Online Resource 1: Supplementary Information

Topic-specific public-facing attention and inequality in astronomy:
Comparisons with citations and stated priorities

Journal: *Scientometrics*

Authors: Jiwon Park[1] and Yonggi Kim[2] (corresponding)
[1]Department of Astronomy and Astrophysics, Chungbuk National University, Cheongju-si, Republic of Korea
[2]Chungbuk Pro Maker Center, Chungbuk National University, Cheongju-si, Republic of Korea
Corresponding e-mail: ykkim153@cbnu.ac.kr

This document provides the corpus and coverage statistics, the priority codings, and the full robustness analyses referenced in the main text. All quantities are computed on the real source data (ADS metadata retrieved June 2026; Altmetric Explorer export dated 12 June 2026; UAT release current as of June 2026). Section S7 describes the reproducibility package (Online Resource 3), which regenerates every number and figure below. Tables and figures are numbered S1, S2, ... sequentially.

S1 Corpus and journal list

Table S1 lists the core refereed journals of each Science Mission Directorate (SMD) division and their article counts (2013–2022); the counts sum exactly to the 144,973-DOI corpus and the 99,083 astrophysics papers. Figure S1 shows the selection cascade. Because membership is assigned by journal, the corpus is an SMD-aligned, field-level sample of the international literature, not a census of NASA-funded research.

Table S1: Core refereed journals by SMD division and article counts, 2013–2022 (ADS bibstem).

Division	Journals (count)	Total
Astrophysics	MNRAS (35,138), ApJ (30,163), A&A (19,303), ApJL (6,283), AJ (4,760), ApJS (2,228), PASP (1,208)	99,083
Planetary	Icarus (4,131), JGR–Planets (1,843), M&PS (1,473), PSJ (601)	8,048
Heliophysics	JGR–Space Phys. (7,419), Solar Phys. (1,977), Space Weather (1,186)	10,582
Earth science	GRL (14,902), JGR–Atmos. (7,979), Remote Sens. Environ. (4,379)	27,260
Corpus total		144,973

S2 Coverage and representativeness

Keywords are resolved to the Unified Astronomy Thesaurus (2,312 concepts, polyhierarchy, eleven top-level categories) by mapping numeric identifiers and labels (including alternative labels) to canonical concepts. Table S2 compares UAT-tagged and untagged astrophysics papers across the full citation distribution, not only the mean: the groups are near-identical except in recency. Table S3 shows that coverage rises sharply after the 2019–2020 UAT adoption, which motivates the year fixed effects and the 2019–2022 and AAS-only sensitivity analyses. Table S4 shows that within-year topic shares track the pooled shares closely; Table S5 gives journal-level tagging rates, which are high for AAS journals and lower for the non-AAS venues.

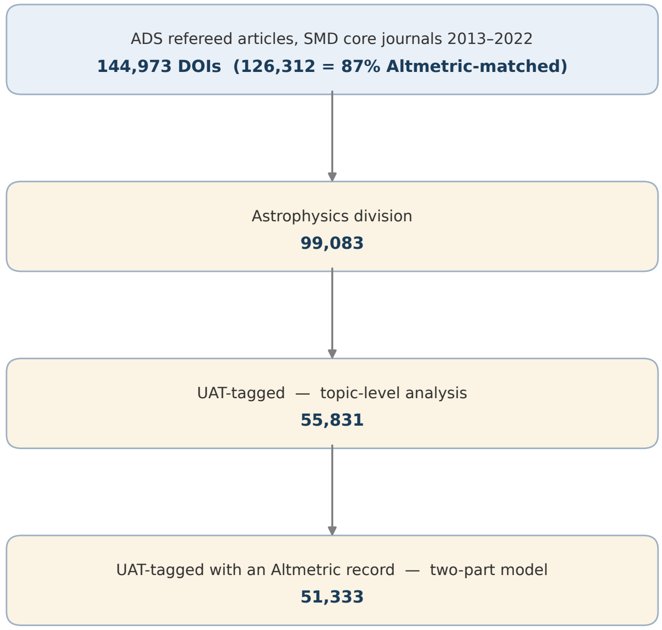

Figure S1: Selection cascade. The astrophysics analysis is a strict subset funnel; the Altmetric-match rate (87%) is a corpus-wide attention-coverage fact reported in the top box.

