Title: At a Physics/InfoSci Intersection P. Ginsparg, Physics and InfoSci

Date: Apr 15, 2015  02:00 PM

URL: http://pirsa.org/15040090

Abstract: <p>Over Twenty-five years into the internet era, over twenty years into the WorldWideWeb era, fifteen years into the Google era, and a few years past the Facebook/Twitter era, we've yet to converge on a new long-term methodology for scholarly research communication. I will provide a sociological overview of our current metastable state, and then a technical discussion of the practical implications of literature and usage data considered as computable objects, using arXiv as exemplar. From the physics standpoint, there is a surprising amount of statistical mechanics in text-mining and machine learning.</p>
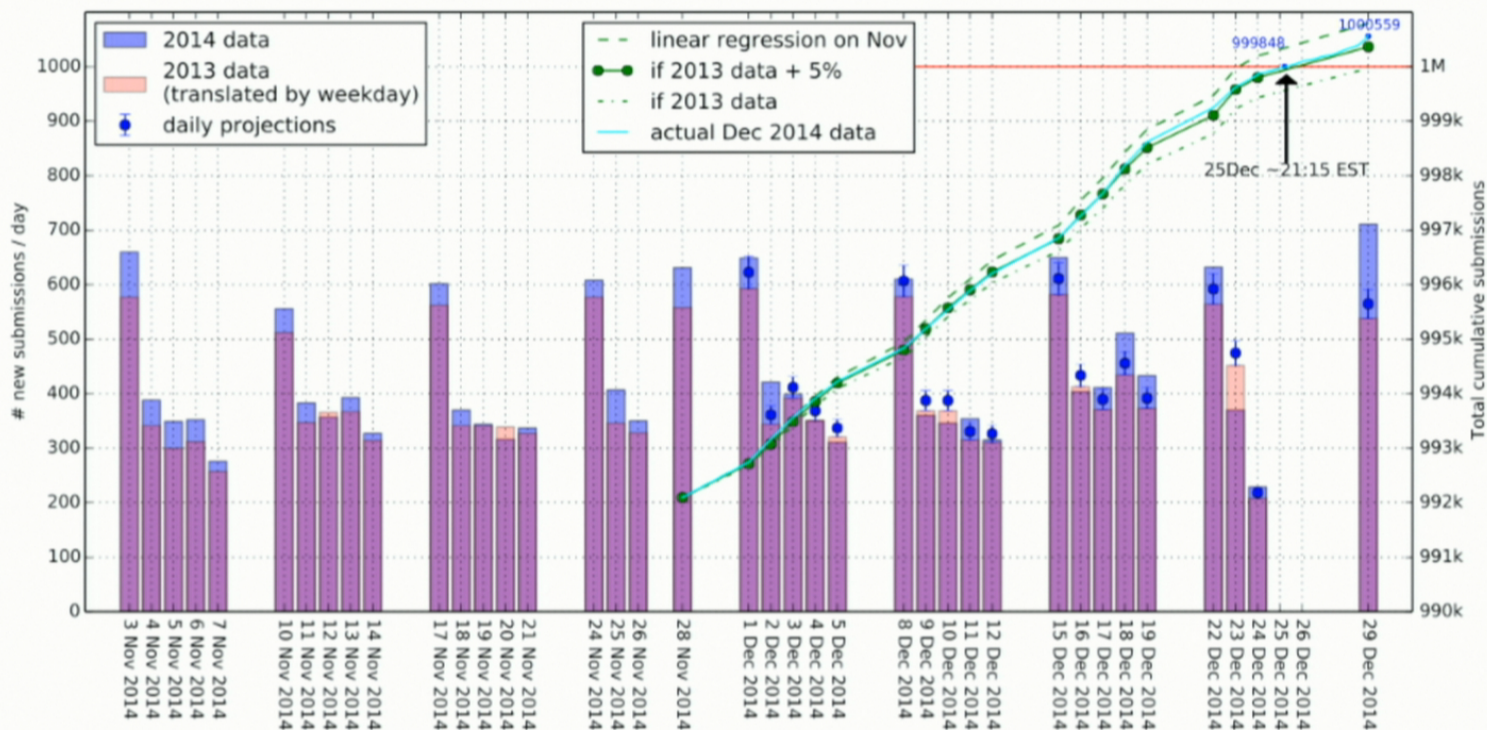
Submissions / month, '91 - '15

