# Disruption indices and their calculation using web-of-science data

*Indicators of historical developments or evolutionary dynamics?*

Leydesdorff, L.; Bornmann, L.

**DOI**
10.1016/j.joi.2021.101219

**Publication date**
2021

**Document Version**
Final published version

**Published in**
Journal of Informetrics

**License**
CC BY

Link to publication

**Citation for published version (APA):**
Leydesdorff, L., & Bornmann, L. (2021). Disruption indices and their calculation using web-of-science data: Indicators of historical developments or evolutionary dynamics? *Journal of Informetrics*, *15*(4), [101219]. https://doi.org/10.1016/j.joi.2021.101219

Journal of Informetrics 15 (2021) 101219

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

# Disruption indices and their calculation using web-of-science data: Indicators of historical developments or evolutionary dynamics?

Loet Leydesdorff[a,*], Lutz Bornmann[b]

[a] Amsterdam School of Communication Research (ASCoR), University of Amsterdam, PO Box 15793, 1001 NG Amsterdam, the Netherlands
[b] Science Policy and Strategy Department, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany

ARTICLE INFO

*Keywords:*
Disruption
Evolutionary regime
Historical trajectory
Kullback-Leibler
Redundancy

ABSTRACT

Science and technology develop not only along historical trajectories, but also as next-order regimes that periodically change the landscape. Regimes can incur on trajectories which are then disrupted. Using citations and references for the operationalization, we discuss and quantify both the recently proposed "disruption indicator" and the older indicator for "critical transitions" among reference lists as changes which may necessitate a rewriting of history. We elaborate this with three examples in order to provide a proof of concept. We shall show how the indicators can be calculated using Web-of-Science data. The routine is automated (available at < http://www.leydesdorff.net/software/di/index.htm >) so that it can be upscaled in future research. We suggest that "critical transitions" can be used to indicate disruption at the regime level, whereas disruption is developed at the trajectory level. Both conceptually and empirically, however, continuity is grasped more easily than disruption.

## 1. Introduction

Generations of citations can be used as indicators of historical trajectories in scholarly literature or analogously among patents (Garfield, 1963). Garfield et al. (2003) introduced the concept of "algorithmic historiography" for such an indicator-based and computer-assisted reconstruction of historical developments. These authors developed HistCite™ for the analysis and visualization of citation networks over time. However, at that time the resulting graphs were not sufficiently elaborated for statistical analysis (Carley et al., 1993; Hummon & Doreian,1989; Lucio-Arias & Leydesdorff, 2008). Van Eck and Waltman (2014) developed CitNet-Explorer (available at https://www.citnetexplorer.nl/) as a software tool for visualizing the development of citation networks over time.

The time lines show both continuities and changes; the focus in visualizations is on historical continuities along trajectories. Disruptions are negatively visible as a lack of continuity. However, science and technology develop not only along historical trajectories, but also as next-order regimes that periodically change the landscape (e.g., Dosi, 1982; Geels, 2002). One sees the footprints as trajectories and infers to a more composed dynamic at the regime level. As the evolutionary economist (Schumpeter, 1943) formulated: "The problem that is usually being visualized is how capitalism administers existing structures, whereas the relevant problem is how it creates and destroys them" (p. 84).

Regime changes incur on historical trajectories as selection pressures. When agents, for example, are competitively hill-climbing along trajectories, it makes a difference whether one is climbing in the daylight or in the dark. When the light is turned on or off, the

* Corresponding author.
*E-mail addresses:* l.a.leydesdorff@uva.nl (L. Leydesdorff), bornmann@gv.mpg.de (L. Bornmann).

https://doi.org/10.1016/j.joi.2021.101219
Received 1 February 2021; Received in revised form 4 September 2021; Accepted 8 September 2021
Available online 27 September 2021

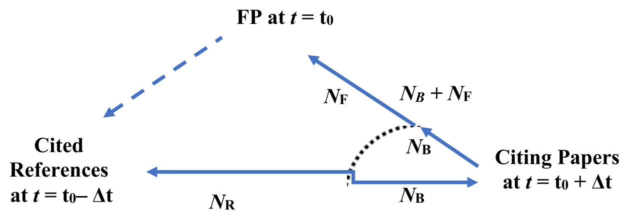


**Fig. 1.** Scheme of disruption in terms of types of citations and their couplings.

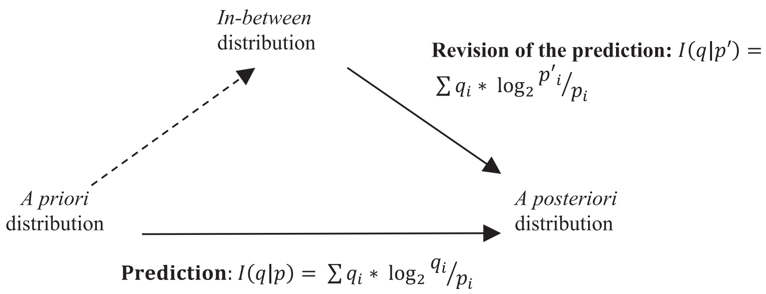


**Fig. 2.** Revision of the prediction at an in-between station.

selection environments change and thus also the historically available options. For example, the risk involved when deviating from the trajectory may be different in the dark than in the light.

Changes in selection environments are part of an evolutionary or systems dynamics which manifests itself historically in instantiations (Giddens, 1979) which couple the historical development to the evolutionary dynamics in (series of) events. However, the selection mechanisms are also dynamic (Hayami & Ruttan, 1970; Nelson & Winter, 1977). Thus, the dynamics are dually layered: both the generation of phenotypical variation and the interactions among "genotypical" selection environments can generate both change and continuity. Unlike "natural selection" in biology, these selection mechanisms are (co-)constructed along with the variation. The selection environments therefore have the status of hypotheses (Langton, 1989, p. 6).

## 2. Disruption and its operationalization

Developments at the regime level can disrupt the historical continuity at the trajectory level. Recently, Funk and Owen-Smith (2017) proposed a disruption indicator (DI) for patent citations; Wu et al. (2019) elaborated this indicator for scholarly literature. These authors defined the DI of a focal paper (FP) as follows:

$$DI = \frac{N_F - N_B}{N_F + N_B + N_R} \quad (1)$$

We follow the notation of Wu and Wu (2019):

- $N_B$ is the number of papers citing both the FP and at least one of its references;
- $N_F$ is the number of papers citing exclusively the FP, but none of its references;
- $N_R$ is the number of papers citing the references of FP, but not FP itself.