S3 Decadal Survey priority codings

Table S6 gives the ordinal tiers for both surveys under the single explicit rule of Section 3.5 of the main text. Tier 3 is assigned to the topics named by a 2010 science-frontier discovery area (Cosmic Dawn → galactic/extragalactic; New Worlds → exoplanet; Physics of the Universe → cosmology and high-energy). Heliophysics is out of scope (a separate Solar and Space Physics Decadal Survey). The two tier vectors correlate at $\rho = 0.93$. The coding was performed by the author against the published documents using this rule; the codebook and per-topic scores are in the reproducibility package. No independent inter-coder reliability is claimed (main-text Section 7, limitation 5).

S4 Attention sources and fine concepts

Table S7 reports total mentions by source. The public-facing composite (news, blogs, Wikipedia, Reddit, Facebook) is 40.6% of all mentions; microblogging (X, Bluesky) dominates the remainder. Table S8 lists the top fine concepts by each signal.

S5 Robustness analyses

Level, intensity, and compositional correlations. Table S9 shows that the attention–citation share correlation ($\rho = 0.71$) is a paper-volume artefact: it persists in the Pearson correlation between centred-log-ratio coordinates but vanishes at the per-paper intensity level and is more

Table S2: UAT-tagged versus untagged astrophysics papers. SMD = standardised mean difference; KS = Kolmogorov–Smirnov distance between the two citation distributions.

Characteristic	Tagged (n = 55,831)	Untagged (n = 43,252)	SMD / KS
Median publication year	2019	2017	SMD 0.38
Mean citations	46.0	45.3	SMD 0.00
Median citations	24	25	—
Citation IQR (Q1–Q3)	11–48	12–50	—
90th-percentile citations	93	94	—
Citation distribution	—		KS 0.02
Altmetric-match rate	91.9%	93.6%	—
Open-access rate	90.6%	92.1%	—
Any public-facing attention	62.0%	58.3%	—

Table S3: Share of astrophysics-division articles carrying at least one resolvable UAT concept, by publication year.

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
UAT-tagged (%)	48.1	48.2	47.5	48.1	47.8	49.1	52.4	70.8	72.6	72.1

than halved by conditioning on paper share. (We do not report the centred-log-ratio Spearman correlation, which equals the raw one by construction.)

Two-part model, per topic. Table S10 reports the coverage odds ratio and intensity rate ratio for every topic (matched papers, n = 51,333; adjusted for log citations, publication year, journal, open access). Topic-specific Altmetric-record rates (final column) vary from 99% (exoplanet) to 80% (heliophysics), which the record stage of the main-text cascade accounts for.

Per-source differential. Table S11 re-estimates the citation-adjusted exoplanet-versus-cosmology differential source by source, including X and Bluesky, with the same controls. It is broad-based—largest on X and Wikipedia, present on news and blogs, small on Reddit, Facebook, and Bluesky—and so is not confined to any single medium.

Topic-by-source contingency. A topic-by-source contingency test on fractional mention counts (11 topics × 6 channels) rejects independence (χ^2 = 11,150, df = 50), but at this volume that is near-automatic. The effect size is modest: Cramér's $V = 0.071$ (paper-bootstrap 95% CI $[0.065, 0.085]$, $B = 300$); the structure is concentrated in the topic–channel pairs of main-text Figure 4.

Exploratory dynamic panel. Table S12 reports the corrected topic-by-year growth model (integer primary-category counts; Poisson pseudo-maximum likelihood; exact permutation of the 11 topic-tier labels refitting the full model). Priority does not predict citation-intensity or output growth. The nominal attention-growth association is fragile: it is coding-sensitive, exoplanet-driven, and disappears in the AAS-only subset; a direct test that citation and attention growth differ is not significant. Figure S2 plots the per-topic growth slopes.

