The Rise of Preprints in Physics, Mathematics, and Computer Sciences

Paul Ginsparg

Physics and InfoSci, Cornell University

I will attempt to provide a series of informative answers to ill-posed questions, and to dispel the usual misconceptions, misinformation, and misgivings.

ASAP BIO at HHMI - 16 Feb 2016

Déjà vu all over again

PG, HMS Beagle: The BioMedNet Magazine, Issue 61 (3 Sep 1999):

"My own involvement in what evolved to become the current NIH proposal was a talk I gave in December '98 at the Banbury Center in Cold Spring Harbor, where I encouraged the biological and life scientists present to move in the direction of broader global archiving systems. (And frankly the participants at this meeting, many very dissatisfied with the current system, needed vanishingly little encouragement -- it's thrilling if the biomedical people are ready to join the 1990s, better late than never...) I described how the Los Alamos e-print archives, since their inception in '91 (where "e-print" denotes self-archiving by the author), have become a major forum for dissemination of results in physics and mathematics, and suggested some of what we foresee as the advantages of a unified global archive for research in these fields. I also pointed out that these e-print archives are entirely scientist driven, and are flexible enough either to co-exist with the pre-existing publication system, or help it evolve to something better optimized for researcher needs. In particular, the rapid dissemination they provide is not in the least inconsistent with concurrent or post facto peer review, and in the long run provides a possible framework for a far more functional archival structuring of the literature than is provided by current peer review methodology. The subsequent direct NIH involvement became an enormous opportunity to build on the existing resources at NCBI, and a potential model for other funding agencies, provided major miscues could be avoided. My primary comments since have been simply to encourage direct communication to the target scientists, trying to ensure their direct support and participation, rather than dealing through intermediaries whose vested interests might hinder more rapid progress. …"

… [every two years or so] …

Geneticists eye the potential of arXiv, Nature 488,19(2012), 31 July 2012

"It's wonderful if biologists are belatedly joining the late twentieth century," he quips. "Welcome to the party; better late than never."

arXiv.org e-Print archive

Automated e-print archives [physics] [Search] [Form Interface] [Catchup] [Help]

11 Nov 2004: New CoRR interface introduced for our cs users.
29 Sep 2004: Search engine for user help pages installed.
For more info, see cumulative "What's New" pages.
Robots Beware: indiscriminate automated downloads from this site are *not* permitted.

Physics

- Astrophysics (**astro-ph** new, recent, abs, find)
- Condensed Matter (**cond-mat** new, recent, abs, find)
  includes: Disordered Systems and Neural Networks; Materials Science; Mesoscopic Systems and Quantum Hall Effect; Other; Soft Condensed Matter; Statistical Mechanics; Strongly Correlated Electrons; Superconductivity
- General Relativity and Quantum Cosmology (**gr-qc** new, recent, abs, find)
- High Energy Physics - Experiment (**hep-ex** new, recent, abs, find)
- High Energy Physics - Lattice (**hep-lat** new, recent, abs, find)
- High Energy Physics - Phenomenology (**hep-ph** new, recent, abs, find)
- High Energy Physics - Theory (**hep-th** new, recent, abs, find)
- Mathematical Physics (**math-ph** new, recent, abs, find)
- Nuclear Experiment (**nucl-ex** new, recent, abs, find)
- Nuclear Theory (**nucl-th** new, recent, abs, find)
- Physics (**physics** new, recent, abs, find)
  includes (see detailed description): Accelerator Physics; Atmospheric and Oceanic Physics; Atomic Physics; Atomic and Molecular Clusters; Biological Physics; Chemical Physics; Classical Physics; Computational Physics; Data Analysis, Statistics and Probability; Fluid Dynamics; General Physics; Geophysics; History of Physics; Instrumentation and Detectors; Medical Physics; Optics; Physics Education; Physics and Society; Plasma Physics; Popular Physics; Space Physics
- Quantum Physics (**quant-ph** new, recent, abs, find)

Mathematics

- Mathematics (**math** new, recent, abs, find)
  includes (see detailed description): Algebraic Geometry; Algebraic Topology; Analysis of PDEs; Category Theory; Classical Analysis and ODEs; Combinatorics; Commutative Algebra; Complex Variables; Differential Geometry; Dynamical Systems; Functional Analysis; General Mathematics; General Topology; Geometric Topology; Group Theory; History and Overview; K-Theory and Homology; Logic; Mathematical Physics; Metric Geometry; Number Theory; Numerical Analysis; Operator Algebras; Optimization and Control; Probability; Quantum Algebra; Representation Theory; Rings and Algebras; Spectral Theory; Statistics; Symplectic Geometry

Nonlinear Sciences

- Nonlinear Sciences (**nlin** new, recent, abs, find)
  includes (see detailed description): Adaptation and Self-Organizing Systems; Cellular Automata and Lattice Gases; Chaotic Dynamics; Exactly Solvable and Integrable Systems; Pattern Formation and Solitons

Computer Science

- Computing Research Repository (**CoRR** new, recent, abs, find)
  includes (see detailed description): Architecture; Artificial Intelligence; Computation and Language; Computational Complexity; Computational Engineering, Finance, and Science; Computational Geometry; Computer Science and Game Theory; Computer Vision and Pattern Recognition; Computers and Society; Cryptography and Security; Data Structures and Algorithms; Databases; Digital Libraries; Discrete Mathematics; Distributed, Parallel, and Cluster Computing; General Literature; Graphics; Human-Computer Interaction; Information Retrieval; Information Theory; Learning; Logic in Computer Science; Mathematical Software; Multiagent Systems; Multimedia; Networking and Internet Architecture; Neural and Evolutionary Computing; Numerical Analysis; Operating Systems; Other; Performance; Programming Languages; Robotics; Software Engineering; Sound; Symbolic Computation

Quantitative Biology

- Quantitative Biology (**q-bio** new, recent, abs, find)
  includes (see detailed description): Biomolecules; Cell Behavior; Genomics; Molecular Networks; Neurons and Cognition; Other; Populations and Evolution; Quantitative Methods; Subcellular Processes; Tissues and Organs

---

About arXiv

- some related and unrelated servers (including arXiv **mirror** sites)
- RSS feeds are now available for individual archives and categories.
- today's usage for arXiv.org (not including mirrors)
- some info on delivery type [src] and potential problems
- arXiv Advisory Board
- available macros and brief description
- available help on submitting and retrieving papers
- some background blurb, including invited talk at UNESCO HQ (Paris, 21 Feb '96), update Sep '96
- some info on hypertex