Fig. 1 visualizes the model. The indicator varies from +1 (disruption) to –1 (continuity). Consequently, disruption and continuity are considered as a trade-off on a single dimension: more disruption means less continuity and vice versa. Alternatively, one could model disruption and continuity as two (e.g., orthogonal) dimensions. The DI operationalizes the construction of continuity by bibliographically coupling the FP under study to its references in later citations (Kessler, 1963). These latter—that is, citing—papers (re-)construct the context of validation from the perspective of hindsight by means of bibliographic couplings between FPs and their references. Without this reorganization, the FPs remain variations in knowledge claims made in the context of discovery (Popper, 1959 and 1972).

Three generations are thus distinguished along the time line: (i) the FP(s) under study in the middle, (ii) the knowledge base of the paper(s) in terms of the references cited—from the previous generation—and (iii) citations of the FP and its references in the next-generation literature ($t = t_0 + \Delta t$; see Fig. 1). This bibliographical coupling can vary between zero and one hundred percent. However, disruption is defined as the absence of continuity (to be measured below as $N_B \rightarrow 0$). Whereas the mechanism for continuity generation is thus operationalized as bibliographic coupling, disruption is operationalized negatively as the absence of thus observable continuity.

Let us compare the triangle in Fig. 1 with a similar one for "critical transitions" in Fig. 2. "Critical transitions" were introduced by one of us as an information-theoretical measure of continuity and potentially disruptive change (Leydesdorff, 1991). Information-theoretically, the in-between (focal) paper can be considered as an auxiliary station in the transmission of a signal from a sender to a receiver. If this in-between station boosts the signal for the receiver, the history of what happened before the boosting no longer matters from the later perspective of the receiver. As in the case of DI, the in-between layer of FPs redirects the signal from its cited references to the FP(s) itself. The past is thus incorporated (Price, 1965; cf. Garfield, 1975). The history of the signal before this passage can be rewritten after this uncoupling from the past. A discontinuity can thus be generated.

Kullback and Leibler's (1951) divergence can be used to quantify changes among probability distributions of events (e.g., Frenken & Leydesdorff, 2000; Leydesdorff, 1991; 1995, at p. 341; Lucio-Arias & Leydesdorff, 2008). Whereas DI indicates disruption by specific events, critical transitions operate on distributions of events. We use the distributions of cited journals and the bibliographic couplings among them.

The differences between the models depicted in Fig. 1 and 2 may help to clarify and operationalize the differences and tensions between historical developments—as a series of events—and evolutionary dynamics as a series of configurations (probability distributions) in terms of scientometric indicators. Trajectories relate events, whereas systems evolve in terms of distributions of events (e.g., in Markov chains). The two indicators have in common that after the FP, the history written before the event may be outdated. In Fig. 1, the arrows follow the referencing direction from citing to cited; in Fig. 2, the arrows point in the opposite direction following the information flow. However, the intellectual distance between these theoretical notions and their scientometric operationalization has hitherto remained considerable. Bridging this gap is one of the objectives of this study.

## 3. Evolutionary dynamics and historical developments

The evolutionary process is not necessarily punctuated by one or more FPs as events. When restructurations are postponed, tensions can be expected which are then relaxed, as in the case of avalanches. Kuhn (1977) considered the tensions between novelty generation and historical embedding as "essential" for the sciences. Such changes can be radical or gradual: one would expect large changes to be rare and smaller ones more common. Crucial is the point that the dynamics of selection environments are different from the historical generation of variants along trajectories. Continuity is constructed in research practices (e.g., Edge, 1979); for example, by the intellectual or bibliographic coupling of sources. However, Fujigaki (1998, p. 11) added that evolutionary change in socio-cognitive systems is generated by the discursive development of distinctions. Tensions between the historical and the analytical orders can be expected, since the recursive process of distinguishing tends to generate meta-stable states—at the edge of chaos.

It can be shown that a dynamic operating as a feedback on historical traces leads to pink or 1/f noise—operationalized as a linear relation between the logarithm of the frequency and the logarithm of the size of the events (Bak et al., 1987). The system retains previous states as footprints along the trajectory of its history. The tensions in the system lead to adjustments via "avalanches." Freeman and Perez (1988) modeled retention as a dialectic between innovation in the production forces and structural adjustments in the carrying networks.

A diffusion dynamic coupled to a production flow at the level of the system can also be considered as a reaction-diffusion dynamic (Rashevsky, 1940; Rosen, 1968; Turing, 1952): the production of manuscripts is selectively coupled to their diffusion, for example, by peer review. The textual carriers serve a cognitive diffusion dynamic in the evolution of the sciences that is different from the underlying production dynamic among agents. However, the more complex dynamics (including diffusion) should not be reduced to their genesis.

Historical developments along trajectories and evolutionary change at the regime level do not stand orthogonally to each other, but are co-constructed (unlike models of evolution in biology, where variations and selections act independently). In the history and philosophy of science, the two processes have been conceptualized in terms of a logic of discovery, in which new knowledge claims are generated as variation, versus a logic of justification insuring quality control by making selections (Popper, 1972, [1935] 1959). Note that selection mechanisms are structural and deterministic, whereas variation can be stochastic.

As against these two logics operating on parallel planes, the coupling between variation and selection is here considered as a feedback loop: each citation relation is part of a citation network (although this network may initially be relatively empty). As the network is increasingly populated, it develops a structure that feeds back top-down on the bottom-up submissions. Such a feedback, for example, can bend or disturb a sigmoid growth curve to a plateau or lead to a bifurcation.

From an evolutionary perspective, variation is phenotypical; but only genotypes can evolve. The phenotypical networks serve the historical retention of the self-organizing flows (Hodgson & Knudsen, 2011). In a critique of the so-called "neo-Schumpeterian" school in evolutionary economics (e.g., Dosi et al., 1988; Lee, 2013), Andersen (1992 and 1994) suggested that one had focused on historical reconstructions without sufficient specification of "what is evolving?" Andersen quoted Boulding (1978, p. 33), who argued that "what evolves is something very much like knowledge." The knowledge base of an economy can evolve by selecting and coding distinctions (Leydesdorff, 2021). The textual archives offer stepping stones for new articulations.

## 4. Algorithms and methods

The sciences develop evolutionarily in terms of selection mechanisms (Popper, 1972), while the developments leave their footprints historically. How can one model and measure developments along trajectories and under regimes in terms of scientometric indicators? Let us further specify the two indicators under study in more detail.

