We’ve got issues

Understanding the current strain on scientific publishing

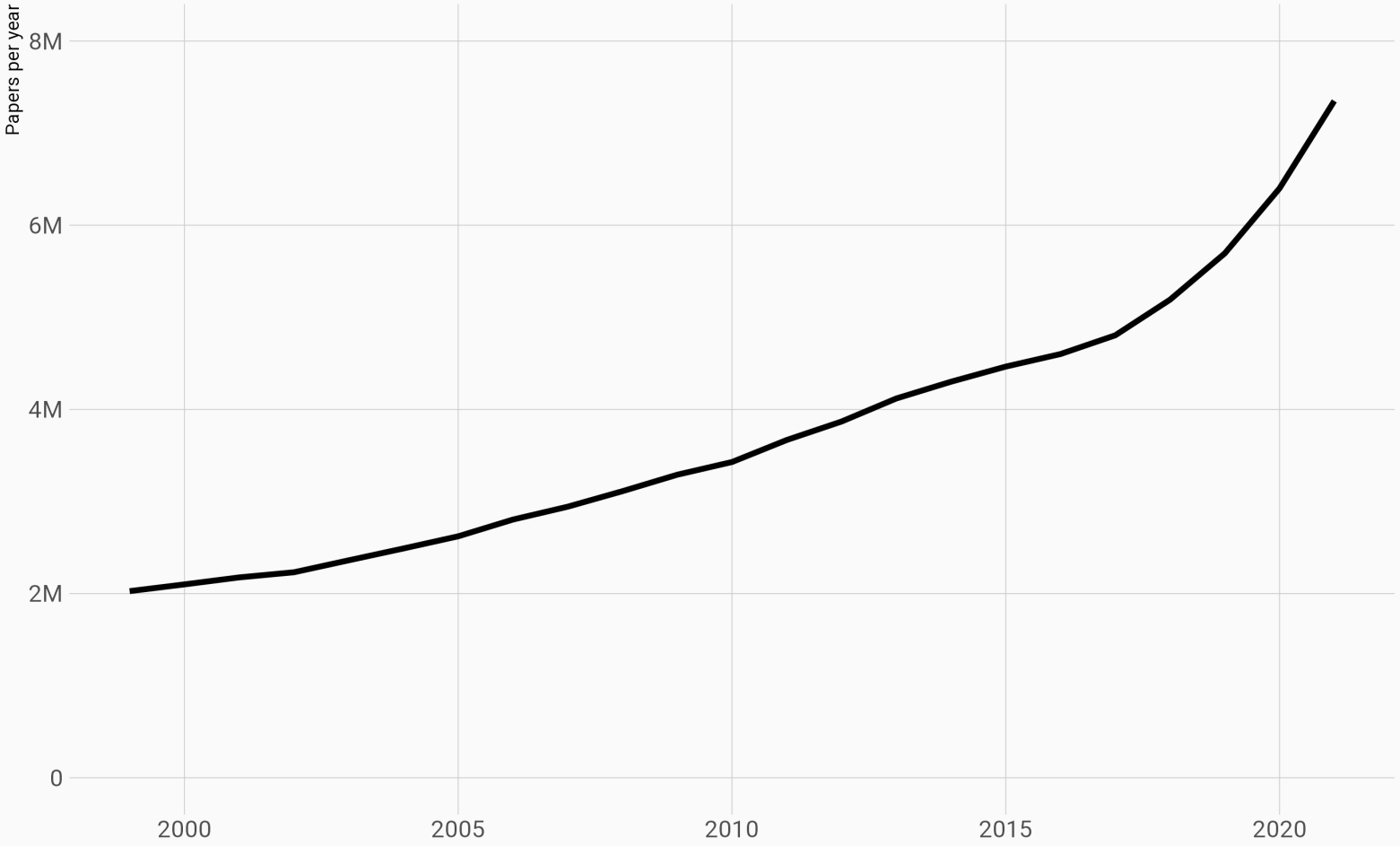
M. A. Hanson, P. Gómez Barreiro, **P. Crosetto**, D. Brockington

SNSF – April 19th 2023

Academic publishing is undergoing an exponential growth

New papers published each year in the world

Scimago database, all publishers



SJR data -- analysis DB, PC, PGB, MH

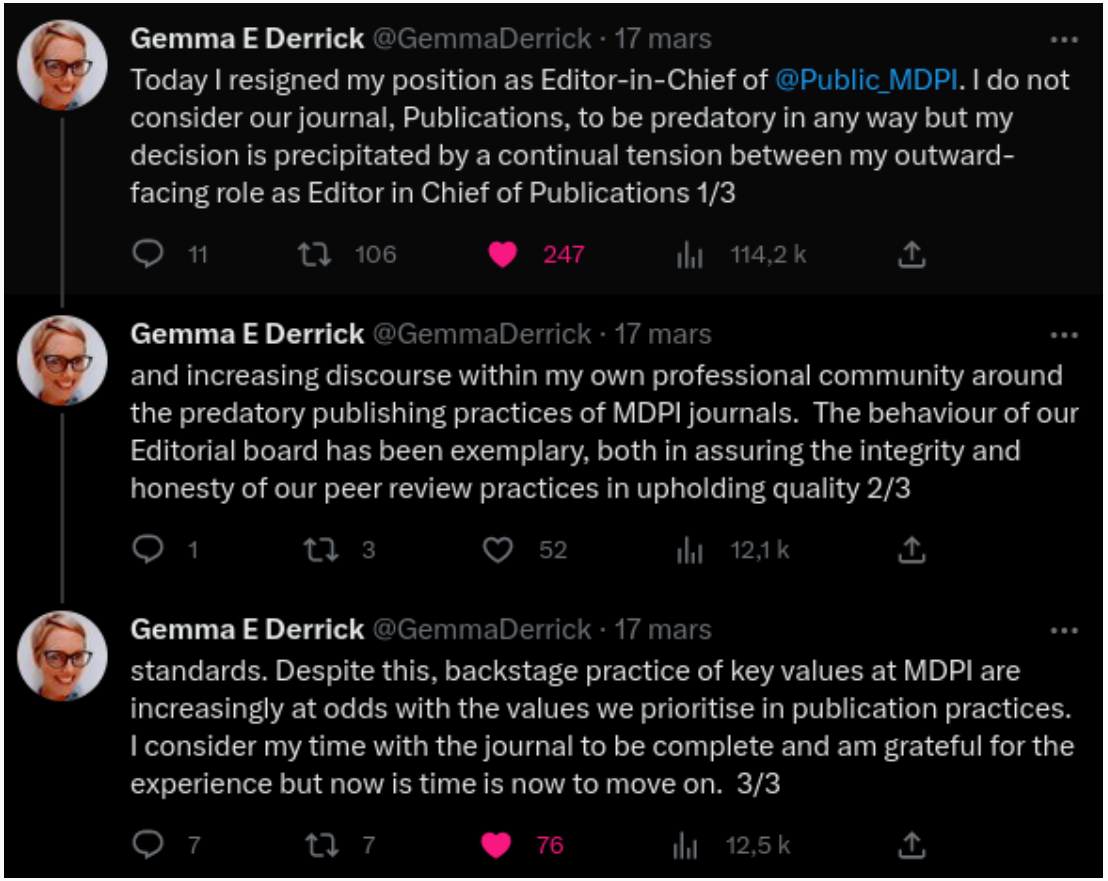
This is mostly a good thing

- More scientists around
- More funds for research
- Open Access: more results available to anyone
- Web tools: faster dissemination of ideas
- Lower file drawer effects
- More replications, robustness, reviews, meta-analyses

Yet we've got issues

Editors resigning over high fees

Yet we've got issues



Gemma E Derrick @GemmaDerrick · 17 mars

Today I resigned my position as Editor-in-Chief of @Public_MDPI. I do not consider our journal, Publications, to be predatory in any way but my decision is precipitated by a continual tension between my outward-facing role as Editor in Chief of Publications 1/3

11 106 247 114,2 k

Gemma E Derrick @GemmaDerrick · 17 mars

and increasing discourse within my own professional community around the predatory publishing practices of MDPI journals. The behaviour of our Editorial board has been exemplary, both in assuring the integrity and honesty of our peer review practices in upholding quality 2/3

1 3 52 12,1 k

Gemma E Derrick @GemmaDerrick · 17 mars

standards. Despite this, backstage practice of key values at MDPI are increasingly at odds with the values we prioritise in publication practices. I consider my time with the journal to be complete and am grateful for the experience but now is time is now to move on. 3/3

7 7 76 12,5 k

Editors resigning
over bad publisher practices

Yet we've got issues

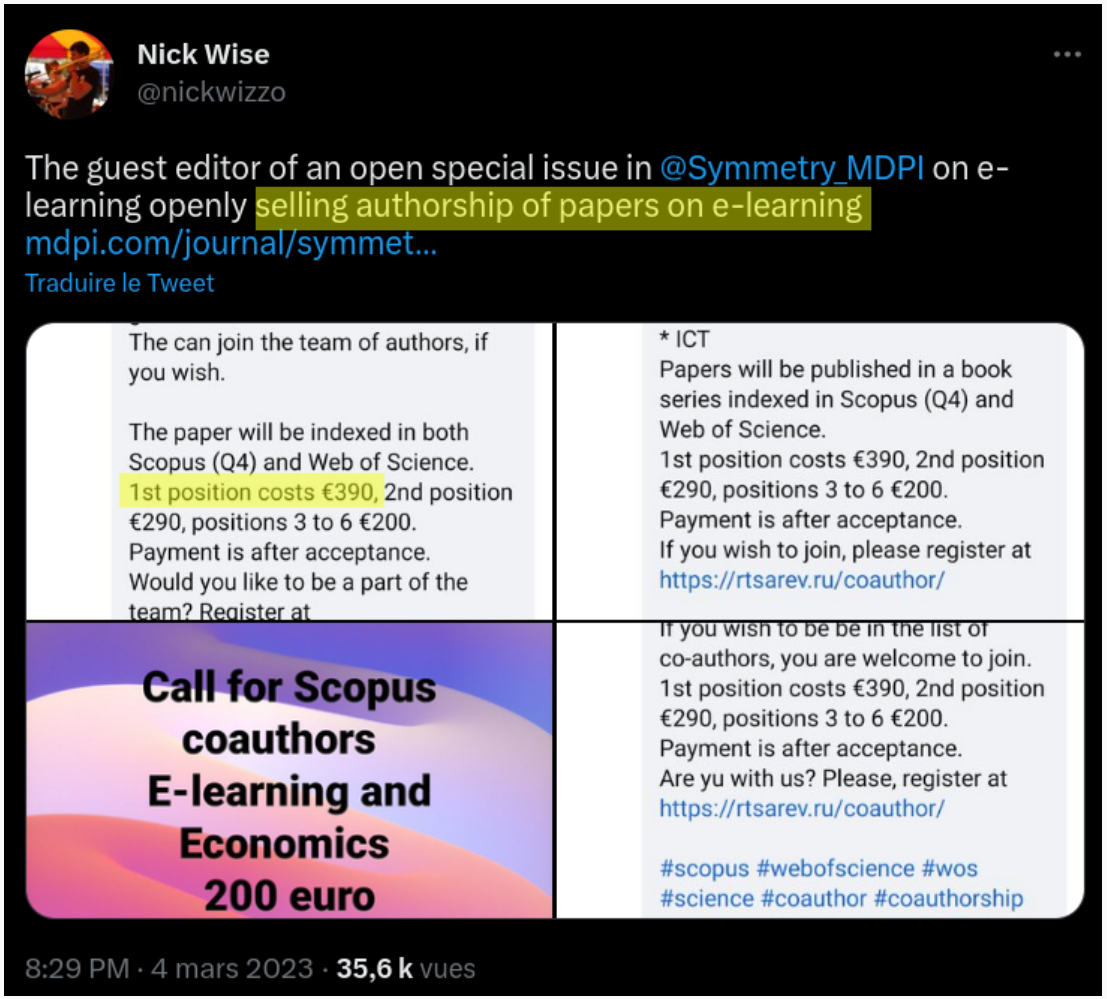
Paper mills
mass producing
fake articles

NEWS FEATURE | 23 March 2021

The fight against fake-paper factories that churn out sham science

Some publishers say they are battling industrialized cheating. A *Nature* analysis examines the 'paper mill' problem – and how editors are trying to cope.

