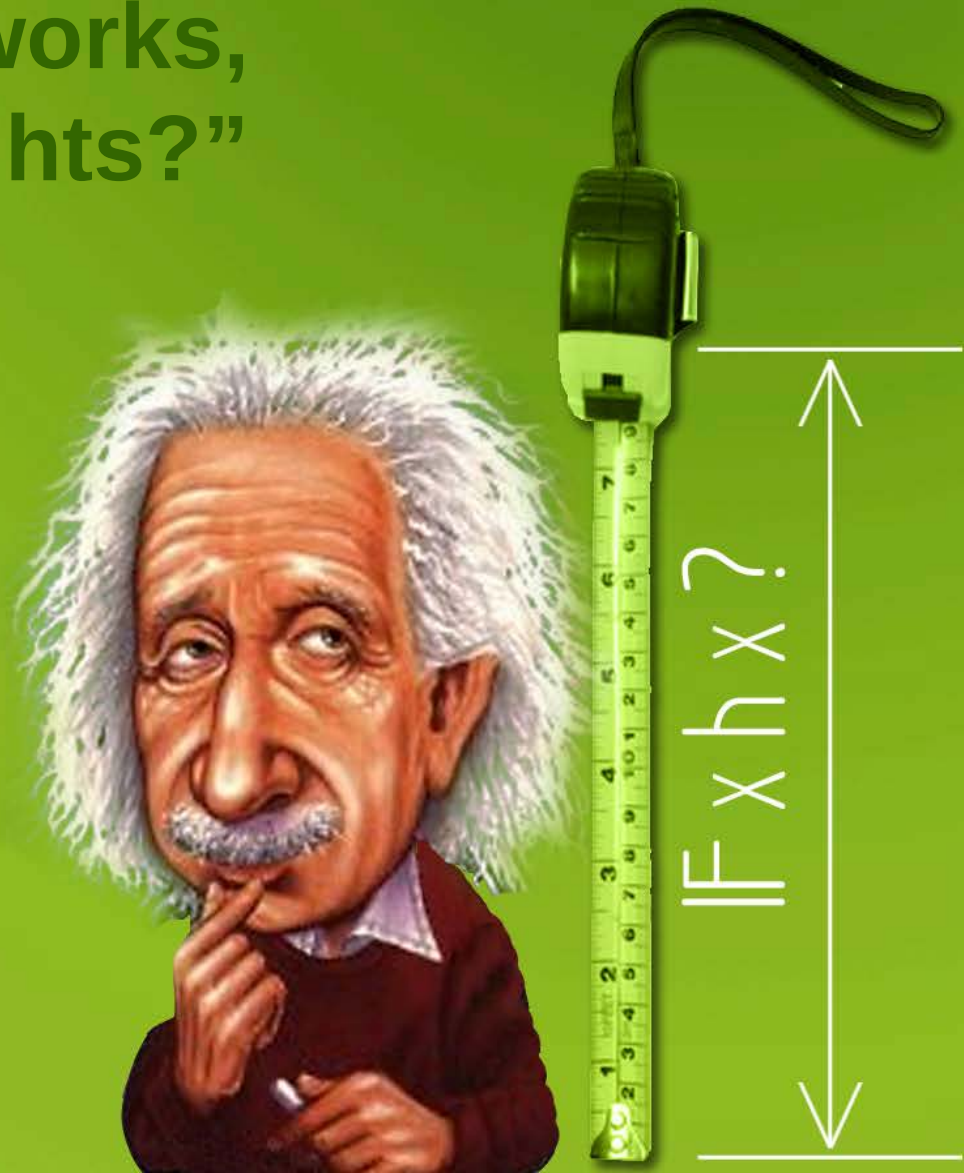
“Quantifying scientific impact: networks, measures, insights?”

The downloads as a measure of **attractiveness** of scientific publication

Olesya Mryglod



Yurij Holovatch

Laboratory for Statistical Physics of Complex Systems, Institute for Condensed Matter Physics of the Nat. Acad. Sci. Ukraine

Ralph Kenna

Applied Mathematics Research Centre, Coventry University, Coventry, England

The downloads as a measure of **attractiveness** of scientific publication

Olesya Mryglod

[MO, Kenna R., Holovatch. EPL. - 2014. - 108. – 50011; ArXiv: 1409.7889]

What is scientific impact?

…Quality

Efficiency

Performance

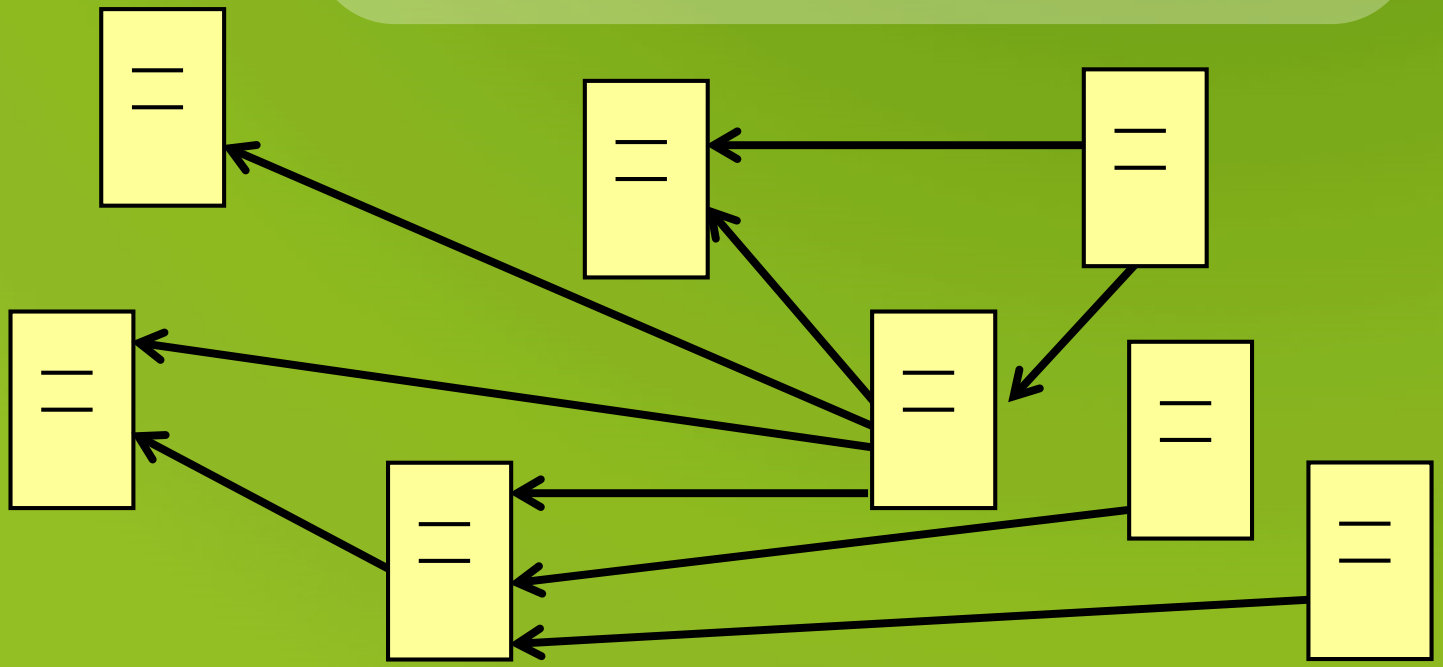
Popularity

…

Citations – indirect indicator of research **quality**

- Goodhart's law: "When a measure becomes a target, it ceases to be a good measure."
- Other nuances: technical & methodological imperfections, "sleeping beauties", "Matthew effect", negative citations, etc.

Moreover…



What is scientific impact?