### 4.1. Disruption

As noted, Wu et al. (2019) formulated the DI of an FP as follows:

$$DI = \frac{N_F - N_B}{N_F + N_B + N_R} \tag{1}$$

Wu and Wu (2019) noted that the numerator $(N_F - N_B)$ can be negative when the majority of the citations of FP couples FP bibliographically to its references. This leads to a value between minus one and zero. Increases in the value $N_R$ (other references) lead to less disruption in the case that $N_F - N_B > 0$, but enhance disruption when $N_F - N_B < 0$. The authors considered this effect as "inconsistent," and called for a revision of the indicator.

In a previous communication, Leydesdorff et al. (2021), we argued that the choice of $[N_F - N_B]$ in the numerator of Eq. (1), subtracts $N_B$ twice from the total citations of the FP which is $[N_F + N_B]$ (see Fig. 1). $N_F$ in the numerator itself can also be used as a measure of uncoupling and therefore disruption, whereas $N_B$ counts and consolidates the number of bibliographic couplings. Using $N_F$ in the numerator solves the noted inconsistency. Following Wu and Wu (2019), Leydesdorff et al. (2021) revised this indicator $DI^*$ as follows:

$$DI^* = \frac{N_F}{N_F + N_B + N_R} \tag{2}$$

Analogously, one can define

$$DI^\# = \frac{N_B}{N_F + N_B + N_R} \tag{2}$$

as a measure of consolidation.

Increases in the number of other (uncoupled) references ($N_R$) may proliferate in empirical cases. Since $N_R$ adds to the denominator, the value of DI tends towards zero by this effect on the denominator. Noting the prevalence of values of almost zero, Bornmann & Tekles (2021) and Bornmann et al. (2020) extended DI to $DI_n$, where n denotes the minimum threshold value for counting the bibliographic couplings between FP and its references in a single citing paper. Only papers which cite *n* or more references among the cited references of FP are counted in $M_B^n$, which is then used in Eq. (3) instead of $N_B$:

$$DI_n = \frac{N_F - M_B^n}{N_F + M_B^n + N_R} \tag{3}$$

It follows analytically that the original indicator is the same as $DI_1$. By adding thresholds, these authors aimed to adjust the indicator in order to focus on identifying disruptive research by discarding weak signals as noise. Focusing on articles published in *Scientometrics*, Bornmann et al. (2020) suggested that a field-specific variant of $DI_5$ may be a promising approach for improving existing disruption indicators. However, further exploration on a variety of datasets, journals, and disciplines would be needed to elucidate this possibility to identify disruptive papers which concentrate the attention of researchers within a discipline. This suggests that focusing on the citation links in a specific field (here: scientometrics as represented by the journal *Scientometrics*) is a promising approach for improving existing disruption indicators. Further exploration on a variety of datasets, journals, and disciplines is certainly needed to elucidate this possibility.

### 4.2. Critical transitions

The other indicator under study— "critical transitions" (Leydesdorff, 1991) —builds on Shannon's (1948) definition of the uncertainty in a probability distribution $[H_i = -\sum_i p_i \log_2 p_i;\ \sum_i p_i = 1]$. Kullback and Leibler (1951) formalized the dynamic extension of Shannon's measure as follows:

$$I_{(q:p)} = \Sigma_i q_i \log_2 (q_i/p_i) \tag{4}$$

where $I_{(q:p)}$ measures the expected information of the message that the prior distribution ($\Sigma_i\ p_i$) was turned into the posterior distribution ($\Sigma_i\ q_i$). (When the two-base of the logarithm is used, I is expressed in bits of information.)[1] In the case of a perfect prediction, $I_{q:p} = 0$ and the two distributions are similar. However, the information content $I_{(q:p)}$ goes to infinity if one of the terms in the prior distribution is zero (in the denominator and therefore generating a division by zero). The interpretation is that a zero in the prior distribution ($\sum_i p_i$) cannot predict a non-zero ex post. Novelty has a surprise value!

Theil (1972, at p. 77) added that the prediction of the posterior distribution can be improved or worsened by a distribution at an in-between moment of time (see Fig. 2 above). This improvement of the prediction of the a posteriori probability distribution ($\Sigma_i\ q_i$) on the basis of an in-between probability distribution ($\Sigma_i\ p_i'$) compared with the original prediction ($\Sigma_i\ p_i$) can be formulated as follows:

$$I(q:p) - I(q:p') = \sum_i q_i \log_2(q_i/p_i) - \sum_i q_i \log_2(q_i/p'_i) = \sum_i q_i \log_2(p'_i/p_i) \tag{5}$$

[1] Theil (1972, at pp. 59 f.) has proven that *I* is necessarily equal to or larger than zero. However, the non-negative *aggregated* value for *I* allows for local entropy-changes as contributions that are negative. Note furthermore that *I* is *asymmetrical* in *p* and *q*: the information content of a change along the arrow of time is different from one in the reverse (backward) direction.

**Table 1**
Six papers selected as the domain for the exploration
(see Annex I for full bibliographic references).

| Nr | Abbreviations | Times cited | Number of references | DOI |
|---|---|---|---|---|
| n1 | Bornmann L, 2018, J INFORMETR, V12, P931 | 3 | 60 | 10.1016/j.joi.2018.07.009 |
| n2 | Leydesdorff et al., 2018, SCIENTOMETRICS, V116, P623 | 3 | 57 | 10.1007/s11192-018-2734-6 |
| n3 | Bornmann L, 2018, SCIENTOMETRICS, V115, P1119 | 17 | 22 | 10.1007/s11192-018-2682-1 |
| n4 | Bornmann L, 2018, PLOS ONE, V13, P | 4 | 53 | 10.1371/journal.pone.0194805 |
| n5 | Bornmann L, 2018, SCIENTOMETRICS, V114, P439 | 13 | 31 | 10.1007/s11192-017-2608-3 |
| n6 | Leydesdorff et al., 2018, SCIENTOMETRICS, V114, P567 | 22 | 72 | 10.1007/s11192-017-2528-2 |
| | | 62 | 295 | |

z

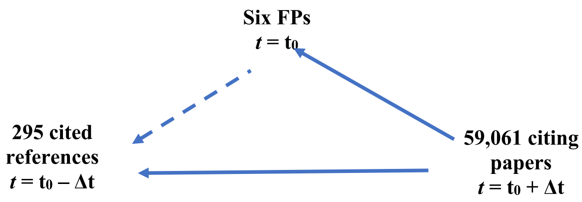


**Fig. 3.** Schematic representation of the searches needed for measuring disruptiveness in the six papers listed in Table 1 (source: WoS).