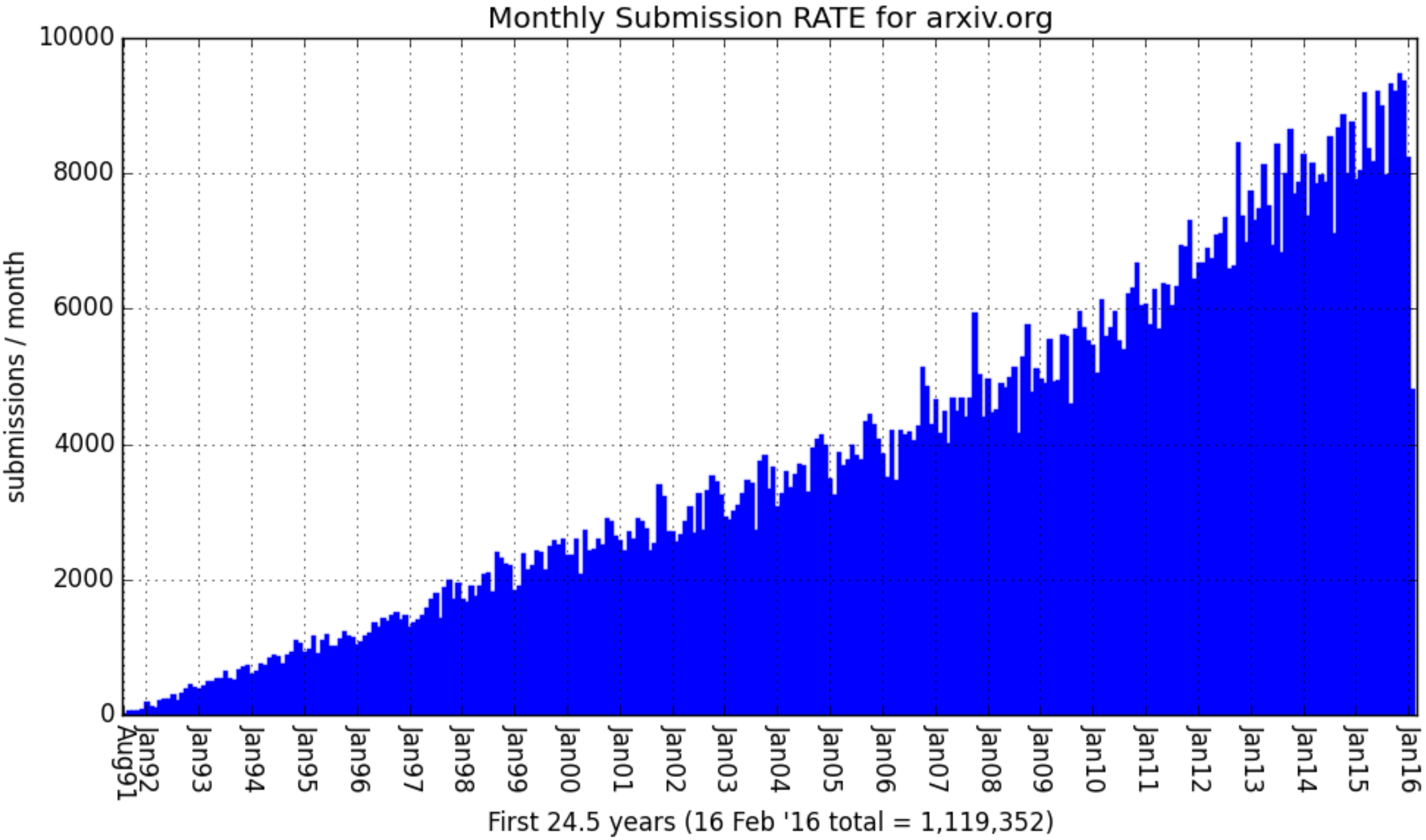
Cornell University Library

arXiv is an e-print service in the fields of physics, mathematics, non-linear science, computer science, and quantitative biology. The contents of arXiv conform to Cornell University academic standards. arXiv is owned, operated and funded by Cornell University, a private not-for-profit educational institution. arXiv is also partially funded by the National Science Foundation.

The Cornell University Library acknowledges the support of Sun Microsystems and U.S. Department of Energy's Office of Scientific and Technical Information (providers of the E-Print Alert Service, which automatically notifies users of the latest information posted on arXiv and other related databases).

