Measuring disruptiveness and continuity of research by using the Disruption Index (DI) – A Bayesian statistical approach

Rüdiger Mutz[*] and Lutz Bornmann[**]

[*]*ruediger.mutz@uzh.ch*
ORCID 0000-0003-3345-6090
Center for Higher Education and Science Studies, University of Zurich, Switzerland

[**]*lutz.bornmann@gv.mpg.de*
Policy and Strategy Department, Administrative Headquarters of the Max Planck Society, Germany

Abstract

L. Wu, Wang, and Evans (2019) introduced the disruption index (DI) which has been designed to capture disruptiveness of individual publications based on dynamic citation networks of publications. In this study, we propose a statistical modelling approach to tackle open questions with the DI: (1) how to consider uncertainty in the calculation of DI values, (2) how to aggregate DI values for paper sets, (3) how to predict DI values using covariates, and (4) how to unambiguously classify papers into either disruptive or not disruptive. A Bayesian multilevel logistic approach is suggested that extends an approach of Figueiredo and Andrade (2019). A reanalysis of sample data from Bornmann and Tekles (2021) and Bittmann, Tekles, and Bornmann (2022) shows that the Bayesian approach is helpful in tackling the open questions. For example, the modelling approach is able to predict disruptive papers (milestone papers in physics) in a good way.

1. Introduction

Bibliometrics is one of the most frequently used method of research evaluation. Most users of bibliometrics are not aware yet that advanced metrics have been proposed, for example, to measure citation impact field- and time-normalized or to measure interdisciplinarity, internationality or novelty of research. The most recent development in bibliometrics are approaches to measure originality, novelty, disruptiveness, innovativeness, and similar characteristics of research (Winnink, Tijssen, & van Raan, 2016). L. Wu et al. (2019) introduced in this context the disruption index (DI) which has been designed to capture disruptiveness (and continuity) of individual publications based on dynamic citation networks of publications. The initial idea for the indicator has been proposed for patent data (Funk & Owen-Smith, 2017). L. Wu et al. (2019) transferred the idea to bibliometrics. The DI has experienced great popularity in the scientific community, since the identification of disruptiveness is highly regarded in the recognition system of science (L. Wu et al., 2019). In a widely recognized study, Park, Leahey, and Funk (2023) analyzed the DI of 45 million papers from several databases and found that papers are becoming less disruptive in the history of science.

Although Park et al. (2023) tested the reliability and validity of the indicator, the indicator has not been without criticism in recent years. For example, L. Wu et al. (2019) found that the effect of one parameter (N_R) in the DI formula (see the formula in Figure 1) is contradictory to what is being measured by the index (disruptiveness). Other possible limitations of the index are as follows:

First, probabilities enter the DI calculation as statistical quantities whose measurement accuracy or precision in turn depends on the number of papers (e.g., by the consideration of N_R in the formula, see Figure 1) used for the calculation. Thus, older focal papers (FPs) can be expected to have a higher DI precision than younger FPs (S. Wu & Wu, 2019). Older disruptive papers tend to have a higher number of citing papers than younger disruptive papers. The higher the

precision, the more robust the DI is to dynamic changes in time. Such aspects of measurement accuracy or standard error should be considered in the calculation of DI values. The statistical distribution of DI values or their basic parameters across all publications in a set provides important additional information that could improve the DI estimation for a focal paper. In an empirical Bayes estimate, two statistical estimators of the DI or the underlying probabilities appear. If all information for a FP is available, the DI can be estimated. If sparse data are available instead, the sample mean is still the best estimate for the DI. In the end, the best DI estimate is a weighted sum of both estimators (estimated DI and sample mean). The used weight is denoted as precision. If there is a large number of citing and cited papers for calculating the DI, its precision is high, and the estimate is largely determined by the citation information of the FP. However, if the precision is low, the overall mean of the sample is weighted more heavily. A shrinkage of the DI value toward the sample mean could happen in the case of low precision.

Second, covariates may affect the DI value, e.g., the scientific field of the FP or its number of authors. We know that citations depend on many factors (Tahamtan & Bornmann, 2019). It is therefore interesting to determine the influence of certain covariates and to predict DI values.