Yet we've got issues



Authorship sales rings

Yet we've got issues

Stunningly prolific authors

EL PAÍS

ce & Tech

SILICON VALLEY · YOUTUBE ·

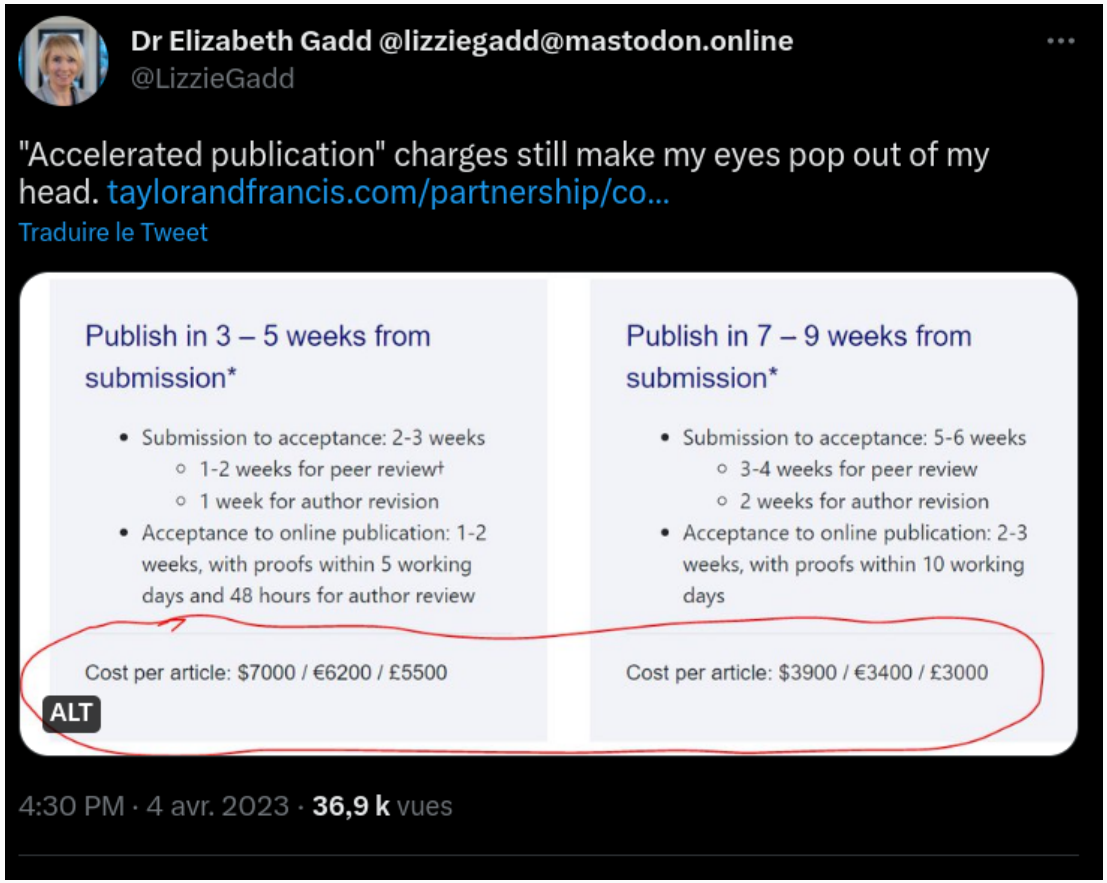
SCIENTIFIC ETHICS >

One of the world's most cited scientists, Rafael Luque, suspended without pay for 13 years

The prolific chemist, who has published a study every 37 hours this year, has been sanctioned by the University of Córdoba over his research work for other institutions in Russia and Saudi Arabia

Yet we've got issues

Pay to get faster through peer-review



Yet we've got issues

PHR Public Health Reviews

CiteScore 9.6 How to publish Submit

EDITORIAL

Public Health Rev, 17 November 2022
https://doi.org/10.3389/phrs.2022.1605407

Check for updates

«I Do Not Have Time»—Is This the End of Peer Review in Public Health Sciences?

Nino Künzli[1,2,3*], Anke Berger[1,3], Katarzyna Czabanowska[4], Raquel Lucas[5], Andrea Madarasova Geckova[6], Sarah Mantwill[7] and Olaf von dem Knesebeck[8]

Editors unable to find referees

Yet we've got issues

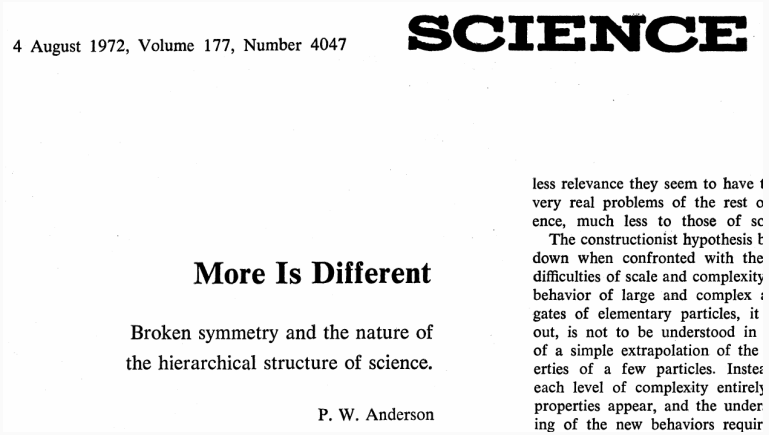
Mega-journals being delisted from WoS

What’s going on?

More is different

Growth is not more of the same:
growth means change.

- new practices
- new business strategies
- new incentives
- new constraints
- new meanings



4 August 1972, Volume 177, Number 4047 SCIENCE

More Is Different

Broken symmetry and the nature of the hierarchical structure of science.

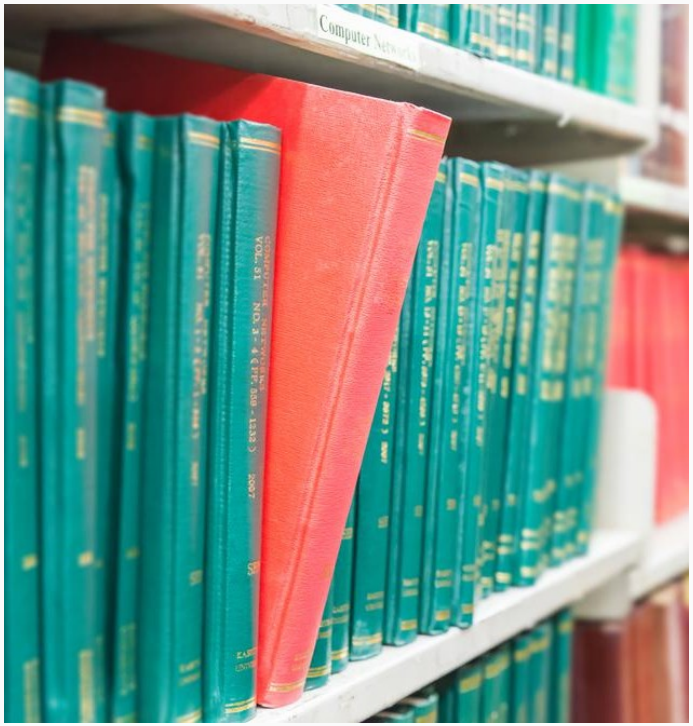
P. W. Anderson

less relevance they seem to have t
very real problems of the rest o
ence, much less to those of sc
The constructionist hypothesis b
down when confronted with the
difficulties of scale and complexity
behavior of large and complex
gates of elementary particles, it
out, is not to be understood in
of a simple extrapolation of the
erties of a few particles. Instea
each level of complexity entirely
properties appear, and the under
ing of the new behaviors requir

A semantic shift

"**Journal**"

used to mean

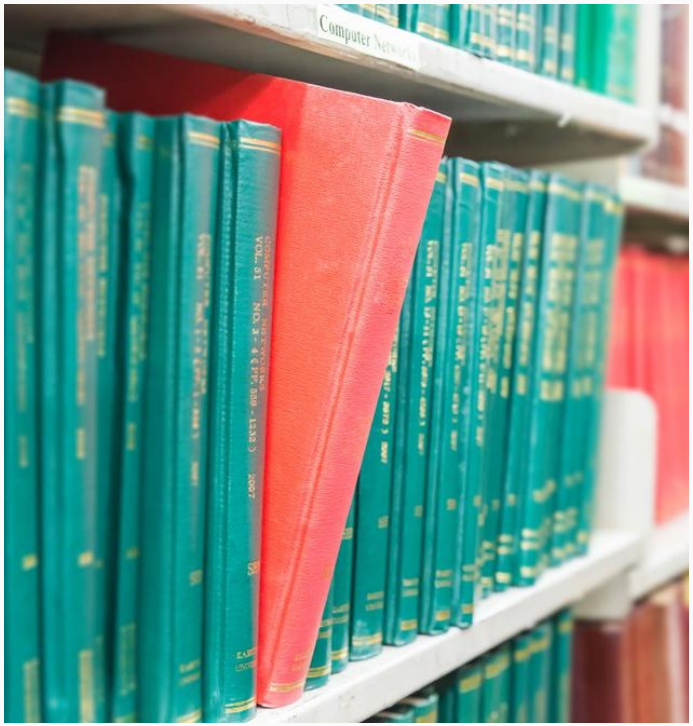


A physical object with
limited available space

A semantic shift

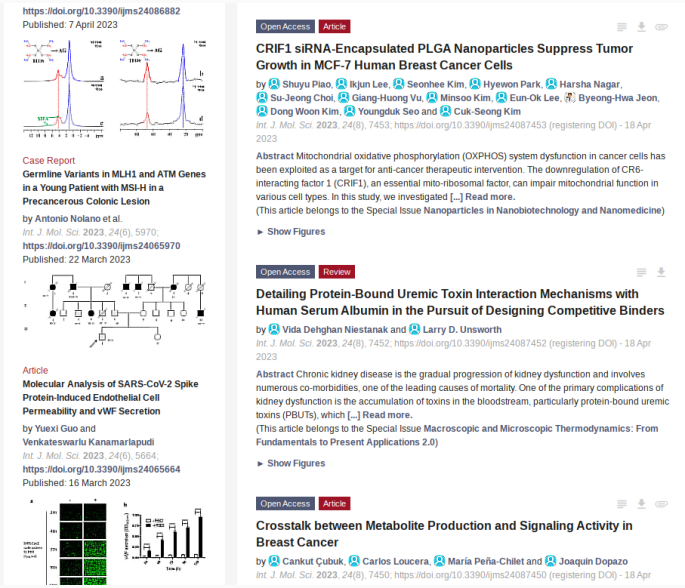
"**Journal**"

used to mean



A physical object with limited available space

now it also means



A limitless electronic repository with a name

A semantic shift

"**Publication**"

used to mean

- a handful of journals
- long delays
- low acceptance rates
- free for authors
- do it and thrive

⇒ *good science rejected?*

A semantic shift

"**Publication**"

used to mean

- a handful of journals
- long delays
- low acceptance rates
- free for authors
- do it and thrive

⇒ *good science rejected?*

now it also means

- thousands of journals
- short delays
- high acceptance rates
- authors pay
- don’t do it and die

⇒ *bad science accepted?*

A semantic shift

"**Special issue**"

used to mean

- A once-in-a-while issue
- About a special topic
- Strict editor control
- regular > special

A semantic shift

"**Special issue**"

used to mean

- A once-in-a-while issue
- About a special topic
- Strict editor control
- regular > special

now it also means

- A many-a-day issue
- About any topic
- Relaxed editor control
- special > regular

A semantic shift

"**Publisher business model**"

used to mean

- Many small journals
- Readers pay
- $ through subscription
- "*Polish your gems*"

Incentive to ⇑⇑ quality,
quantity? . . .

