Journal of Informetrics 19 (2025) 101678

Contents lists available at ScienceDirect

Journal of Informetrics

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

Team size and its negative impact on the disruption index

A B S T R A C T

As science transitions from the age of lone geniuses to an era of collaborative teams, the question of whether large teams can sustain the creativity of individuals and continue driving innovation has become increasingly important. Our previous research first revealed a negative relationship between team size and the Disruption Index—a citation network-based metric of innovation—by analyzing 65 million projects across papers, patents, and software over half a century. This work has sparked lively debates within the scientific community about the robustness of the Disruption Index in capturing the impact of team size on innovation. Here, we present additional evidence that the negative link between team size and disruption holds, even when accounting for factors such as reference length, citation impact, and historical time. We further show how a narrow 5-year window for measuring disruption can misrepresent this relationship as positive, underestimating the long-term disruptive potential of small teams. Like "sleeping beauties," small teams need a decade or more to reveal their transformative contributions to science.

1. Main

As science transitions from the age of lone geniuses to an era of collaborative teams, a critical question arises: Can large teams sustain individual creativity and continue driving innovation? Our previous research (Wu et al., 2019) was the first to uncover a negative relationship between team size and innovation performance, measured by the Disruption Index—a network-based metric of innovation—through an analysis of over 65 million papers, patents, and software products, spanning 1954 to 2014. We further demonstrated that this negative relationship is remarkably robust across thirteen controlled variables for research articles, summarized below (analyses for patents and software are omitted, as this study focuses on research articles):

Recent research reported a positive marginal effect of team size on the D-index while accounting for a set of control variables (Petersen et al., 2024b, 2024a), contrary to our findings (Wu et al., 2019). This raises questions about whether the negative relationship we observed could be attributed to a specific combination of confounders not tested in our 2019 study. One concern raised in Petersen et al. relates to self-citations, which they suggest might inflate the observed disruption scores. However, our 2019 study already addressed this issue by explicitly removing self-citations, a control summarized in Table 1 (Index 11). Given this, we focus on a more consequential difference between our study and Petersen et al.: the choice of the citation window.

Upon closely examining their model, we find that the key difference lies not in the confounders but in the dependent variable—the D-index. While we used the longest available citation window for the D-index across all papers in our dataset, their reliance on a shorter 5-year window likely introduces bias. We explain this in detail below.

Small teams act as "sleeping beauties," requiring more time to accumulate citations compared to large teams. Extended Data Fig. 7 in Wu et al. (2019) illustrates the difference in long-term citation dynamics between small and large teams and demonstrates a positive correlation between the Disruption Index and the Sleeping Beauty Index (Ke et al. 2015). Since the D-index stabilizes only once a paper stops receiving citations, using a short citation window overestimates the disruptive impact of large teams while underestimating it for small teams. This leads to a misleading observation of a positive effect of team size on the Disruption Index. In the text below, we first replicate their results and then demonstrate how the negative relationship between team size and the Disruption Index reappears as the citation window length increases.

The same authors published two models (Petersen et al., 2024b, 2024a), but their specifications and sample selection criteria appear to vary across studies. We begin by revisiting their first model on the team size effect, denoted as Eq. 3 in Petersen et al. (2024b), which we cite and re-index below to assess its consistency and suitability for evaluating the relationship between team size and disruptiveness.

https://doi.org/10.1016/j.joi.2025.101678
Received 6 December 2024; Received in revised form 10 March 2025; Accepted 29 April 2025
Available online 13 May 2025
1751-1577/

Table 1
Robustness Tests for the Negative Relationship Between Team Size and the Disruption Index.

Index	Controlled Variable	Explanation	Results in Wu et al. 2019
1	Citation Impacts	High-impact vs. low-impact articles, divided into six percentile groups	Main Fig. 3
2	Academic Disciplines	Nine disciplines, including Physical Sciences, Biology, Medicine, and others	Main Fig. 3; Method: "Fields, subfields, and journals of WOS papers"
3	Publication Years	Articles spanning 1954–2014	Main Fig. 3
4	Authors	Numeric index for disambiguated scholars among 38,000,470 unique authors, controlled using fixed-effect regression models	Main Fig. 3; Supplementary Information Table 4
5	Article topics	100-dimensional vectors generated using a doc2vec model (Gensim Python library)	Method: "Modeling topics of WOS papers using Doc2vec"
6	Time Periods	Five distinct historical decades	Extended Data Fig. 3
7	Time Window Length	Five distinct disruption measurement time windows	Extended Data Fig. 3
8	Academic Journals	15,146 academic journals in the Web of Science	Extended Data Fig. 4
9	Article Types I	1502 theoretical articles vs. 2756 methodological articles	Extended Data Fig. 5
10	Article Types II	22,672 reviewing articles vs. 1,338,808 reviewed articles	Extended Data Fig. 5
11	Self-citation removal	Citations referencing papers with at least one common author are removed (account for 10.2 % of all 615,697,434 citations)	Extended Data Fig. 5; Method: "Removing self-citations from WOS papers"
12	Alternative Metrics	Five alternative definitions of disruption	Extended Data Fig. 5
13	Awards	877 Nobel Prize-winning papers compared to a control group	Extended Data Fig. 10