Monthly Submission RATE for arxiv.org

First 23.7 years (15 Apr '15 total = 1,029,852)
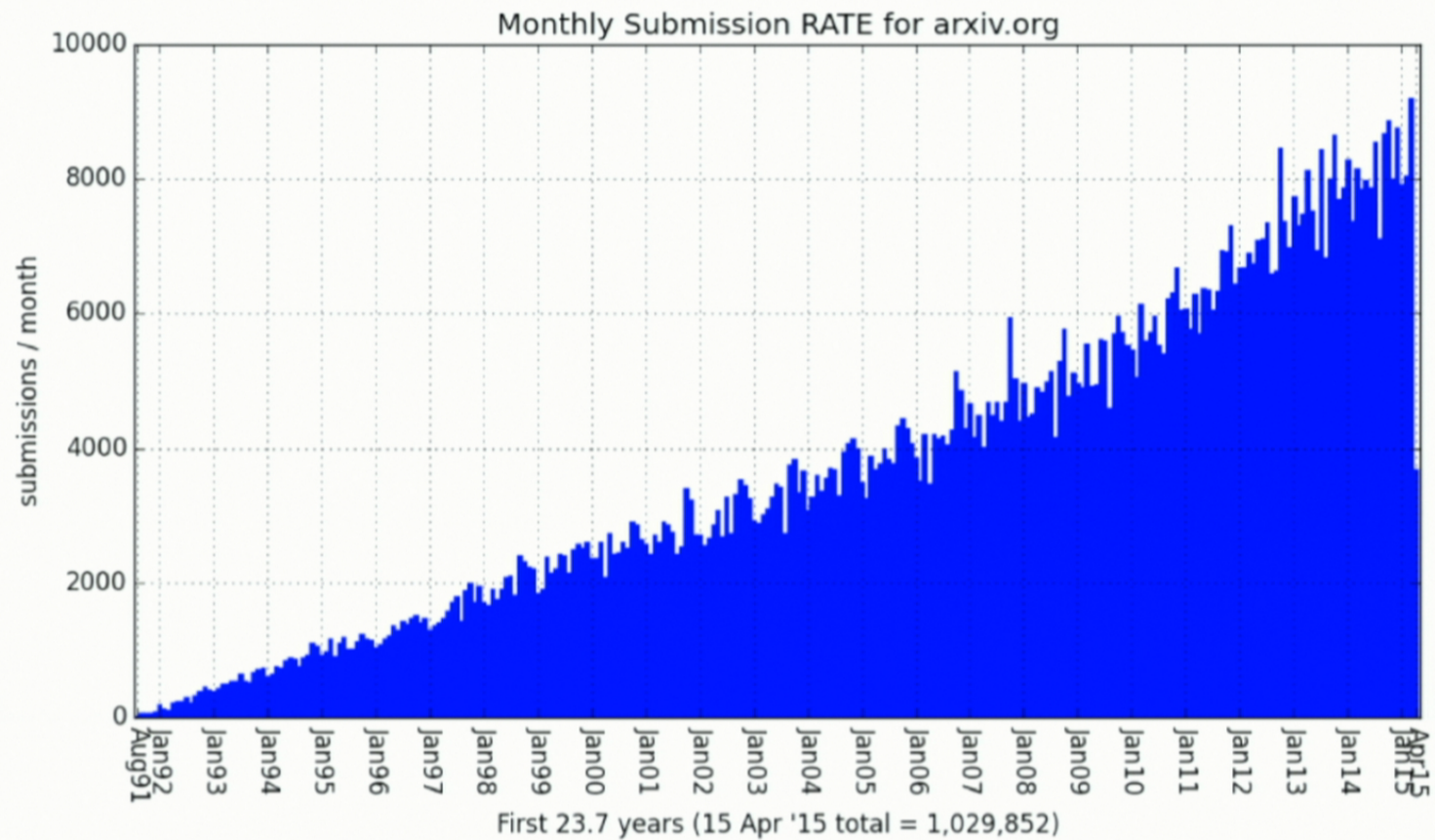
# arXiv.org