www-admin@arxiv.org

Submissions / month, ’91 - ’16

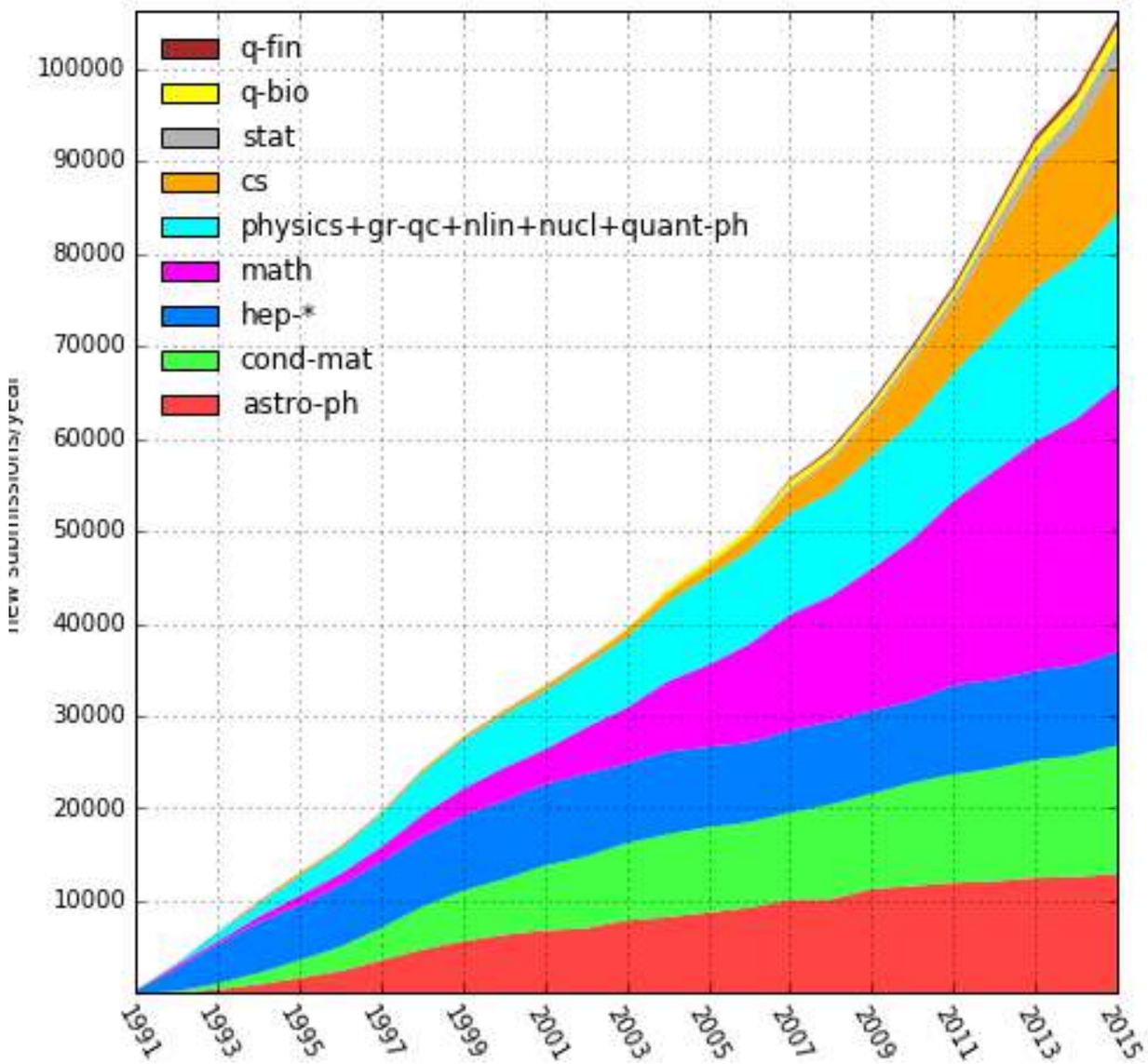


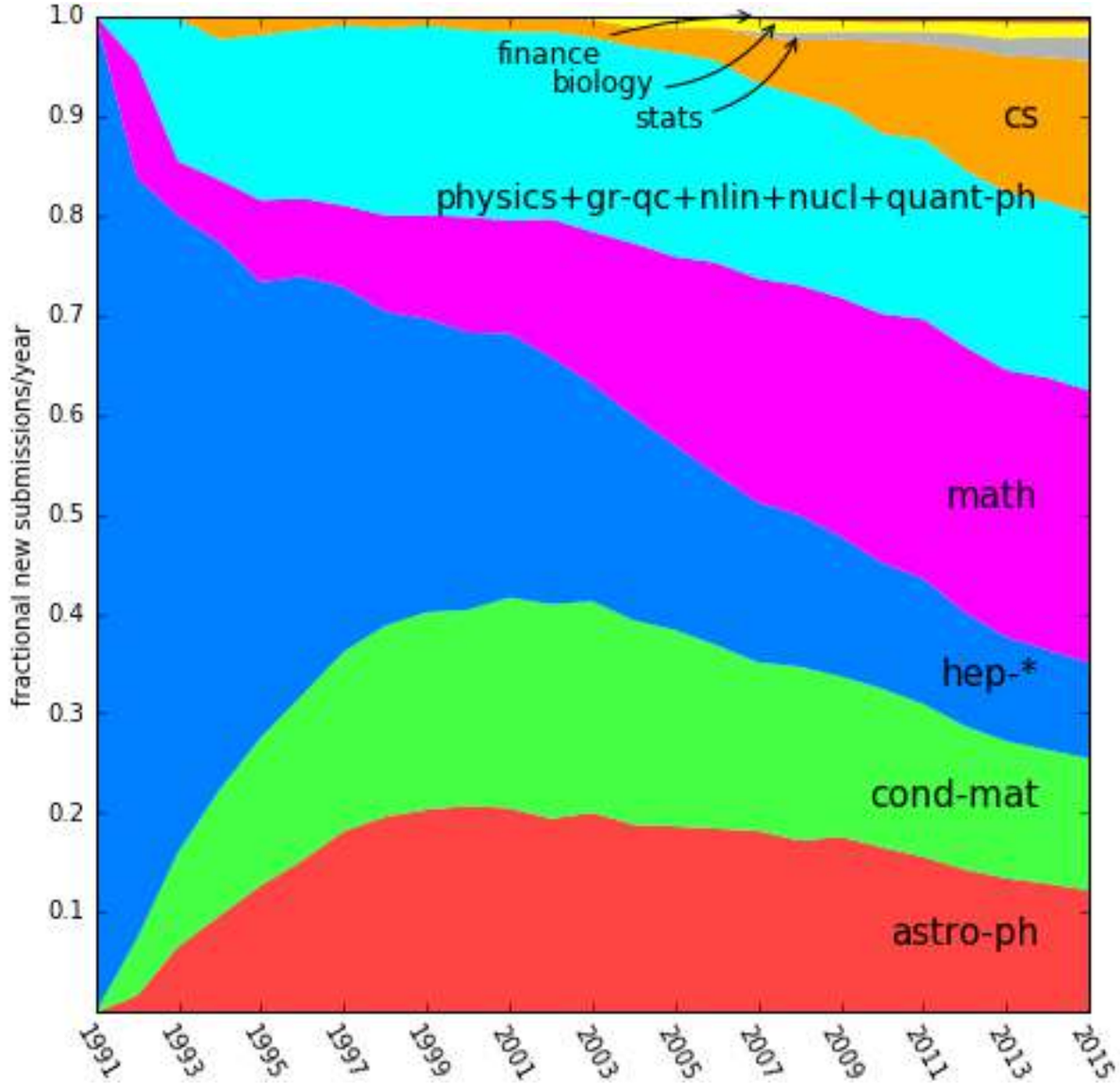
arXiv.org

- **e-mail interface started August 1991**
  - **download data available from start**
  - **WWW usage logs starting from 1993**
- **1,120,000 full text documents (with full graphics), 15 Feb 2016**
  - **physics, mathematics, q-bio, non-linear, computer science**
  - **growing at >110,000 new submissions per year**

    (est. ⇒ > 1,220,000 at end of 2016, 1.75M by end 2020)
- **hundreds of millions of full text downloads per year**
- **hundreds of thousands of distinct users per day**