...Quality

Efficiency

Performance

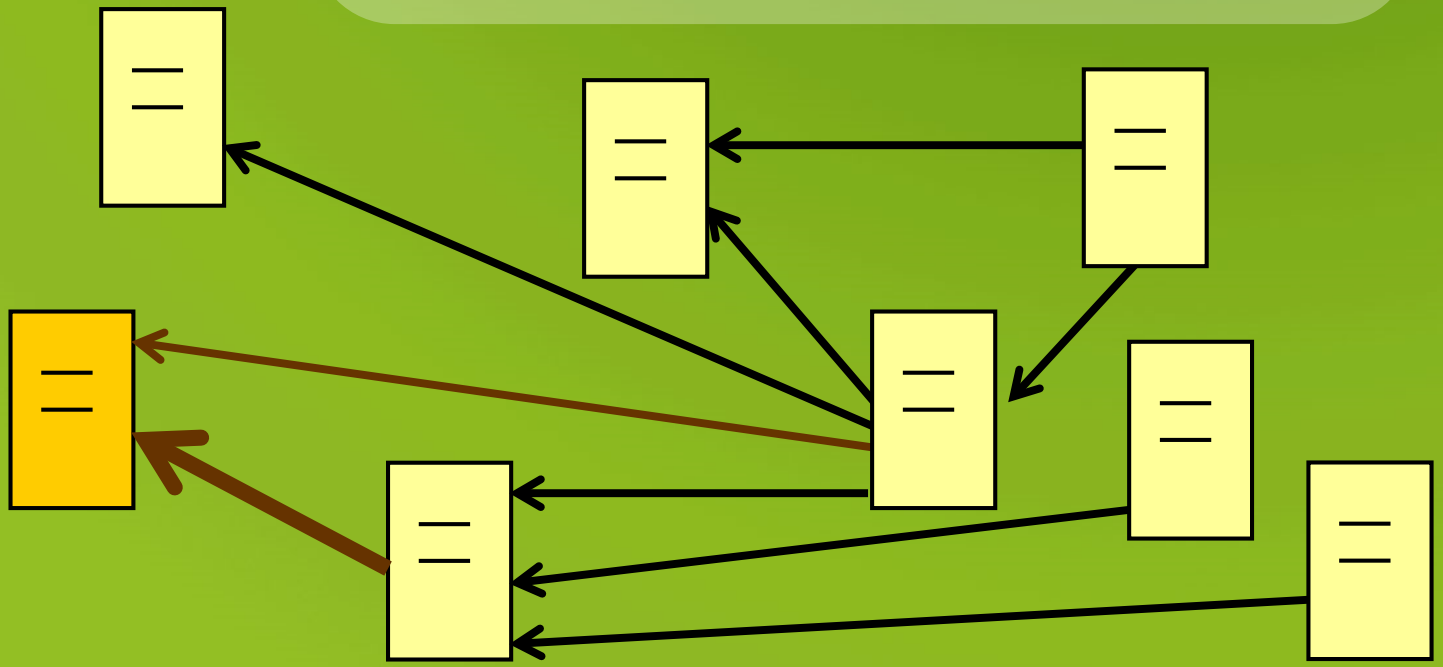
Popularity

...

Citations – indirect indicator of research **quality**

- Goodhart's law: "When a measure becomes a target, it ceases to be a good measure."
- Other nuances: technical & methodological imperfections, "sleeping beauties", "Matthew effect", negative citing, etc.

Moreover…



Attractiveness

altmetrics?
(downloads)

attention
(acquaintances)

Popularity

citations

friends

Prestige

“important”
citations

authoritative
friends

Popularity Prestige

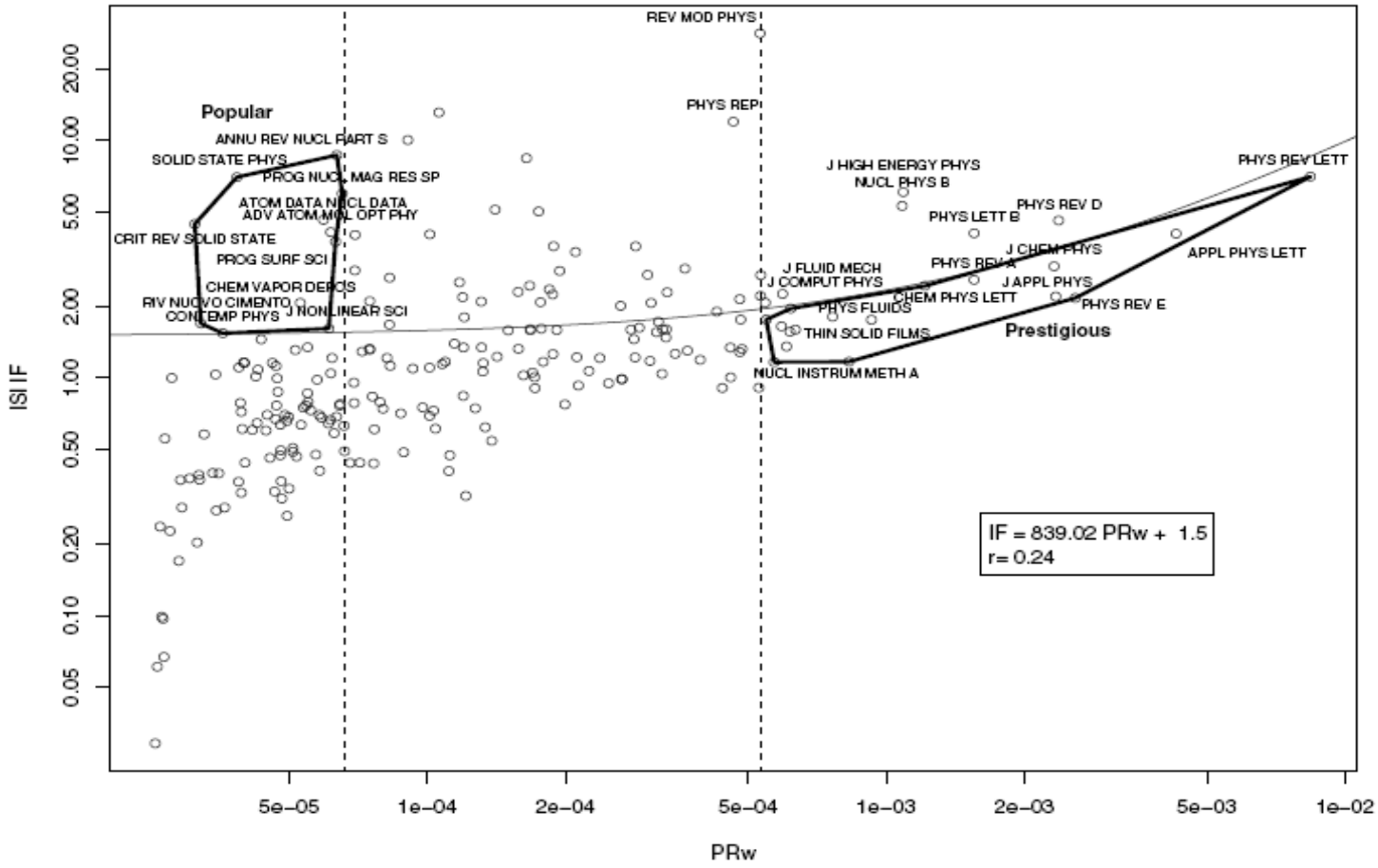
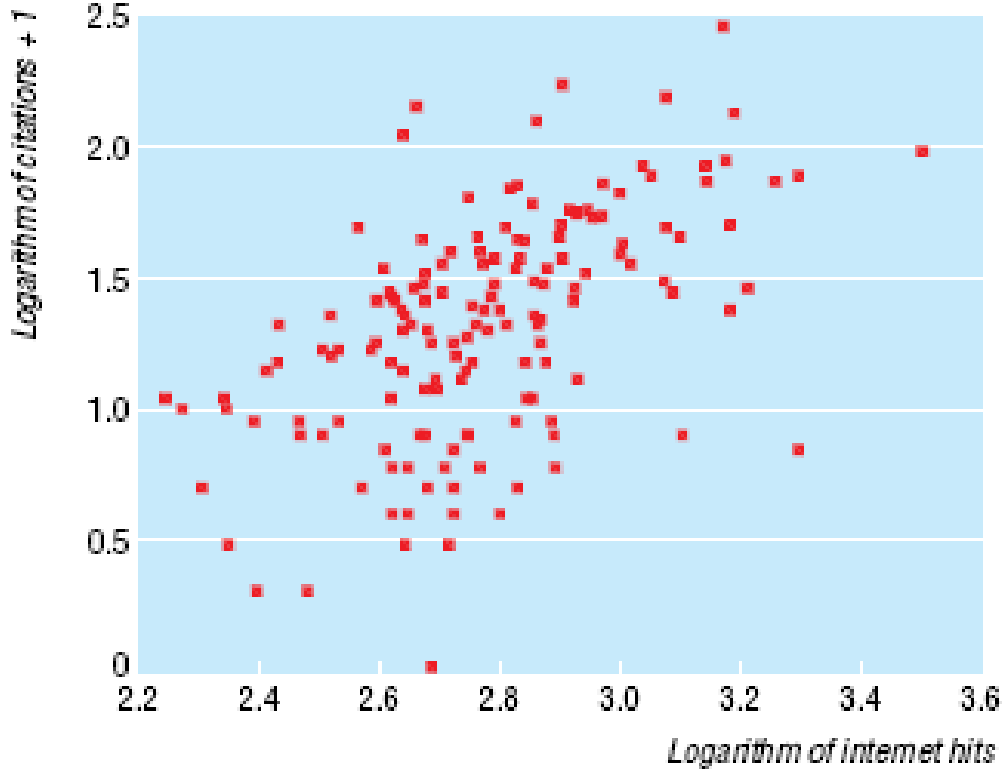


Figure 2: Popular and Prestigious Journals in Physics.

[Bollen, J., Rodriguez, M. A., de Sompel, H. V., 2006. Journal status. Scientometrics 69 (3), 669–687.]