- **e-mail interface started August 1991**
  - **download data available from start**
  - **WWW usage logs starting from 1993**
- **1,030,000 full text documents (with full graphics), 15 Apr 2015**
  - **physics, mathematics, q-bio, non-linear, computer science**
  - **growing at 100,000 new submissions per year**

    **(est. ⇒ > 1,100,000 at end of 2015, 1.75M by end 2020)**
- **hundreds of millions of full text downloads per year**
- **hundreds of thousands of distinct users per day**

# Projection (28 Nov 2014)

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | Fo

News & Comment › News › 2014 › December › Article

NATURE | NEWS

## The arXiv preprint server hits 1 million articles

Website where scientists flock to upload manuscripts before peer review has doubled its holdings in six years.

Richard Van Noorden

30 December 2014

The popular preprint server arXiv.org, where physicists, mathematicians and computer scientists routinely upload manuscripts to publicly share their findings before peer review, now holds more than 1 million research articles.

The repository, launched as an 'electronic bulletin board' in August 1991, just before the dawn of the World Wide Web, took 17 years to accumulate half a million manuscripts, but has taken just 6 more to double its holdings.

Researchers now submit around 8,000 articles to the arXiv each month — more than 250 a day, on average. The site's administrators make the raw, non-peer-reviewed manuscripts available in batches after a brief quality-control check, such as a cursory glance for appropriateness by one of 130 volunteer moderators, and automated filtering to check for text overlap with existing papers.

JE BENT HIER: HOME / NIEUWS / MILJOENSTE ARTIKEL TOEGEVO

## Miljoenste artikel toegevoegd aan arXiv

31 DECEMBER 2014 DOOR ARIE NOUWEN • REAGEER

Op de populaire preprint server arXiv.org, waar wetenschappers h
door iedereen kunnen worden gelezen, is onlangs het miljoenste a
gestart door Paul Ginsparg als een 'electronic bulletin board', nog
werden een half miljoen artikelen toegevoegd aan de arXiv, de vol

**30 DICEMBRE 2014**

## I milioni di arXiv

Ci sono voluti 17 anni per accumulare mezzo milione di mano
'electronic bulletin board' nell'agosto 1991 un po' prima dell'alba
raddoppiare i suoi averi.

Era una una bacheca elettronica per un centinaio di fisici delle

## TheScientist
EXPLORING LIFE, INSPIRING INNOVATION

News ▾ | Magazine ▾ | Multimedia ▾ | Subjects ▾ | Surveys ▾ | Careers ▾

The Scientist » News & Opinion » Daily News

## Q&A: One Million Preprints and Counting

A conversation with ArXiv founder Paul Ginsparg

By Tracy Vence | December 29, 2014

Since 1991, scientists from a variety of fields have published their research to the preprint server ArXiv, to quickly share data and to stake intellectual claim on new discoveries.

Today (December 29), the preprint server clocked its one-millionth upload. In anticipation of this milestone, The Scientist spoke with ArXiv founder Paul Ginsparg of Cornell University about sharing data, peer review, and what's next for the resource.