If I(q:p) > [I(q:p') + I(p':p)], the detour via the revision (in Fig. 2) can provide more of a path for the communication between sender and receiver in terms of bits of information than their direct link. Contrary to the geometry of Fig. 2, the sum of the information distances via the intermediate station is then shorter than the direct information path between the sender and the receiver. In other words, a negative entropy is generated (Krippendorff, 2009).

## 5. Operationalizations and measurements

### 5.1. Sample choices

Hirsch (2005)—with the paper in which the h-index was introduced—fulfills the criteria for disruption. These are:

1 A new research field was developed using and referencing this source. Since the introduction of the h-index by Hirsch (2005), a large number of variants of this indicator have been proposed with the goal of improving or complementing the indicator (e.g., Bornmann et al., 2011; Burrell, 2013). Based on the large number of papers dealing with the h-index itself or its variants, one can entertain the hypothesis that a subfield has emerged in bibliometric research.
2 The paper by Hirsch (2005) is not rooted in the relevant literature (Tahamtan & Bornmann, 2018): the development and discussion of the h-index in the paper was not grounded in the relevant literature; the paper contained only six references. At the time of introducing the h-index, Jorge Hirsch was not publishing as a scientometrician, but as a physicist. What is interesting here is whether the paper is disruptive if it does not cite previous literature in its field. The answer is 'yes' for Hirsch's (2005) paper. However, if this is true through different disciplines and through different years, the disruption index will be doubted.

We will compare Hirsch (2005) as a single paper first with the following six papers which we ourselves co-authored during 2018 (Table 1; see Annex I for the feull bibliographic references).

An advantage of using our own papers for this exploration is our thorough understanding of the six texts. By choosing a set of similar papers among which we can also appreciate the differences (as authors), we explore here the relevant parameter space before upscaling. The choice for one of these papers as also a single case is motivated by the idea to work bottom-up from the level of single cases. The initial choice among of the single papers is arbitrary.

The six FPs were cited 62 times at the time of the download (9 September 2020) and together contain 295 cited references Fig. 3). The sum total of citations of the FP(s) and these cited references, however, was 59,061 papers at this date. In other words, $N_R$—that is, the papers citing the cited references to the FP but not the FP itself—can "explode" and drive the values of DI, DI*, and DI# to zero by increasing the denominators in Eqs. (1) and ((2).

### 5.2. Retrieval of linked citations in the Web-of-Science (WoS)

Whereas it is notoriously difficult to disambiguate the results of a search on the basis of cited references in WoS (Thor et al., 2016), the Cited-Reference Search in WoS nowadays includes an alternative option to search with the digital object identifiers (DOI) of both a paper and its references. Only the so-called "linked" citations—listed in source publications of WoS—have DOIs indicated in the cited-references field as an additional sub-field. Of the 295 cited references in the sample of six papers in Table 1, 194 had

**Table 2**
DI, DI*, DI#, and $DI_5$ values for the papers investigated in this study.

| | | Hirsch (2005) | Six papers in Table 1 | Bornmann et al. (2018) |
|---|---|---|---|---|
| | $N_F$ | 3,734 | 0 | 0 |
| | $N_B$ | 151 | 72 | 3 |
| | $N_R$ | 1,320 | 59,840 | 1,390 |
| DI | Numerator | 3,582 | -72 | -3 |
| | Denominator | 5,185 | 59,912 | 1,393 |
| | DI | 0.69 | -0.0012 | -0.0022 |
| DI* | Numerator | 3,734 | 0 | 0 |
| | Denominator | 5,185 | 59,912 | 1,393 |
| | DI* | 0.72 | 0 | 0 |
| DI# | Numerator | 151 | 72 | 3 |
| | Denominator | 5,185 | 59,912 | 1,393 |
| | DI# | 0.03 | 0.00 | 0.00 |
| $DI_5$ | $\boldsymbol{M_B^5}$ | 0 | 24 | 0 |
| | Denominator | 5,054 | 59,864 | 1,390 |
| | $DI_5$ | 0.74 | -0.0004 | 0.0000 |

DOIs; 171 of these DOIs were unique references. In the remainder of this study, we use these linked citations for the retrieval in WoS. For the local analysis, however, we continue to use all cited references.

One can search the DOIs as sets by concatenating up to 1,000 DOIs into a single search string using the OR-operator. One can thus collect in a single search the 58,578 papers citing the 171 unique DOIs in our sample (as of 9 September 2020). The citations of the six FPs themselves resulted in an additional retrieval of 483 citing papers, so that the total of citing records is 59,061. We use here below a retrieval of 60,085 citing papers on 17 November 2020, containing 3,457,091 references.

As a third case, we zoom into the first reference in Table 1—Bornmann et al. (2018)—as another single paper. These three sets enable us thus to compare between two single papers and also to analyze the effects of aggregation.

## 6. Results

Table 2 provides descriptive statistics of the recall using (a) Hirsch (2005), (b) the six papers listed in Table 1, and (c) the first of these six papers.

As of 9 September 2020, Hirsch (2005) had been cited 4,568 times, of which the linked citations are $(N_F + N_B) = (total\ citations) =$ $3734 + 151 = 3885$; $N_F = 3734$, $N_B = 151$; $N_R = 1320$. The following values of DI, DI*, DI#, and $DI_5$ can thus be derived:

$$DI = \frac{N_F - N_B}{N_F + N_B + N_R} = \frac{3734 - 151}{3734 + 151 + 1320} = \frac{3583}{5205} = 0.688 \tag{6}$$

$$DI* = \frac{N_F}{N_F + N_B + N_R} = \frac{3734}{3734 + 151 + 1320} = \frac{3734}{5205} = 0.717 \tag{7}$$

$$DI^{\#} = \frac{N_B}{N_F + N_B + N_R} = \frac{151}{3734 + 151 + 1320} = \frac{151}{5205} = 0.029 \tag{8}$$

$$DI_5 = \frac{N_F - M_B^5}{N_F + M_B^5 + N_R} = \frac{3734 - 0}{3734 + 0 + 1320} = \frac{3734}{5054} = 0.739 \tag{9}$$

Since none of the citing papers meet Bornmann and Tekles' (2021) additional requirement of five bibliographic couplings between the cited references and the FP within a single citing document ($DI_5$), the value of $M_B^5$ is zero and therefore $DI_5 > DI^*$. However, the choice for n = 5 of $DI_5$ was ad hoc.

### 6.2. Automation of the computation

For the computation and automation of these indicators, one needs two downloads from WoS: (1) the references in the papers citing FP(s), and (2) the cited references in the FP(s). Three sets of references are needed for computing critical transitions. We have developed software for automating these two analyses; see Annexes 2 and 3, respectively. The software routines have been made available at http://www.leydesdorff.net/software/di/index.htm .