A semantic shift

"**Publisher business model**"

used to mean

- Many small journals
- Readers pay
- $ through subscription
- "*Polish your gems*"

Incentive to ↑↑ quality, quantity? . . .

now it also means

- Few mega-journals
- Authors pay
- $ through publication
- "*Get authors on board*"

Incentive to ↑↑ quantity, quality? . . .

Our aim:

understanding the strain on publishing

A *caveat*: no need for "predatory" labels

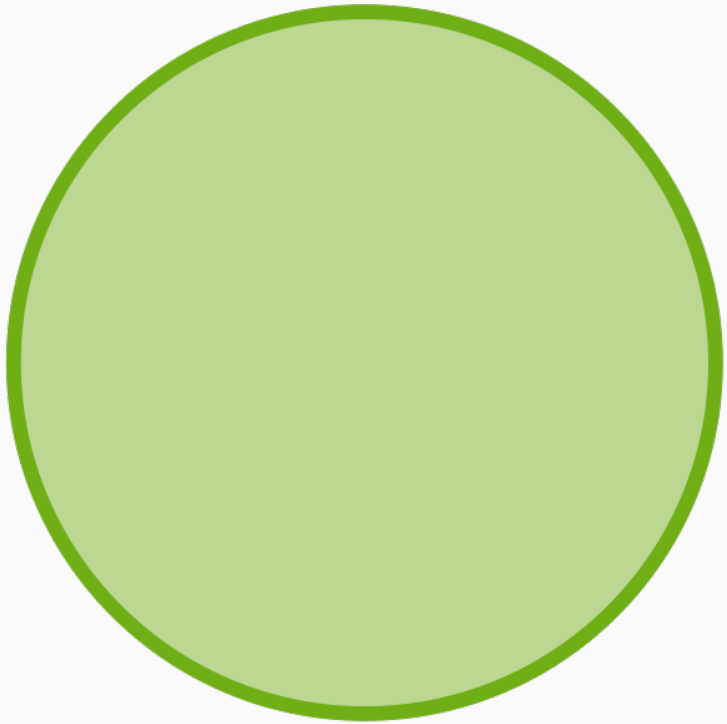
We don’t think binary labels improve our understanding

There’ll be no "predatory" judgments here

- outright fraudsters do exist (publishers *and* authors)
- agents just follow their interest
- market rules generate outcomes
- outcomes can be good or bad
  - for the different actors
  - for the public good that is science

Behold the scientific publishing system

Publishers



Researchers

Funders

What does the system do?

What are the functions the system fulfills...

for Scientists

- dissemination
- reputation
- sorting

for Publishers

- profits
- dissemination
- sustainability

for Funders

- selection
- prioritization
- public access

What do the different actors want?

What do different actors want from the system?

Scientists

- high reputation
- low effort
- stable reputation

Publishers

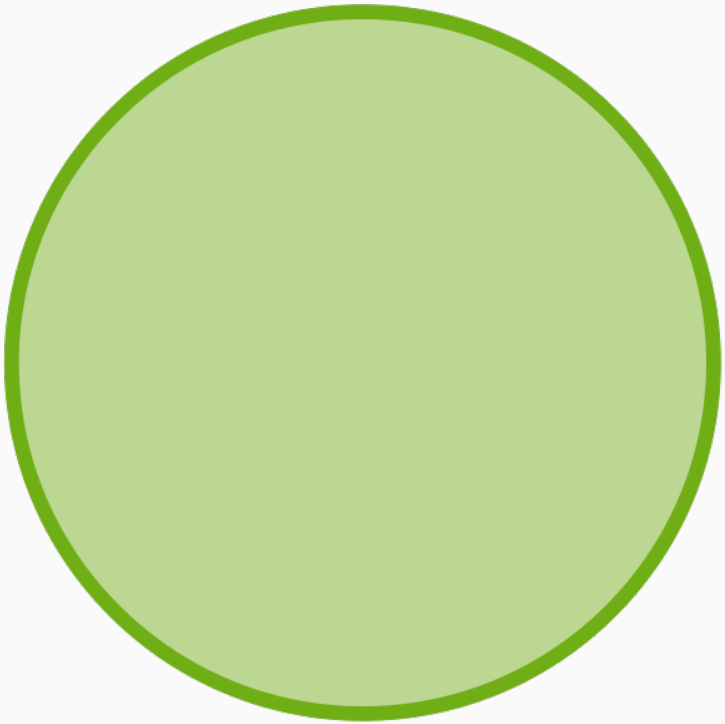
- high reputation
- high quantity
- high revenue

Funders

- stability
- true signal
- low spending

The system, growing under strain

Publishers

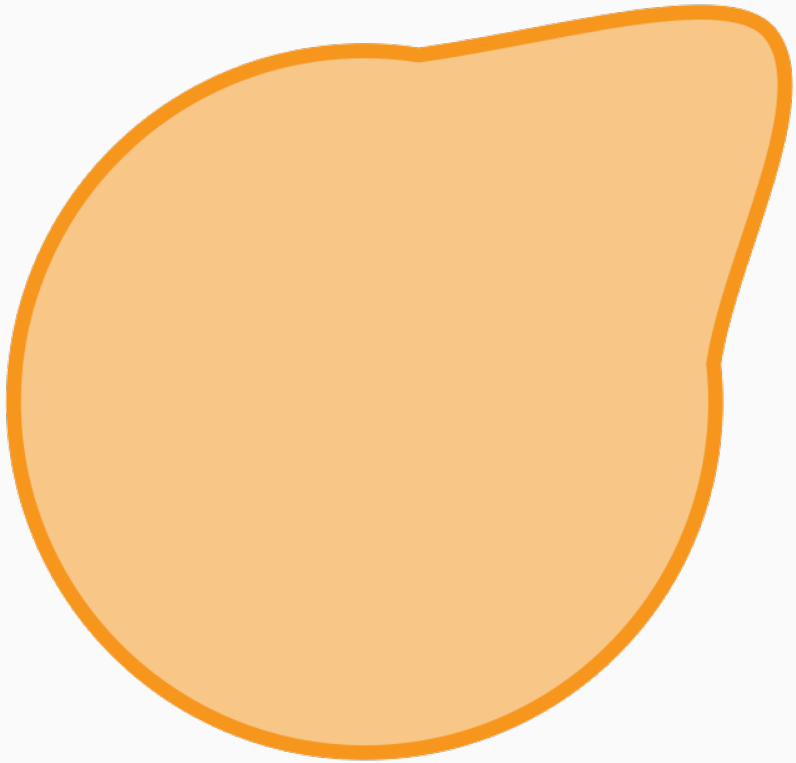


Researchers

Funders

The system, growing under strain

Publishers

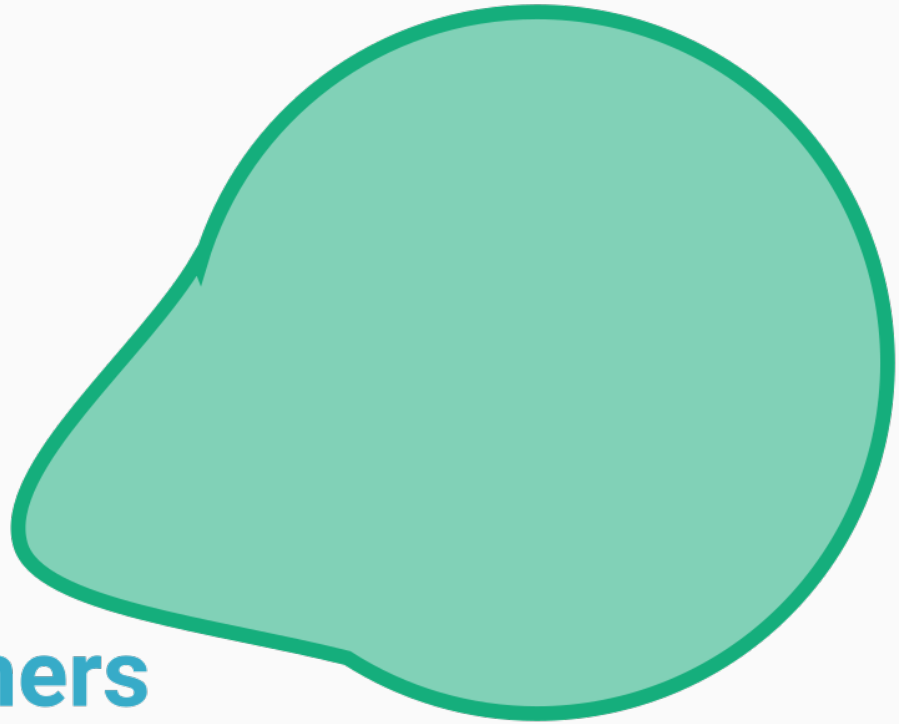


Researchers

Funders

The system, growing under strain

Publishers

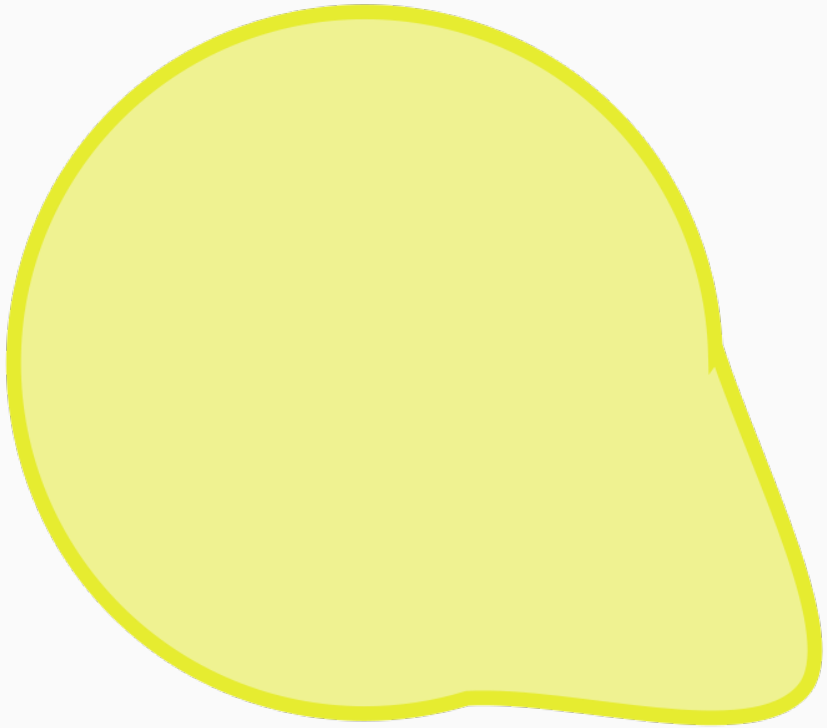


Researchers

Funders

The system, growing under strain

Publishers



Researchers

Funders

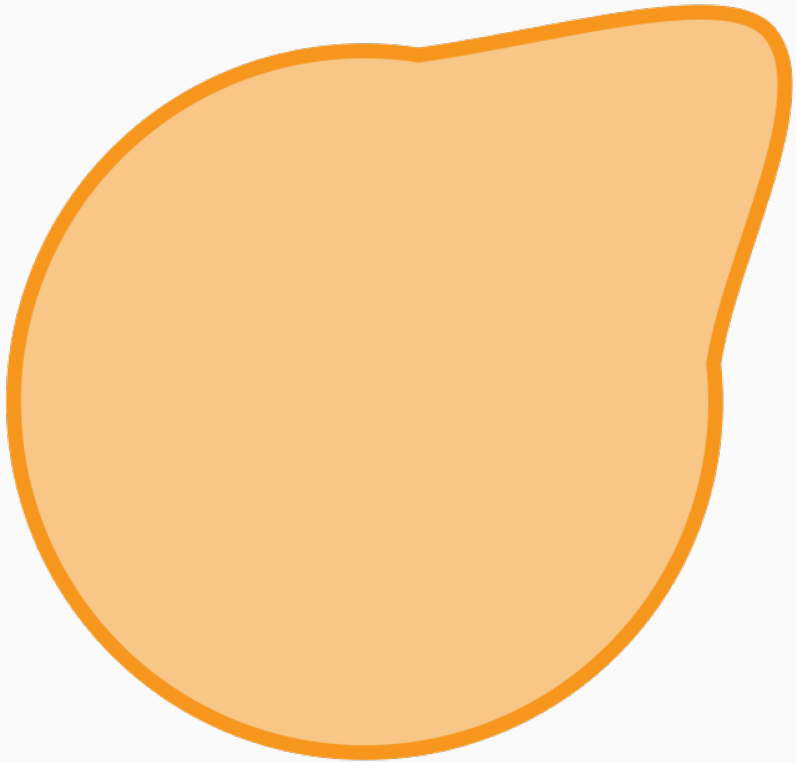
Our analysis:

Understanding the strain put on the system

by evolving publishers practices

So, this

Publishers



Researchers

Funders

Data sources

We exploit data coming from various sources:

- A full scrape of the Scimago Journal Rankings database

  used for: comparisons across publishers, IF, SJR rank...

