc. Phys.	24014	18733	125252	60018

In INSPIREHEP some indexed papers have exceptionally large lists of authors, which sometimes even exceed 1,000 authors. Such large-scale coauthorships were termed *hyperauthorships* in Cronin (2001). Concerns were raised that it is unclear which authors actually made substantial contributions to such papers (Cronin, 2001), and that the coauthorship network is not an accurate representation of the social network of authors (Newman, 2004). Indeed, every author in such a hyperauthorship gets possibly thousands of collaborators from just a single paper, despite likely not having collaborated with all of them personally. This introduces a bias for measures involving coauthorships, and thus for our study. It was found that hyperauthorships usually occur in papers from large experiments (Newman, 2001), such as the ATLAS experiment at CERN. To avoid this bias we remove experimental papers from the database. To identify experimental papers we used meta-tags that INSPIREHEP provides, so-called XML-tags. These are essentially labels for papers that provide additional information, such as arXiv identifiers, author affiliations, or sometimes estimates of whether a paper is experimental or theoretical. We removed all papers from the database that are explicitly tagged as *experimental*. But because this tag might be unavailable for a paper, we further removed all papers that are not explicitly tagged as *theoretical work* or *work in general physics*.

To summarize, Table 1 provides summary statistics of the nine journals. It further also shows how large these journals actually are. For example, there is only one journal, MP, which contains fewer than 10,000 authors, and there are more than 400,000 citations between papers in PRA.

3. SOCIAL INFLUENCE ON CITATION RATE

3.1. Multilayer Network Representation

3.1.1. Combining information about papers and authors

Our aim is to combine the information about collective attention, as proxied by the citation rate, with information about the social relations between authors. For the latter, we specifically focus on *coauthorship*, because this is the most objective and best documented relation. Again, this is a proxy because it neglects other forms of social relationships, such as friendship, personal

encounters (e.g., during conferences), electronic communication, or relations in social media. But we do not have this type of information available for all authors over long periods. Therefore we restrict our analysis to the coauthorship network that can be constructed from the available data, as described below.

To relate information about authorship and about papers in a tractable manner, multilayer networks come into play, because they allow us to represent such separate information in different layers. The nodes on the first layer correspond to papers and the (directed) links to their citations. Different from this, the nodes in the second layer correspond to the authors and the links to their coauthorships (i.e., there is a link between two authors if they wrote at least one paper together). Then, there are links that connect nodes on the first layer with nodes on the second layer. These links correspond to the authorship relations (i.e., for every author, there is exactly one such link to each of her papers). We construct such a two-layer network for each of the nine journals in our data set to represent the information about citations between papers as well as about the authorships.

To summarize the above, Figure 2 illustrates the two layers of citation and coauthorship networks and their coupling. It further displays the temporal dimension: The multilayer network evolves over time because new papers are published, and hence new coauthors appear. As the timeline indicates, paper i is published at time δ_i and then accumulates citations in the future, at times $\delta > \delta_i$. The publication layer allows us to define the degree of a *paper* i as the number $k_i^{\text{in}}(\delta)$ of papers that cite i until time δ (see Eq. 1 and Figure 2). Specifically, it is the *in-degree*, because the publication network is directed. The question is now how the citation rate of this paper evolves over time, conditional on the social information about its authors at time δ_i, which is the publication time of paper i. In other words, we analyze the impact of information from *before* this publication.

3.1.2. *Quantifying authors' social relations*

The coauthorship layer allows us to define the *degree* of an *author* n as the total number of distinct coauthors $k_n(\delta_i)$ that the author had *before* time δ_i. Degree is the simplest centrality measure for

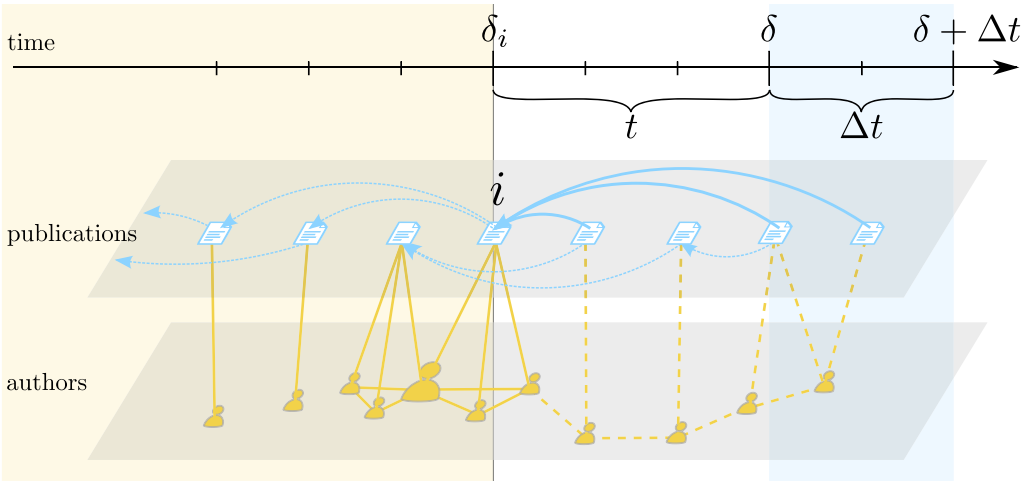

Figure 2. Multilayer network illustrating the coupling between the coauthorship network and the citation network. Links between the two layers represent the relation between authors and papers. The timeline on top indicates that links within the citation layer are directed and point to papers already existing at the time when a paper is published.