$$D_{p,5} = b_0 + b_k lnk_p + b_r lnr_p + b_c lnc_p + D_t + \epsilon_t \quad (1)$$

Here, $D_{p,5}$ represents the Disruption Index of paper p, calculated using a five-year citation window; k_p is the team size (number of coauthors); r_p is the reference length; and c_p is the number of citations. D_t denotes the yearly fixed effects of the Disruption Index. The independent variables were modeled using a logarithmic transform to address their right-skewed distributions.

To fit this model, we use the same dataset—the archived version of Microsoft Academic Graph (now integrated into OpenAlex)—as Petersen et al., and apply the same selection criteria: $1 \le k_p \le 10$ for coauthors, $5 \le r_p \le 50$ for references, and $10 \le c_p \le 1000$ for citations. While Petersen et al. calculated the 5-year disruption index for all 3 million papers published between 1990 and 2009 that met these criteria, we selected six cohorts of papers, totaling 1.9 million, published in 1995, 2000, 2010, 2015, 2017, and 2019. This selection allows us to calculate the disruption index for these papers up to 2020, representing citation windows of 25, 20, 10, 5, 3, and 1 year. We chose this approach because we hypothesize that citation windows moderate the relationship between team size and disruption index and aim to test this hypothesis.

Like Petersen et al., we performed ordinary least squares (OLS) estimation using the same model, dataset criteria, and software. We successfully replicated their Fig. 2 g (Petersen et al., 2024b), which shows a positive marginal effect of team size on the 5-year D-index (Fig. 1c; fixed effects were not applied because each regression is limited to a single publication year, naturally accounting for temporal variation).

However, this result appears to be influenced by the use of a short citation window. When applying the same regression model to other cohorts, we observe that the negative impact of team size on the Disruption Index clearly re-emerges as the citation window increases, with the turning point occurring at a 10-year citation window (Fig. 1a-f).

Next, we re-analyze their second model, which includes additional nonlinear effects and is denoted as Eq. 5 in Petersen et al. (2024a), cited and re-indexed below:

$$D_{p,5} = b_t + b_r lnr_p + b_c lnc_p + b_{r2}\left(lnr_p\right)^2 + b_{c2}\left(lnc_p\right)^2 + \gamma_k + \epsilon_t \quad (2)$$

Building on the variables in Eq. 1—k_p for team size, r_p for reference length, and c_p for the number of citations—Eq. (2) introduces two key changes. First, it adds quadratic terms for r_p and c_p to better capture their nonlinear impacts on the D-index. Second, it represents team size as a set of dummy variables (e.g., whether the team size is 3, 4, or 5), denoted by γ_k, where k indexes team size groups, to more accurately estimate the marginal effect of team size on the D-index at each value.

To fit this model, we apply the same selection criteria as in Petersen et al. (2024a) to select papers from the Microsoft Academic Graph: $1 \le k_p \le 25$ for coauthors, $10 \le r_p \le 200$ for references, and $1 \le c_p \le 1000$ for citations, and re-run ordinary least squares (OLS) estimation using the STATA 13.0, focusing on two cohorts of papers, 2015 and 2000, corresponding to 5-year and 20-year citation windows, respectively.

Despite the added complexity of this model, we find that the pattern observed in our previous analysis remains consistent. Specifically, while we successfully replicate the positive effect of team size, which appears biased by the use of a short citation window (Fig. 2a)—the negative impact of team size on the Disruption Index re-emerges as the citation window lengthens (Fig. 2b). To further validate the robustness of our findings, we conduct an additional analysis using the same cohort of 676,831 papers published in 2005, comparing the five-year and twenty-year versions of the Disruption Index. This cohort-based analysis produces results nearly identical to Fig. 2a and 2b, confirming that the observed pattern holds and reinforcing that the observed effect is not an artifact of sample composition differences.