Table S4: Spearman correlation of within-year topic shares with the pooled 2013–2022 shares.

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Citation-share ρ	0.99	0.95	0.95	0.97	0.96	0.98	0.97	0.94	0.92	0.91
Attention-share ρ	0.88	0.86	0.92	0.90	0.79	0.85	0.79	0.71	0.74	0.94

Table S5: UAT tagging rate by astrophysics journal (share of articles with a resolvable UAT concept). AAS-published journals (ApJ, ApJL, ApJS, AJ, PASP) adopted the UAT earliest.

Journal	ApJL	ApJ	AJ	ApJS	PASP	MNRAS	A&A
Tagged (%)	78	76	71	68	63	44	41

Scope sensitivity. Table S13 shows that the (already non-significant) priority–intensity correlations weaken further as out-of-scope topics are removed.

Inequality by year and topic. Tables S14 and S15 show that the attention–citation concentration gap holds in every cohort and every topic.

Counterfactual coverage decomposition. Because the Gini coefficient is not additively decomposable, the contribution of differential coverage to the attention–citation gap is quantified counterfactually (Table S16): each signal is re-assigned the other's zero mass while its positive-value shape is held fixed. The two counterfactuals sum to more than the observed gap, which is precisely why an additive "x% coverage / y% intensity" split would be misleading.

Citation covariate: raw versus field-normalised. The primary two-part models condition on log(1 + raw ADS citations). Re-fitting both margins on ADS field-normalised citations leaves the topic contrasts essentially unchanged (Table S17), so the reported differential does not depend on which citation scale is used.

Figures S2, S3, S4, and S5 give the exploratory growth slopes, the single-equation specification curve, the multi-index inequality comparison, and the leave-one-out check.

S6 Cross-division comparison (descriptive)

Because divisional shares depend on the journal selection, this comparison is descriptive. Astrophysics accounts for the majority of papers, citations, and attention but a minority of the budget; planetary science holds the largest budget share (mission-cost driven) yet small shares of papers, citations, and attention (Figure S6a). Policy-document attention is almost entirely an Earth-science phenomenon (97.7% of all policy mentions; Figure S6b).

S7 Reproducibility package

The analysis is released as a documented package (Online Resource 3), to be deposited in Zenodo under an MIT licence upon acceptance and available to reviewers on request. It contains:

Table S6: Ordinal Decadal priority tiers by UAT top-level category (3 = science-frontier / flagship primary science; 2 = theme-core; 1 = cross-cutting; out = out of the astronomy survey's scope).

UAT top-level category	2010 tier	2020 tier	Rationale (2010 primary benchmark)
Cosmology	3	3	Physics of the Universe; WFIRST dark energy
Galactic & extragalactic	3	3	Cosmic Dawn (first galaxies, reionisation)
High-energy astrophysics	3	3	Physics of the Universe; LSST transients
Exoplanet astronomy	3	3	New Worlds science-frontier discovery area
Stellar astronomy	2	2	Star formation; Milky Way stellar populations
Interstellar medium	2	2	Theme-core (star-formation environments)
Astrophysical processes	2	2	Theme-core (physical mechanisms)
Interdisciplinary astronomy	2	2	Theme-core (methods, instrumentation science)
Observational astronomy	1	1	Survey infrastructure (cross-cutting)
Solar-system astronomy	1	1	Partly governed by the Planetary Science decadal
Heliophysics	out	out	Separate Solar & Space Physics decadal

Table S7: Total mentions by Altmetric source (UAT-tagged astrophysics; sources with zero mentions omitted). "% of composite" is the share within the five-source public-facing composite.