Attractiveness Popularity



Relation between citations and internet hits for 153 papers in volume 318 of the *BMJ* (1999)

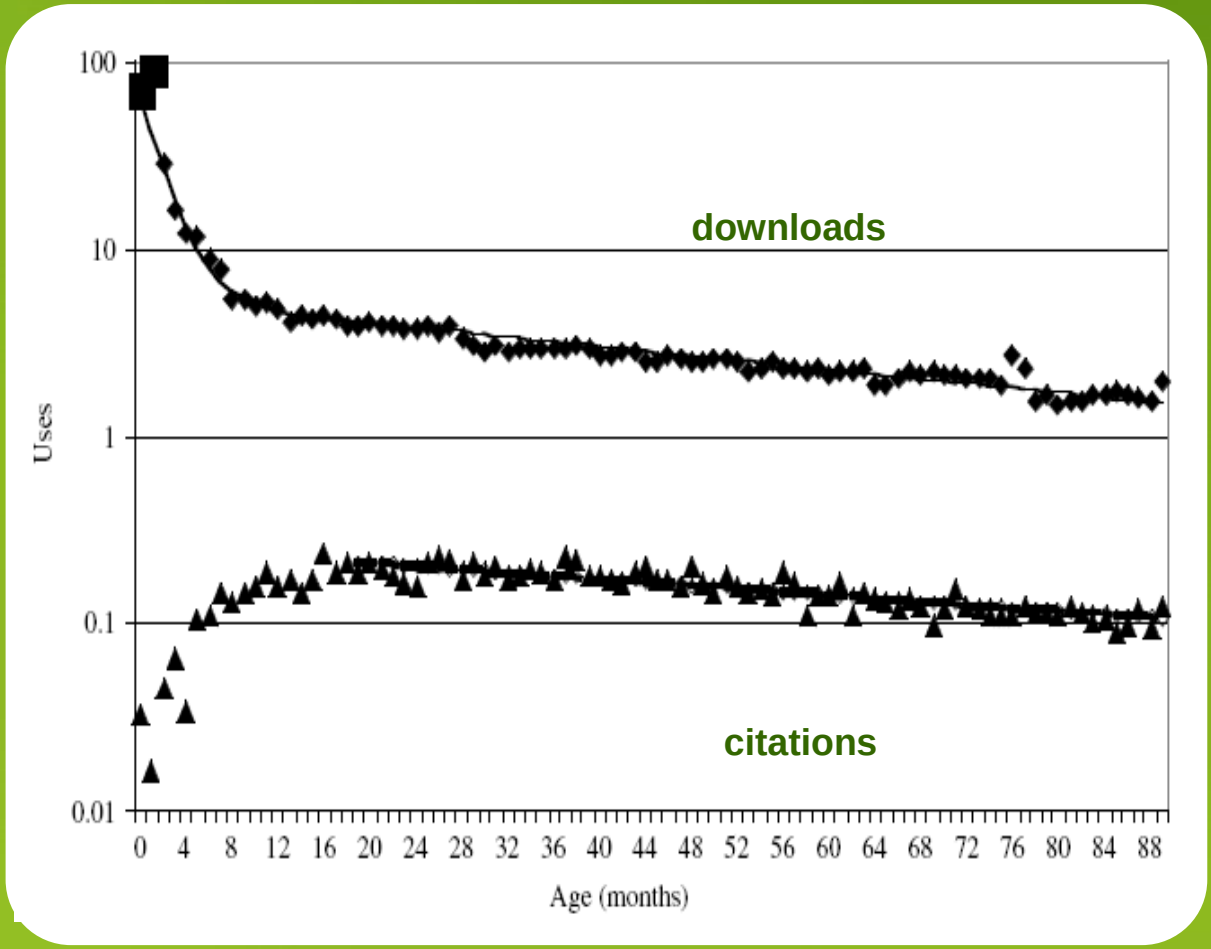
- To download ≠ to read
- To read ≠ to cite
- Citations – interest after reading, downloads – interest before reading
- Different motives for downloads, wider audience of readers
- …

[T. Perneger. “Relation between online “hit counts” and subsequent citations: prospective study of research papers in the BMJ” BMJ 2004;329:546–7]

“Downloads influence citations and that citations influence downloads”

“To examine the relationships between downloads and citations, insight into their age distributions is crucial”

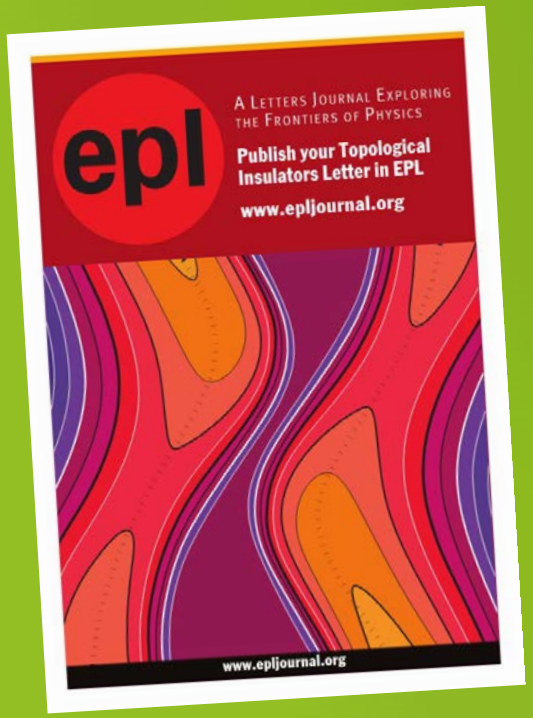
[Moed HF. Statistical relationships between downloads and citations at the level of individual documents within a single journal. J AmSoc Inf Sci Technol 2005;56:1088-97]



Motivation

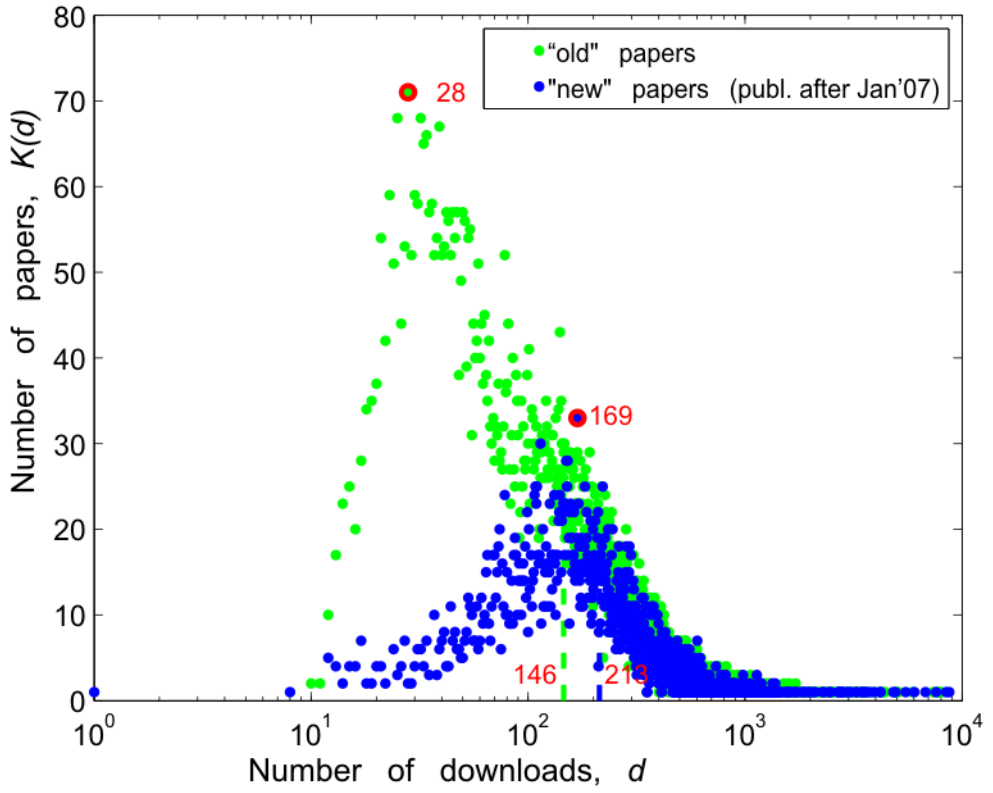
- Is attractiveness a complementary feature of scientific publications (e.g., better correlated with mass media “impact”)?
- How it correlates with other measures/metrics?
- Do typical download patterns exist? What are the reasons? How to describe downloads ageing?