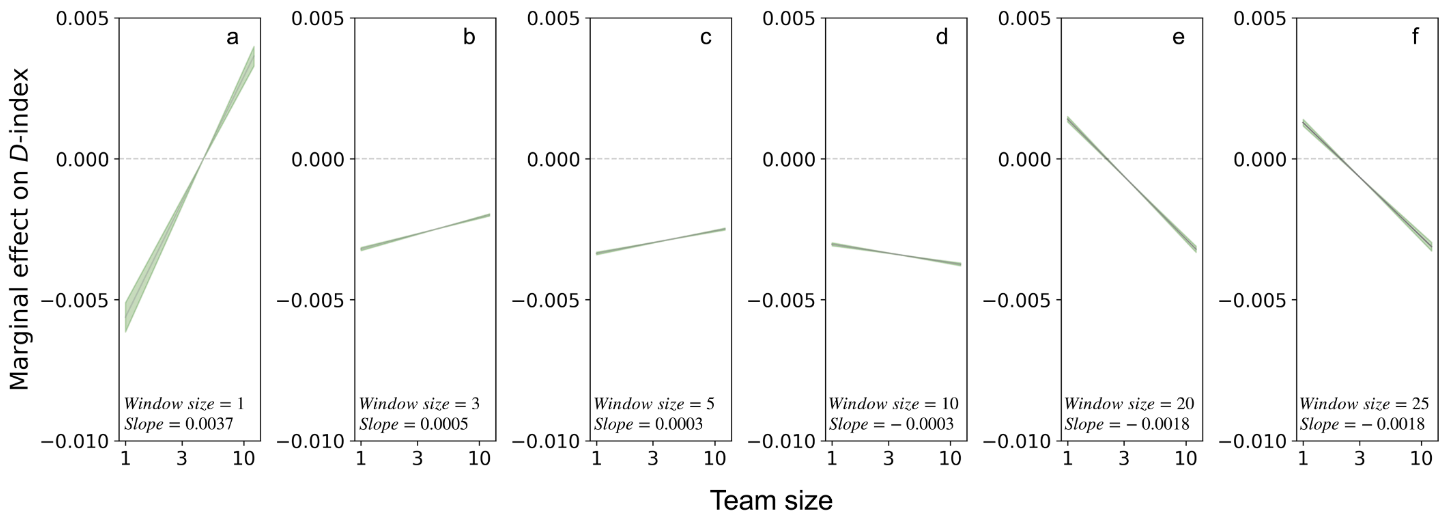

Fig. 1. The negative impact of team size on the D-index reappears over longer citation windows. To examine how citation window length moderates the relationship between team size and the D-index, we analyzed six annual cohorts of papers, each receiving citations from subsequent papers published through 2020. Our dataset includes 47,129 papers from 2019, 271,496 from 2017, 444,675 from 2015, 536,463 from 2010, 344,582 from 2000, and 226,358 from 1995, corresponding to citation windows of 1, 3, 5, 10, 20, and 25 years, respectively. The regression coefficients (slopes) estimated from Eq. (1) are presented, with marginal effects calculated while holding all other covariates at their mean values. Shaded green confidence intervals are shown around the regression lines.