networks and reflects the *local* information about the embedding of an author in the social network. We use it here because it was shown recently (Nanumyan et al., 2020) that this measure is a particularly good predictor for the future citation rate.

We characterize a paper i at time δ_i as the number of distinct coauthors of its authors *before* time δ_i. This means that we sum over the authors' individual degrees k_r

$$s_i^{NC}(\delta_i) = \sum_r k_r(\delta_i) - C(\delta_i) \tag{4}$$

and subtract a correction term $C(\delta_i)$ to not count multiple times those coauthors who collaborated with more than one author in the past. The derivation of this correction term can be found in the Supplementary Material, Section S1. The index *NC* refers to number of coauthors. Furthermore, the paper i published at time δ_i is *not* counted in $s_i^{NC}(\delta_i)$.

We also make use of the coupling between the two layers to define a second measure, which we can later compare with $s_i^{NC}(\delta_i)$. First we define the *interlayer-degree* $\tilde{k}_n(\delta_i)$ of an author n as the total number of distinct papers written by n *before* time δ_i. This measure allows us to quantify the experience of author n that she gained before a given point in time. To characterize a paper i at time δ_i by using this information about its authors r before time δ_i, we compute

$$s_i^{NP}(\delta_i) = \sum_r \tilde{k}_r(\delta_i) - \tilde{C}(\delta_i) \tag{5}$$

by analogy with Eq. 4. Here $\tilde{C}(\delta_i)$ is again a correction term used to only count unique papers (if some authors had written a paper together in the past already). Its derivation can again be found in the Supplementary Material, Section S1.

3.1.3. Parametrizing citation rates

The quantities $s_i^{NC}(\delta_i)$ and $s_i^{NP}(\delta_i)$ are based on the information of the *authors* of paper i. Our goal is to determine how they influence the *citation dynamics* of paper i (i.e., we need an analytically tractable parametrization of the citation rates). To parametrize the citation dynamics we resort to the two characteristic phases of citation histories mentioned in section 2.1. The first phase corresponds to increasing citation rates, and we parametrize by its duration t_i^{peak}, because we have no more precise knowledge about a general functional form of this phase. The second phase corresponds to an exponential decay, and we parametrize it as the parameter τ_i in Eq. 3 (i.e., the so-called lifetime). Both parameters, t_i^{peak} and τ_i, are illustrated in Figure 1.

We now have four parameters to summarize the information about paper i. The first two parameters are $s_i^{NC}(\delta_i)$ and $s_i^{NP}(\delta_i)$, which characterize the *authors* of paper i. The other two parameters are t_i^{peak} and τ_i, which characterize the *citation history*.

3.1.4. Excluding incomplete citation histories

Obviously, our data sets only contain papers published before the release date of the respective database. Hence, the time-span on which we can compute a given paper's citation history is also limited by this date. This introduces an issue, especially for recent papers: The observable period of the citation history can be so short that the decay phase has not yet started at all. To account for this, we omitted all papers that were published within the last 5 years before the release of the respective database. Hence, for all papers in our study the citation histories are covered over at

least 5 years. In addition, we also removed those papers whose citation rate is nondecreasing in the latest year, as this is a sign that the respective paper has not yet reached its decay phase.

3.2. Time to the Peak Citation Rate

3.2.1. Regressions

Our aim is to study the dependence between peak-delays, t_i^{peak}, and the number of previous coauthors, s_i^{NC}, or publications, s_i^{NP}. At first, linear regression seems applicable to determine such a dependence. It would allow a straightforward interpretation of fitted coefficients. However, peak-delays are essentially counts, because we count in which year after publication the peak citation rate occurs (i.e., whether this is in year 0, or in year 1, or in year 2). For such data, classical linear regression can give wrong conclusions, for example, because it can predict negative values, which are impossible for counts. Instead, we apply a negative binomial regression, which is a standard model for count data (Hilbe, 2011). We chose this model over the simpler Poisson regression, because we found that the variance of peak-delays across papers is larger than their mean. We test this so-called overdispersion for our data in the Supplementary Material, Section S3. Overdispersion violates an assumption of Poisson regression, while the negative binomial regression becomes applicable. Hence, the model we fit is

$$t_i^{\text{peak}} = \text{negbin}(\alpha + \beta \cdot s_i) \tag{6}$$

where s_i is the number of previous coauthors, s_i^{NC}, or the number of previous publications, s_i^{NP}, and t_i^{peak} is measured in years (Venables & Ripley, 2002). negbin stands for a negative binomial regression. The parameters α and β are to be fitted. We use the function *glm.nb* in the R-package *MASS* to fit them.