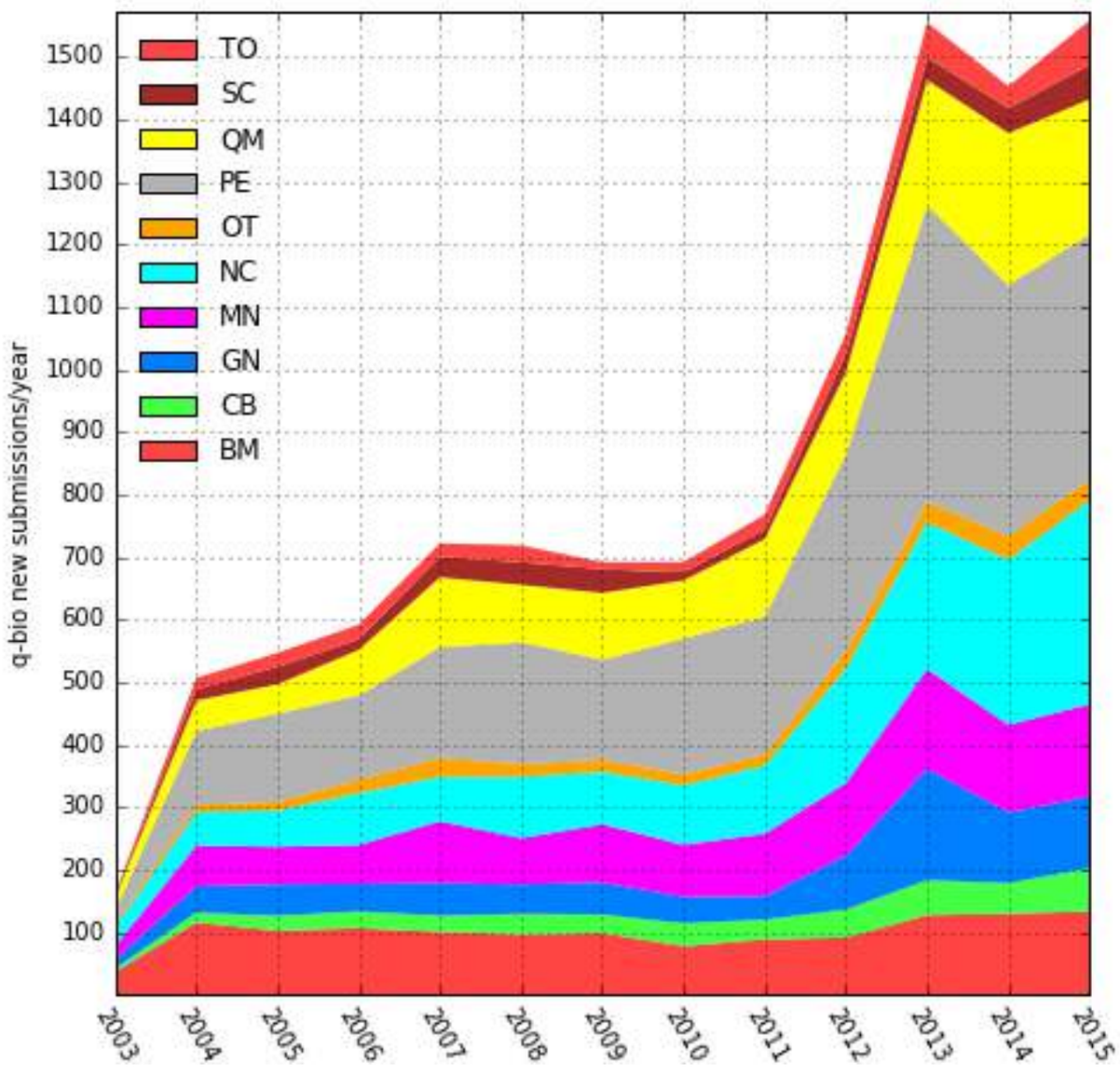
Submissions / year

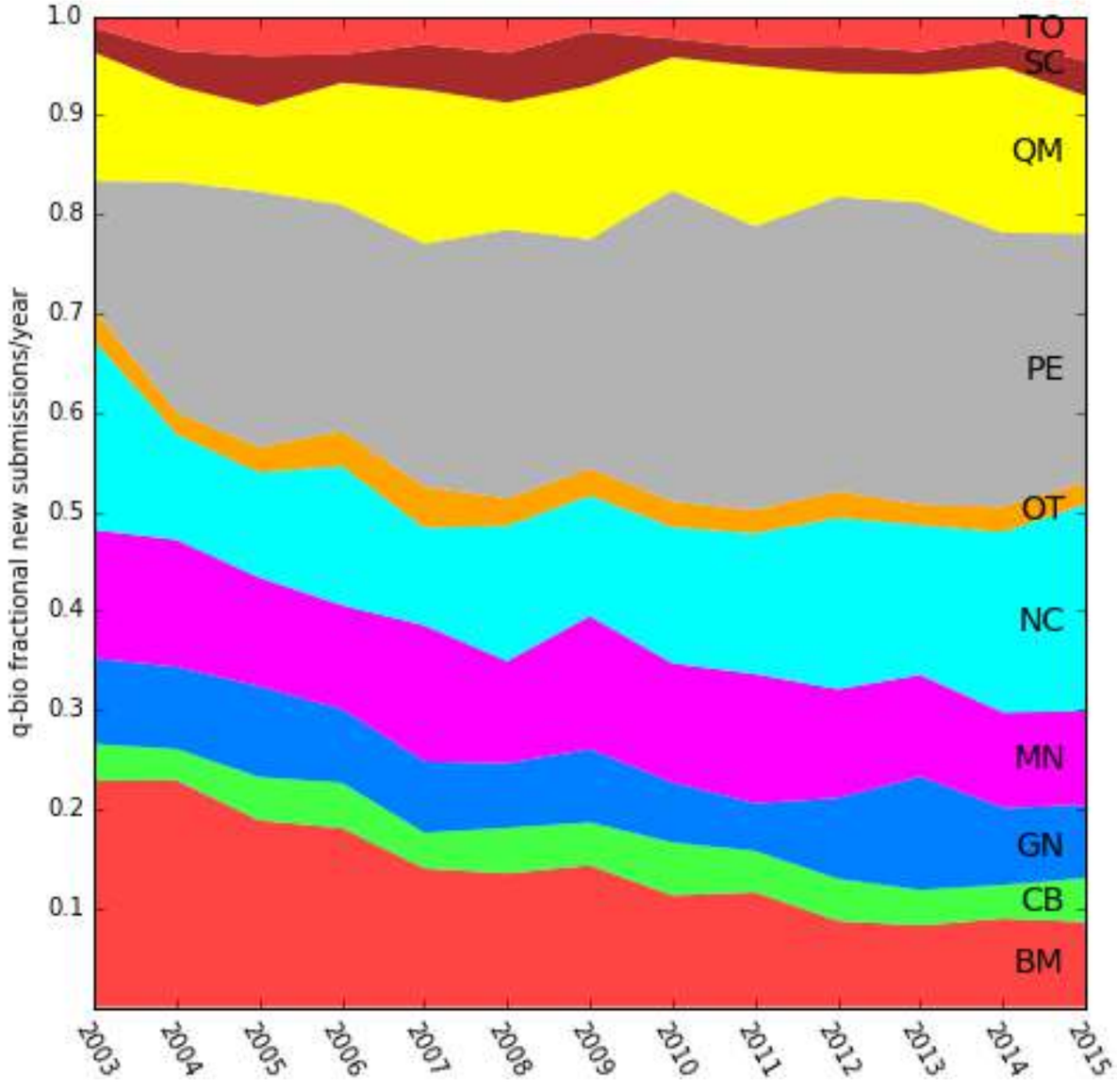




Yes, arXiv has had a biology sector since 2003

http://arxiv.org/archive/q-bio





(over 11,000 submissions + recent growth)

Now taken for granted

But once cutting edge:

abstract page as hub

author names linked to search index

compressed ps and later pdf as network transmission format

foreshadowed web 2.0

cloud service

. . .