Our data



Download statistics (Jan 2007 – Jun 2013): 15 431 papers

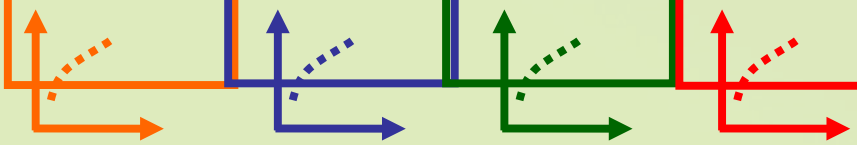
- 9 895 – “old” (published earlier and available online later, after Jan 2007)
- 5 536 – “new” (published online instantly); ~7.6% are open access



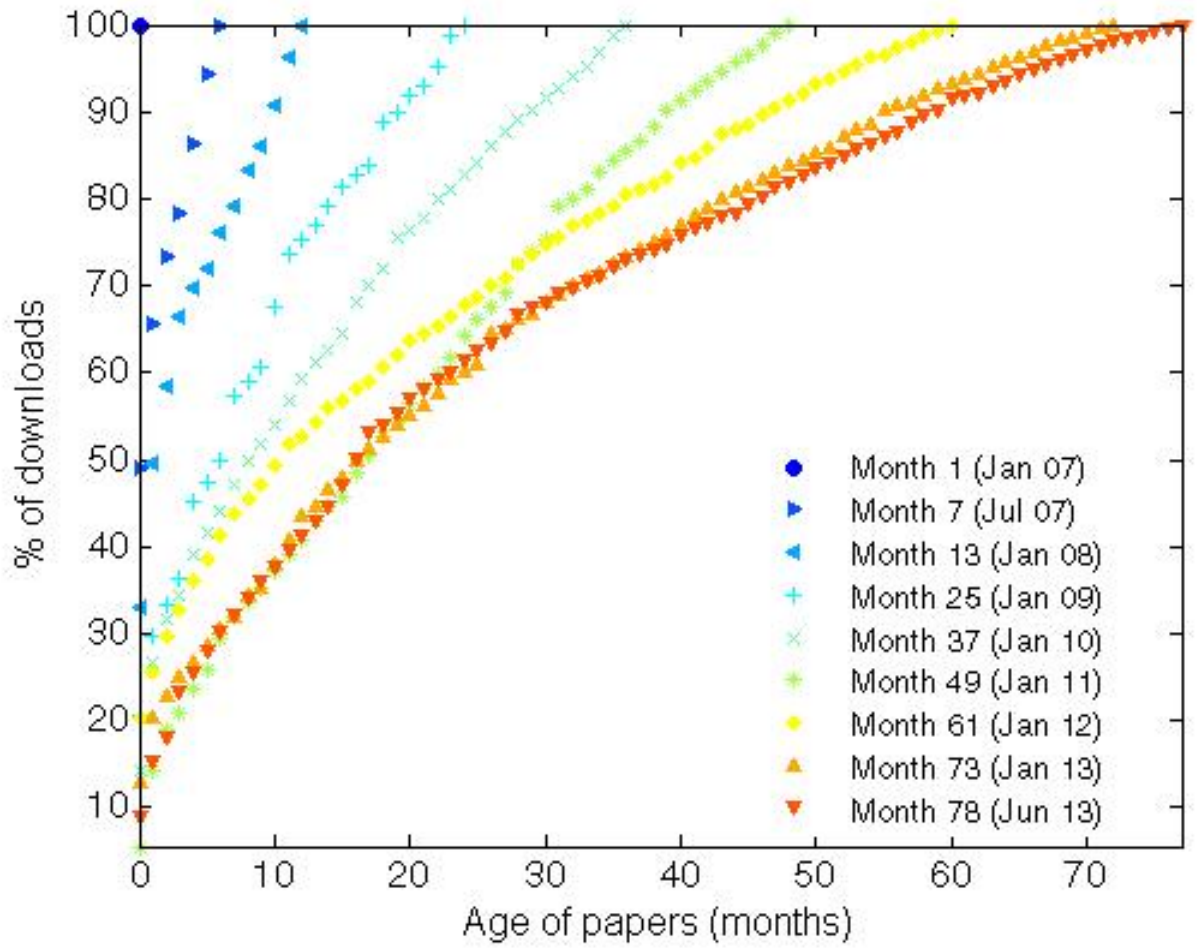
Synchronous approach

Jan'07 | Feb'07 | Mar'07 | Apr'07 | ...

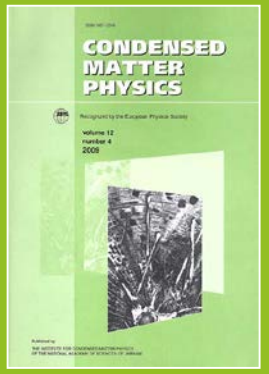
	1	2	3	4	5	6	7	8
722	0	65	37	34	41	13	5	8
723	0	43	27	39	8	8	5	1
724	49	69	27	29	14	7	13	4
725	28	70	30	42	28	12	5	7
726	0	66	16	15	6	2	4	2
727	0	100	46	29	28	14	12	17
728	0	111	79	20	7	1	0	3
729	0	82	99	35	25	17	6	8
730	0	92	131	27	21	2	5	3
731	0	61	69	26	16	2	3	0
732	0	49	85	59	27	15	9	15
733	0	50	79	66	40	6	8	8
734	0	0	80	36	3	4	1	1
735	0	0	73	29	9	10	6	5
736	0	23	49	29	10	13	5	5
737	0	23	53	37	16	8	3	6
738	0	50	44	35	18	4	0	1
739	0	45	39	29	22	4	2	9
740	0	39	29	22	13	3	2	0
741	0	58	86	67	24	11	4	6
742	0	54	51	28	16	8	7	13
743	0	88	90	38	46	18	16	14

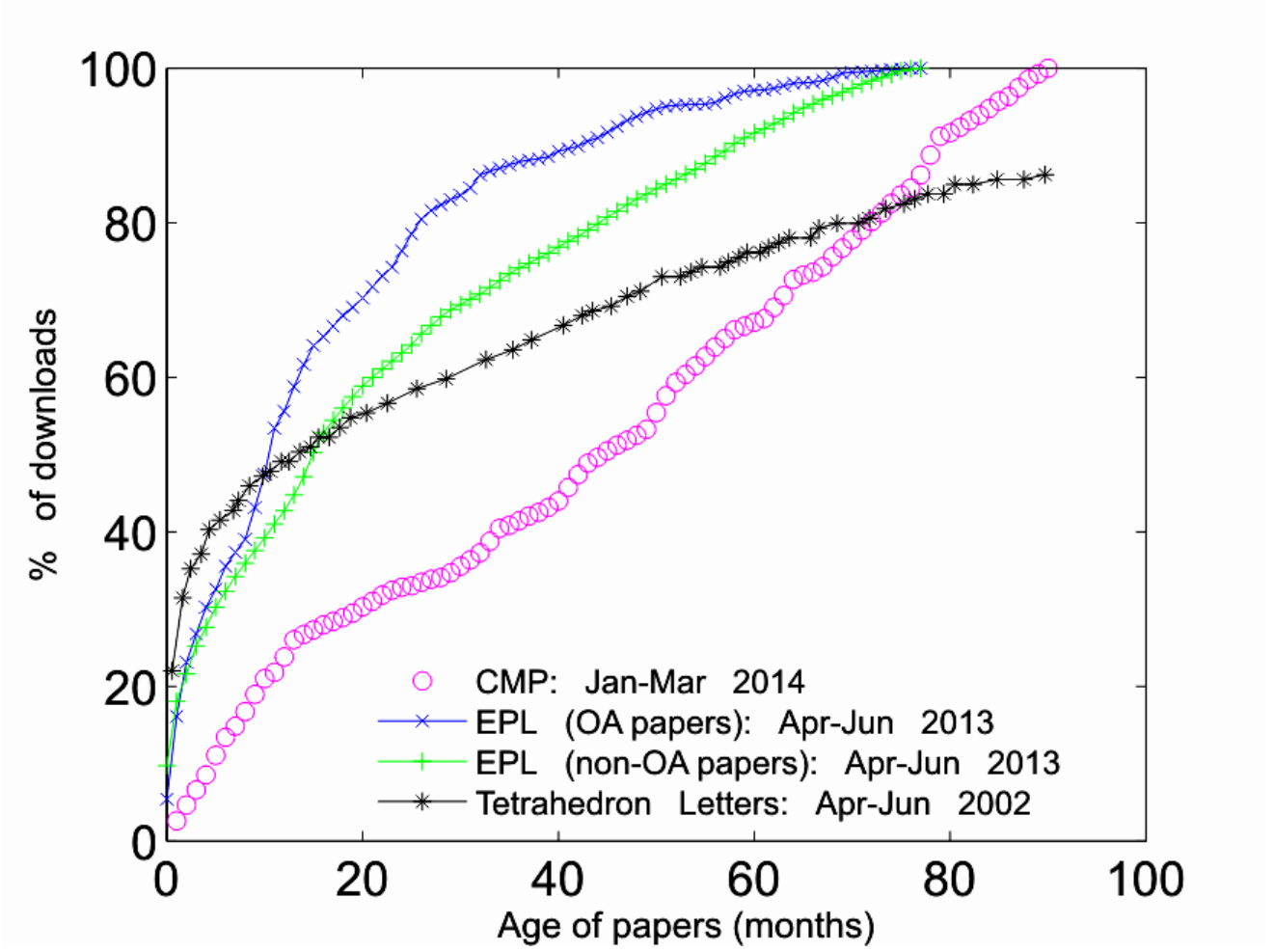


What minimal paper age is required to get *y*% of downloads within a particular time window?