Source	Total mentions	% of public-facing composite
X (microblogging)	258,619	—
News	82,171	45.6
Blogs	43,665	24.2
Wikipedia	36,907	20.5
Reddit	9,222	5.1
Facebook	8,277	4.6
Q&A	2,103	—
Video	1,633	—
Bluesky (microblogging)	1,088	—
Public-facing composite	180,242	100.0

- `reproduce_aggregate.py` — regenerates the exact permutation tests, Bonett–Wright intervals, Kendall's τ, and the alternative-coding and scope correlations (main-text Tables 1–2; Tables S9, S13).
- `reproduce_from_realdata.py` — reconstructs the UAT roll-up (validated against a per-journal-year smoke test and the tagged-paper count of 55,831), and computes the topic shares and intensities, the two-part and single-equation differential models with marginal effects (main-text Tables 3–4; Tables S10, S11), the inequality indices (main-text Table 5; Tables S14–S15), and the exploratory Poisson-PML growth model with exact topic-label permutation (Table S12).
- `make_figures.py` — regenerates all main and supplementary figures from the deposited source data.
- `environment.yml`, `requirements.txt` — pin the versions (Python 3, numpy 1.26, pandas, scipy, statsmodels 0.14, sympy); a fixed random seed is set in every script.
- Aggregated data: the unrounded eleven-category share/intensity matrix; the topic-by-year panel; the 2010 and 2020 priority codings with per-topic scores and codebook; the attention-source totals; the topic-by-source matrix; the per-topic two-part estimates; and machine-

Table S8: Top twelve UAT fine concepts by fractional public-facing attention and by fractional citations.

Rank	Most attention	Most citations
1	Surveys	Surveys
2	Circumstellar matter	Accretion
3	Astrobiology	Dark matter
4	Black-hole physics	Gravitational waves
5	Accretion	Hydrodynamics
6	Dark matter	Black-hole physics
7	Gravitational waves	Cosmological parameters
8	Protoplanetary disks	Catalogs
9	Catalogs	Protoplanetary disks
10	Celestial mechanics	Radiative transfer
11	Brown dwarfs	Extinction
12	Astrometry	Cosmic background radiation

Table S9: Attention–citation association across representations ($n = 11$ topics).

Representation	Correlation
Topic shares (raw), Spearman	$\rho = 0.71$
Centred-log-ratio coordinates, Pearson	$r = 0.73$
Partial (conditioning on paper share), Spearman	$\rho = 0.14$
Per-paper intensities, Spearman	$\rho = 0.01$

readable source data for every table and figure.

- Documentation: the ADS query, the Altmetric export column schema and matching procedure, the UAT release identifier, and the primary-category rule.

Paper-level Altmetric records cannot be redistributed under Altmetric's terms of use; the specific derived fields used here are shared to the extent the licence permits, and the manuscript documents the export procedure (main-text Section 3.4). Because Altmetric counts change over time, a future export will not exactly reproduce the June 2026 snapshot; the archived aggregates therefore preserve every published statistic.

Table S10: Per-topic two-part attention margins relative to cosmology, and topic Altmetric-record rate.

Topic	Coverage OR (95% CI)	Intensity RR (95% CI)	Record rate
Exoplanet astronomy	2.53 (2.03–3.16)	3.01 (2.78–3.27)	98.7%
Stellar astronomy	1.75 (1.61–1.90)	1.55 (1.48–1.62)	95.1%
Solar-system astronomy	1.68 (1.49–1.90)	3.33 (3.12–3.56)	85.7%
Interstellar medium	1.37 (1.22–1.53)	1.31 (1.22–1.40)	93.5%
Observational astronomy	1.34 (1.24–1.46)	1.56 (1.48–1.63)	93.9%
Galactic & extragalactic	1.31 (1.21–1.41)	1.49 (1.42–1.56)	96.9%
High-energy astrophysics	1.25 (1.12–1.38)	1.33 (1.25–1.41)	95.3%
Astrophysical processes	1.13 (1.05–1.21)	1.22 (1.17–1.28)	92.1%
Interdisciplinary astronomy	1.09 (1.00–1.18)	1.48 (1.40–1.56)	83.6%
Cosmology (reference)	1.00	1.00	93.2%
Heliophysics	0.65 (0.59–0.72)	0.84 (0.79–0.90)	80.2%

Table S11: Citation-adjusted exoplanet-versus-cosmology attention differential and concentration, by source (controls: log citations, year, journal, open access).