Surprises along the way

Google, Wikipedia, Facebook, Twitter

- **power of crowdsourcing**

We're still using TeX ?!?

- **slow move to article formats and capabilities better adapted to network transmission**

Scholarly publishing as a whole still remains in transition

- **(no consensus on the best way to implement quality control, how to fund it, and how to integrate data and other tools needed for scientific reproducibility, and still metastable w.r.t. arXiv/open access)**

It's a commercial network.

But Why?

It is the primary mode of communication
for many of these fields

For some, crucial at multiple stages:
from early dissemination, to improvement,
to medium term visibility,
to archival findability

Web 2.0

In 2005 Congress Passed Law Requiring Users to Post to Youtube ...

Used to Stake Priority Claim

Can't get scooped if Public

Oldenburg, Philosophical Transactions (1665)

(stake property rights)

Vale/Hyman: journal pub dates can obscure priority

preprint and publication equivalent from standpoint of priority claim

physics nobel prize delayed …
fields medal, millenium prize

can't be scooped. let's agree it's non sequitur

Metastable Equilibrium?

Still using both systems in parallel
(Not just inertia — why not?)

sometimes pre-, sometimes post-, sometimes simul

’07-’14 data (HEP): > 80% with journal ref
vast majority of rest subject to some form of review
(conf proceedings, theses, lecture notes)

(so of course it’s not the great unwashed masses)

But is it safe?

arXiv was on-line long before any journals,
so would have been impossible to forbid back then

and hasn't happened since:
remember journals can't risk alienating authors
so have to work with rather than dictate community practice

used in CV for jobs and grants
as evidence of recent productivity

(regularly used, e.g., for NSF grants [search the site]
and as part of press releases, …)

Irreproducibility?

(i.e., even worse than currently?)

in general, no decrease in quality

have to be even more careful…

(risk of embarrassment even greater)

Obligatory added Thomas Jefferson quote:

"The price of freedom is eternal vigilance."

applied machine learning…

Jul ’14 - Jun ’15 (1407-1506)

published arXiv submissions: 100904

physics: 55667 (55.2% of total)

physics.gen-ph: 302 (from 246 submitters)

removals: 556 (from 420 submitters)

Sensitivity:

gen-ph: 266 / 293 = 91%

removals: 471 / 537 = 88%

(flagged by text analyzer, confirmed by moderators; removed non-english / too short)

Specificity:

47 strongly flagged, 128 secondary flagged

gives 175 / 100904 < 0.2% false positive rate

(i.e., (266+471) / 912 = 81% acceptance rate)

Submitted to:

gr-qc 31%
hep-th 17%
quant-ph 11%
astro-ph.CO 7%
math-ph 5%
hep-ph 5%
physics.class-ph 3%
physics.atom-ph 2%
physics.ed-ph 2%
cond-mat.stat-mech 1%
physics.hist-ph 1%
astro-ph.GA 1%
…

Can even improve Quality

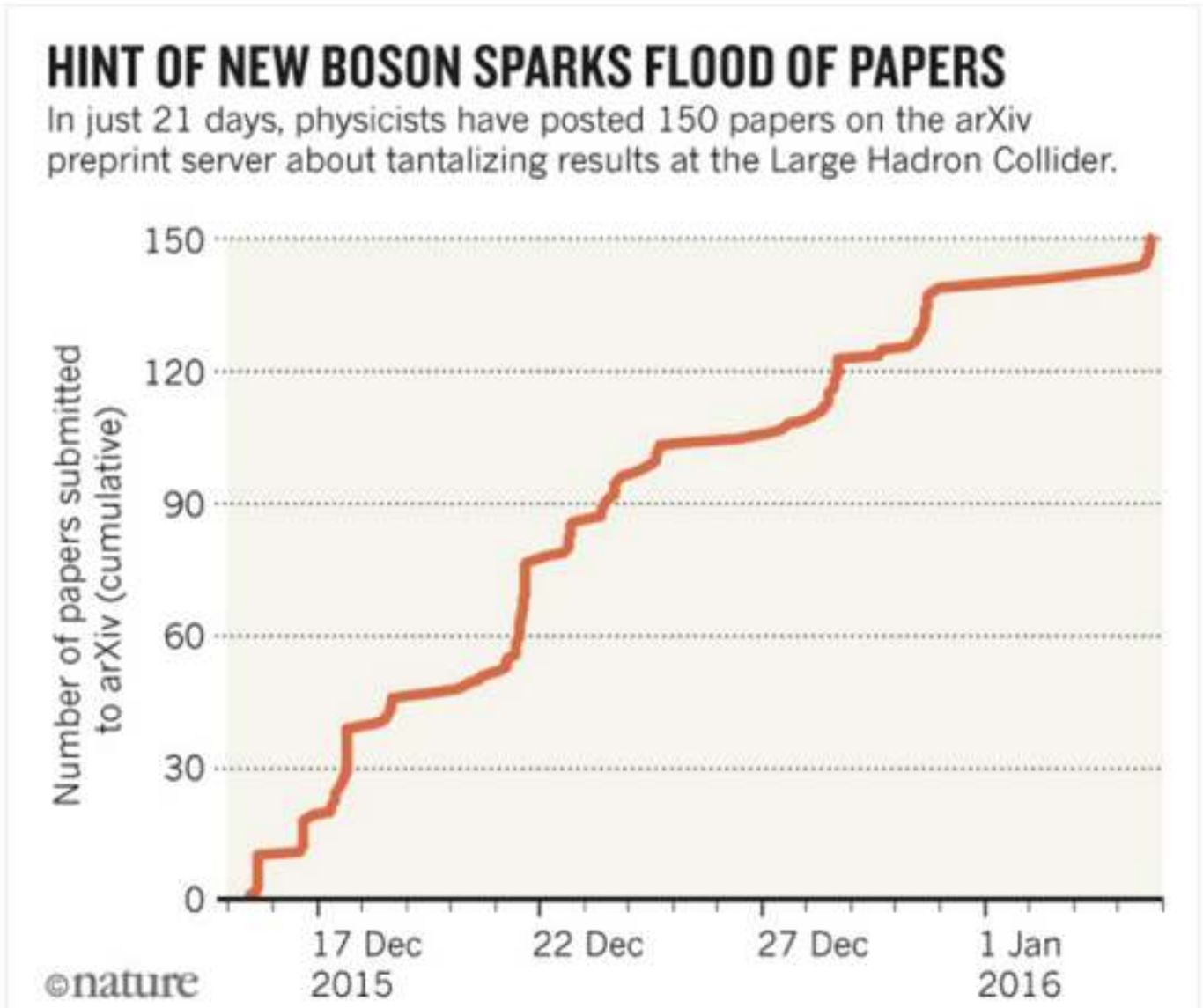
discussion with R. Hoffman today at Ithaca airport
(and arXiv study of v2 vs v1:
refs added and other 'crowdsourced' improvements)