3.2.2. Fitted parameters

In Table 2 we show the fitted parameters for all journals. Except for one coefficient, all parameters β are negative, which means that peak delays get smaller for increasing numbers of previous coauthors or publications. The exception is JHEP, which has a positive β for the number of previous publications. However, this coefficient is not significant, meaning that it is likely not different from zero, and therefore does not contradict the discovered trend. To conclude, we find that the larger the number of previous coauthors or publications is, the sooner the peak citation rate is reached.

3.2.3. Size of the effect

We also study the size of the dependence between a paper's peak-delay, t_i^{peak}, and the number of previous coauthors, s_i^{NC}, or publications, s_i^{NP}. To this end, we use our fitted models to predict the average peak-delay for given s_i^{NC} and s_i^{NP} for each journal. Figure 3 shows these predictions. Let us first focus on the number of previous coauthors s_i^{NC} in Figure 3 (left). We see that for all journals except RMP the predicted average t_i^{peak} is always less than 4 years, irrespective of the number of previous coauthors. For RMP, papers with no authors take around 7.5 years on average to reach the peak, but this number then also decreases to 4 years at roughly 150 previous coauthors.

We further point out the differences in speed across journals at which the peak-delays decrease for increasing numbers of previous coauthors. For example, papers in the journal PR-HEP reach the peak citation rate on average after 3.75 years for zero previous coauthors. This duration changes to roughly 2.5 years for papers with 100 previous coauthors. This is different from the journal PRE. There, a paper reaches the peak citation rate on average after

Table 2. Fitted parameters β for the negative binomial regression in Eq. 6, computed for each journal individually. Four fits are displayed for each journal, depending on whether the predictor is s_i^{NC} (NC) or s_i^{NP} (NP), and whether time is measured in years (cf. section 3.2) or in publications (cf. section 3.4). The stated significance levels of the estimated parameter β are given as *** (< 0.001), ** (< 0.01), * (< 0.05)

s_i	Time	α	β
PR			
NC	years	0.768	−0.022***
	pubs	0.903	−0.003*
NP	years	0.721	−0.010***
	pubs	0.922	−0.006***
PRA			
NC	years	1.055	−0.001***
	pubs	0.637	0.000
NP	years	1.095	−0.005***
	pubs	0.625	0.001
PRC			
NC	years	1.141	−0.000*
	pubs	1.132	0.000
NP	years	1.130	−0.000
	pubs	1.141	−0.000
PRE			
NC	years	0.907	−0.001***
	pubs	0.971	−0.000
NP	years	0.941	−0.004***
	pubs	0.992	−0.002**
RMP			
NC	years	1.987	−0.004*
	pubs	1.482	−0.003
NP	years	2.022	−0.008*
	pubs	1.534	−0.008*
JHEP			
NC	years	0.050	−0.002
	pubs	−0.251	−0.001

Table 2. (*continued*)

s_i	Time	α	β
NP	years	0.035	0.000
	pubs	−0.286	−0.000
PR-HEP			
NC	years	1.292	−0.005***
	pubs	1.870	−0.000*
NP	years	1.321	−0.004***
	pubs	1.907	−0.001***
Phys. Lett.			
NC	years	0.813	−0.007***
	pubs	1.088	−0.009***
NP	years	0.824	−0.004***
	pubs	1.094	−0.005***
Nuc. Phys.			
NC	years	1.156	−0.007***
	pubs	1.199	−0.005***
NP	years	1.211	−0.005***
	pubs	1.239	−0.004***