Source	Exoplanet differential (95% CI)	Source Gini (tagged)
Public-facing composite	1.92 (1.78–2.07)	0.83
X (microblogging)	1.98 (1.86–2.10)	0.69
Wikipedia	1.68 (1.59–1.78)	0.95
News	1.47 (1.36–1.59)	0.96
Blogs	1.20 (1.16–1.25)	0.65
Reddit	1.08 (1.05–1.11)	0.97
Facebook	1.03 (1.01–1.05)	0.94
Bluesky (microblogging)	1.01 (1.00–1.02)	0.99

Table S12: Exploratory topic-by-year growth model: coefficient on (2010 priority tier $\times$ year-trend), Poisson PML with topic and year fixed effects and an exposure offset (11 topics $\times$ 10 years). Exact p from permuting topic-tier labels and refitting.

Outcome	$\exp(\beta)$ per tier-year	Exact perm. p	Clustered p	AAS-only exact p
Citation intensity	1.00	0.87	0.90	—
Public attention	1.03	0.04	0.04	0.15
Publication output	1.00	0.97	0.98	—

Direct contrast (citation-growth minus attention-growth slope) vs priority: exact $p = 0.16$–0.21.
Attention-growth sensitivity to Exoplanet tier (2 vs 3): exact $p = 0.12$ vs 0.04.

Table S13: Priority–signal intensity correlations by scope subset (2010 coding; exact two-sided p in parentheses).

Scope	n	Attention–priority ρ (p)	Citation–priority ρ (p)
All topics	11	0.17 (0.62)	0.42 (0.20)
In-scope (no heliophysics)	10	−0.13 (0.73)	0.21 (0.57)
Astronomy-core (no helio/solar)	9	0.08 (0.83)	0.08 (0.83)

Table S14: Gini coefficient of public-facing attention and citations, by publication year (astrophysics division, unmatched as zero).

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Attention Gini	0.89	0.91	0.91	0.83	0.80	0.79	0.73	0.81	0.77	0.80
Citation Gini	0.59	0.57	0.56	0.57	0.56	0.58	0.57	0.57	0.55	0.54

Table S15: Gini coefficient by primary UAT topic (public-facing attention vs citations; tagged astrophysics papers).

Topic	Attention Gini	Citation Gini	n
Solar-system astronomy	0.87	0.50	2,277
High-energy astrophysics	0.83	0.56	2,813
Interdisciplinary astronomy	0.83	0.56	5,538
Observational astronomy	0.83	0.67	6,033
Astrophysical processes	0.82	0.59	9,999
Cosmology	0.81	0.63	8,111
Galactic & extragalactic	0.81	0.56	7,380
Interstellar medium	0.80	0.56	2,301
Heliophysics	0.79	0.50	3,467
Exoplanet astronomy	0.78	0.49	1,029
Stellar astronomy	0.78	0.52	6,883

Table S16: Counterfactual decomposition of the attention–citation Gini gap (astrophysics division, $n = 99{,}083$).

Distribution	Gini	Share of the observed gap
Public-facing attention (observed)	0.83	—
Citations (observed)	0.58	—
Observed gap	*0.24*	*100%*
Citations given attention's zero mass (40%)	0.75	68%
Attention given citations' zero mass (0.6%)	0.71	53%

Table S17: Exoplanet-versus-cosmology two-part margins under raw and field-normalised citation covariates.