- Web scrape of MDPI, Frontiers, Hindawi, PLoS

  used for: turnaround times, special issues

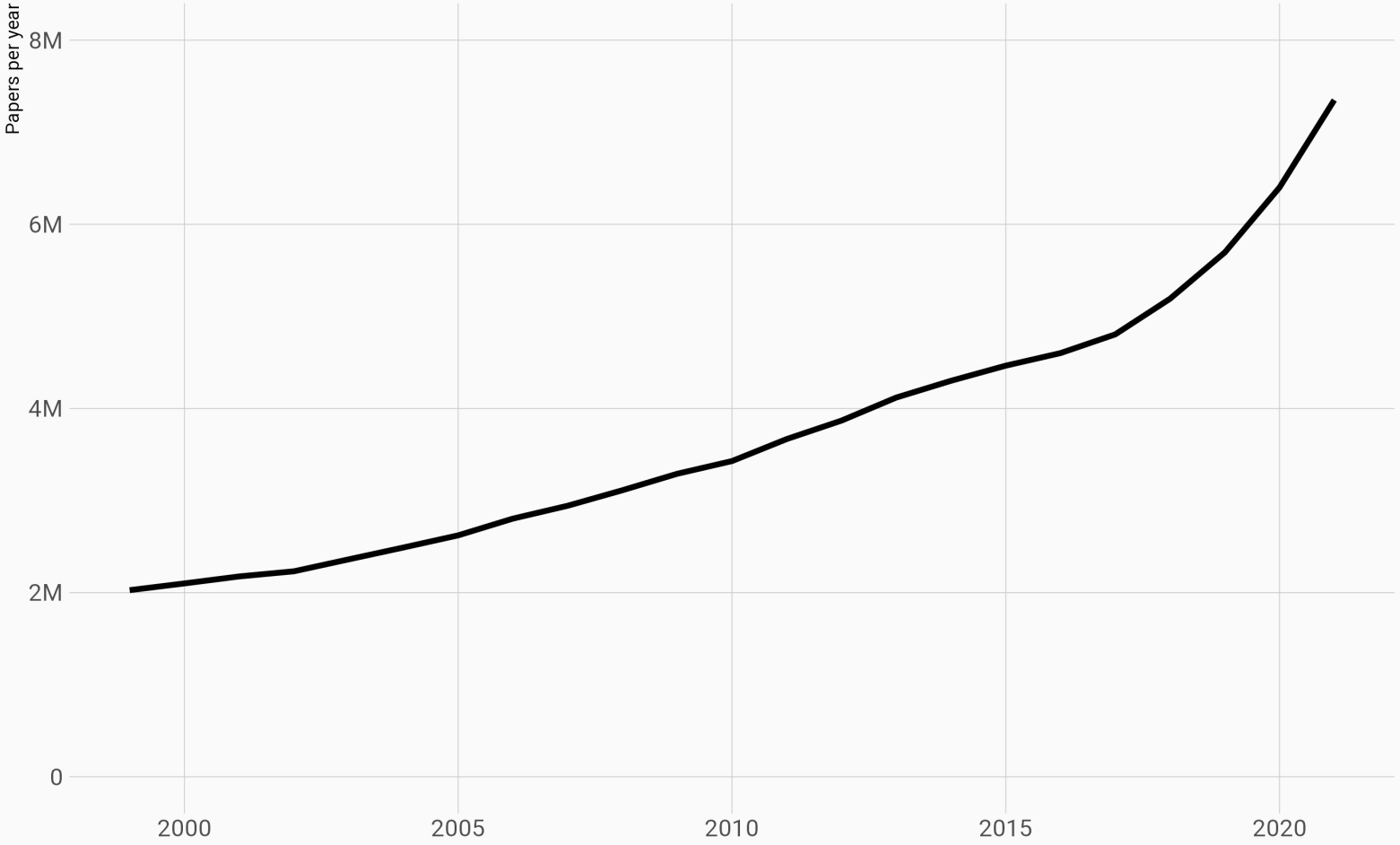
- First hand data from publisher reports and websites

  used for: rejection rates

Which trends and threats are hidden by this exceptional growth?

New papers published each year in the world

Scimago database, all publishers



SJR data -- analysis DB, PC, PGB, MH

Analysis plan

We single out five indicators of strain on the system:

- Number and size of journals
- Number and role of Special Issues
- Turnaround times
- Rejection rates
- Impact Factor inflation