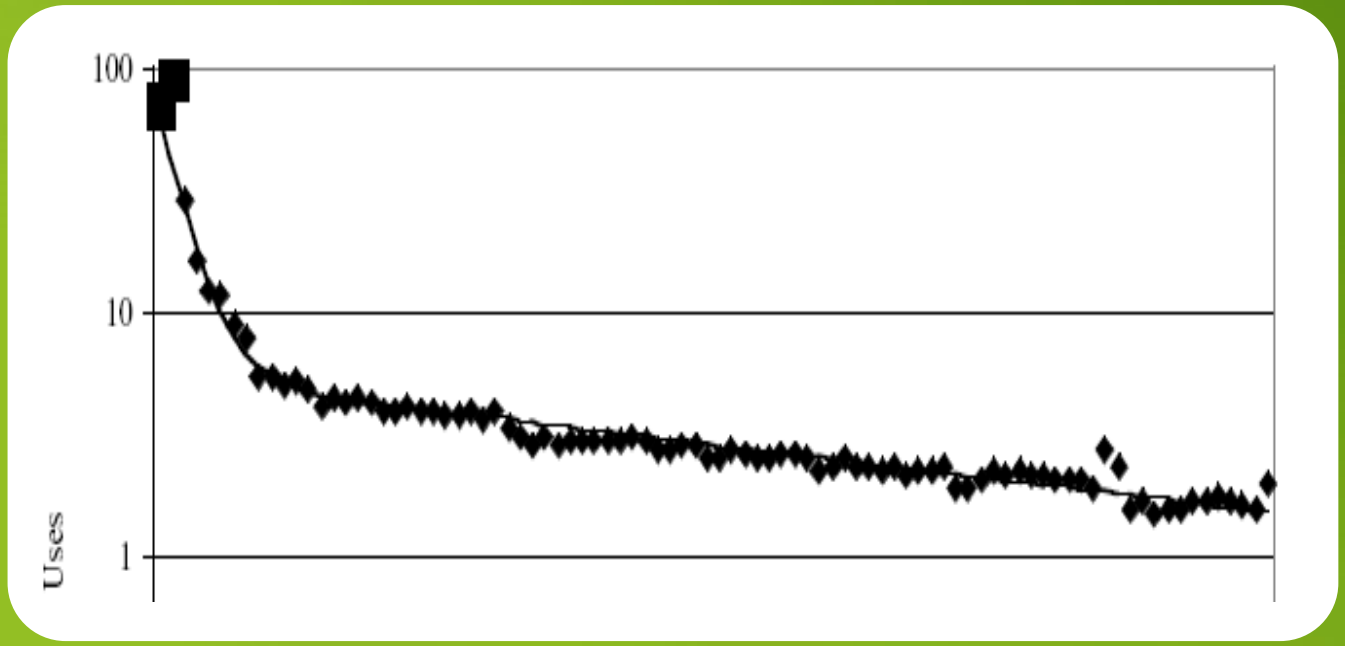


~5000 of non-OA EPL papers
~400 of OA EPL papers
~1200 of CMP papers
~1200 of Tetrahedron letters papers





Model of downloads obsolescence



$$U(t) = U(0) \cdot \left(A\mathrm{e}^{-b_1 t} + (1 - A)\mathrm{e}^{-b_2 t} \right)$$

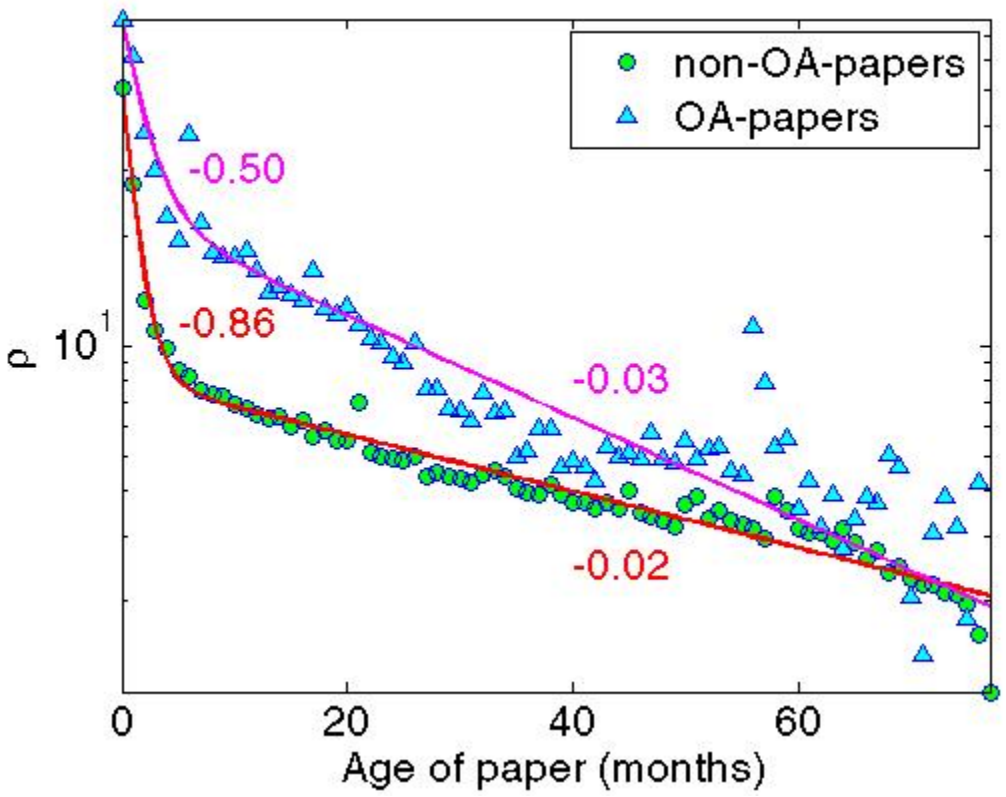
$$0 \le A \le 1 \qquad A \approx 0.92$$

$$b_1 > 0 \qquad b_1 \approx 0.50$$

$$b_2 > 0 \qquad b_2 \approx 0.014$$

[Moed HF. Statistical relationships between downloads and citations at the level of individual documents within a single journal. J AmSoc Inf Sci Technol 2005;56:1088-97]

Model of downloads obsolescence



Non-OA papers:

$A \approx 0.84$

$b_1 \approx 0.86$

$b_2 \approx 0.02$

OA papers:

$A \approx 0.71$

$b_1 \approx 0.50$

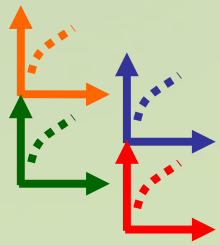
$b_2 \approx 0.03$

[Moed 2005] $A \approx 0.92$ $b_1 \approx 0.50$ $b_2 \approx 0.014$

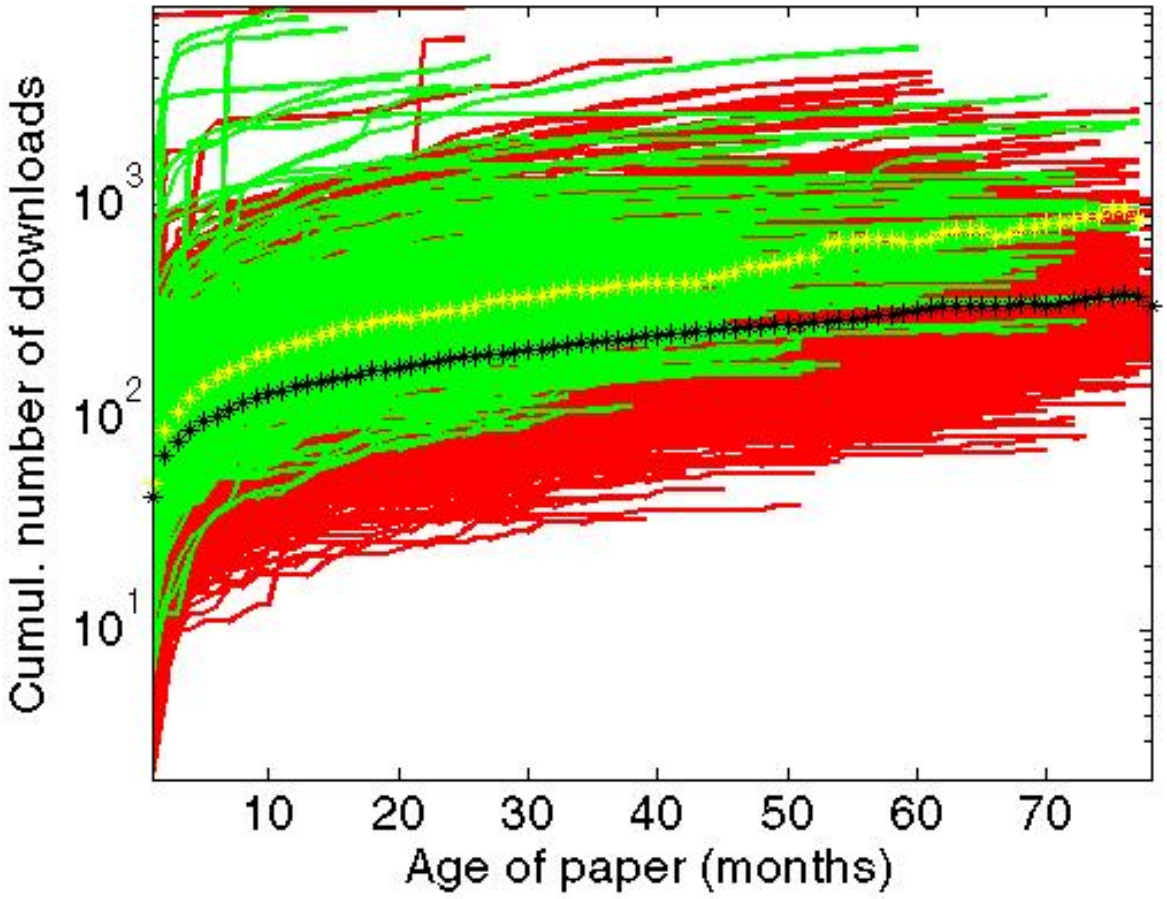
Diachronous approach

Jan'07 Feb'07 Mar'07 Apr'07 ...