Third, it would be interesting not only to calculate the DI, but also to determine at the same time whether a FP can be denoted as disruptive or not. The DI value does not provide clear labels whether the FP is disruptive or not.

The aim of this paper is to tackle limitations of (open questions with) the DI by using a statistical Bayesian approach to estimate DI values as a statistical parameter of a sample. Our approach extends the approach of Figueiredo and Andrade (2019). The authors proposed a Bayesian approach to estimate DI values for single FPs. We extend the approach for an entire sample (a set of publications). Our approach is illustrated with a reanalysis of publication data from Bornmann and Tekles (2021) and Bittmann, Tekles, and Bornmann (2022) regarding the two journals *Physical Review Letters* and *Physical Review E*. The data are especially interesting, since assessments by experts are available whether the publications in the sets can be denoted as landmark papers (that are disruptive in all likelihood) or not.

2. Methodological Approach

Disruptiveness of a FP is defined by L. Wu et al. (2019) as a weighted sum index DI (see **Figure 1**), where "the difference between the number of papers citing the FP without citing any of its cited references (N_i) and the number of papers citing both the FP and at least one of its cited references (N^1_j) is divided by the sum of N_i, N^1_j and N_k. N_k is the number of papers citing at least one of the FP's cited references without citing FP itself" (Bornmann, Devarakonda, Tekles, & Chacko, 2020a, pp. 1244-1245). A FP can be disruptive (DI=1), developing (DI=-1) or neutral (DI=0).

Figueiredo and Andrade (2019) developed a Bayesian model to statistically estimate DI for a single FP, which refers to the original DI concept suggested by Funk and Owen-Smith (2017). As in L. Wu et al. (2019) DI is defined as difference in proportions or probabilities (Eq. 1, **Figure 1**):

$$DI = \frac{N_i - N^1_j}{N_i + N^1_j + N_k} = \frac{N_i}{N_i + N^1_j + N_k} - \frac{N^1_j}{N_i + N^1_j + N_k} = p_i - p_j \qquad (1)$$
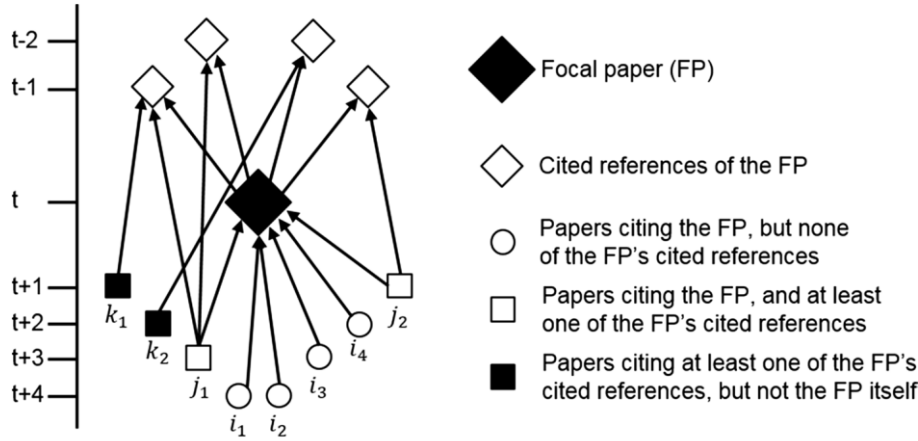
$$DI_l = \frac{N_i - N_j^l}{N_i + N_j^l + N_k}$$

$$N_i = |\{i_1, i_2, \dots\}| \qquad N_j^l = |\{j_m | BC(j_m, FP) \geq l\}| \qquad N_k = |\{k_1, k_2, \dots\}|$$

$$BC(p_1, p_2) = No. \, of \, bibliographic \, coupling \, links \, between \, p_1 \, and \, p_2$$

Figure 1. Definition of the Disruptive Index (DI) (Bornmann, Devarakonda, Tekles, & Chacko, 2020b, p. 1150)

In order for the probabilities to add up to 1.00, a third probability p_k is required ($p_k = 1 - p_i - p_j$), which represents the proportion of papers citing at least one of the FPs' cited references without citing the FP itself ("unused"). Figueiredo and Andrade (2019, p. 3) assumed for a *single FP* that the probabilities are multinomial distributed with $N = N_i + N_j^1 + N_k$:

$$N_i, N_j^1, N_k \sim Multinomial(p_i, p_j, p_k, N) \qquad (2)$$

A multinomial distribution is a discrete distribution, which generalizes the binomial distribution to more than two categories.