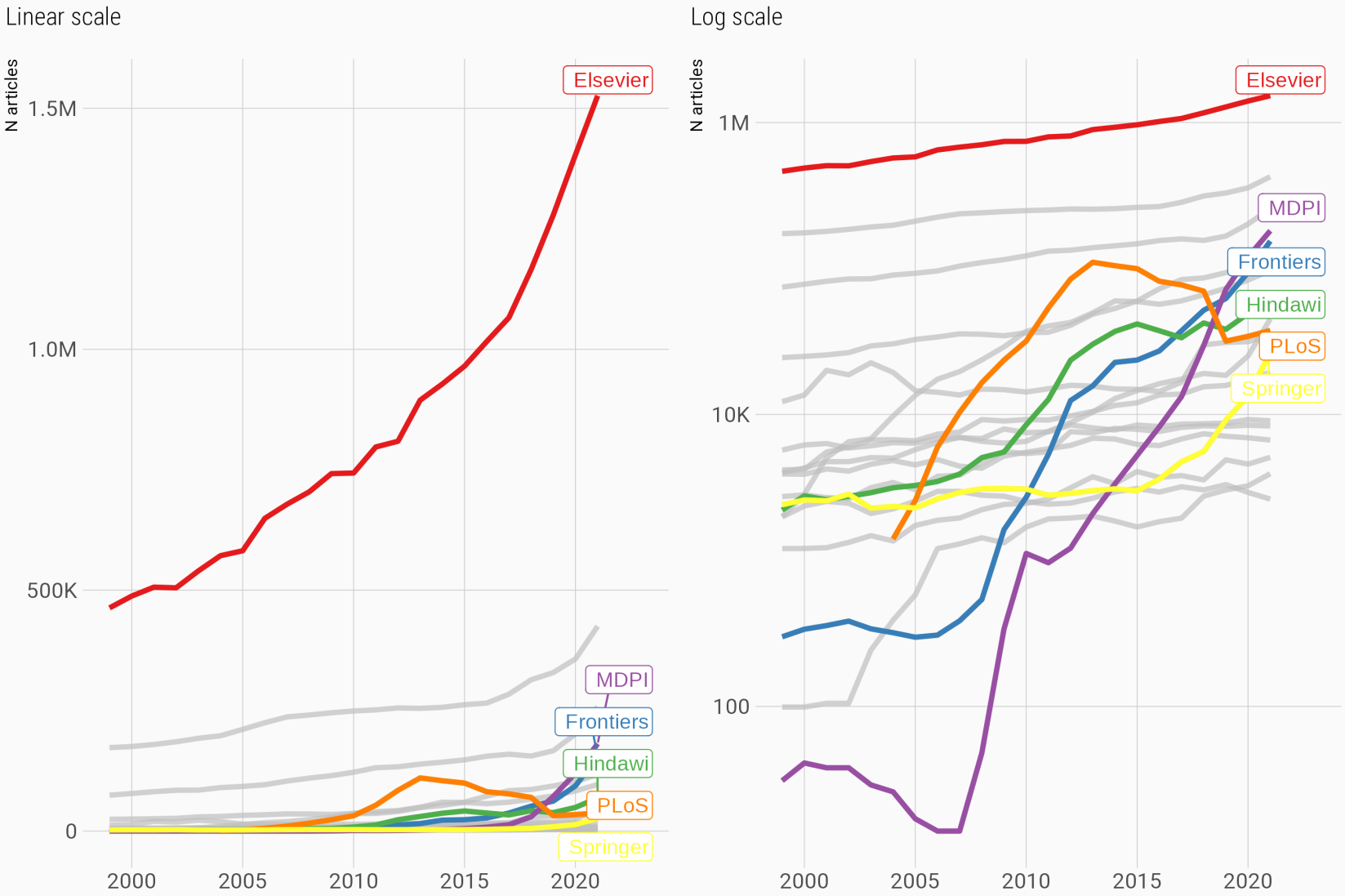
None of them is critical *per se*

together they indicate strain imposed by publishers

Number of articles & journal size

The rise of new publishers

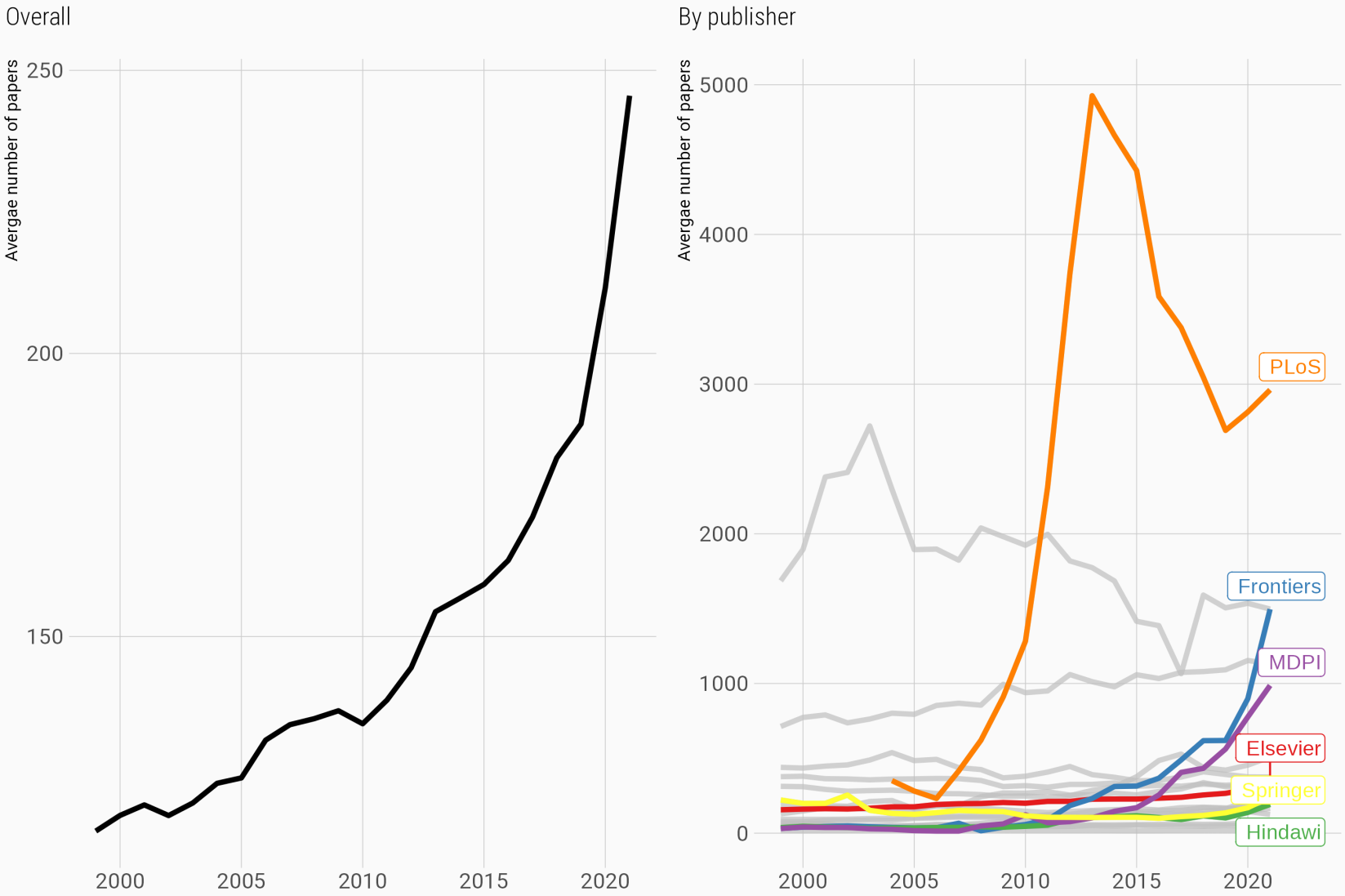
Article growth by publisher, 1999-2021



Scimago data -- analysis MH, PC, PGB, DB

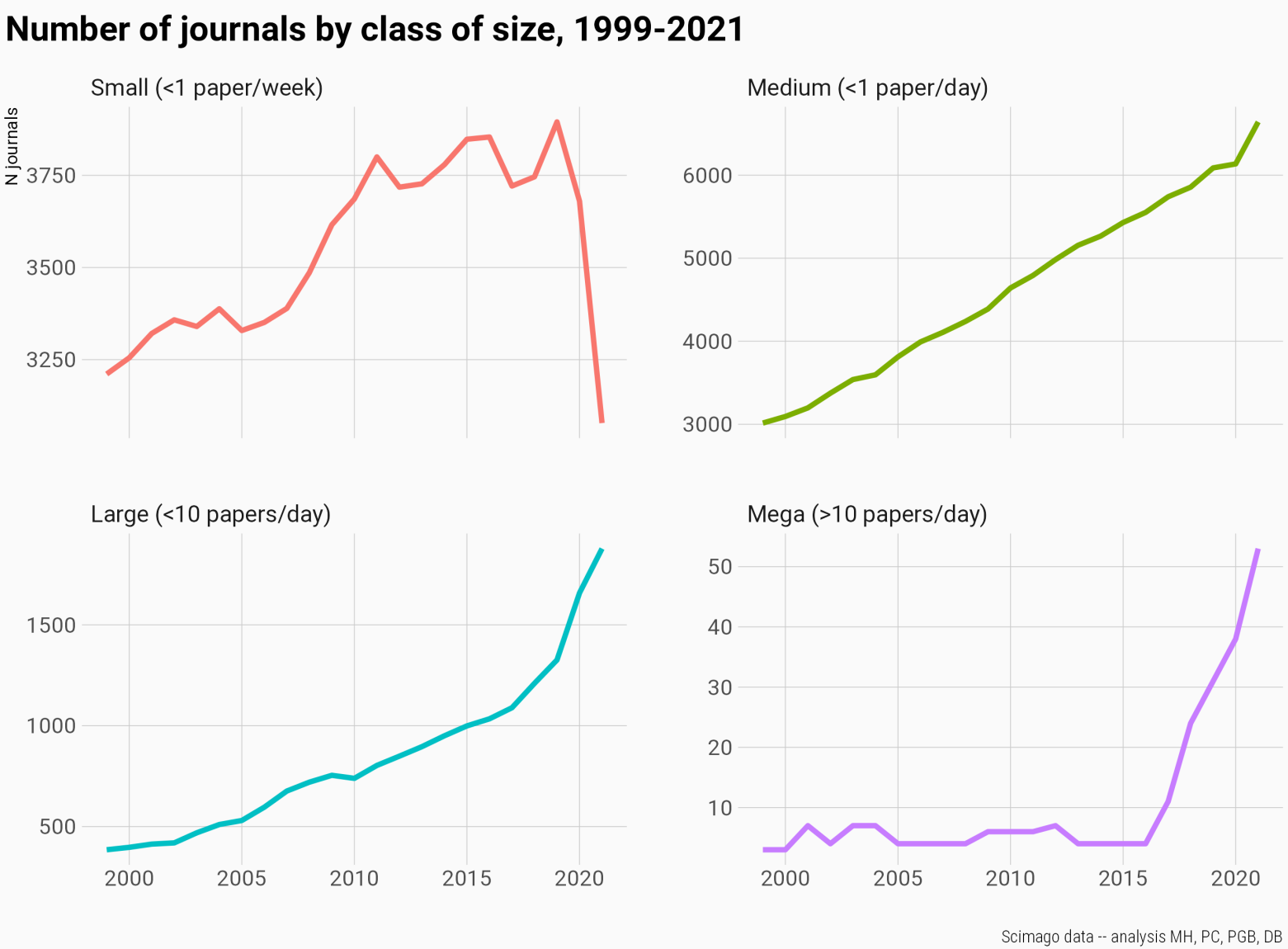
Bigger journals

Average number of papers per journal, 1999-2021



Scimago data -- analysis MH, PC, PGB, DB

The rise of mega-journals



Scimago data -- analysis MH, PC, PGB, DB

What's going on?

Trends:

- Growth means concentration, especially for new players

Why?

- Scientists tend to flock to journals with high reputation
- Hard to set up, but if you have one, exploit it

Threats

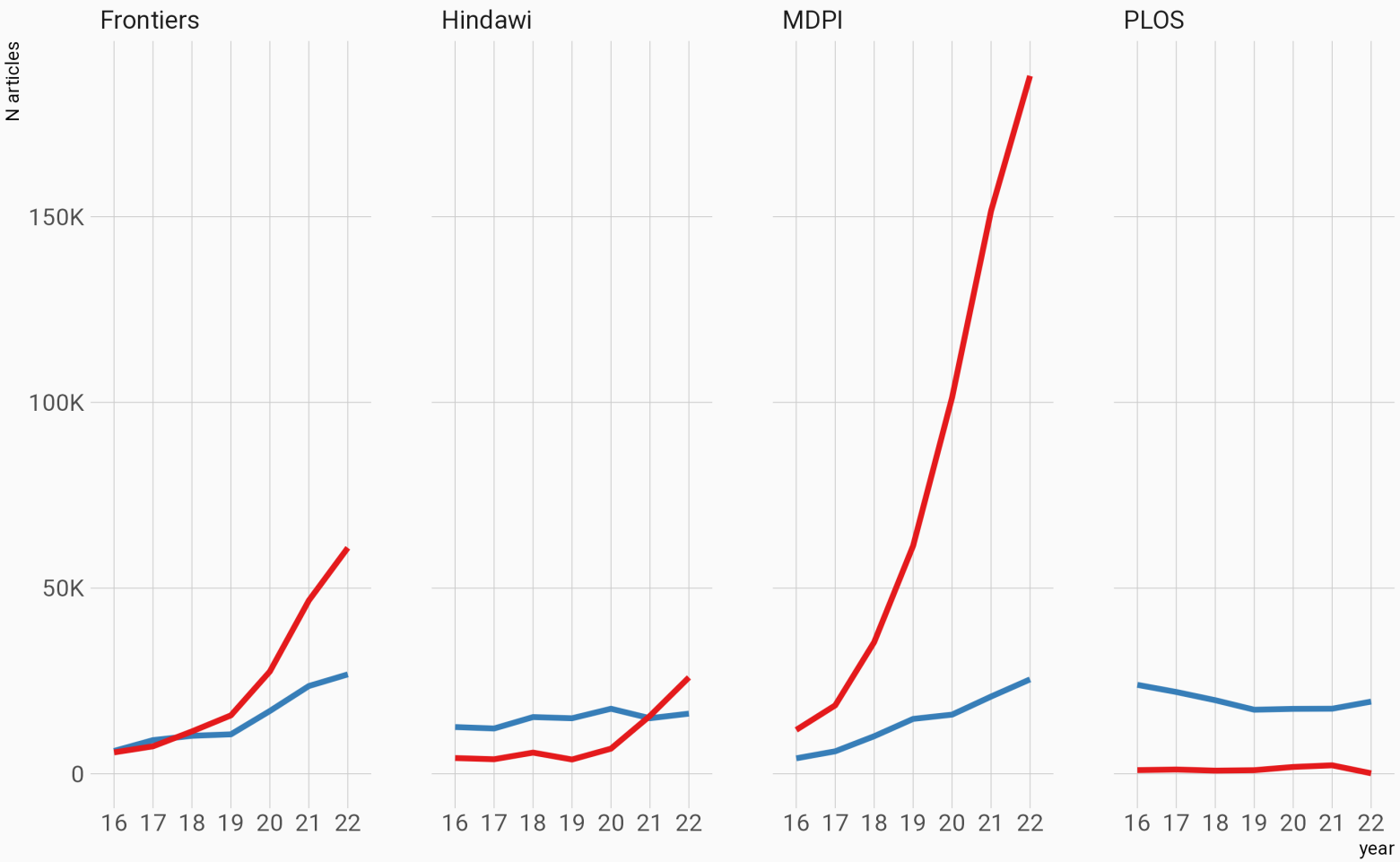
- How much can a journal inflate before it loses reputation?
- Risk of instability of quality signals