And certainly increases speed

www.nature.com/news/hint-of-new-boson-at-lhc-sparks-flood-of-papers-1.19098

Apps twspeed MathJax bm meta aq h trn ac hr ao

Experimenters revealed their observations in a 15 December announcement at CERN, the European laboratory of particle physics that hosts the LHC near Geneva. Since then, 150 research manuscripts have been posted to the preprint server arXiv discussing the hypothetical particle, even though the statistical significance of the findings is low (see 'Hint of new boson sparks flood of papers').



The surge of interest was anticipated; Tiziano Camporesi, a spokesperson for the LHC's CMS experiment, told *Nature* just after the webcast announcement that he expected to see hundreds of preprints in the next two weeks. "I am extremely curious to see what

Turkish s
supportir
Peace pet
investigati

Big biolo

Ecology's
The Unite
observato
budget ov

Ebola va

Unusual
Drugmake
and begin

Nature P

or just last week:

(GW150914)

arXiv.org Search Results

Back to Search form | Next 11 results

The URL for this search is http://arxiv.org/find/all/1/all:+GW150914/0/1/0/all/0/1

Showing results 1 through 25 (of 36 total) for all:GW150914

1. arXiv:1602.05882 [pdf, other]

Testing the speed of gravitational waves over cosmological distances with strong gravitational lensing

Thomas E. Collett, David Bacon (ICG, Portsmouth)

Comments: 3 Pages. Submitted to ApJL. Comments welcome

Subjects: **High Energy Astrophysical Phenomena (astro-ph.HE)**; Cosmology and Nongalactic Astrophysics (astro-ph.CO); General Relativity and Quantum Cosmology (gr-qc); High Energy Physics - Experiment (hep-ex)

2. arXiv:1602.05554 [pdf, ps, other]

Binary Black Hole Merger Rates Inferred from Luminosity Function of Ultra-Luminous X-ray Sources: Implications to the Origin of GW150914

Yoshiyuki Inoue (ISAS/JAXA), Yasuyuki T. Tanaka (Hiroshima), Naoki Isobe (ISAS/JAXA)

Comments: 4 pages, 1 figure

Subjects: **High Energy Astrophysical Phenomena (astro-ph.HE)**; Solar and Stellar Astrophysics (astro-ph.SR); General Relativity and Quantum Cosmology (gr-qc)

3. arXiv:1602.05529 [pdf, ps, other]

Modeling the Afterglow of GW150914-GBM

Brian J. Morsony, Jared C. Workman, Dominic M. Ryan

Comments: 5 pages, 8 figures, 1 table. Submitted to MNRAS

Subjects: **High Energy Astrophysical Phenomena (astro-ph.HE)**

4. arXiv:1602.05411 [pdf, other]

High-energy Neutrino follow-up search of Gravitational Wave Event GW150914 with ANTARES and IceCube

ANTARES Collaboration: S. Adrián-Martínez, A. Albert, M. André, G. Anton, M. Ardid, J.-J. Aubert, T. Avgitas, B. Baret, J. Barrios-Martí, S. Basa, V. Bertin, S. Biagi, R. Bormuth, M.C. Bouwhuis, R. Bruijn, J. Brunner, J. Busto, A. Capone, L. Caramete, J. Carr, S. Celli, T. Chiarusi, M. Circella, A. Coleiro, R. Coniglione, H. Costantini, P. Coyle, A. Creusot, A. Deschamps, G. De Bonis, C. Distefano, C. Donzaud, D. Dornic, D. Drouhin, T. Eberl, I. El Bojaddaini, D. Elsässer, A. Enzenhöfer, K. Fehn, I. Felis, L.A. Fusco, S. Galatà, P. Gay, S. Geißelsöder, K. Geyer, V. Giordano, V. Giordano, A. Gleixner, H. Glotin, R. Gracia-Ruiz, K. Graf, S. Hallmann, H. van Haren, A.J. Heijboer, Y. Hello, J.J. Hernández-Rey, J. Hößl, J. Hofestädt, et al. (1342 additional authors not shown)

Comments: 13 pages, 2 figures

Subjects: **High Energy Astrophysical Phenomena (astro-ph.HE)**

5. arXiv:1602.05140 [pdf, other]

Short Gamma-Ray Bursts from the Merger of Two Black Holes

Rosalba Perna, Davide Lazzati, Bruno Giacomazzo

Comments: 2 figures, submitted to ApJL, comments welcome

Subjects: **High Energy Astrophysical Phenomena (astro-ph.HE)**

6. arXiv:1602.05050 [pdf, ps, other]

Electromagnetic Afterglows Associated with Gamma-Ray Emission Coincident with Binary Black Hole Merger Event GW150914

Ryo Yamazaki, Katsuaki Asano, Yutaka Ohira

Comments: 9 pages, 2 figures

Subjects: **High Energy Astrophysical Phenomena (astro-ph.HE)**

7. arXiv:1602.04782 [pdf, ps, other]

Testing local Lorentz invariance with gravitational waves

Alan Kostelecky, Matthew Mewes

Comments: 12 pages

Subjects: **General Relativity and Quantum Cosmology (gr-qc)**; High Energy Astrophysical Phenomena (astro-ph.HE); High Energy Physics - Phenomenology (hep-ph)

8. arXiv:1602.04779 [pdf, other]