We now extend the approach of Figueiredo and Andrade (2019) for a sample with r=1 to R FPs. Data are available on N_{ri}, N_{rj}, N_{rk} and N_r (= $N_{ri} + N_{rj} + N_{rk}$) for r FPs, which are combined into a matrix $\mathbf{Y}_{R*3}$, where $\mathbf{Y}$ is multinomially distributed, as follows:

$$\mathbf{Y} \sim Multinomial(\mathbf{p}_i, \mathbf{p}_j, \mathbf{p}_k, \mathbf{N}), \qquad (3)$$

where $\mathbf{p}_i$, $\mathbf{p}_j$, $\mathbf{p}_k$ and $\mathbf{N}$ are r-dimensional vectors.

We assume a Bayesian mixed-effects model (Mutz, 2022, p. 7407). The three probabilities or proportions can be expressed in logistic terms (Eq. 4).

$$p_{ri} = \frac{\exp(\eta_{ri})}{\exp(\eta_{ri}) + \exp(\eta_{rj}) + 1}$$

$$p_{rj} = \frac{\exp(\eta_{rj})}{\exp(\eta_{ri}) + \exp(\eta_{rj}) + 1}$$

$$p_{rk} = \frac{1}{\exp(\eta_{ri}) + \exp(\eta_{rj}) + 1}, \tag{4}$$

where η_{ri} and η_{rj} are random effects, which are normally distributed across R FPs with expected values μ_{i0} and μ_{j0} and standard deviations σ_{i0} and σ_{j0}:

$$\eta_{ri} \sim Normal(\mu_{i0}, \sigma_{i0})$$

$$\eta_{rj} \sim Normal(\mu_{j0}, \sigma_{j0})$$

The basic model can be extended by including covariates x_{lr} and x^*_{lr} within a regression approach:

$$\mu_{i0} = b_{i0} + \sum_{l=1}^{L} b_{il} x_{lr}$$

$$\mu_{j0} = b_{j0} + \sum_{l^*=1}^{L^*} b_{jl} x^*_{lr} \tag{5}$$

If FPs are assigned to publication sets (e.g., universities), another level can be added to the statistical model in sense of a two-level mixed-effects model (e.g., level 1: FPs, level 2: universities). This allows the estimation of DI values on aggregated levels, and comparisons on the aggregated level (e.g., universities) are possible.

In a Bayesian approach as opposed to a frequency approach, probabilities are interpreted as uncertainties rather than frequencies. In the Bayesian estimation process, initial uncertainties (priors) about parameters are (hopefully) reduced in the light of the data (posterior estimates). In the absence of prior information of the parameters, the following non-informative priors are chosen to represent maximal uncertainty about the parameter values in advance (Eq. 6):

$$b_{i0}, b_{ik} \sim Normal(0, \sigma_i = 10)$$

$$b_{j0}, b_{jk} \sim Normal(0, \sigma_j = 10) \tag{6}$$

Not only the degree of disruptiveness (DI_r) can be calculated for single FPs with credible intervals, but it is also possible to determine whether a FP is disruptive or not ($DI_{cat, r}$). Thus, probabilities can be transformed into z-values using a probit function. If the difference of the z-values is larger than a certain criterion (here 4), then the paper is disruptive (Eq. 6):

$$DI_{cat,r} = (probit(p_{ri}) - probit(p_{rj})) > 4 \tag{7}$$

3. Data and Methods

3.1. Data

Data from Bornmann and Tekles (2021) and Bittmann et al. (2022) were re-analysed. The datasets include all papers published in *Physical Review Letters* and *Physical Review E*. Editors of both journals identified several milestone papers. In the data analyses of this study, papers (with the document type "article") published between 1980 and 2002 (*Physical Review Letters*) and between 1993 and 2004 (*Physical Review E*) were included as FPs. The datasets consist of 44,806 papers published in *Physical Review Letters* and 22,084 papers published in *Physical Review E*.