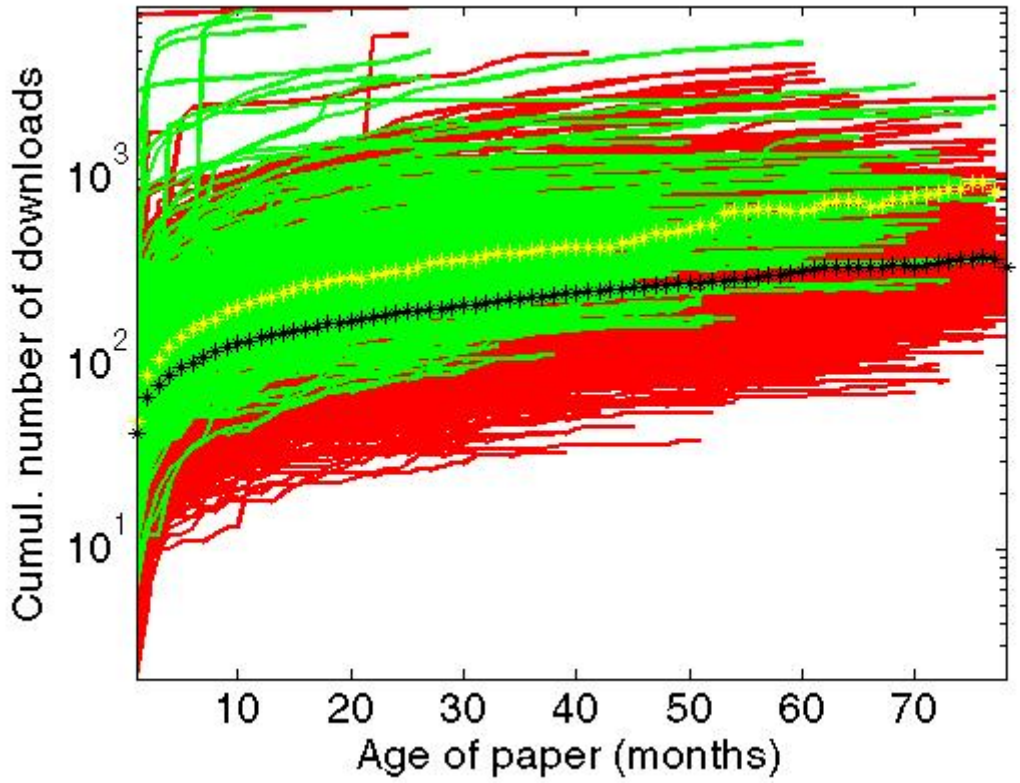
	1	2	3	4	5	6	7
722	0	65	37	34	41	13	5
723	0	43	27	39	8	6	5
724	49	69	27	29	14	7	13
725	28	78	38	42	28	12	5
726	0	66	16	15	6	2	1
727	0	100	46	29	28	14	12
728	0	111	79	20	7	1	0
729	0	82	99	35	25	17	6
730	0	92	131	27	21	2	5
731	0	61	69	26	16	2	3
732	0	49	85	59	27	15	9
733	0	50	79	66	40	6	8
734	0	0	80	36	3	4	1
735	0	0	73	29	9	10	6
736	0	23	49	29	10	13	5
737	0	23	53	37	16	8	3
738	0	50	44	35	18	4	0
739	0	45	39	29	22	4	2
740	0	39	29	22	13	3	2
741	0	58	86	67	24	11	4
742	0	54	51	28	16	8	7
743	0	88	90	38	46	18	16



Diachronous approach



"Bursty" (exotic?) papers detection



"Bursty papers" – with strong (how much?) deviations from an individual paper's typical download pattern

$\sigma_i(T)$

– STD of T-months-old paper *i* ("red curve")

$\sigma_{\mathbf{med}}(T)$

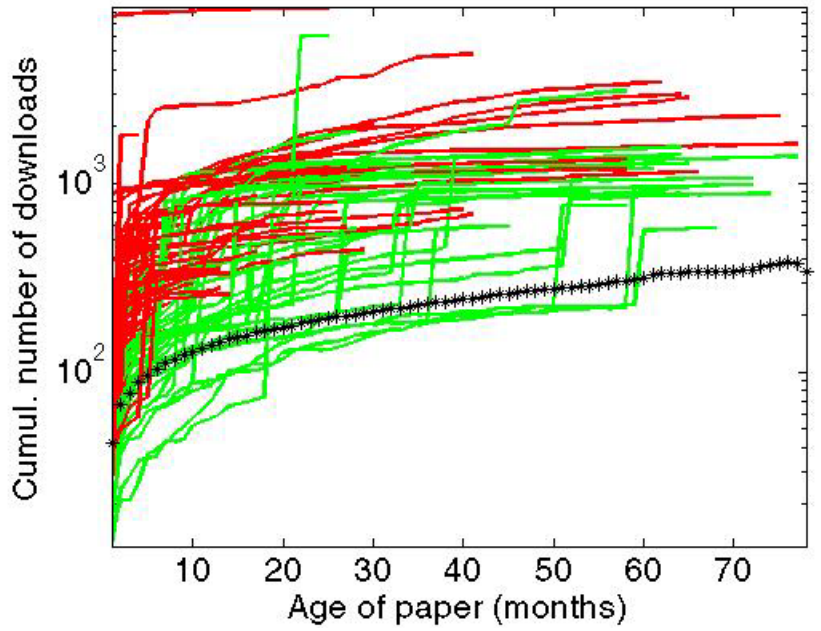
– STD of typical T-months-old paper

$$\Delta_i(T) = \left|\sigma_i(T) - \sigma_{\mathbf{med}}(T)\right|$$

$$\frac{\Delta_i(T)}{\left\langle \Delta_i(T) \right\rangle} = 5$$

for ~**2%** of papers this value ≥ 5

“Bursty” (exotic?) papers detection



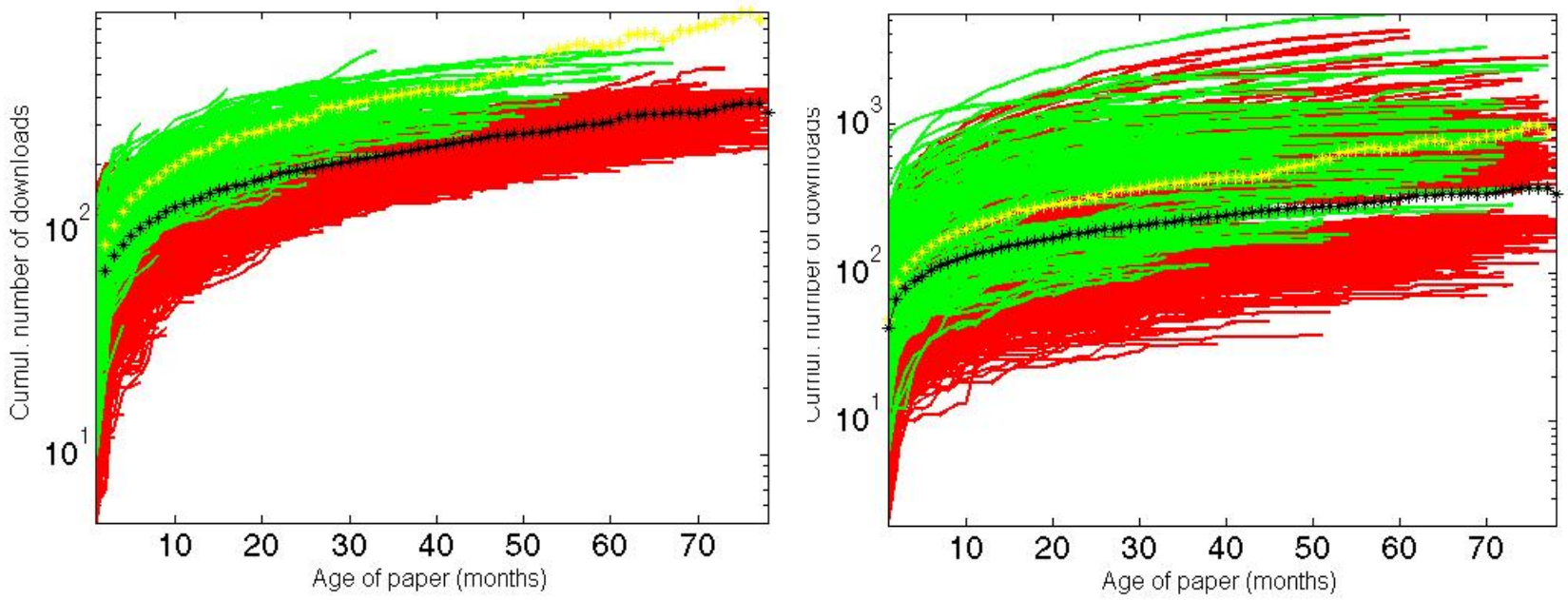
“Sleeping beauties” – papers with “delayed” bursts