The role of special issues

Not so special after all

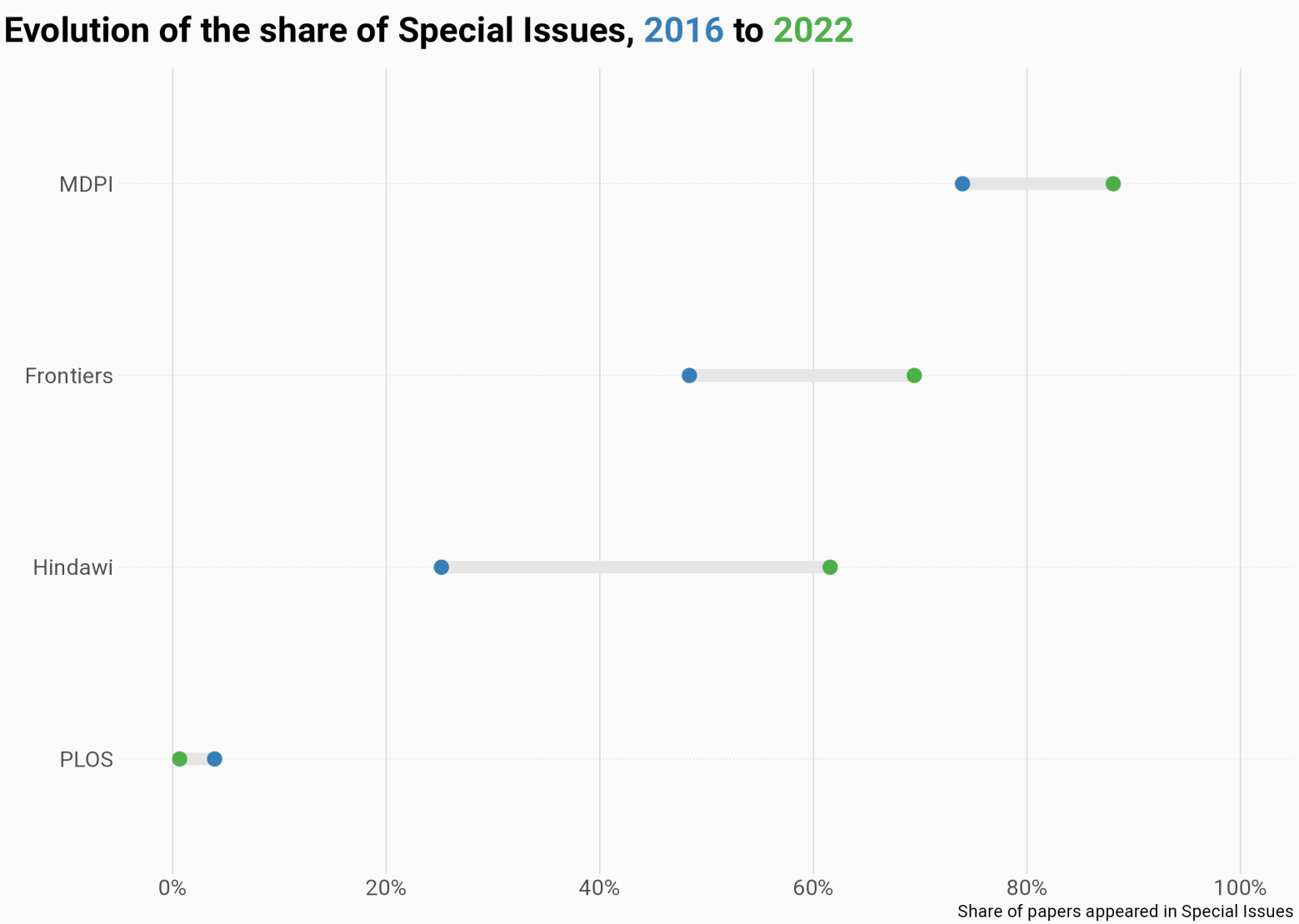
Regular and Special issues growth, 2016-22

Special issues are called Collections at PLOS and Topics at Frontiers



Source: data scraped by by @paolocrosetto & @pagomba -- analysis MH, DB, PGB and PC

Journals at most big OA publishers are mostly special issues



Source: data scraped by by @paolocrosetto & @pagomba -- analysis MH, PC, PGB, DB

What's going on?

Trends:

- SI as a fantastic engine of growth for big OA publishers

Why?

- Mobilization of an army of guest editors & their networks

Threats

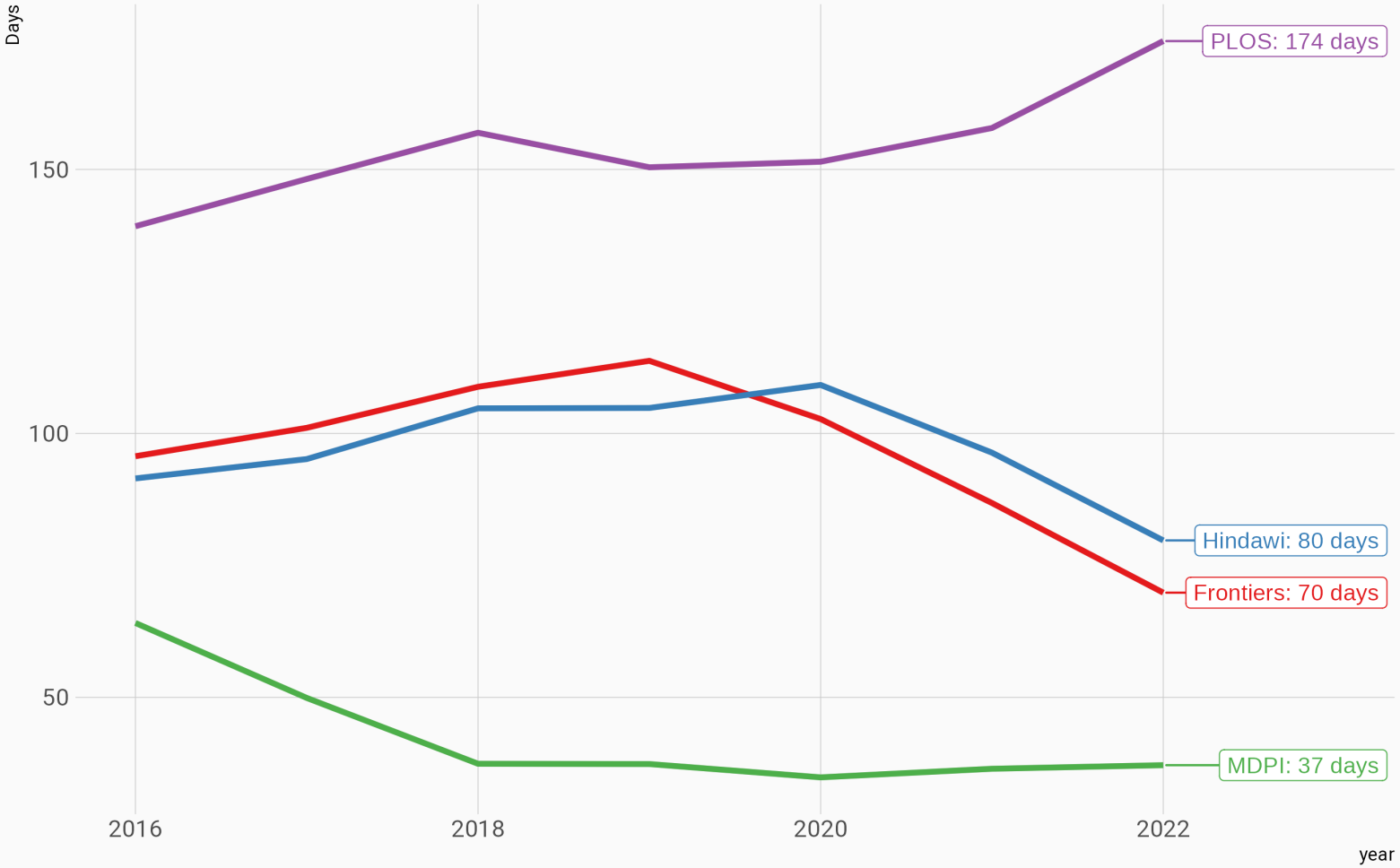
- Less control increases chance of exploitation by authors
- Potential crisis of the SI model (Hindawi, IJERPH delisting)

Turnaround times

Turnaround times have decreased for all for-profit OA publishers

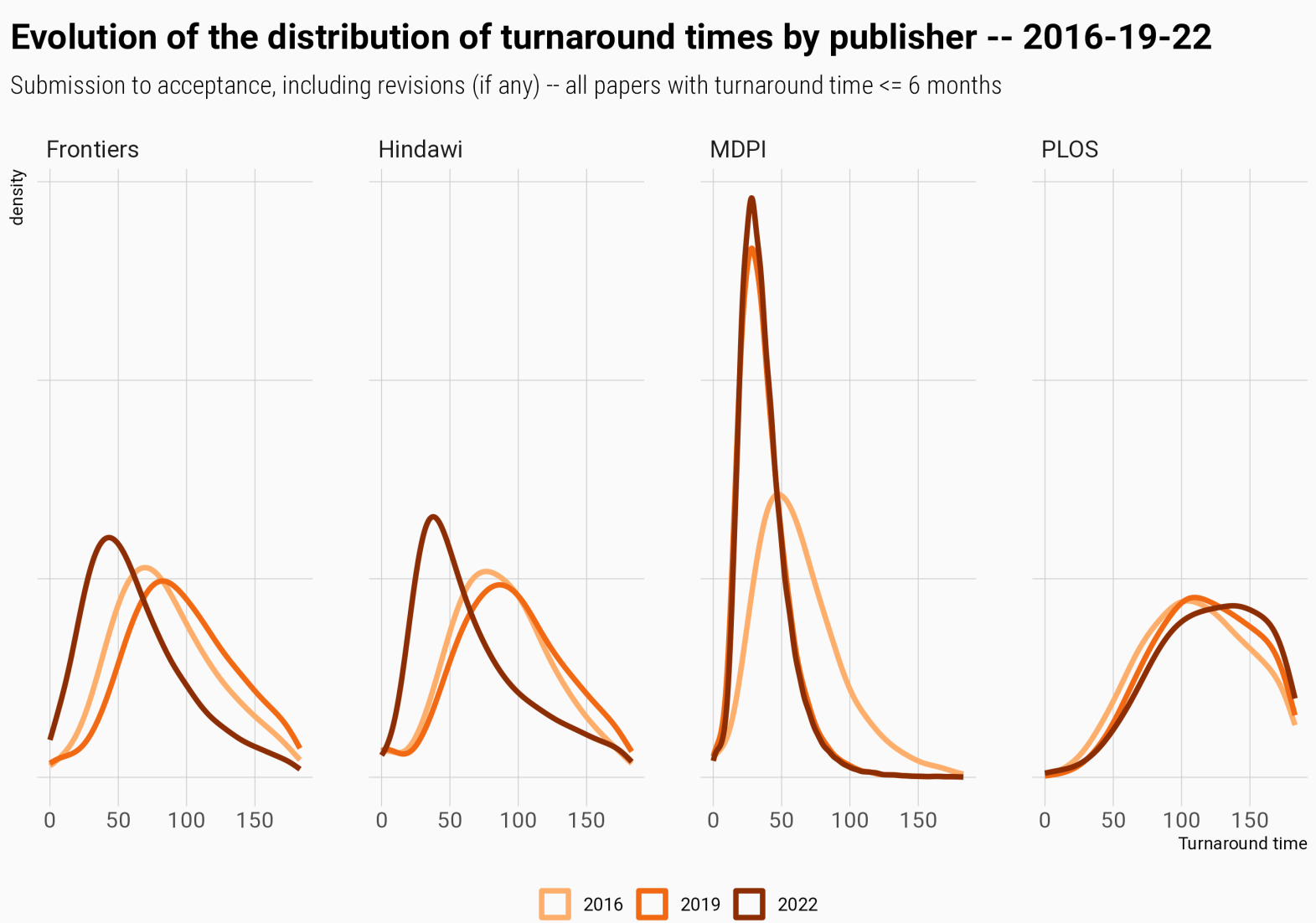
Mean turnaround times by publisher, 2016-22

Submission to acceptance, including revisions (if any) -- all papers with turnaround time <= 1 year



Source: data scraped on the publisher's website by @paolocrosetto & @pagomba, analysis MH, DB, PC, PGB

Turnaround times are getting more homogeneous



What's going on?

Trends:

- TAT can be due to inefficiencies – good that they go down

Why?