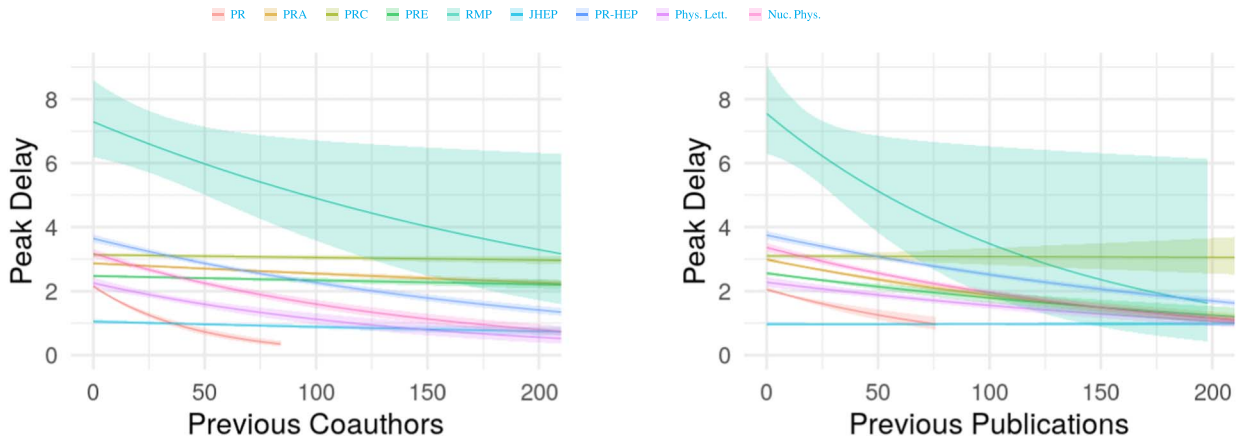

Figure 3. Relation between peak-delays t_i^{peak} measured in years, and the number of previous coauthors s_i^{NC} (left) or publications s_i^{NP} (right) according to Eq. 6. The solid lines are the estimated responses, and the respective colored areas are 95% confidence bands from the negative binomial regressions. Estimated responses are plotted at most until the largest observed number of previous coauthors or publications in the respective journal.

2.5 years for zero previous coauthors, which stays almost the same even at 100 previous coauthors. This means that journals have a large impact on the time *when* citations occur, especially with respect to the prospective decrease as the number of coauthors grows. Figure 3 also shows confidence bands for the predicted average t_i^{peak}. These are narrow for all journals except one, because of the large numbers of papers used in the model fits. For the exception, RMP, only 214 papers were used, which is why its confidence bands are wider.

Figure 3 (right) shows the average t_i^{peak} predicted by the number of previous publications, s_i^{NP}. The main difference from Figure 3 (left) is that now also the peak-delays for the journals PRA and PRE decrease noticeably for increasing numbers of previous publications. For example, t_i^{peak} is on average equal to 3 years for zero previous publications, but this number drops to 1 year for 200 previous publications. This means that, to receive citations earlier in these journals, increasing the number of publications appears to be a more successful strategy than increasing the number of coauthors.

To summarize, the negative binomial regression models show that for *increasing* numbers of previous coauthors or publications the highest citation rate is reached *sooner*. They also identify differences in the benefit of high numbers of coauthors or publications across journals: For journals such as PRC there is almost no decrease in peak delay, even with 200 previous coauthors. But for journals such as PR, papers that already have 50 previous coauthors reach their peak on average in less than half the time of papers with zero previous coauthors.

3.3. Characteristic Decay Time

3.3.1. Regressions

We now analyze the relationship between characteristic decay time τ_i of paper i and the social relations of its authors. To find whether there is a significant relationship, we perform a linear analysis for log-transformed variables:

$$\log_{10} \tau_i = \alpha^{\tau} + \beta^{\tau} \cdot \log_{10} s_i \tag{7}$$

where again s_i is the number of previous coauthors s_i^{NC} or the number of previous publications s_i^{NP}, and the time unit is again chosen as years. In the Supplementary Material, Section S2, we show that Eq. 7 reasonably fulfils the assumptions of linear regression models.

3.3.2. Fitted parameters

These are presented in Table 3. There, we see that all fitted parameters β^{τ} are negative and significantly different from 0 (on a significance level of 0.05). To interpret the effect of s_i on τ_i, one can exponentiate Eq. 7 to obtain

$$\tau_i \sim [s_i]^{\beta^{\tau}}. \tag{8}$$

Because β^{τ} is negative, this means that the more previous coauthors the authors have, the smaller the value of τ_i becomes. From Eq. 3 we know that the smaller τ_i is, the faster the decay of the normalized citation rate $\tilde{c}_i(t)$. This in turn means that such a paper faces a quicker and stronger shortage in new citations. Again, we also find significantly negative parameters β^{τ} when using the number of previous publications s_i^{NP} in Eq. 7. To conclude, we find that the larger the number of previous coauthors or publications is, the quicker and stronger the shortage in new citations after the peak.

Table 3. Parameters β^{τ} fitted according to Eq. 7 for each examined journal in our APS data set (left) and in our INSPIREHEP data set (right). Four fits are displayed for each journal, depending on whether the predictor is s_i^{NC} (NC) or s_i^{NP} (NP), and whether time is measured in years (cf. section 3.3) or in publications (cf. section 3.4). The significance levels of the *p*-values for β^{τ} are encoded as *** (< 0.001), ** (< 0.01), * (< 0.05)

s_i	Time	α^{τ}	β^{τ}
PR			
NC	years	0.714	−0.082***
	pubs	0.859	−0.013
NP	years	0.687	−0.032***
	pubs	0.866	−0.020*
PRA			
NC	years	0.978	−0.138***
	pubs	0.833	−0.056***
NP	years	0.964	−0.142***
	pubs	0.827	−0.058***
PRC			
NC	years	0.996	−0.071***
	pubs	1.017	−0.052***
NP	years	0.986	−0.083***
	pubs	1.013	−0.063***
PRE			
NC	years	0.779	−0.041***
	pubs	0.808	−0.049***
NP	years	0.768	−0.036***
	pubs	0.789	−0.038***
RMP			
NC	years	1.328	−0.272***
	pubs	1.281	−0.337***
NP	years	1.305	−0.247***
	pubs	1.224	−0.281***
JHEP			
NC	years	0.573	−0.061***
	pubs	0.512	−0.002