1. "Observing different quantum trajectories in cavity QED"
2. "Large anisotropic magnetoresistance across the Schottky interface in all oxide ferromagnet/semiconductor heterostructures"
3. "Maintain the structural controllability under malicious attacks on directed networks"
4. "Anomalous molecular dynamics in the vicinity of a conical intersection"
5. "Soccer: Is scoring goals a predictable Poissonian process?"
6. "Entropy generation in biophysical systems"
7. "Preferential attachment in the interaction between dynamically generated interdependent networks"
8. "Shock-driven jamming and periodic fracture of particulate rafts"
9. "Generation of pure spin currents by superconducting proximity effect in quantum dots"
10. "Coupling of magnetic edge states in Li-intercalated bilayer and multilayer zigzag graphene nanoribbons"

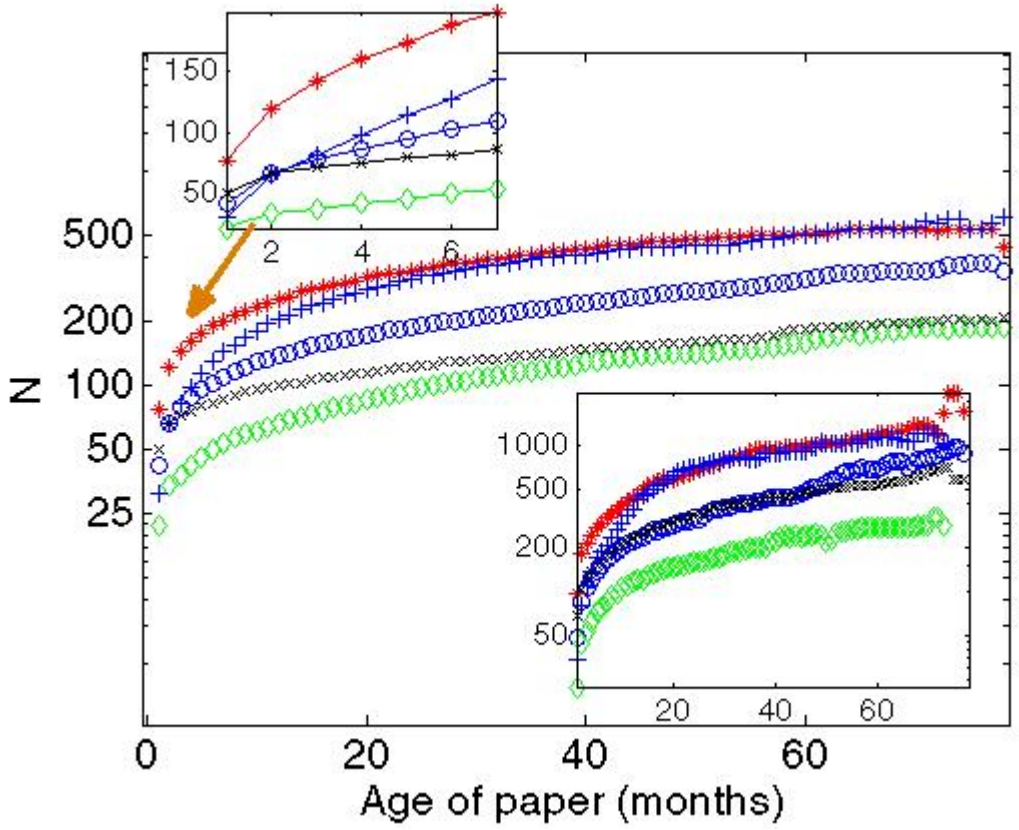
Overall attractiveness of papers

60% - “typical” papers (RMSD < $RMSD_c$)

40% - “non-typical”



Overall attractiveness of papers



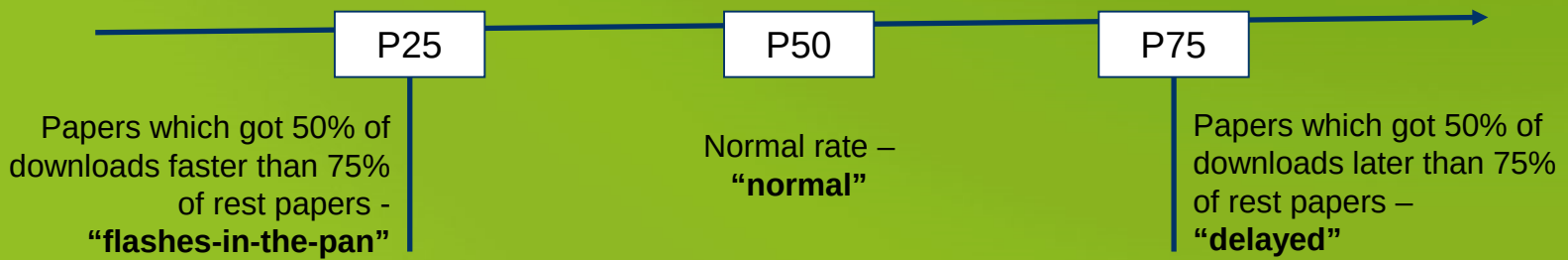
- **~18-22% (of all papers) – persistently higher cumulative download values**
- **~12-15% (of all papers) – persistently less downloaded**
- **~10% (of all paper) – attractiveness is changing with time**

Diversity of download patterns: durability-based classification

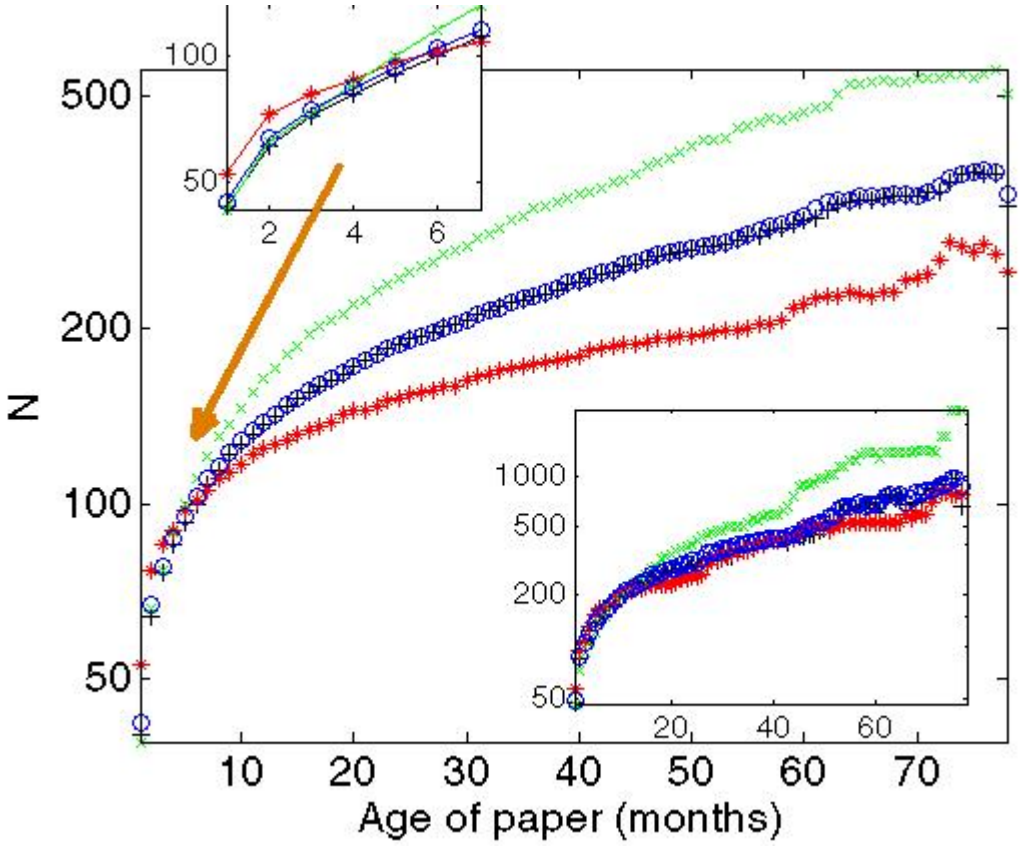
[Costas R., van Leeuwen T.N., van Raan A.F.J. J Am Soc Inf Sci Technol 010;61:329-339; Scientometrics 2011; 89, No. 1, 177-205]

Notion of "durability"

1. The half-life of downloads $M^{50}(t)$ is defined for each paper: i.e., number of months by which it achieves 50% of its total downloads
2. The percentiles P25 and P75 of $M^{50}(t)$ for papers of different age *t* are calculated



Diversity of download patterns: durability-based classification



~62–65% – “normal”

~18% – “flashes-in-the-pan”

~20% – “delayed”

Categorisation by burstiness		
4986 (377) papers	98% are "non-bursty" papers	
	2% (3%) are "bursty" papers	1% (1%) are "sleeping beauties" 1% (2%) burst early