- Convergence of authors & OA publishers incentives

Threats

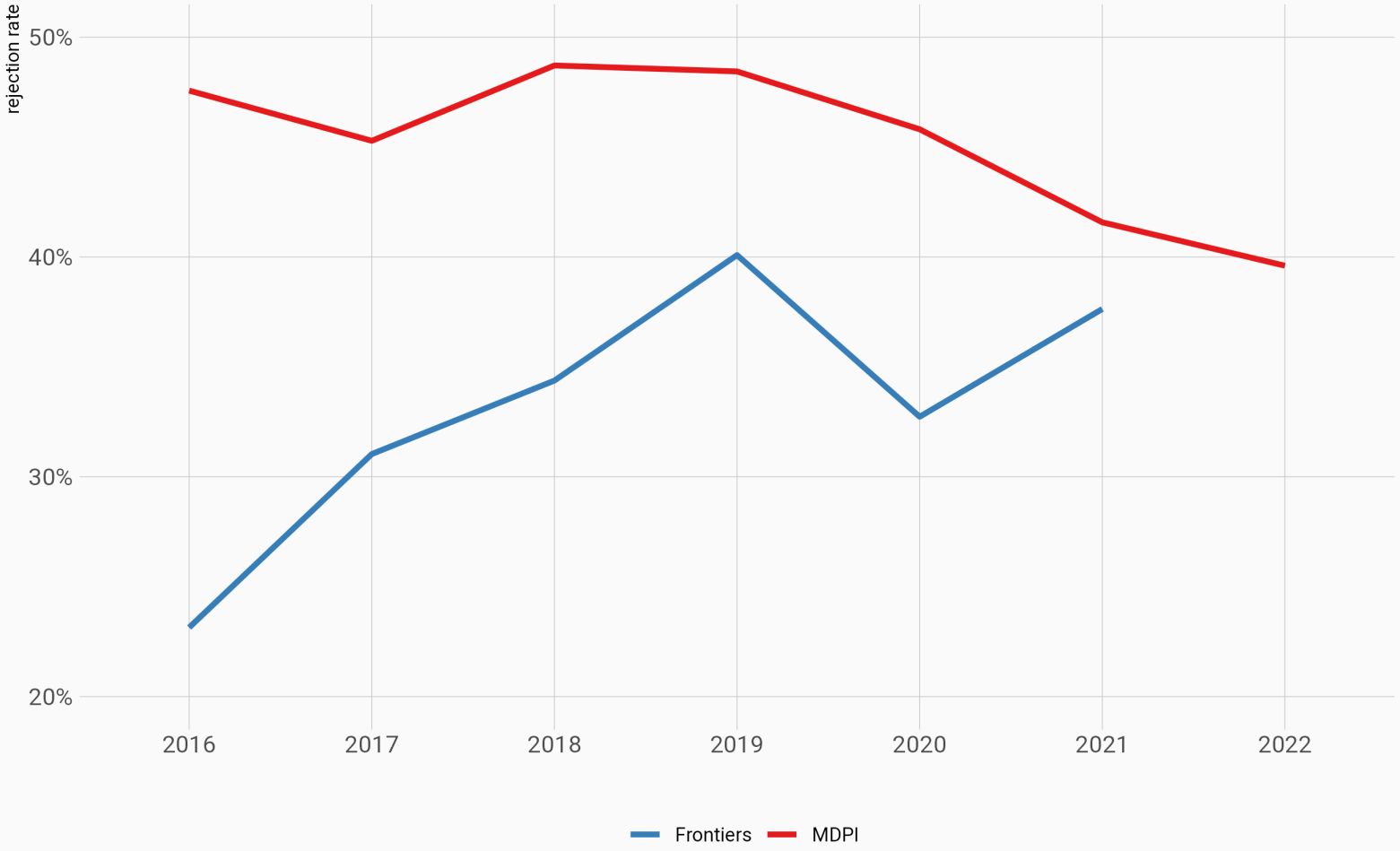
- Lower TAT must still allow for proper peer review
- Some TAT so low, it casts doubts on quality

Rejection rates

Rejection rates at MDPI and Frontiers

Evolution of rejection rates, 2016-22

Frontiers: aggregate rates only; MDPI: computed from rejection rates at each journal

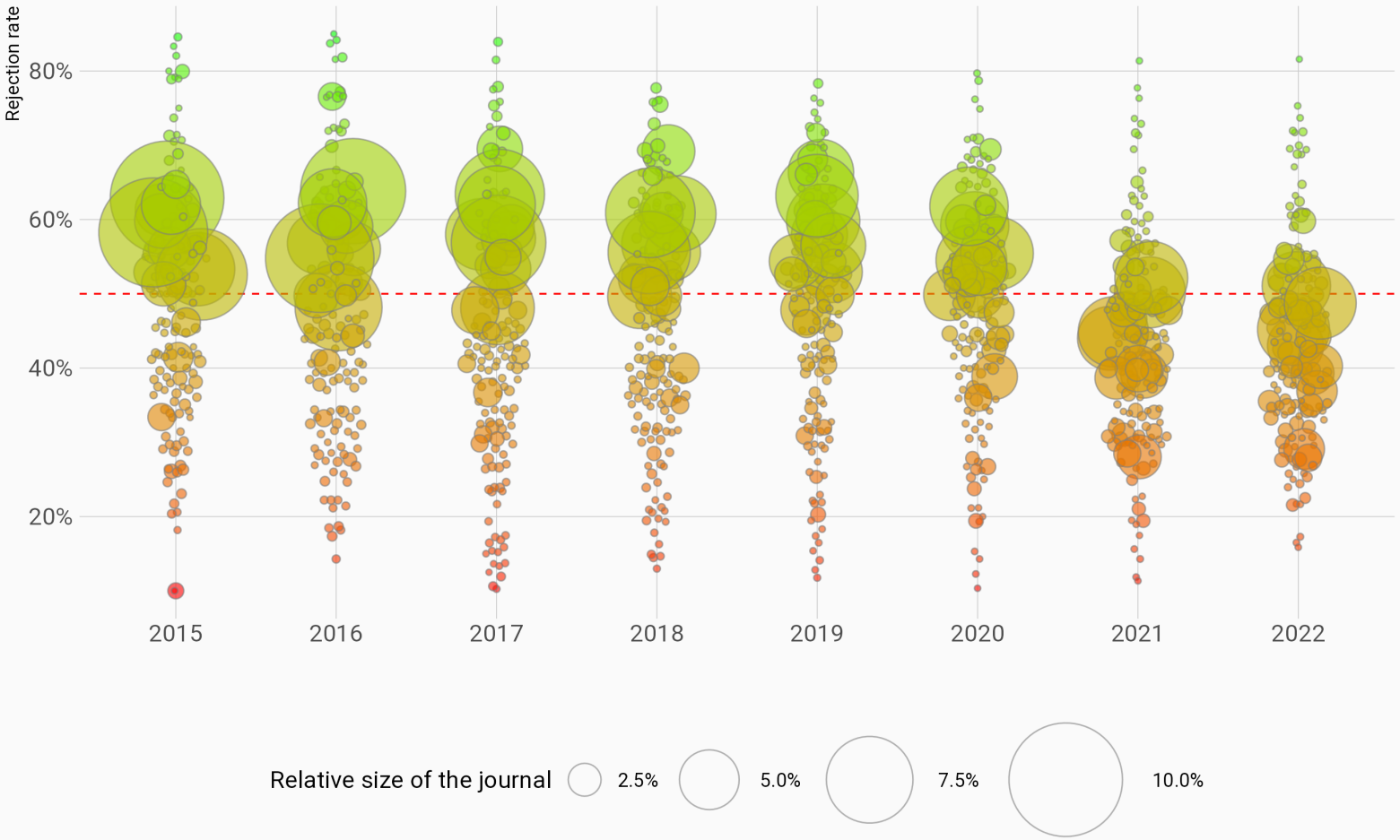


Source: data scraped on the publisherss website, analysis by MH, PC, DB, PGB

A focus on MDPI

Evolution of rejection rates by relative size of the journal at MDPI, 2015-22

Only journals existing in 2015, size relative to MDPI total publications in a given year



Source: data scraped on the publisherss website, analysis by MH, DB, PGB, PC

What's going on?

Trends:

- Rejection rates are decreasing at some key publishers
- Increasing at others
- Very little data

Why?

- Convergence of authors & OA publishers incentives

Threats

- Lower rejection rates might mean lower quality
- Risk of instability of quality signals

Impact Factor inflation

Indicators of impact: Impact factor, Scimago Journal Rank

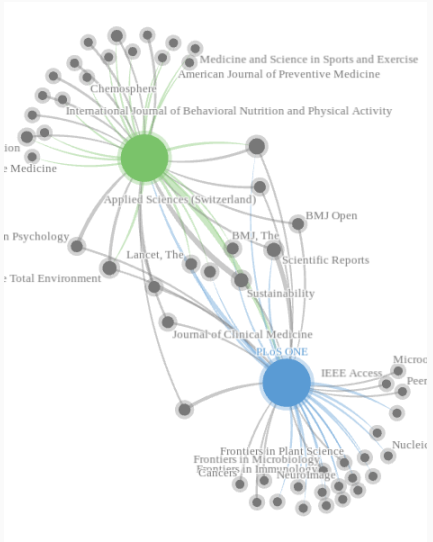
We measure Impact Factor Inflation as the ratio of IF to SJR

Impact Factor:

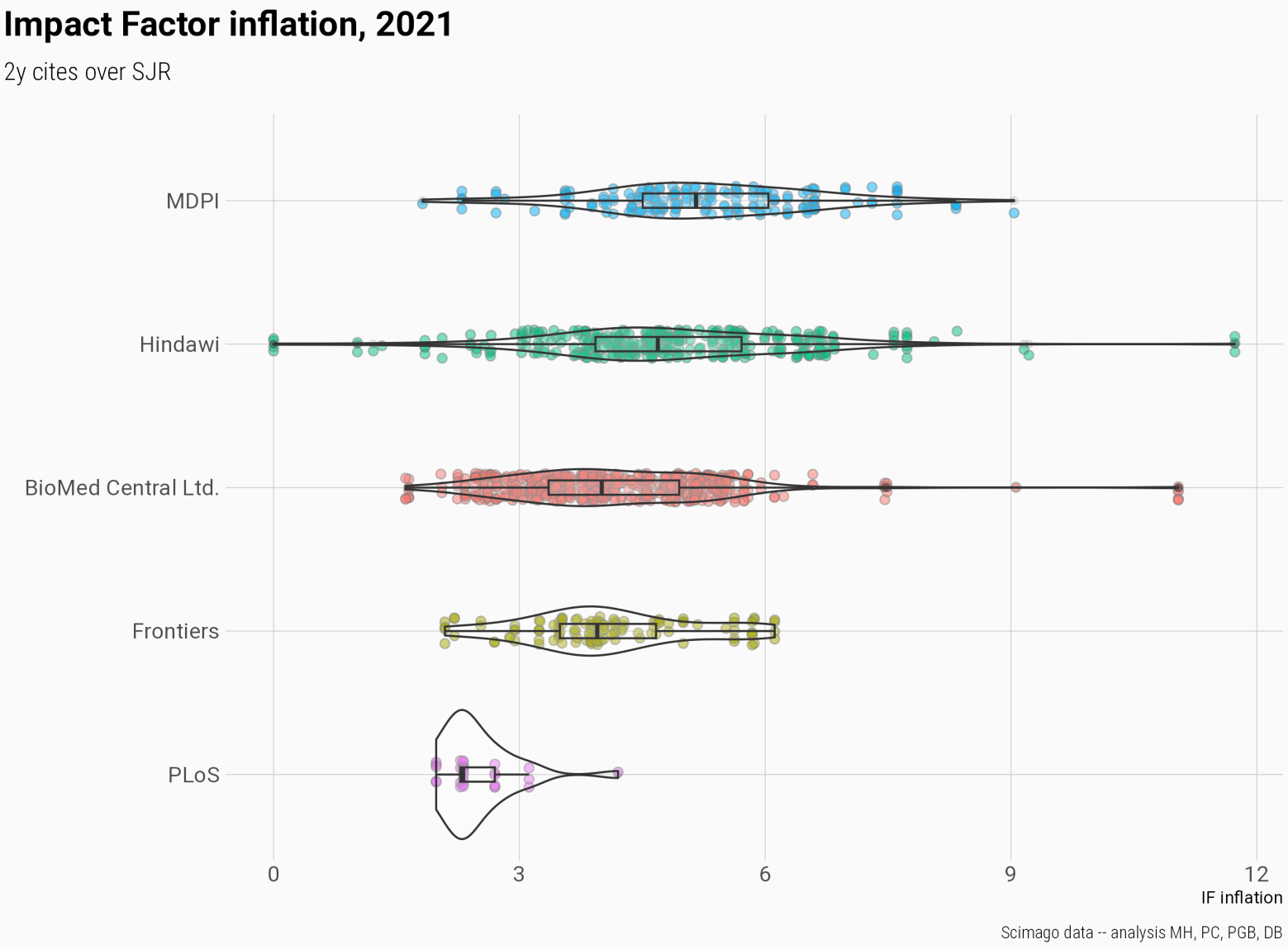
- cites/document at N years
- easily gamed