Table 3. (*continued*)

s_i	Time	α^τ	β^τ
NP	years	0.524	−0.026***
	pubs	0.490	0.011
PR-HEP			
NC	years	0.917	−0.159***
	pubs	1.284	−0.054***
NP	years	0.902	−0.135***
	pubs	1.269	−0.039***
Phys. Lett.			
NC	years	0.846	−0.080***
	pubs	0.918	−0.123***
NP	years	0.849	−0.070***
	pubs	0.916	−0.102***
Nuc. Phys.			
NC	years	0.971	−0.117***
	pubs	0.997	−0.116***
NP	years	0.997	−0.119***
	pubs	1.012	−0.111***

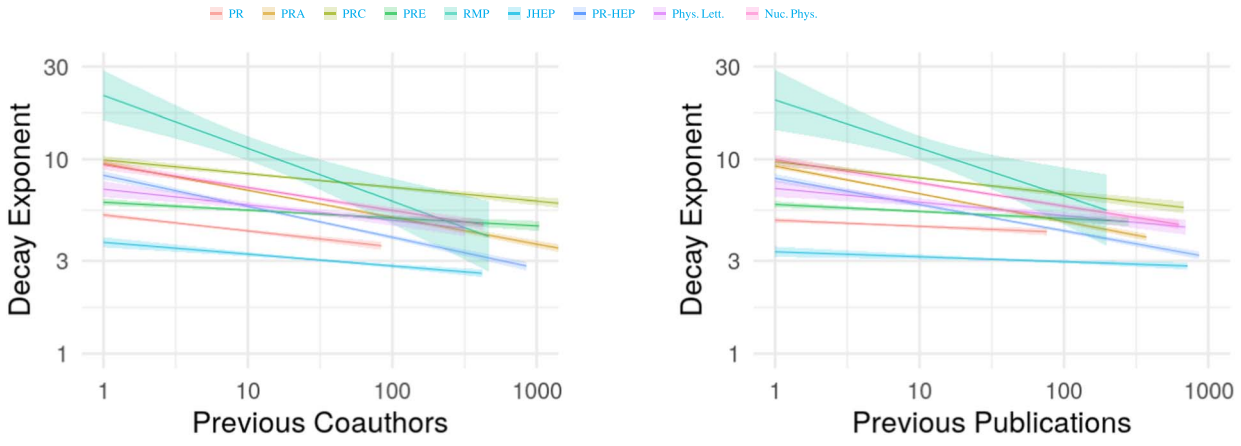

Figure 4. Relation between decay exponents τ_i and the number of previous coauthors s_i^{NC} (left) or the number of previous publications s_i^{NP} (right) according to Eq. 7. The solid lines are the estimated responses, and the respective colored areas are 95% confidence bands derived from the standard errors. Estimated responses are plotted at most until the largest observed number of previous coauthors or publications in the respective journal.

3.3.3. Size of the effect

We also intend to study the size of the dependence between decay exponents τ_i and the number of previous coauthors s_i^{NC} or publications s_i^{NP}. To this end, we visualize the estimated average decay parameters for the different journals in Figure 4. We focus on the description of the number of previous coauthors, Figure 4 (left), because overall both plots convey a similar message. We see that for papers with zero previous coauthors, the decay exponents are below 10 for all journals, except for RMP, which attains a decay exponent below 30. We further point out that papers in the journal JHEP have the smallest decay exponents even for up to 1,000 previous coauthors. This in turn means that decays in this journal tend to be particularly fast compared to the other journals.

3.4. Rescaling Time by Counting Publications

3.4.1. Effect of the growing scientific output

It is known that the number of papers published every year grows exponentially over time (Price, 1951). This means that in recent years there are more papers published in a given time interval than was the case longer ago. All of these new publications can potentially cite a given paper. This time dependence likely affects our regression results by confounding the respective response (t_i^{peak} or τ_i) and predictor variable (s_i^{NC} or s_i^{NP}). In the past it was suggested that the dependence of the citation rate on the publication year of a paper can be weakened by counting time in terms of the number of published papers instead of absolute time (days, weeks, years, etc.; Parolo et al., 2015). Therefore we repeat our regressions from section 3.2 and 3.2, and while measuring time on this alternative timescale. Thereby we assess whether such a bias from the publication year of a paper is present in the relations that we found.