For each paper in the sets, the necessary information for calculating the DI (see Eq. 1) was available. The status of papers as milestone paper or not (see Bornmann & Tekles, 2021, p. 2, for further information) can be used in the study to investigate whether (qualitative) judgements by experienced editors correspond with the results of quantitative analyses (i.e., DI values).

3.2. Statistical Methods

Bayesian models can be realized in different software applications, e.g., with the R-package R2WinBUGS (Sturtz, Ligges, & Gelman, 2005). We used the procedure PROC MCMC from SAS (SAS Institute Inc., 2018) with a Metropolis-Algorithm. To speed up the estimation process, the samples were divided into 20 random subsamples to estimate the DI parameters.

3. Results

In the *first step* of the statistical analyses, the distribution of the estimated DI values for the two journals were visually inspected (**Figure 2**, **Figure 3**). The estimated DI values (black curve) for a random sample of about 10% of all publications of each of the two journals with 95% credible intervals were depicted.
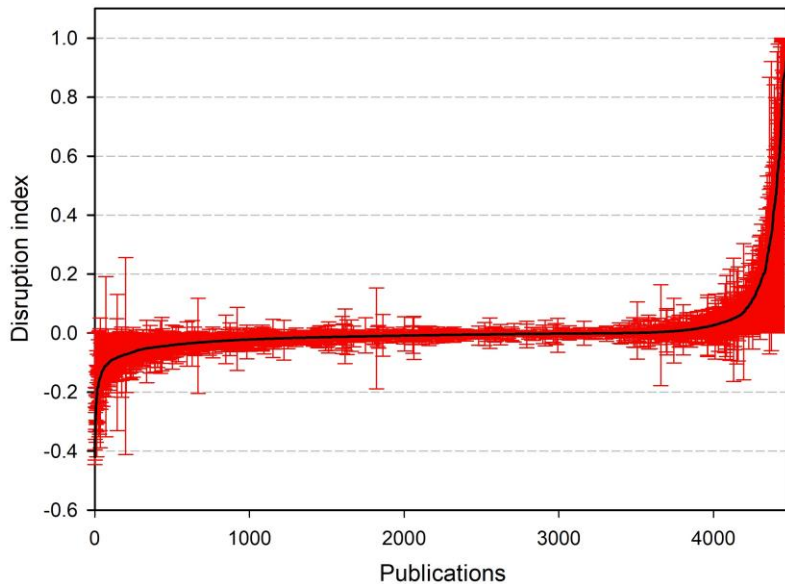

Figure 2.   Bayesian estimated DI values (**black line**) for a random sample of 10% of all publications of *Physical Review Letters* with 95% credible interval (red error bar), sorted by increasing DI values.
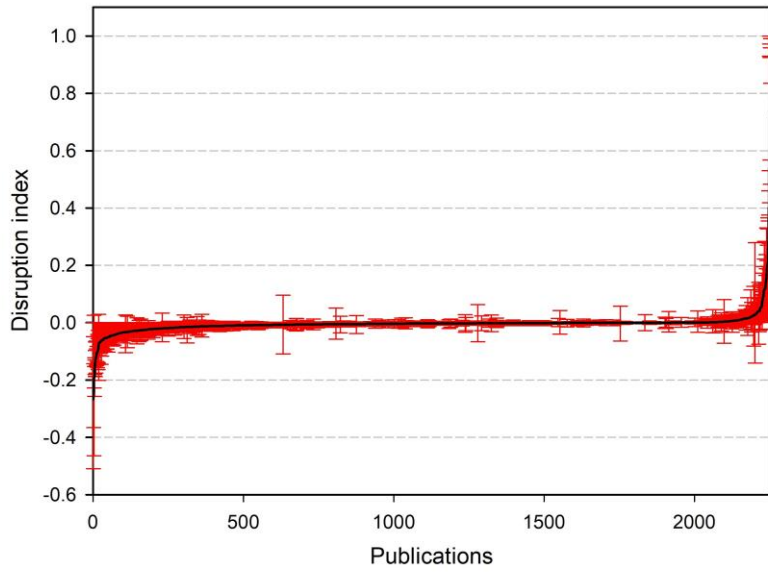
Figure 3. Bayesian estimated DI values (**black line**) for a random sample of 10% of all publications of *Physical Review E* with 95% credible interval (red error bar), sorted by increasing DI values.