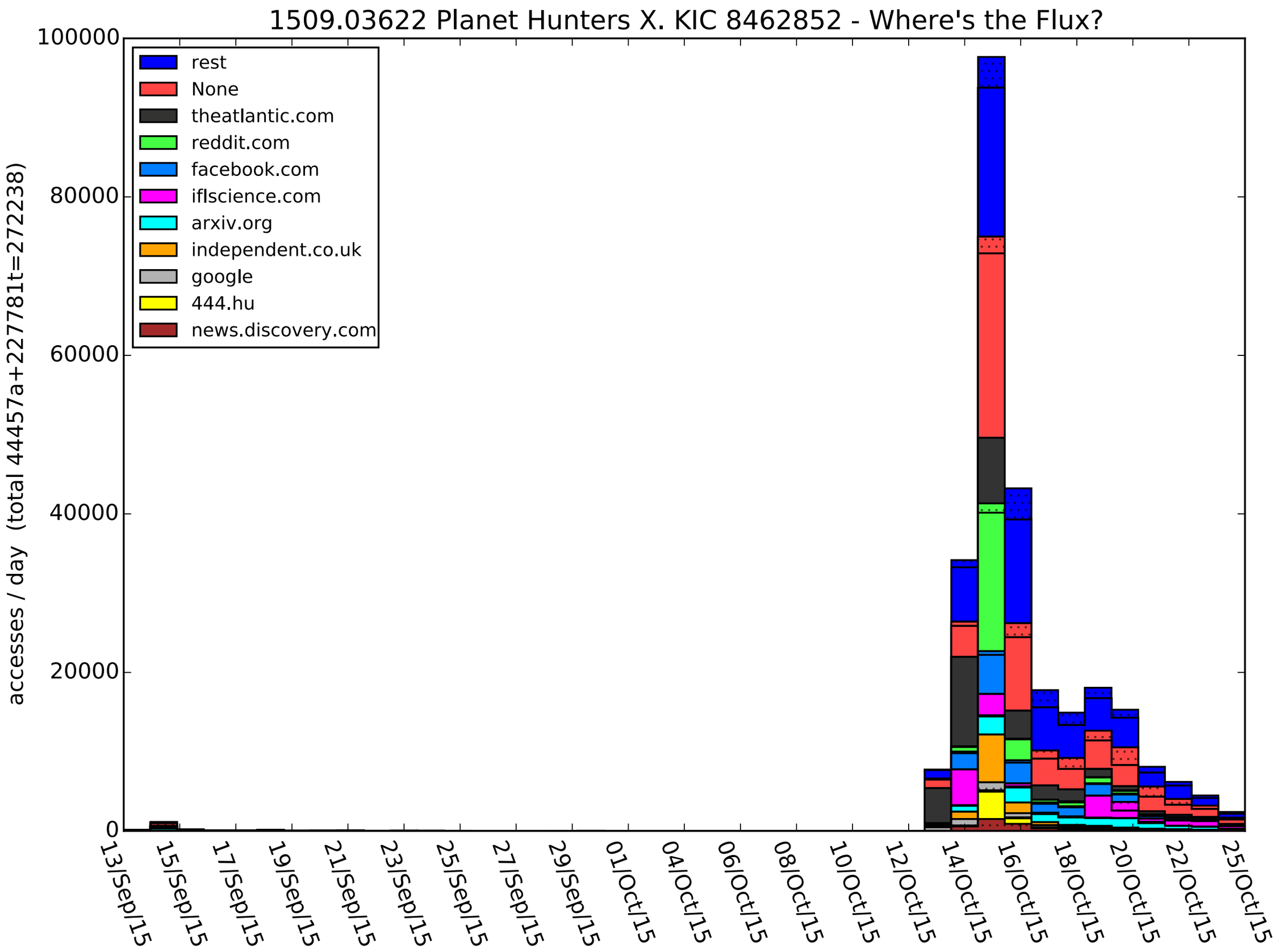
Constraints on frequency-dependent violations of Shapiro delay from GW150914

Emre O. Kahya, Shantanu Desai

Comments: 3 pages. This article is dedicated to the memory of Prof. Steven Detweiler

Subjects: **General Relativity and Quantum Cosmology (gr-qc)**; Cosmology and Nongalactic Astrophysics (astro-ph.CO); High Energy Astrophysical Phenomena (astro-ph.HE)

and it engages the public:


1509.03622 Planet Hunters X. KIC 8462852 - Where's the Flux?
accesses / day (total 44457a+227781t=272238)
100000
80000
60000
40000
20000
0
rest
None
theatlantic.com
reddit.com
facebook.com
iflscience.com
arxiv.org
independent.co.uk
google
444.hu
news.discovery.com
13/Sep/15
15/Sep/15
17/Sep/15
19/Sep/15
21/Sep/15
23/Sep/15
25/Sep/15
27/Sep/15
29/Sep/15
01/Oct/15
04/Oct/15
06/Oct/15
08/Oct/15
10/Oct/15
12/Oct/15
14/Oct/15
16/Oct/15
18/Oct/15
20/Oct/15
22/Oct/15
25/Oct/15

Astrophysics > Solar and Stellar Astrophysics

Planet Hunters X. KIC 8462852 - Where's the Flux?

T. S. Boyajian, D. M. LaCourse, S. A. Rappaport, D. Fabrycky, D. A. Fischer, D. Gandolfi, G. M. Kennedy, M. C. Liu, A. Moor, K. Olah, K. Vida, M. C. Wyatt, W. M. J. Best, F. Ciesla, B. Csak, T. J. Dupuy, G. Handler, K. Heng, H. Korhonen, J. Kovacs, T. Kozakis, L. Kriskovics, J. R. Schmitt, Gy. Szabo, R. Szabo, J. Wang, S. Goodman, A. Hoekstra, K. J. Jek

(Submitted on 11 Sep 2015)

Over the duration of the Kepler mission, KIC 8462852 was observed to undergo irregularly shaped, aperiodic dips in flux down to below the 20% level. The dipping activity can last for between 5 and 80 days. We characterize the object with high-resolution spectroscopy, spectral energy distribution fitting, and Fourier analyses of the Kepler light curve. We determine thatKIC 8462852 is a main-sequence F3 V/IV star, with a rotation period ~0.88 d, that exhibits no significant IR excess. In this paper, we describe various scenarios to explain the mysterious events in the Kepler light curve, most of which have problems explaining the data in hand. By considering the observational constraints on dust clumps orbiting a normal main-sequence star, we conclude that the scenario most consistent with the data is the passage of a family of exocomet fragments, all of which are associated with a single previous breakup event. We discuss the necessity of future observations to help interpret the system.