3.4.2. Results for the alternative timescale

The fitted parameters are listed in the *pubs* rows in Table 2 for the peak-delay models and in Table 3 for the decay models. They remain smaller than 0, except for three journals: PRA, PRC, and JHEP. For PRA and PRC the fitted parameters β are positive for the peak-delay models with the number of previous coauthors, s_i^{NC}, as predictor. However, neither of these parameters is significantly different from 0. For JHEP the fitted parameter, β_1^{τ}, is positive for the decay model with the previous number of publications, s_i^{NP}, as predictor. However, this parameter is also not significantly different from 0. Only one significantly positive parameter occurs in the whole study, namely for PRA with the number of previous publications, s_i^{NP}, as predictor. The fitted parameters for all other journals are either negative or insignificantly different from zero, as was the case when measuring time in years. This means that, also according to the alternative timescale, for most journals the citation rate peak is reached faster for papers by authors with more previous coauthors or publications. Accordingly, the decay becomes steeper for papers by such authors.

4. CONCLUSIONS

In this paper, we address the question of how the attention towards an academic publication is accumulated over time, depending on the social relations of its authors, as expressed in the coauthorship network. For example, does the attention mostly occur in an early phase right after publication? Or is it rather spread uniformly over time? Or might it even happen only after a long time has passed since publication? To obtain a tractable, objective characterization of attention, we proxy attention by the citation rate of a paper (i.e., the number of new citations obtained in a particular time interval). We argue that, in order for a citation to occur, the authors of the citing paper have to be aware of the cited paper.

To study the time when this attention occurs, we compute the change in the number of citations over a time interval (i.e., the citation rate). It is known that the citation rates of most papers have two characteristic phases over time, namely an increasing phase followed by a decay phase. We found that the first phase tends to get shorter and the decay in the second phase tends to get faster for papers written by authors who have many previous coauthors. We also found that for some journals the time to the peak citation rate is almost halved within the first 100 previous coauthors, while for other journals it stays almost unchanged. Such a difference is also present in the decay exponents for different journals.

In terms of attention, our findings mean that papers written by authors with more previous coauthors attract attention faster, but are then also forgotten sooner. We also found this effect when measuring the number of previous papers of the authors instead of the number of previous coauthors. Furthermore, this effect also persisted when we controlled for the time when a paper was published. But most importantly, we found this effect in *nine journals*, based on *hundreds of thousands of authors and papers* and *far more than a million citations*. A study on such a large scale is a strong sign that we have uncovered a general trend that is not limited to the analyzed data sets.

4.1. A Speculative Explanation

Which mechanisms could be responsible for this? One way how authors learn about the papers which they cite is through communication with other scientists. Hence, authors can use their (few or many) social contacts, proxied by coauthors, to "advertise" a paper. Our findings indicate that authors with many previous coauthors or papers tend to do so within a short period of time after publication. When a new publication is made, the authors "advertise" it to the scientific community by presenting it in conferences and seminars, by sharing it on social media, etc. This behaviour happens within a finite time period, after which the authors stop actively promoting the given publication. However, this explanation is merely speculative at this point.

4.2. Regressions Not Suitable for Predictions

Our performed regressions have low predictive power, as indicated by extremely small coefficients of determination, R^2. For instance, for some regressions the R^2 is as low as 0.001, meaning that only 0.1% of the variance in the dependent variable is explained. However, while our regression models are not useful for prediction, our inferred relations are significant. In particular our regressions show that the time to the peak citation rate and the subsequent decay are not independent of the authors.

4.3. No Causal Relations Studied

In our study, we focus on the detection of the dependence between citation rate and social relations of the authors. However, we do not (yet) aim to understand the actual mechanisms behind it. In other words, we study *associations* between measures of social relations and citation histories, but we do not aim to detect causal relationships between them. For example, our study does not guarantee that a paper gets scientific attention faster simply by replacing its authors by scientists with larger publication or coauthor counts. Instead, we observe such faster attention among papers whose authors were not actively chosen based on their past social relations.

4.4. Future Work

In the future, we also intend to study causal relationships. Such a study will allow us to determine *why* authors with many previous publications or coauthors tend to write papers that receive scientific attention faster. To this end, we can use generative modeling to learn more about these

underlying mechanisms. For instance, hypotheses can be formulated and tested using the framework of coupled growth models presented in Nanumyan et al. (2020).

We find that a paper receives attention from the scientific community faster, the more coauthors the authors had prior to its publication. But we find as well that such a paper is also forgotten sooner again afterwards. Our findings indeed highlight that the citations of a paper can have substantially different dynamics depending on the social relations of the authors. Furthermore, our approach illustrates how such coupled dynamics can be studied by representing scientific collaborations in a multilayer network.

ACKNOWLEDGMENTS

The authors would like to thank all reviewers for their comments and Luca Verginer and Giacomo Vaccario for discussions concerning the negative binomial regression models.

AUTHOR CONTRIBUTIONS

Christian Zingg: Conceptualization, Data curation, Formal analysis, Software, Validation, Visualization, Writing—original draft, Writing—review & editing. Vahan Nanumyan: Conceptualization, Data curation, Formal analysis, Software, Validation, Visualization, Writing—original draft, Writing—review & editing. Frank Schweitzer: Conceptualization, Formal analysis, Project administration, Supervision, Visualization, Writing—original draft, Writing—review & editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

No funding has been received for this research.