Both figures reveal the following results:

- "Disruptive" papers (DI→1) are very rare events, and most papers are "neutral" papers (DI=0). L. Wu et al. (2019, p. 379) suggested to denote papers which are neither "disruptive" nor "developing" as "neutral". For both journals, papers are missing that can be clearly identified as "developing" papers (DI→-1). Only a value of -.60 is reached as minimum.
- The credible interval indicates that high uncertainty about a DI value is not necessarily unique to the "neutral" papers, but also occurs in "disruptive" or "developing" papers.
- The credible intervals and thus the uncertainty about DI are slightly higher on average for the journal *Physical Review Letters* than for the journal *Physical Review E*. Overall, the sample sizes N_r (= N_{ri} + N_{rj} + N_{rk}) are on the average lower for *Physical Review Letters* (N_{median}=1754) with higher uncertainty of DI than for *Physical Review E* (N_{median}=2452).

In the *second step* of the statistical analyses, dichotomized calculated DI_{cat} (DI>.90) values were cross-tabulated with Bayesian estimated DI_{cat} values (see Eq. 7) (**Table 1**). Regarding the degree of agreement (calculated with Cramer`s V), there is a high but not fully perfect correspondence between the calculated and the estimated indicator based on the same data. For *Physical Review Letters*, the agreement is very high (Cohen`s Kappa=.98). The proportion of disruptive papers is higher in *Physical Review Letters* (2.3%) than in *Physical Review* E (0.13%). While 96.6% (1,022/1,058) of the *Physical Review Letters* papers that were classified as disruptive (according to the calculated DI_{cat}) can be denoted as also disruptive according to the Bayesian estimated DI_{cat}, the same applies to only 60.8% (28/46) of the *Physical Review E* papers. The Pearson-Bravais-correlation between the continuous versions of the DI values (that were calculated versus Bayesian estimations) amounts to .99 for *Physical Review Letters* and .94 for *Physical Review Letters*. Thus, the results reveal that the Bayesian estimations of DI (which takes into account both the distribution and the sample sizes of the probabilities) are similar but not quite identical to the calculated DI_{cat}.

Table 1. Cross table of calculated DI_{cat} with Bayesian estimated DI_{cat} (0=non-disruptive, 1=disruptive paper)

Frequency Column %		Bayesian estimated DI_{cat} (0/1)					
		Physical Review Letters			*Physical Review E*		
		0	1	Total	0	1	Total
Calculated DI_{cat}	0	43,741	7	43,748	22,038	0	22,038
		99.92%	0.68%	97.64%	99.92%	0.00%	99.79%
	1	36	1,022	1,058	18	28	46
		0.08%	99.32%	2.36%	0.08%	100%	0.21%
	Total	43,777	1,029	44,806	22,056	28	22,084
		97.70%	2.30%	(100%)	99.87%	0.13%	100%
	Cramer`s V		.98			.78	

Table 2. Proportion of detected milestone papers by the calculated DI_{cat} in comparison to the Bayesian estimated DI_{cat}.