Comments: Submitted to MNRAS. 15 pages, 12 figures
Subjects: **Solar and Stellar Astrophysics (astro-ph.SR)**; Earth and Planetary Astrophysics (astro-ph.EP)
Cite as: arXiv:1509.03622 [astro-ph.SR]
(or arXiv:1509.03622v1 [astro-ph.SR] for this version)

Trackbacks for 1509.03622

The Most Mysterious Star in Our Galaxy [The Atlantic @ www.theatlantic.com/science] [trackback posted Tue Oct 13 09:16:00 2015]

Sure, This Star Is Weird. But Aliens? [Slate Magazine @ www.slate.com/blogs] [trackback posted Wed Oct 14 09:00:00 2015]

Has Kepler spotted an 'alien megastructure'? [USA TODAY @ www.usatoday.com/story] [trackback posted Thu Oct 15 13:41:09 2015]

The strange star that has serious scientists talking about an alien megastructure [Washington Post @ www.washingtonpost.com/ne

Has Kepler Discovered an Alien Megastructure? [DNews @ news.discovery.com/space] [trackback posted Wed Oct 14 14:23:00 2015]

Have researchers found an alien MEGASTRUCTURE? [Mail Online @ www.dailymail.co.uk/sciencetec...] [trackback posted Tue Oct 13 19:5

Astronomers think they have found an alien megastructure [The Independent @ www.independent.co.uk/news] [trackback posted Fri C

Kepler ha individuato una costruzione aliena? [Focus.it @ www.focus.it/scienza] [trackback posted Thu Oct 15 12:00:00 2015]

Spekulation uber Alien-Strukturen um mysteriosen Stern [heise online @ www.heise.de/newsticker] [trackback posted Thu Oct 15 08:20:

Astronomers Have Spotted Something Very, Very Strange Surrounding A Distant Star [IFLScience @ www.iflscience.com/space] [trac

KIC 8462852, l'etoile la plus mysterieuse de la galaxie ? [Passeur de sciences @ passeurdesciences.blog.lemonde...] [trackback pos

Have We Detected Megastructures Built By Aliens Around A Distant Star? [Popular Science @ www.popsci.com/have-we-detecte...] [

Mysteriously Variable Star Causes Speculation About Dyson Sphere [slashdot @ science.slashdot.org/story] [trackback posted Wed Oct

Before the media blows up reporting on how we've found a Dyson Sphere what does the paper for KIC 8462852 actually say? [redd

Astronomers may have found giant alien 'megastructures' orbiting star near the Milky Way [reddit /r/Futurology @ www.reddit.com

The strange star that has serious scientists talking about an alien megastructure [reddit /r/worldnews @ www.reddit.com/r] [trackb

"Ilyet epitene egy idegen civilizacio" - hatalmas targyakat fedeztek fel egy titokzatos csillag, a KIC 8462852 korul [444 @ 444.hu/2

SETI: On the Verge of a Breakthrough? [AboveTopSecret.com @ www.abovetopsecret.com/forum] [trackback posted Tue Oct 13 17:04:00

Are Aliens Building Structures Around KIC 8462852? No. (Yes?) [MetaFilter @ www.metafilter.com/153817] [trackback posted Tue Oct 1

PRVI ZNAK POSTOJANJA NAPREDNE CIVILIZACIJE IZVAN ZEMLJE? Znanstvenike uzbudili jedinstveni nalazi svemirskog teleskopa Kepl

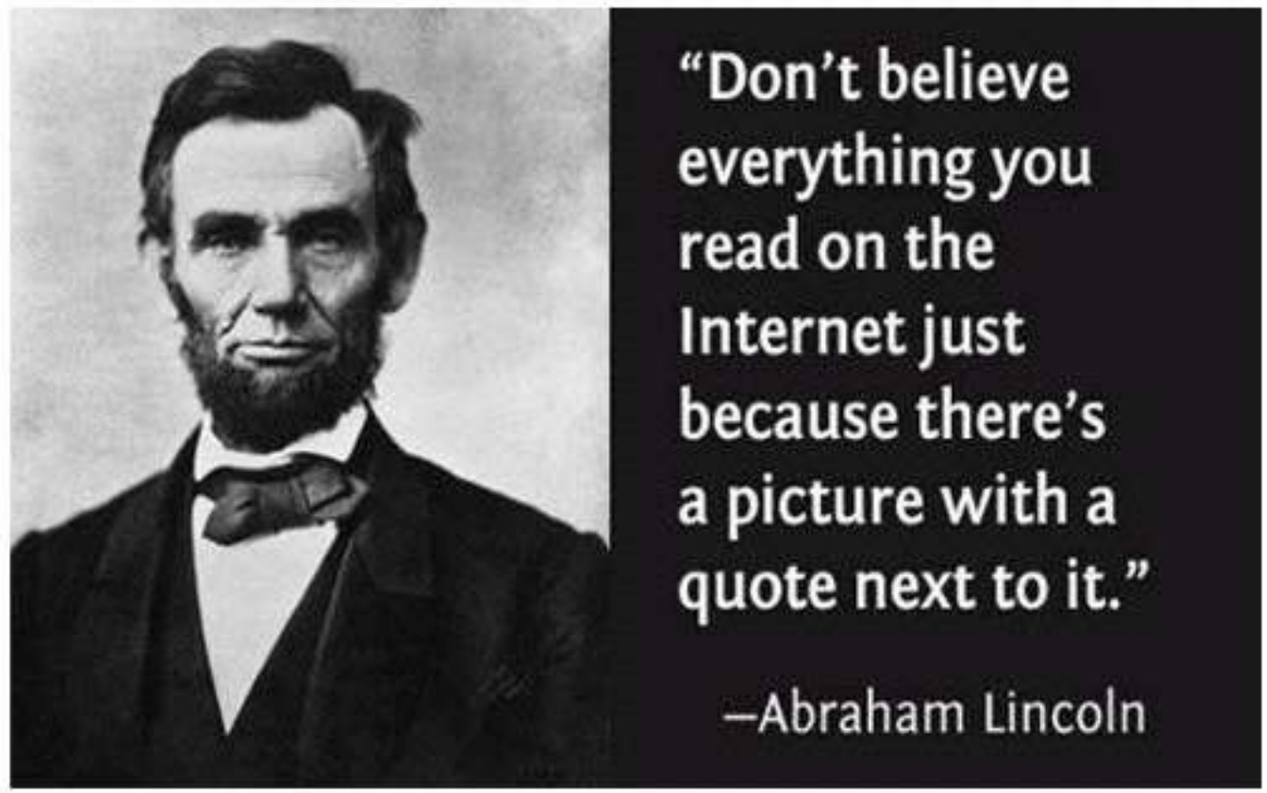
Star exhibits strange light patterns which could be a sign of alien activity [NeoGAF @ www.neogaf.com/forum] [trackback posted Wed

Exaggerated Differences

no, all physicists do not work in
experimental groups of thousands of people

and there is less competition ?!?!?!?

(another obligatory quote from a former US president)



Heterogeneous Community

no, "physics" is not a monolithic community
mixture of completely different cultures
from low temp experiment
(adopted more slowly,
but then new materials)

to astrophysics,
to …

to mathematics

1509.05363 Erdős Discrepancy Problem

and

computer science

1512.03385 Deep Learning for Image Recognition

and finance

1511.09054 It's a Trap: Emperor Palpatine's Poison Pill

to …

Single-stop shopping?

useful to have aggregated

various data-mining applications

in principle, many could be federated via overlay
(e.g., OAI for metadata)

Technical Issues

library support on the edge

missing cutting edge dev

goal: amazon warehouse.
(comprehensive, better tools)