DATA AVAILABILITY

We use two large bibliographic databases, APS and INSPIREHEP. Access to the APS database can be requested for research purposes at https://journals.aps.org/datasets. Access to the INSPIREHEP database is possible either as a download or through an API as explained on its website https://inspirehep.net/. For this paper, we downloaded the INSPIREHEP database.

REFERENCES

Avramescu, A. (1979). Actuality and obsolescence of scientific literature. *Journal of the American Society for Information Science*, *30*(5), 296–303. **DOI:** https://doi.org/10.1002/asi.4630300509

Battiston, F., Iacovacci, J., Nicosia, V., Bianconi, G., & Latora, V. (2016). Emergence of multiplex communities in collaboration networks. *PLOS ONE*, *11*(1), e0147451. **DOI:** https://doi.org/10.1371/journal.pone.0147451, **PMID:** 26815700, **PMCID:** PMC4731389

Burrell, Q. L. (2005). Are "sleeping beauties" to be expected? *Scientometrics*, *65*(3), 381–389. **DOI:** https://doi.org/10.1007/s11192-005-0280-5

Cetina, K. (2009). *Epistemic cultures: How the sciences make knowledge*. Harvard University Press. **DOI:** https://doi.org/10.2307/j.ctvxw3q7f

Ciotti, V., Bonaventura, M., Nicosia, V., Panzarasa, P., & Latora, V. (2016). Homophily and missing links in citation networks. *EPJ Data Science*, *5*(1), 7. **DOI:** https://doi.org/10.1140/epjds/s13688-016-0068-2, **PMID:** 32355597, **PMCID:** PMC7175687

Claudel, M., Massaro, E., Santi, P., Murray, F., & Ratti, C. (2017). An exploration of collaborative scientific production at MIT through spatial organization and institutional affiliation. *PLOS ONE*, *12*(6), e0179334. **DOI:** https://doi.org/10.1371/journal.pone.0179334, **PMID:** 28640829, **PMCID:** PMC5480888

Clauset, A., Larremore, D. B., & Sinatra, R. (2017). Data-driven predictions in the science of science. *Science*, *355*(6324), 477–480. **DOI:** https://doi.org/10.1126/science.aal4217, **PMID:** 28154048

Colavizza, G., & Franceschet, M. (2016). Clustering citation histories in the *Physical Review*. *Journal of Informetrics*, *10*(4), 1037–1051. **DOI:** https://doi.org/10.1016/j.joi.2016.07.009

Costas, R., van Leeuwen, T. N., & van Raan, A. F. (2010). Is scientific literature subject to a "Sell-By-Date"? A general methodology to analyze the 'durability' of scientific documents. *Journal of the*

American Society for Information Science and Technology, *61*(2), 329–339. **DOI:** https://doi.org/10.1002/asi.21244

Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, *52*(7), 558–569. **DOI:** https://doi.org/10.1002/asi.1097

Eom, Y.-H., & Fortunato, S. (2011). Characterizing and modeling citation dynamics. *PLOS ONE*, *6*(9), e24926. **DOI:** https://doi.org/10.1371/journal.pone.0024926, **PMID:** 21966387, **PMCID:** PMC3178574

Fister, I., Fister, I., & Perc, M. (2016). Toward the discovery of citation cartels in citation networks. *Frontiers in Physics*, *4*, 49. **DOI:** https://doi.org/10.3389/fphy.2016.00049

Guimera, R. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, *308*(5722), 697–702. **DOI:** https://doi.org/10.1126/science.1106340, **PMID:** 15860629, **PMCID:** PMC2128751

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge: Cambridge University Press. **DOI:** https://doi.org/10.1017/CBO9780511973420

Jadidi, M., Karimi, F., Lietz, H., & Wagner, C. (2018). Gender disparities in science? Dropout, productivity, collaborations and success of male and female computer scientists. *Advances in Complex Systems*, *21*(03–04), 1750011. **DOI:** https://doi.org/10.1142/S0219525917500114

Jeong, H., Néda, Z., & Barabási, A. L. (2003). Measuring preferential attachment in evolving networks. *Europhysics Letters (EPL)*, *61*(4), 567–572. **DOI:** https://doi.org/10.1209/epl/i2003-00166-9

Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying Sleeping Beauties in science. *Proceedings of the National Academy of Sciences*, *112*, 7426–7431. **DOI:** https://doi.org/10.1073/pnas.1424329112, **PMID:** 26015563, **PMCID:** PMC4475978

Lachance, C., & Larivière, V. (2014). On the citation lifecycle of papers with delayed recognition. *Journal of Informetrics*, *8*(4), 863–872. **DOI:** https://doi.org/10.1016/j.joi.2014.08.002

Leydesdorff, L., Carley, S., & Rafols, I. (2013). Global maps of science based on the new Web-of-Science categories. *Scientometrics*, *94*(2), 589–593. **DOI:** https://doi.org/10.1007/s11192-012-0784-8, **PMID:** 23335826, **PMCID:** PMC3547244

Li, J., & Ye, F. Y. (2014). A probe into the citation patterns of high-quality and high-impact publications. *Malaysian Journal of Library and Information Science*, *19*(2), 17–33.