	Physical Review Letters			*Physical Review E*		
	Milestone papers	Number of detected papers	Proportion of detected papers	Milestone papers	Number of detected papers	Proportion of detected papers
Calculated DI_{cat}	39	2	5.13%	21	0	0%
Bayesian estimated DI_{cat}	39	2	5.13%	21	0	0%

In the *third step* of the statistical analyses, a logistic regression was conducted for *Physical Review Letters* to predict the milestone status of a FP from the editor (**Table 2**). Odds ratios for the disrupted papers were calculated. For calculated and estimated DI, the odds ratio deviates statistically significant from 1.0 (=no effect). The odds ratio for disrupted papers of 8.22 for the Bayesian estimated DI was higher than the odds ratio of 6.70 for the calculated DI. For *Physical Review E*, the difference was even higher: the odds ratio for the calculated DI amounted to 6.72; the odds ratio for the Bayesian estimated DI amounted to 17.11. As disruptiveness increases the odds of milestone papers occurring is more than twice as high for the Bayesian estimated DI than for the calculated DI in terms of a one unit change in DI.

4. Discussion

In this study, we dealt with several open questions regarding the DI. There is the danger in quantitative research evaluation that the DI is applied in evaluation studies without clarifying open questions with the indicator itself. This study was intended to target some of these questions, such as how to consider uncertainty in the calculation of DI values.

A Bayesian multilevel logistic approach was suggested in this study to address the open questions. The approach considers not only the variability of DI values in the sample as additional information, but also the different sample sizes (e.g., N_{ri}, N_{rj}) to calculate the probabilities. A DI value for each FP can be estimated as well as a credible interval. The credible interval can be used to assess the stability of the DI values over time or to compare two publications for "significant" differences in DI values. The approach also allows to include covariates.

Our empirical results (based on the papers published in two physical journals) show that the Bayesian estimates of the DI values are not redundant with the DI values calculated from the data: we did not find perfect correlations. We also tested in this study whether DI values are

able to identify milestone papers in the paper sets. In terms of the categorical milestone paper prediction (yes or no), the Bayesian estimator did not outperform the calculated DI values. The Bayesian variant is superior, however, to the continuous DI values.

References

Bittmann, F., Tekles, A., & Bornmann, L. (2022). Applied usage and performance of statistical matching in bibliometrics: The comparison of milestone and regular papers with multiple measurements of disruptiveness as an empirical example. *Quantitative Science Studies, 2*(4), 1246-1270. doi:10.1162/qss_a_00158

Bornmann, L., Devarakonda, S., Tekles, A., & Chacko, G. (2020a). Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers. *Quantitative Science Studies, 1*(3), 1242-1259. doi:10.1162/qss_a_00068

Bornmann, L., Devarakonda, S., Tekles, A., & Chacko, G. (2020b). 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*(2), 1149-1155. doi: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). doi:10.1016/j.joi.2021.101159

Figueiredo, F., & Andrade, N. (2019). *Quantifying disruptive influence in the allmusic guide.* Paper presented at the Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019.

Funk, R. J., & Owen-Smith, J. (2017). A Dynamic Network Measure of Technological Change. *Management Science, 63*(3), 791-817. doi:10.1287/mnsc.2015.2366

Mutz, R. (2022). Diversity and interdisciplinarity: Should variety, balance and disparity be combined as a product or better as a sum? An information-theoretical and statistical estimation approach. *Scientometrics, 127*(12), 7397-7414. doi:10.1007/s11192-022-04336-3

Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature, 613*, 138–144.

SAS Institute Inc. (2018). *SAS/STAT® 15.1 User's Guide*. Cary, NC: SAS Institute Inc.

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software, 12*, 1-16. doi:10.18637/jss.v012.i03

Tahamtan, I., & Bornmann, L. (2019). What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. *Scientometrics, 121*(3), 1635–1684. doi:10.1007/s11192-019-03243-4

Winnink, J. J., Tijssen, R. J. W., & van Raan, A. F. J. (2016). Breakout discoveries in science: What do they have in common? In I. Ràfols, J. Molas-Gallart, E. Castro-Martínez, & R. Woolley (Eds.), *Proceedings of the 21. International Conference on Science and Technology Indicator*. València, Spain: Universitat Politècnica de València.

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature, 566*(7744), 378-382. doi:10.1038/s41586-019-0941-9

Wu, S., & Wu, Q. (2019). A confusing definition of disruption. Retrieved from https://osf.io/preprints/socarxiv/d3wpk/