Citation covariate	Coverage OR (95% CI)	Intensity RR (95% CI)
Raw ADS citations (primary)	2.53 (2.03–3.16)	3.01 (2.78–3.27)
Field-normalised citations	2.65 (2.13–3.31)	2.82 (2.60–3.06)

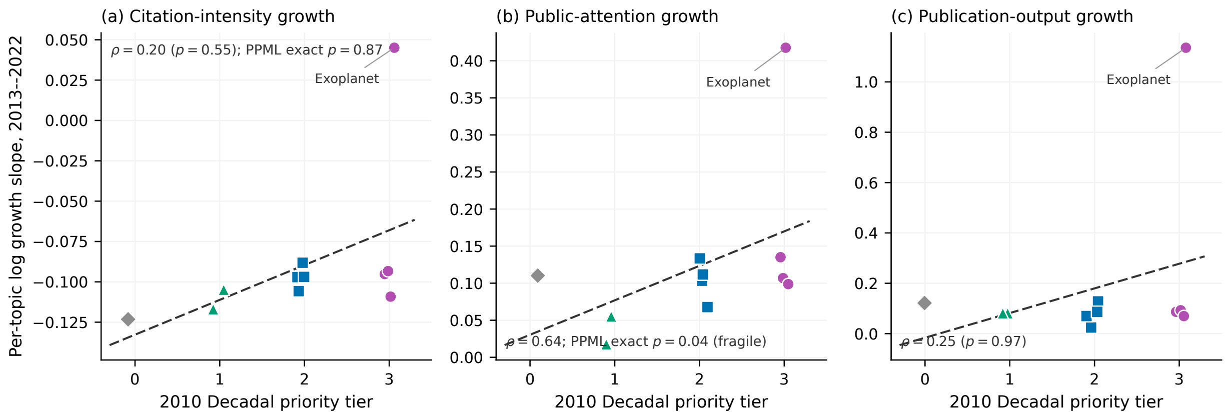

Figure S2: Exploratory per-topic growth slopes (2013–2022) against 2010 priority tier. Citation-intensity and output growth show no association; the attention-growth association is nominal and driven by the exoplanet topic (main-text Section 4.6).

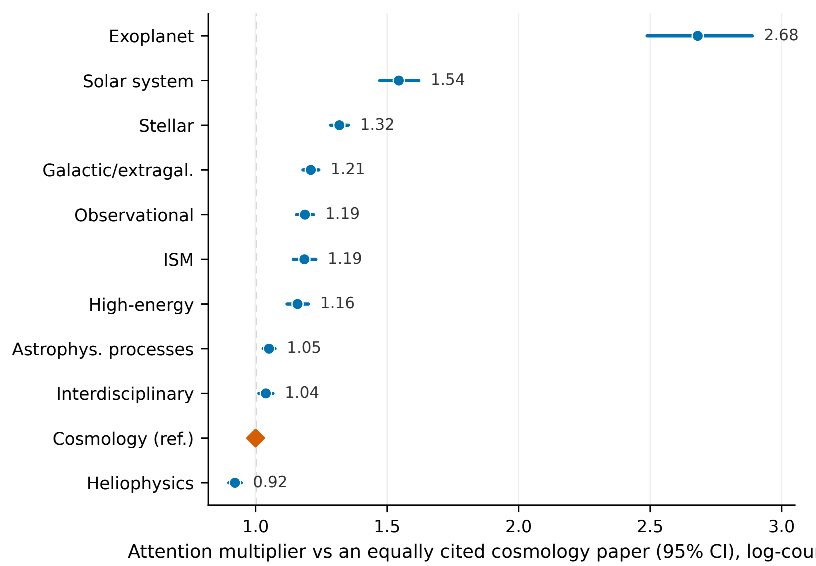

Figure S3: Single-equation log-count attention differential by topic (HC3 CIs; $n = 55{,}831$), retained for continuity with the prior literature; the two-part margins are in main-text Figure 2 and Table S10.