# Submissions / year
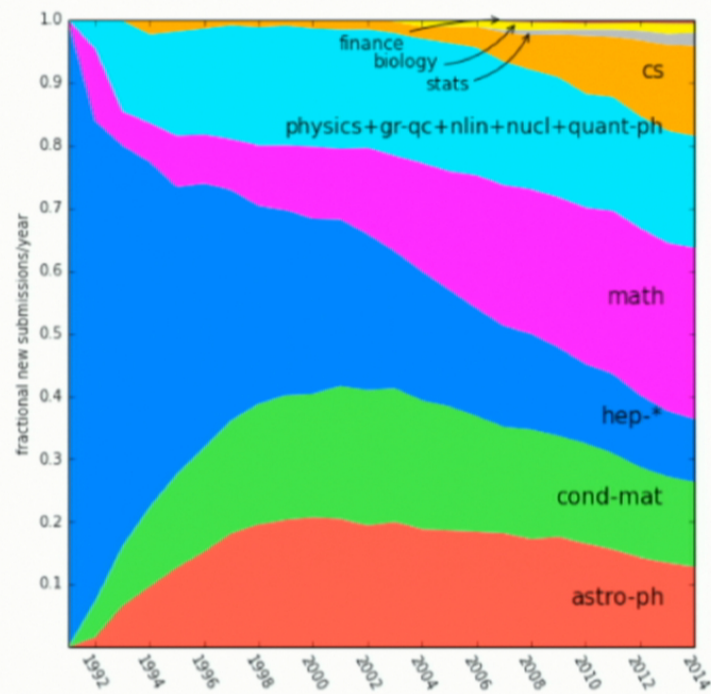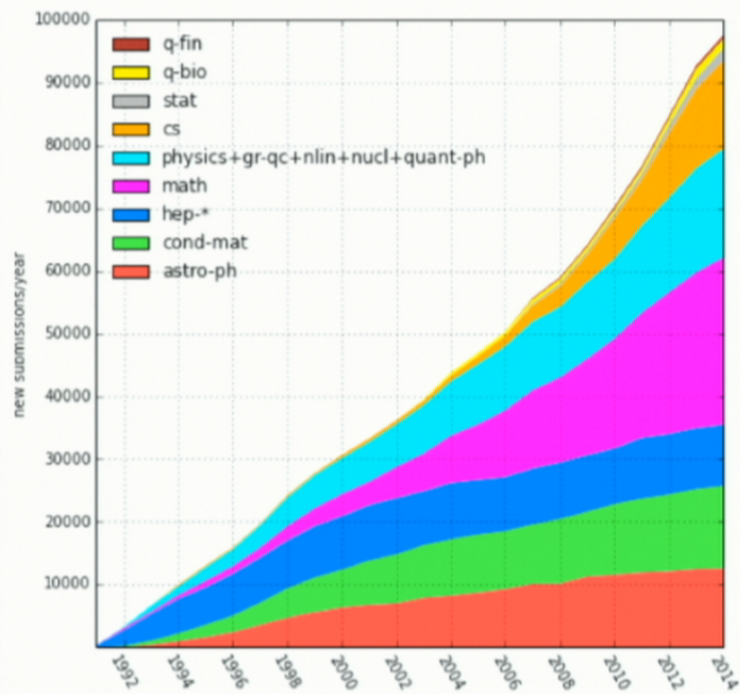
# Surprises along the way

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.

# What is Science?

guarding the perimeter

text classifier, multi-grams

machine learning for suspects

would we have invented journals just to filter the non-scientists?
(N.B. it's a jungle out there)

plagiarism, hashes fit in ram

"information geneology"

# naive bayes

**Bayes:** $p(C|w) = p(w|C)p(C)/p(w)$

**Naive:** $p(\{w_i\}|C) = \prod_i p(w_i|C)$

- **spam filter** $(p(S|\{w_i\})/p(\overline{S}|\{w_i\}))$

- **text classification (on arXiv $> 95\%$ now)**

- **spell correction**

- **voice recognition**

- . . .

simplest algorithm works better with more data.

for arXiv use multigram vocab: genetic_algorithm, black_hole

# Publishers withdraw more than 120 gibberish papers

Nature | News 24 Feb 2014

The publishers Springer and IEEE are removing more than 120 papers from their subscription services after a French researcher discovered that the works were computer-generated nonsense.

http://pdos.csail.mit.edu/scigen/ (D.Aguayo, M.Krohn, J.Stribling)

# SCIgen - An Automatic CS Paper Generator

**About  Generate  Examples  Talks  Code  Donations  Related  People  Blog**

## About

SCIgen is a program that generates random Computer Science research papers, including graphs, figures, and citations. It uses a hand-written **context-free grammar** to form all elements of the papers. Our aim here is to maximize amusement, rather than coherence.

One useful purpose for such a program is to auto-generate submissions to conferences that you suspect might have very low submission standards. A prime example, which you may recognize from spam in your inbox, is SCI/IIIS and its dozens of co-located conferences (check out the very broad conference description on the **WMSCI 2005** website). There's also a list of **known bogus conferences**. Using SCIgen to generate submissions for conferences like this gives us pleasure to no end. In fact, one of our papers was accepted to SCI 2005! See **Examples** for more details.