SJR: citation network counts

- Limits prestige from single source
- More prestige if cited by relevant journals
- Normalizes for field size
- Less easily gamed



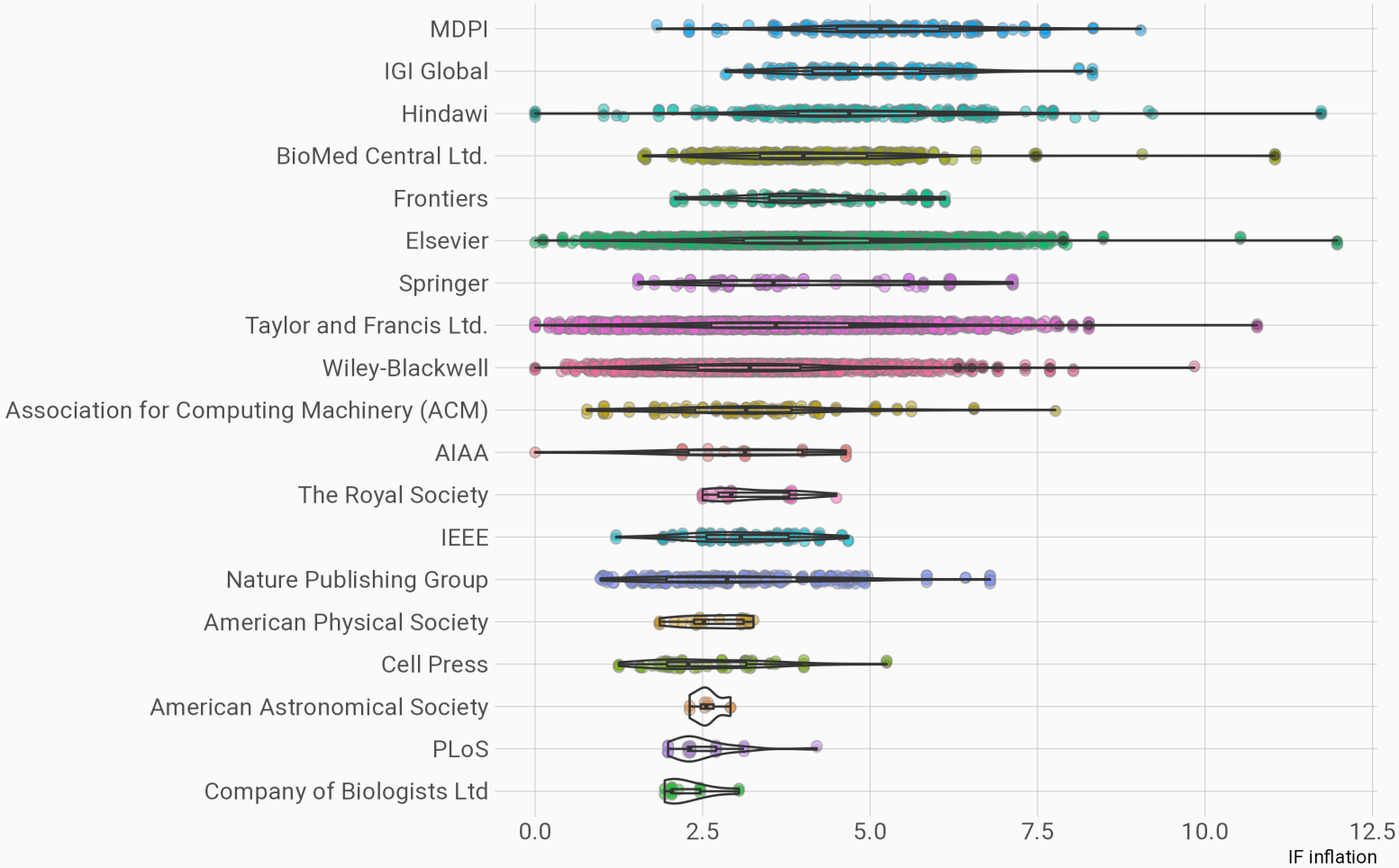
IF inflation 2021: some publishers



IF inflation 2021: more publishers

Impact Factor inflation, 2021

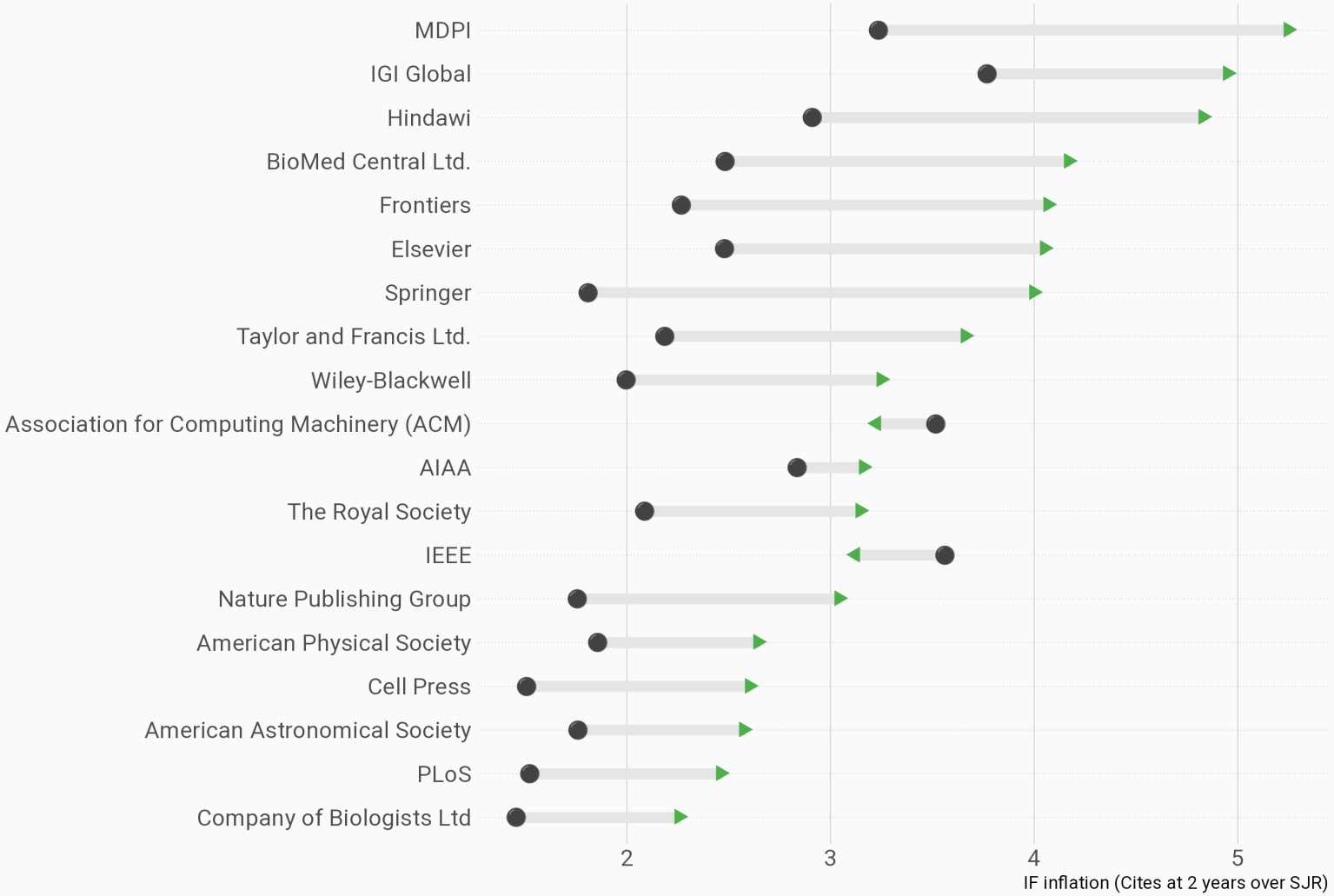
2y cites over SJR



Scimago data -- analysis MH, PC, PGB, DB

Evolution of IF inflation

Evolution of Impact Factor inflation: 2015 to 2021



Scimago data -- analysis MH, PC, PGB, DB

What's going on?

Trends:

- IF is inflating across the board – more so at some publishers

Why?

- Goodhart's law: *When a measure becomes a target, it ceases to be a good measure*

Threats

- Risk of instability of quality signals

A focus on MDPI

Strain indicators for MDPI

Strain indicators for MDPI: 2022 and evolution 2016-22

MDPI 20 largest journals as of 2022

	2022				CHANGE 2016-22			
JOURNAL	N	SHARE SI	TAT	REJECTION %	N	SHARE SI	TAT	REJECTION %
Int. J. Environ. Res. Public Health	17445	78%	42 days	45%	14.4x	+28pp	-27 days	-12pp
Sustainability	17394	77%	42 days	49%	12.8x	+10pp	-30 days	-13pp
Int. J. Mol. Sci.	16482	100%	35 days	51%	7.6x	+1pp	-19 days	-13pp
Appl. Sci.	13229	84%	38 days	43%	28x	+19pp	-23 days	-34pp
Sensors	10149	100%	38 days	40%	4.5x	+1pp	-30 days	-15pp
Energies	9843	80%	37 days	39%	8.9x	+10pp	-38 days	-22pp
Materials	9184	78%	37 days	29%	8.9x	+6pp	-17 days	-27pp
Molecules	9144	86%	34 days	37%	5.2x	+1pp	-11 days	-11pp
J. Clin. Med.	7641	99%	39 days	44%	65.9x	+17pp	-27 days	-2pp
Remote Sens.	6479	83%	43 days	55%	6.3x	+27pp	-43 days	-4pp
Cancers	6359	87%	39 days	52%	57.8x	+2pp	-29 days	18pp
Polymers	5625	100%	33 days	28%	12.7x	+23pp	-17 days	-13pp
Nutrients	5405	100%	34 days	47%	6.4x	+42pp	-27 days	-2pp
Mathematics	4931	86%	36 days	60%	71.5x	+35pp	-52 days	-22pp
Nanomaterials	4540	84%	32 days	35%	18.5x	+1pp	-20 days	-30pp
Electronics	4319	93%	35 days	42%	44.5x	+11pp	-36 days	-34pp
Water	4245	98%	40 days	40%	7x	+36pp	-36 days	-19pp
Foods	4187	99%	35 days	47%	48.7x	+19pp	-25 days	19pp
Cells	4181	91%	42 days	36%	92.9x	+2pp	-12 days	17pp
Animals	3666	98%	43 days	43%	46.4x	+54pp	-41 days	18pp

Source: data scraped on the publisher's website by @paolocrosetto & @pagomba, analysis DB, MH, PC, PGB

What's going on?

MDPI is the publisher putting by far more strain on the system

- Impressive, exponential growth
- MDPI journals mostly a collection of loosely connected SI
- TAT at the lower bound of the credible interval
- Rejection rates decreasing across the board
- Highest IF inflation

So, what?

Provisional lessons to be learned

- We think we have identified 5 indicators of strain that can guide our understanding of the scientific publishing system.
- We refrain from assigning "predatory" labels, but we see high strain imposed over several dimensions as signals of a critical situation.
- Some publishers, as some *very* prolific authors, might be stretching the system too far.

Thank you!