The results rare provided in Table 2: We compare the first paper among the six listed in Table 1 (that is, Bornmann et al., 2018) with Hirsch's (2005) as a single paper and with the set of six papers. How does the DI scale?

At the time of this search (9 September 2020), Bornmann et al. (2018) had been cited three times. It contains 60 references, of which 35 include a DOI. These DOIs were concatenated with the DOI for the FP (10.1016/j.joi.2018.07.009) into a single search string. The retrieval was 1,390 citing records with 53,079 cited references in total. Using "linked" data, we could retrieve 46,388 references in 13,093 papers.

**Table 3**
DI, DI*, and DI# values for six individual papers (decomposed).

| Focal Paper | NB | NF | NR | DI | DI* | DI# | NREF |
|---|---|---|---|---|---|---|---|
| Bornmann L, 2018, J INFORMETR, V12, P931 | 3 | 0 | 1504 | -0.002 | 0.000 | 0.002 | 36 |
| Leydesdorff et al., 2018, SCIENTOMETRICS, V116, P623 | 2 | 0 | 33956 | 0.000 | 0.000 | 0.000 | 37 |
| Bornmann L, 2018, SCIENTOMETRICS, V115, P1119 | 16 | 0 | 5622 | -0.003 | 0.000 | 0.003 | 17 |
| Bornmann L, 2018, PLOS ONE, V13, P | 6 | 14 | 14832 | 0.001 | 0.001 | 0.000 | 31 |
| Bornmann L, 2018, SCIENTOMETRICS, V114, P439 | 11 | 0 | 4221 | -0.003 | 0.000 | 0.003 | 17 |
| Leydesdorff et al., 2018, SCIENTOMETRICS, V114, P567 | 19 | 1 | 13353 | -0.001 | 0.000 | 0.001 | 56 |

The three citations of FP are all coupled bibliographically with the papers cited by FP, so the numerator of DI is: 0 – 3 = –3.

$$DI = \frac{N_F - N_B}{N_F + N_B + N_R} \frac{0-3}{0+3+1504} = \frac{-3}{1507} = -0.002 \tag{10}$$

$$DI *= \frac{N_F}{N_F + N_B + N_R} = \frac{0}{0+3+1504} = \frac{-0}{1507} = 0 \tag{11}$$

$$DI_5 = \frac{N_F - N_B}{N_F + N_B + N_R} = \frac{0-0}{0+3+1390} = \frac{0}{1507} = 0 \tag{12}$$

$$DI^{\#} = \frac{N_B}{N_F + N_B + N_R} = \frac{3}{0+3+1390} = \frac{3}{1507} = 0.002 \tag{13}$$

Using Eq. (1), it follows that DI = – 0.002. The paper is slightly conservative; that is, the DI is almost zero. Using a large number of papers, Bornmann et al. (2020) found that values close to zero are prevalent in a majority of the cases.

The values of DI, DI*, and $DI_5$ for the other papers in the group of six papers are provided in Table 3. One can also consider the group of six as a single œuvre and use the aggregated citations and references, summed over the six papers. The result is then as follows:

$$DI = \frac{N_F - N_B}{N_F + N_B + N_R} = \frac{0-72}{0+72+59840} = --0.001$$

$$DI *= \frac{N_F}{N_F + N_B + N_R} = 0$$

$$DI_5 = \frac{M_B^5}{N_F + M_B^5 + N_R} = 0$$

The large values of $N_R$ bring the values of DI closer to zero.

### 6.3. Conclusions about DI, DI*, $DI^{\#}$, and $DI_5$

The paper by Hirsch (2005) has a DI of 0.688. This is a very high value which could further be increased using DI* (= 0.72) or $DI_5$ (= 0.74). However, in the other two cases (see the corresponding columns in Table 2) $DI_5$ = 0, while DI has a negative value close to zero. In other words, the objective of setting the threshold in $DI_5$ – or more generally $DI_n$ – to prevent zeros is not achieved. The zeros will often be generated by large numbers of $N_R$ in the denominators of the respective indicators; that is, referencing by references which are not bibliographically coupled. However, $N_R$ does not play a role in the further analysis.

While the bibliographic coupling of Hirsch (2005) to its six references (of which only four have a DOI) is relatively rare ($N_B$ = 151 or 2.9% of the total citations of this paper), the situation is very different for the examples of scientometric papers in the other two columns of Table 2. In both cases, $N_F$ = 0. In other words, the references are either bibliographically coupled to the FP ($N_B$) or they relate to papers other than the FP(s). However, the references in the FP(s) are never cited without a reference to the FP. Perhaps these results can be considered as a reflection of the strongly inter-reading community structure in scientometrics (Wouters & Leydesdorff, 1994). There is no indication of disruption in this domain except for very low negative values of DI. This suggests that continuity prevails in this domain. Note that DI* cannot be negative, and thus cannot indicate continuity as a negative value. Perhaps continuity should be measured differently; for example, using $DI^{\#}$.

## 7. Critical revisions of the prediction

Critical revisions of the prediction can only be computed for comparisons in which both the prior probability ($p_i$) and the revised prediction ($p'_i$) are larger than zero, since otherwise one risks needing to divide by zero. (When the posterior probability is zero, the contributions to the prediction are also zero, since 0 * log(0) = 0.) Thus, we focus below on the cases where $q_i$, $p_i$ and $p'_i$ are all three larger than zero. In the case of Hirsch (2005), these are only three references in two journals: *European Journal of Physics B* and *Scientometrics.*

**Table 4**
Results and statistics for the three cases under study.

| | Hirsch (2005) | Bornmann et al. (2018) | Six papers 2018 |
|---|---|---|---|
| 1 Improvement of the prediction | –0.296 | –1.910 | –0.102 |
| 2. n of journals | 2 | 3 | 50 |
| 3. I(ex post: ex ante) | 0.664 | 0.002 | 0.962 |
| 4. I(ex post: in-between) | 0.960 | 1.911 | 1.064 |
| 5. I(in-between: ex ante) | 0.027 | 4.076 | 0.278 |
| 6. n of documents, ex ante | 4 | 35 | 142 |
| 7. n of documents, ex post | 5,209 | 1,398 | 60,086 |
| 8. n of citations, ex ante | 94 | 734 | 5,596 |
| 9. n of citations; ex post | 222,207 | 46,388 | 3,457,091 |

**Table 5**
Improvement of the prediction by the cited references in the knowledge base relative to the prediction by the FPs.

| | Hirsch (2005) | Six papers in Table 2 | Bornmann et al. (2018) |
|---|---|---|---|
| Improvement I(p:p') | 0.853 | 0.545 | -1.876 |
| I(p:q) | 0.881 | 0.768 | 0.002 |
| I(q:p') | 0.028 | 0.222 | 1.877 |
| (p':q) | 1.230 | 0.997 | 4.377 |
| N | 2 | 50 | 3 |

In all three columns of Table 4, the improvement of the prediction (row 1 in Table 4) has a negative sign. This negative sign means that the prediction is not improved, but worsened when the in-between distributions are used for the revision. The references in the prior and posterior distributions show a continuity that is disturbed by the single paper. In the case of Bornmann et al. (2018), the Kullback-Leibler divergence between the cited and citing references is only 0.002 bits and thus negligible. The distribution of references (in terms of cited journals) ex ante predicts the ex post ones almost perfectly.