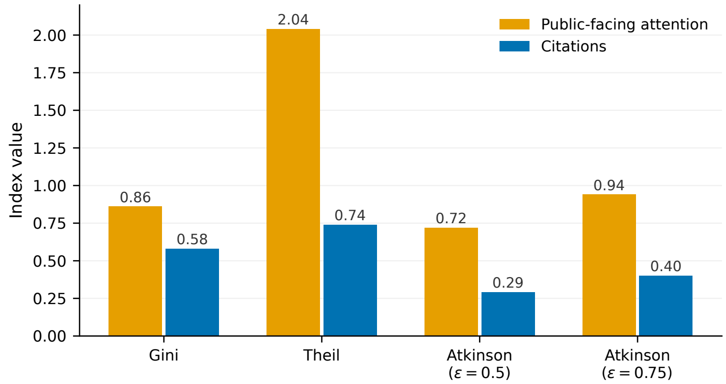

Figure S4: Public-facing attention versus citations on four concentration indices (full corpus).

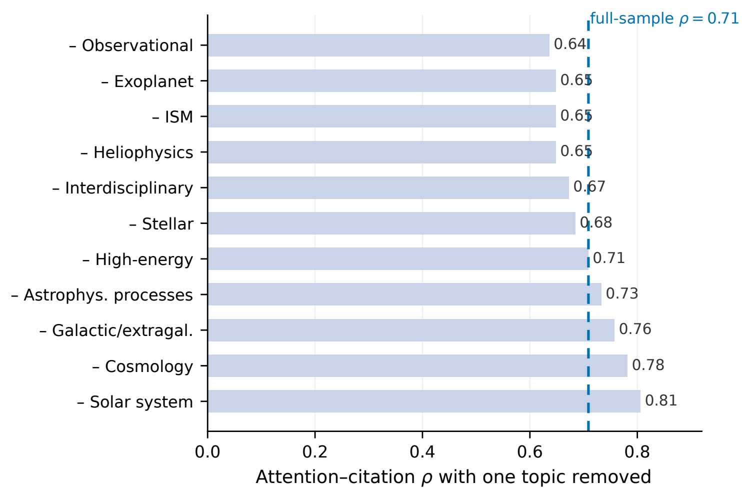

Figure S5: Leave-one-out: the attention–citation share correlation with each topic removed in turn (full-sample $\rho = 0.71$).

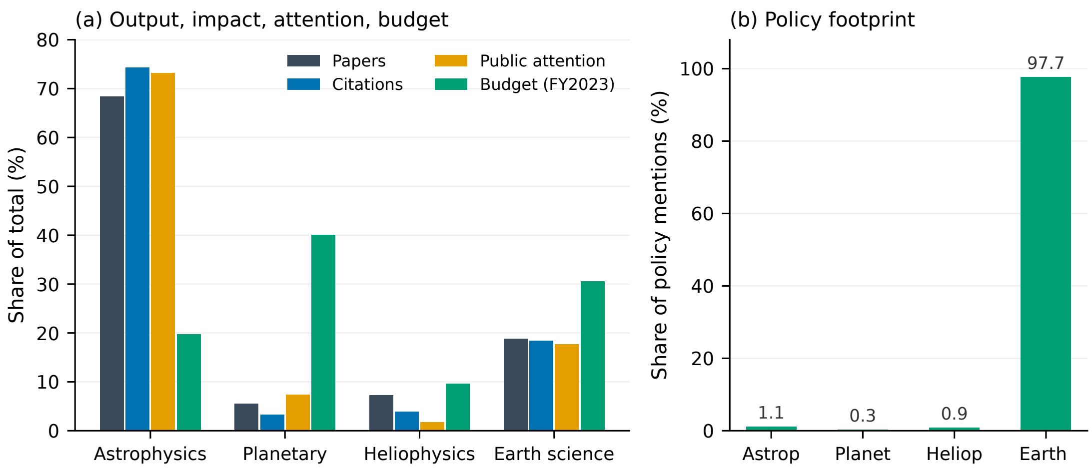

Figure S6: (a) Share of papers, citations, public attention, and budget by division (budget = FY2023 SMD snapshot). (b) Share of policy-document mentions by division. Shares are descriptive and depend on the journal selection; the juxtaposition is not an evaluative comparison.