Categorisation by half-lives (ageing of attractiveness)	
4890 (367) non-bursty papers	62% (65%) exhibit usual ageing behaviour 18% (17.5%) are flashes in the pan 20% (17.5%) exhibit delayed activity

Categorisation by overall attractiveness		
4890 (367) non-bursty papers	60% (50%) have typical overall attractiveness	
	40% (50%) are atypical	18% (22%) are more attractive 12% (15%) are less attractive 10% (13%) change attractiveness

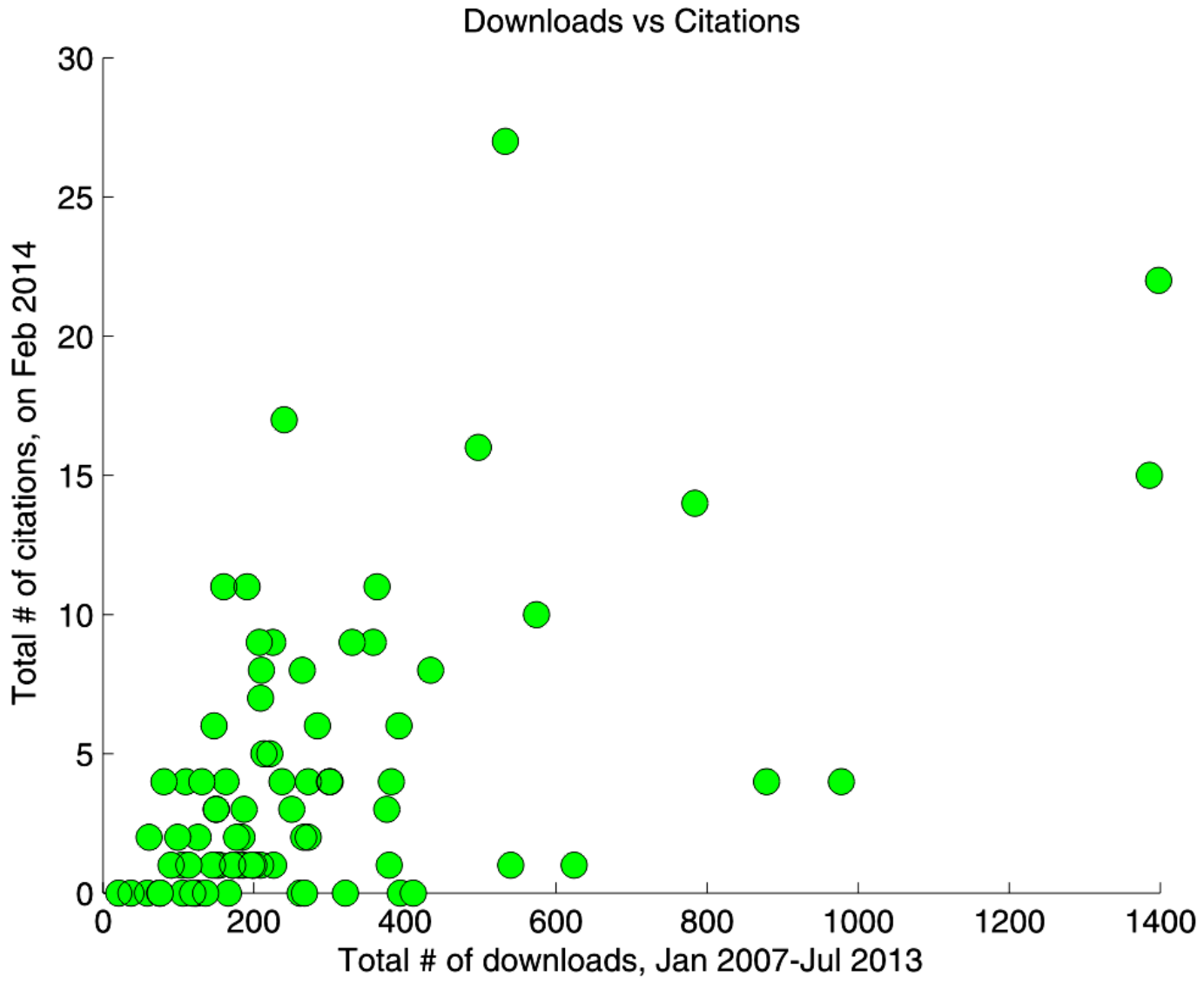
Summary

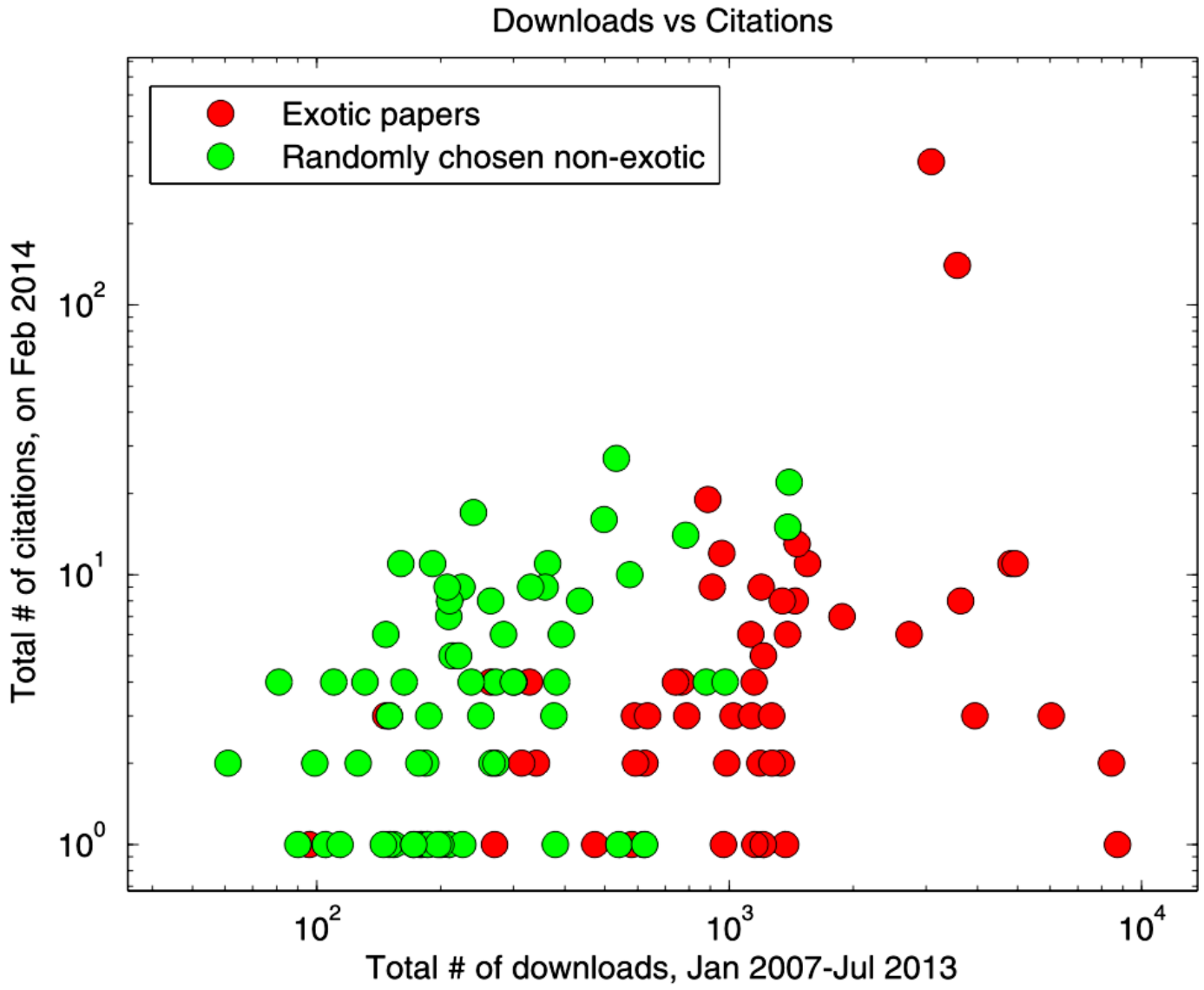
- Several approaches to categorize paper by download patterns were discussed
- The synchronous approach provides a possibility to compare the downloading processes of different journals
- While the download patterns for OA and non-OA papers are slightly different, the corresponding proportions of papers within categories are very similar to those for non-OA papers.
- The two-factor model of download aging has predictive power for the long-term behaviour of paper downloads from the journal.

Perspectives

- Correlation between downloads and citations
- Download analysis for other journals, disciplines, different types of publications
- Analysis of possible factors which influence on the attractiveness of scientific papers
- Topical analysis of publications within different categories

Thank you for attention !


Downloads vs Citations
Total # of citations, on Feb 2014
Total # of downloads, Jan 2007-Jul 2013
0
5
10
15
20
25
30
0
200
400
600
800
1000
1200
1400


Downloads vs Citations
Exotic papers
Randomly chosen non-exotic
Total # of citations, on Feb 2014
10^0
10^1
10^2
10^2
10^3
10^4
Total # of downloads, Jan 2007-Jul 2013