### *7.1. An alternative ordering*

Given the worsening of the prediction in all three cases, one could consider the effects of changing the order: is $\sum_i p_i$ a better revision than $\sum_i p_i'$? Indeed, this is the case for Hirsch (2005) and for the set of six papers (Table 5' ; Fig. 4).

The continuity between the cited and citing references is in these two cases not disturbed by the FPs. In the case of Bornmann et al. (2018), however, the prediction remains worsened in this reversed order Fig. 5. In sum, this single paper is not indicated as part of the evolutionary dynamics; it is part of the historical development.

## 8. Discussion and conclusions

There is no obvious, one-to-one mapping between the theoretical concepts of regimes and trajectories—developed in the first part of this study—and the empirical analyses of the two bibliometric indicators (DI and critical transitions) in the second part. However, the two perspectives inform each other. Critical transitions can be considered as a part of evolutionary dynamics, whereas DI values inform us about historical dynamics. The indicator values for Hirsch (2005) revealed both dynamics: What is interesting here is whether a paper is disruptive if it does not cite previous literature in its field. The answer is “yes” for Hirsch’s (2005) paper. However, is this true through different disciplines and through different years?

In this case, the disruption is relatively large and a revision of the prediction is possible, albeit that the order between the improved and original prediction is different from our initial assumptions about this order. The ex ante distribution of references is a better predictor of the distribution of ex post distributions; this order suggests a Markov chain and thus systemness within the historical layer (Leydesdorff & Oomes, 1999).

In the case of the indices DI, DI*, $DI^{\#}$ and $DI_5$, change and continuity are measured in terms of events; in the case of critical transitions, the measurement is based on comparing the probability distributions of referenced journals between cited and citing journals in a vector space. The papers as events can be compared in terms of the extent to which these papers refer to overlapping or different horizons operationalized as their cited references. The cited references refer to changing and potentially disruptive (or continuous) horizons. The systems of references are intellectual structures which remain latent except for their manifestation in references (Figs. 4 and 5).

Events either happen or don’t happen, whereas probability distributions contain uncertainty. The events are related by historical sequences of citations at the trajectory level. A crucial question is whether or not trajectories are disturbed by the novelty in the knowledge claims. The probability distributions of abbreviated journal names are based on sets of relations, and include also non-relations (that is, zeros). Thus, the discontinuity and potential disruption can be measured at the regime level of changing structures (that is, selection environments). The evolutionary perspective thus enriches our understanding of two types of change. *Plus ça change,*

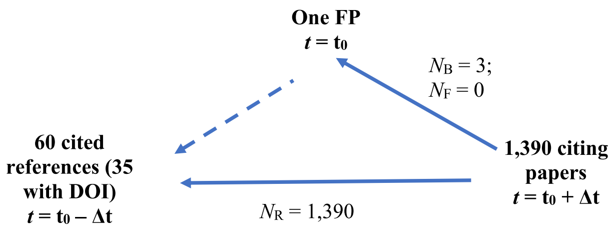


**Fig. 4.** Schematic representation of the search needed for measuring disruptiveness of Bornmann et al. (2018) as a single paper.

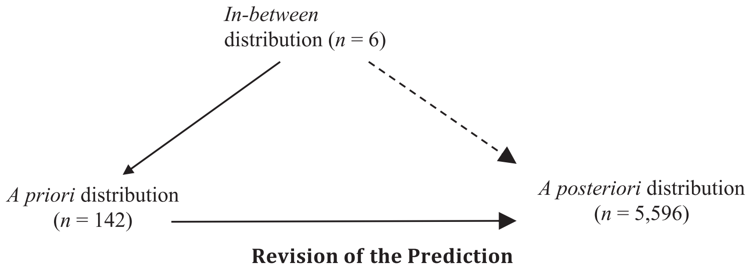


**Fig. 5.** The predictive relations in the six papers co-authored by Bornmann & Leydesdorff in 2018.

*plus ça reste la même chose.* Disruption and change can be phenotypical and observable at the trajectory level or structural in terms of selection mechanisms. (Leydesdorff et al., 2018).

The selection mechanisms for cultural and technological evolutions are not given, but have to be specified with the status of hypotheses. The selection mechanisms of society have become knowledge-intensive and therefore increasingly transparent and available for reconstruction. In an attempt to capture this alternative selection mechanism, Luhmann (1990), for example, suggested the following:

"what is special about the meaningful or meaning-based processing of experience is that it makes possible both the reduction and the preservation of complexity; i.e., it provides a form of selection that prevents the world from shrinking down to just one particular content of consciousness with each act of determining experience" (p. 27).

We suggest that the complexity of the communication evolves, but not the bounded rationality in citation behavior or, mutatis mutandis, the entrepreneurial behavior of firms. Agents make choices and can generate new variants. The bounded rationality of their decisions depends on their capacity to learn reflexively and to recognize opportunities. However, evolution is taking place in terms of what is genotypically binding (the codes in the communication) and not in terms of variation phenotypically bounded by the codes (Hodgson & Knudsen, 2011). The missing links can be related in the vector space (cf. Leydesdorff & Ivanova, 2014).