Martin, T., Ball, B., Karrer, B., & Newman, M. E. J. (2013). Coauthorship and citation patterns in the *Physical Review*. *Physical Review E*, *88*(1), 012814. **DOI:** https://doi.org/10.1103/PhysRevE.88.012814, **PMID:** 23944525

Mazloumian, A. (2012). Predicting scholars' scientific impact. *PLOS ONE*, *7*(11), e49246. **DOI:** https://doi.org/10.1371/journal.pone.0049246, **PMID:** 23185311, **PMCID:** PMC3504022

Nanumyan, V., Gote, C., & Schweitzer, F. (2020). Multilayer network approach to modeling authorship influence on citation dynamics in physics journals. *Physical Review E*, *102*, 032303. **DOI:** https://doi.org/10.1103/PhysRevE.102.032303, **PMID:** 33075907

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*(2), 404–409. **DOI:** https://doi.org/10.1073/pnas.98.2.404

Newman, M. E. J. (2004). Who is the best connected scientist? A study of scientific coauthorship networks. In E. Ben-Naim, H. Frauenfelder, & Z. Toroczkai (Eds.), *Complex networks* (pp. 337–370). Berlin/Heidelberg: Springer. **DOI:** https://doi.org/10.1007/978-3-540-44485-5_16

Parolo, P. D. B., Pan, R. K., Ghosh, R., Huberman, B. A., Kaski, K., & Fortunato, S. (2015). Attention decay in science. *Journal of Informetrics*, *9*(4), 734–745. **DOI:** https://doi.org/10.1016/j.joi.2015.07.006

Petersen, A. M., Fortunato, S., Pan, R. K., Kaski, K., Penner, O., … Pammolli, F. (2014). Reputation and impact in academic careers. *Proceedings of the National Academy of Sciences*, *111*(43), 15316–15321. **DOI:** https://doi.org/10.1073/pnas.1323111111, **PMID:** 25288774, **PMCID:** PMC4217436

Price, D. J. (1951). Quantitative measures of the development of science. *Archives Internationales d'Histoire des Sciences*, *4*(14), 85–93.

Radicchi, F., Weissman, A., & Bollen, J. (2017). Quantifying perceived impact of scientific publications. *Journal of Informetrics*, *11*(3), 704–712. **DOI:** https://doi.org/10.1016/j.joi.2017.05.010

Rinia, E. J., Van Leeuwen, T. N., Bruins, E. E., Van Vuren, H. G., & Van Raan, A. F. (2001). Citation delay in interdisciplinary knowledge exchange. *Scientometrics*, *51*(1), 293–309. **DOI:** https://doi.org/10.1023/A:1010589300829

Sarigol, E., Garcia, D., Scholtes, I., & Schweitzer, F. (2017). Quantifying the effect of editor-author relations on manuscript handling times. *Scientometrics*, *113*(1), 609–631. **DOI:** https://doi.org/10.1007/s11192-017-2309-y, **PMID:** 29056793, **PMCID:** PMC5629258

Sarigol, E., Pfitzner, R., Scholtes, I., Garas, A., & Schweitzer, F. (2014). Predicting scientific success based on coauthorship networks. *EPJ Data Science*, *3*, 9. **DOI:** https://doi.org/10.1140/epjds/s13688-014-0009-x

Schweitzer, F. (2014). Scientific networks and success in science. *EPJ Data Science*, *3*(1), 35. **DOI:** https://doi.org/10.1140/epjds/s13688-014-0035-8

Sinatra, R., & Lambiotte, R. (2018). Editorial. *Advances in Complex Systems*, *21*(03–04), 1802001. **DOI:** https://doi.org/10.1142/S0219525918020010

Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A.-L. (2016). Quantifying the evolution of individual scientific impact. *Science*, *354*(6312). **DOI:** https://doi.org/10.1126/science.aaf5239, **PMID:** 27811240

Tomasello, M. V., Vaccario, G., & Schweitzer, F. (2017). Data-driven modeling of collaboration networks: A cross-domain analysis. *EPJ Data Science*, *6*(1), 22. **DOI:** https://doi.org/10.1140/epjds/s13688-017-0117-5

van Raan, A. F. J. (2004). Sleeping beauties in science. *Scientometrics*, *59*(3), 467–472. **DOI:** https://doi.org/10.1023/B:SCIE.0000018543.82441.f1

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. New York, NY: Springer. **DOI:** https://doi.org/10.1007/978-0-387-21706-2