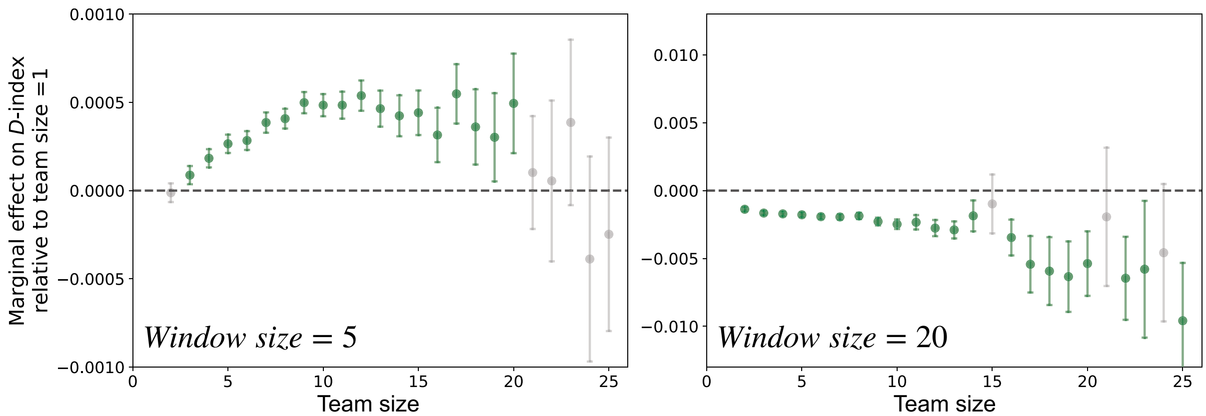

Fig. 2. The negative impact of team size on the D-index reappears with long-term citations, with team sizes modeled as dummy variables. To examine how citation window length moderates the relationship between team size and the D-index, we analyzed two annual cohorts of papers, each receiving citations from subsequent papers published through 2020. Our dataset includes 1,436,276 papers from 2015 to 459,975 from 2000, corresponding to citation windows of 5 and 20 years, respectively. Dots represent the marginal effects of team size, calculated with all other covariates held at their mean values. Bars indicate 95 % confidence intervals, with gray error bars and gray dots denoting results not statistically different from the baseline level ($p > 0.05$), as indicated by the horizontal dashed line.

Based on these analyses across two models (Petersen et al., 2024b, 2024a), the moderating effect of citation window size is clear. Therefore, we did not attempt to replicate nuanced versions of the D-index with a 5-year window, such as the year-journal normalized D-index from Petersen et al. (2024a). We do not anticipate significant differences.

In sum, the Disruption Index (D-index) evolves over time as the focal paper and its references accumulate citations, stabilizing only after citation growth ceases. Since small teams are like "sleeping beauties," taking longer to gather citations compared to their large counterparts, a short citation window provides an incomplete picture of the relationship between team size and the D-index, potentially overemphasizing the impact of large teams while underestimating that of small teams (Leahey et al., 2023).

Our findings reaffirm that the negative relationship between team size and the D-index holds when longer citation windows are applied, with the effect stabilizing at windows of ten years or more (Bornmann & Tekles, 2019; Lin et al., 2022). This underscores the importance of selecting an appropriate citation window for accurate measurements of innovation. Furthermore, the effect of citation window size is not merely a technical issue. The finding that small teams often require a decade or more to reveal their transformative contributions to science underscores the need for science policy and institutional changes to better support small-team science for disruptive innovation. These additional analyses further reinforce the conclusions of our 2019 study on the important role of small teams in driving innovation (Wu et al., 2019).

CRediT authorship contribution statement

Yiling Lin: Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Linzhuo Li:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Lingfei Wu:** Writing – original draft, Supervision, Methodology, Data curation, Conceptualization.

Acknowledgments

We are grateful for support from the National Science Foundation grant SOS: DCI 2239418 (L.W.).

References

Bornmann, Lutz, & Tekles, Alexander (2019). Disruption index depends on length of citation window. *Prof. Inf., 28*(2), Article e280207.
Ke, Qing, Ferrara, Emilio, Radicchi, Filippo, & Flammini, Alessandro (2015). Defining and identifying sleeping beauties in science. *Proc. Natl. Acad. Sci. U S. A, 112*(24), 7426–7431.
Leahey, Erin, Lee, Jina, & Funk, Russell J. (2023). What types of novelty are most disruptive? *Am. Sociol. Rev, 88*(3), 562–597.
Lin, Yiling, Evans, James A., & Wu, Lingfei (2022). New directions in science emerge from disconnection and discord. *J. Inf., 16*(1), Article 101234.
Petersen, Alexander Michael, Arroyave, Felber J., & Pammolli, Fabio (2024a). The disruption index suffers from citation inflation: Re-analysis of temporal CD trend and relationship with team size reveal discrepancies. *J. Inf., 19*(1), Article 101605.

Petersen, Alexander Michael, Arroyave, Felber, & Pammolli, Fabio (2024b). The disruption index is biased by citation inflation. *Quant. Sci. Stud*.
Wu, Lingfei, Wang, Dashun, & Evans, James A. (2019). Large teams develop and small teams disrupt science and technology. *Nature, 566*(7744), 378–382.

Yiling Lin[a], Linzhuo Li[b,*], Lingfei Wu[a,*]
[a] *School of Computing and Information, The University of Pittsburgh, 135 N Bellefield Ave, Pittsburgh, PA 15213, USA*
[b] *Department of Sociology, Zhejiang University, 866 Yuhangtang Road, Hangzhou, Zhejiang, 310058, , PR China*

[*] Corresponding authors.
E-mail address: liw105@pitt.edu (L. Wu).