## Annex 1: Bibliographic information of investigated papers

1. Bornmann, L., Adams, J., & Leydesdorff (2018). The negative effects of citing with a national orientation in terms of recognition: national and international citations in papers from Germany, the Netherlands, and the UK. *Journal of Informetrics*, 12(3), 931-949. doi: 10.1016/j.joi.2018.07.009
2. Leydesdorff, L., Wagner, C., & Bornmann, L. (2018b). Discontinuities in citation relations among journals: Self-organized criticality as a model of scientific revolutions and change. *Scientometrics*, 116(1), 623-644. doi: 10.1007/s11192-018-2734-6
3. Bornmann, L., & Leydesdorff (2018). Count highly-cited papers instead of papers with h citations: use normalized citation counts and compare “like with like”! *Scientometrics*, 115(2), 1119-1123. doi: 10.1007/s11192-018-2682-1
4. Bornmann, L., Wagner, C., & Leydesdorff (2018). The geography of references in elite articles: Which countries contribute to the archives of knowledge? *PloS one*, 13(3), e0194805. doi: 10.1371/journal.pone.0194805
5. Bornmann, L., Haunschild, R., & Leydesdorff (2018). Reference publication year spectroscopy (RPYS) of Eugene Garfield’s publications. *Scientometrics*, 114(2), 439-448. doi: 10.1007/s11192-017-2608-3
6. Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2018). Betweenness and diversity in journal citation networks as measures of interdisciplinarity: A tribute to Eugene Garfield. *Scientometrics*, 114(2), 567-592. doi: 10.1007/s11192-017-2528-2

## Annex II: Computation of DI, DI*, and $DI_5$ using downloads from WoS

DI is based on WoS data (or similar data from Scopus). The routines di1.exe and di13.exe, at http://www.leydesdorff.net/software/di/di1.exe and http://www.leydesdorff.net/software/di/di13.exe, respectively, enable the user to compute values for DI, DI*, and $DI_n$ on the basis of data samples downloaded from WoS.

The order of operations contains two cycles: DI.exe first organizes the DOIs which are used as input in the second cycle for the indication of DI, DI*, and $DI_5$.

1 Download the FPs in the "plain text" format including the cited references; rename the resulting file into data.txt. This is a required file name in the next step. If the download is composed of more than a single file, copy and paste these files in one single file "data.txt."
2 Input this file to isi.exe in order to organize the data in relevant databases; the two files (data.txt and isi.exe) should be in the same folder.
3 Run di1.exe in the same folder; output is the file "doi.txt" with a search-string that can be used as input to WoS for searching with cited DOIs in a so-called Cited-Reference Search. Note that one has to work with batches of not more than 1000 lines because of systems limitations in WoS.
4 In the next screen (with the results of the search), tick "select all" and "finish search" if necessary in batches of 1000.
5 Add the resulting search results to "Marked list" and download in plain-text format including the cited references. Create a new file "data.txt" after renaming the old one.
6 If necessary, collect downloads into a single file "data.txt" after storing the previous file under a different name. (The old files will be overwritten.)
7 For the same reason, rename core.dbf and core.dbt into fp_core.dbf and fp_core.dbt, respectively; rename cr.dbf into fp_cr.dbf.
8 Run isi.exe again; four files should now be available: core.dbf, cr.dbf containing the cited references in the citing papers, fp_core.dbf, and fp_cr.dbf (containing the cited references of the cited papers). Run di13.exe. The resulting file di13.dbf contains the values of DI, $N_F$. $N_B$, and $N_R$. The file di_full.dbf contains detailed information.

## Annex III: Critical revisions measured in WoS

1 For running this routine, one collects three sets in WoS: cited, citing, and focal records. Download in plain text format including cited references and process the three sets of data via isi.exe. Each set will generate a file cr.dbf with the respective cited references. Rename the files cr.dbf, respectively, into crante.dbf (cited records), fp_cr.dbf (focal records), and crpost.dbf (citing records). Files with these names will be further processed.
2 Run jj_cc.exe in the same folder as the three dbf-files and run the routine.
3 Output is written to screen and into the file jj_info.txt. It is also stored in jj_info.dbf. This file is generated if not present; otherwise, a record is added at the time of each run. After the values for the revision, the values of the revision in the reverse order are written (see Section 7.1 above).

For example, for the case of six papers as follows (Table 4):

| | |
|---|---|
| improvement | -0.102 bits |
| I(post: ante) | 0.962 |
| I(post: between) | 1.064 |
| I(between: ante) | 0.278 |
| number of terms | 50 |

reverse the order between ante and between (Table 5)

| | |
|---|---|
| improvement | 0.545 bits |
| I(post: ante) | 0.768 |
| I(post: between) | 0.222 |
| I(between: ante) | 0.997 |
| number of terms | 50 |

## CRediT authorship contribution statement

**Loet Leydesdorff:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft. **Lutz Bornmann:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – review & editing.

## Acknowledgments

We are grateful to comments and suggestions by four anonymous referees. The bibliometric data used in this paper are mainly based on the online installation of the Web-of-Science, but partly retrieved from an in-house database developed and maintained in cooperation with the Max Planck Digital Library (MPDL, Munich) and derived from the Science Citation Index Expanded (SCI-E), Social Sciences Citation Index (SSCI), Arts and Humanities Citation Index (AHCI) prepared by Clarivate Analytics (Philadelphia, Pennsylvania, USA).

## References

Andersen, E. S. (1992). *Artificial economic evolution and Schumpeter. Aalborg: Institute for production*. University of Aalborg.
Andersen, E. S. (1994). *Evolutionary economics: Post-Schumpeterian contributions*. London: Pinter.
Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: An explanation of the 1/f noise. *Physical Review Letters, 59*(4), 381.

Bornmann, L., Mutz, R., Hug, S., & Daniel, H. (2011). A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *Journal of Informetrics, 5*(3), 346–359. 10.1016/j.joi.2011.01.006.
Bornmann, L., Devarakonda, S., Tekles, A., & Chacko, G. (2020). Disruptive papers published in Scientometrics: Meaningful results by using an improved variant of the disruption index originally proposed by Wu, Wang, and Evans (2019). *Scientometrics, 123*, 1149–1155. 10.1007/s11192-020-03406-8.
Bornmann, L., & Tekles, A. (2021). Convergent validity of several indicators measuring disruptiveness with milestone assignments to physics papers by experts. *Journal of Informetrics, 15*(3), Article 101159.
Boulding, K. E. (1978). *Ecodynamics: A new theory of societal evolution*. Beverly Hills, CA: Sage.
Burrell, Q. L. (2013). The h-index: A case of the tail wagging the dog? *Journal of Informetrics, 7*(4), 774–783.
Carley, K. M., Hummon, N. P., & Harty, M. (1993). Scientific influence: An analysis of the main path structure in the journal of conflict resolution. *Science Communication, 14*(4), 417–447.
Dosi, G. (1982). Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Research Policy, 11*(3), 147–162.
Dosi, G., Freeman, C., Nelson, R., Silverberg, G., & Soete, L. (1988). *Technical change and economic theory* (Eds.). London: Pinter.
Edge, D. (1979). Quantitative measures of communication in science: A critical overview. *History of Science, 17*, 102–134.
Freeman, C., & Perez, C. (1988). Structural crises of adjustment, business cycles and investment behaviour. In G. Dosi, C. Freeman, R. Nelson, G. Silverberg, & L. Soete (Eds.), *Technical change and economic theory* (pp. 38–66). London: Pinter.
Frenken, K., & Leydesdorff, L. (2000). Scaling trajectories in civil aircraft (1913-1970). *Research Policy, 29*(3), 331–348.
Fujigaki, Y. (1998). Filling the gap between discussions on science and scientists' everyday activities: Applying the autopoiesis system theory to scientific knowledge. *Social Science Information, 37*(1), 5–22.
Funk, R. J., & Owen-Smith, J. (2017). A dynamic network measure of technological change. *Management Science, 63*(3), 791–817.
Garfield, E. (1963). In *Citation indexes in sociological and historical research: 14* (pp. 289–291). American Documentation.
Garfield, E. (1975). The "obliteration phenomenon" in science—and the advantage of being obliterated. *Current Contents*, 396–398 December 22, #51/52.
Garfield, E., Pudovkin, A. I., & Istomin, V. S. (2003). Why do we need algorithmic historiography? *Journal of the American Society for Information Science and Technology, 54*(5), 400–412.
Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: A multi-level perspective and a case-study. *Research Policy, 31*(8-9), 1257–1274.
Giddens, A. (1979). *Central problems in social theory*. London, etc: Macmillan.
Hayami, Y., & Ruttan, V. W. (1970). Agricultural productivity differences among countries. *The American Economic Review, 60*(5), 895–911.
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the USA, 102*(46), 16569–16572.
Hodgson, G., & Knudsen, T. (2011). *Darwin's conjecture: The search for general principles of social and economic evolution*. Chicago /London: University of Chicago Press.
Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks, 11*(1), 39–63.
Kessler, M. M. (1963). In *Bibliographic coupling between scientific papers: 14* (pp. 10–25). American Documentation.
Krippendorff, K. (2009). Information of interactions in complex systems. *International Journal of General Systems, 38*(6), 669–680.
Kuhn, T. S. (1977). *The essential tension: Selected studies in scientific tradition and change*. Chicago: Chicago University Press.
Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics, 22*(1), 79–86.
Langton, C. G. (1989). Artificial life. In C. G. Langton (Ed.), *Artificial Life (Vol. VI, pp. 1-47)* (Ed.). Redwood, etc.: Addison-Wesley.
Lee, K. (2013). *Schumpeterian analysis of economic catch-up: Knowledge, path-creation, and the middle-income trap*. Cambridge University Press.
Leydesdorff, L. (1991). The static and dynamic analysis of network data using information theory. *Social Networks, 13*(4), 301–345.
Leydesdorff, L. (1995). *The Challenge of Scientometrics: The development, measurement, and self-organization of scientific communications*. Leiden: DSWO Press, Leiden University at http://www.universal-publishers.com/book.php?method=ISBN&book=1581126816 .
Leydesdorff, L. (2021). *The evolutionary dynamics of discursive knowledge: Communication-theoretical perspectives on an empirical philosophy of science*. Cham, Switzerland: Springer.
Leydesdorff, L., & Ivanova, I. A. (2014). Mutual redundancies in interhuman communication systems: Steps toward a calculus of processing meaning. *Journal of the Association for Information Science and Technology, 65*(2), 386–399. 10.1002/asi.22973.
Leydesdorff, L., & Oomes, N. A. (1999). Is the European monetary system converging to integration? *Social Science Information, 38*(1), 57–86.
Leydesdorff, L., Tekles, A., & Bornmann, L. (2021). A proposal to revise the disruption index. *El Profesional de la información (EPI), 30*(1), Article e300121.
Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2018). Discontinuities in citation relations among journals: Self-organized criticality as a model of scientific revolutions and change. *Scientometrics, 116*(1), 623–644. 10.1007/s11192-018-2734-6.
Lucio-Arias, D., & Leydesdorff, L. (2008). Main-path analysis and path-dependent transitions in HistCite (TM)-based historiograms. *Journal of the American Society for Information Science and Technology, 59*(12), 1948–1962. 10.1002/asi.20903.
Luhmann, N. (1990). Meaning as sociology's basic concept. In N. Luhmann (Ed.), *Essays on self-reference (pp. 21-79)* (Ed.). New York /Oxford: Columbia University Press.
Nelson, R. R., & Winter, S. G. (1977). In search of useful theory of innovation. *Research Policy, 6*(1), 35–76.
Popper, K. R. (1959). *The logic of scientific discovery* [1935. London: Hutchinson.
Popper, K. R. (1972). *Objective knowledge: An evolutionary approach*. Oxford: Oxford University Press.
Price, D. J. de Solla (1965). Networks of scientific papers. *Science, 149*(no. 3683), 510–515.
Rashevsky, N. (1940). An approach to the mathematical biophysics of biological self-regulation and of cell polarity. *The Bulletin of Mathematical Biophysics, 1*, 15–25.
Rosen, R. (1968). Turing's morphogens, two-factor systems and active transport. *The Bulletin of Mathematical Biophysics, 30*(3), 493–499.
Schumpeter, J. (1943). *Socialism, capitalism and democracy*. London: Allen & Unwin.
Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(July and October), 379–423 and 623-656.
Tahamtan, I., & Bornmann, L. (2018). Creativity in science and the link to cited references: Is the creative potential of papers reflected in their cited references? *Journal of Informetrics, 12*(3), 906–930.
Theil, H. (1972). *Statistical decomposition analysis*. Amsterdam /London: North-Holland.
Thor, A., Marx, W., Leydesdorff, L., & Bornmann, L. (2016). Introducing CitedReferencesExplorer: A program for Reference Publication Year Spectroscopy with Cited References Disambiguation. *Journal of Informetrics, 10*(2), 503–515. 10.1016/j.joi.2016.02.005.
Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society, Series B, 237*, 37–72.
van Eck, N. J., & Waltman, L. (2014). CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *Journal of Informetrics, 8*(4), 802–823.
Wouters, P., & Leydesdorff, L. (1994). Has price's dream come true: Is scientometrics a hard science? *Scientometrics, 31*, 193–222.
Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature, 566*(7744), 378–382.
Wu, S., & Wu, Q. (2019). A confusing definition of disruption. Retrieved at https://osf.io/preprints/socarxiv/d3wpk/