We went to WMSCI 2005. Check out the **talks and video**. You can find more details in our **blog**.

## Generate a Random Paper

Want to generate a random CS paper of your own? Type in some optional author names below, and click "Generate".

Author 1:   J. Harvey
Author 2:   G. Moore
Author 3:   S. Ramanujan
Author 4:
Author 5:

[Generate] [Reset]

# Arnicin: Visualization of Vacuum Tubes

J. Harvey, G. Moore, S. Ramanujan

## Abstract

The implications of trainable theory have been far-reaching and pervasive. In this paper, we confirm the development of vacuum tubes. Arnicin, our new framework for flip-flop gates, is the solution to all of these obstacles.

## 1 Introduction

Electrical engineers agree that read-write models are an interesting new topic in the field of complexity theory, and steganographers concur. In addition, the drawback of this type of method, however, is that the memory bus can be made concurrent, homogeneous, and peer-to-peer. Furthermore, we emphasize that we allow wide-area networks to visualize highly-available algorithms without the investigation of digital-to-analog converters. To what extent can telephony be analyzed to solve this challenge?

Flexible heuristics are particularly impor-

improves replication, and also Arnicin harnesses constant-time algorithms.

In order to accomplish this purpose, we disconfirm that the seminal signed algorithm for the visualization of the Internet by Sato [35] is Turing complete. To put this in perspective, consider the fact that much-touted leading analysts entirely use context-free grammar to solve this problem. We view complexity theory as following a cycle of four phases: allowance, observation, improvement, and management. The basic tenet of this method is the simulation of RPCs. Existing amphibious and ambimorphic approaches use rasterization to improve the investigation of Byzantine fault tolerance. Despite the fact that similar frameworks visualize the World Wide Web, we accomplish this aim without architecting symbiotic methodologies.

In our research we describe the following contributions in detail. For starters, we concentrate our efforts on verifying that the foremost pervasive algorithm for the construction of model checking by Shastri et al. [11] is NP-

# "Ike Antkare, One of the Great Stars in the Scientific Firmament"

(C. Labbé, ISSI Newsletter, 6(2), 48-52, 2010)

"Since the 8th of April 2010, these tools have allowed a certain Ike Antkare to become one of the most highly cited scientists of the modern world (see Appendix A, Figures 2-6).

"According to Scholarometer, "Ike Antkare" has 102 publications (almost all in 2009) and has an h-index of 94, putting him in the 21st position of the most highly cited scientists.
This score is less than Freud, in 1st position with a h-index of 183, but better than Einstein in 36th position, with a h-index of 84.

"Best of all, with respect to the $h_m$-index, "Ike Antkare" holds the sixth position -- outclassing all scientists in his field (computer science)."

http://www.slate.com/articles/podcasts/lexicon_valley/2012/06/
lexicon_valley_resolving_authorship_controversies
_in_the_federalist_papers_and_the_wizard_of_oz.html

http://www.mhpbooks.com/mapping-the-oz-genome/
Mapping the Oz genome

**http://www.ssc.wisc.edu/~zzeng/soc357/OZ.pdf**

**Who Wrote the 15th Book of Oz?**

**An Application of Multivariate Analysis to Authorship Attribution**

**J. Binongo, Chance vol 16 (2003)**

**L. Frank Baum wrote 14 books starting in 1900, 'til death in 1919** (published: '00, '04, '07, '08–'10, '13–'20). **1918**: gallbladder removed, had written two extra: The Magic of Oz and Glinda of Oz for reserve, then from bed finished:

**#12. The Tin Woodsman of Oz (1918).** Other two published posthumously:

**#13. The Magic of Oz (1919)**

**#14. Glinda of Oz (1920, edited by his son)**

**19 more appeared, one per year from '21-'39**, by 1939 (the movie!) there were **33** by Baum and children's author Ruth Thompson. Burning question:

**#15. The royal book of Oz (1921):** Baum's last or Thompson's first?

Averages (10% stopword depletion)
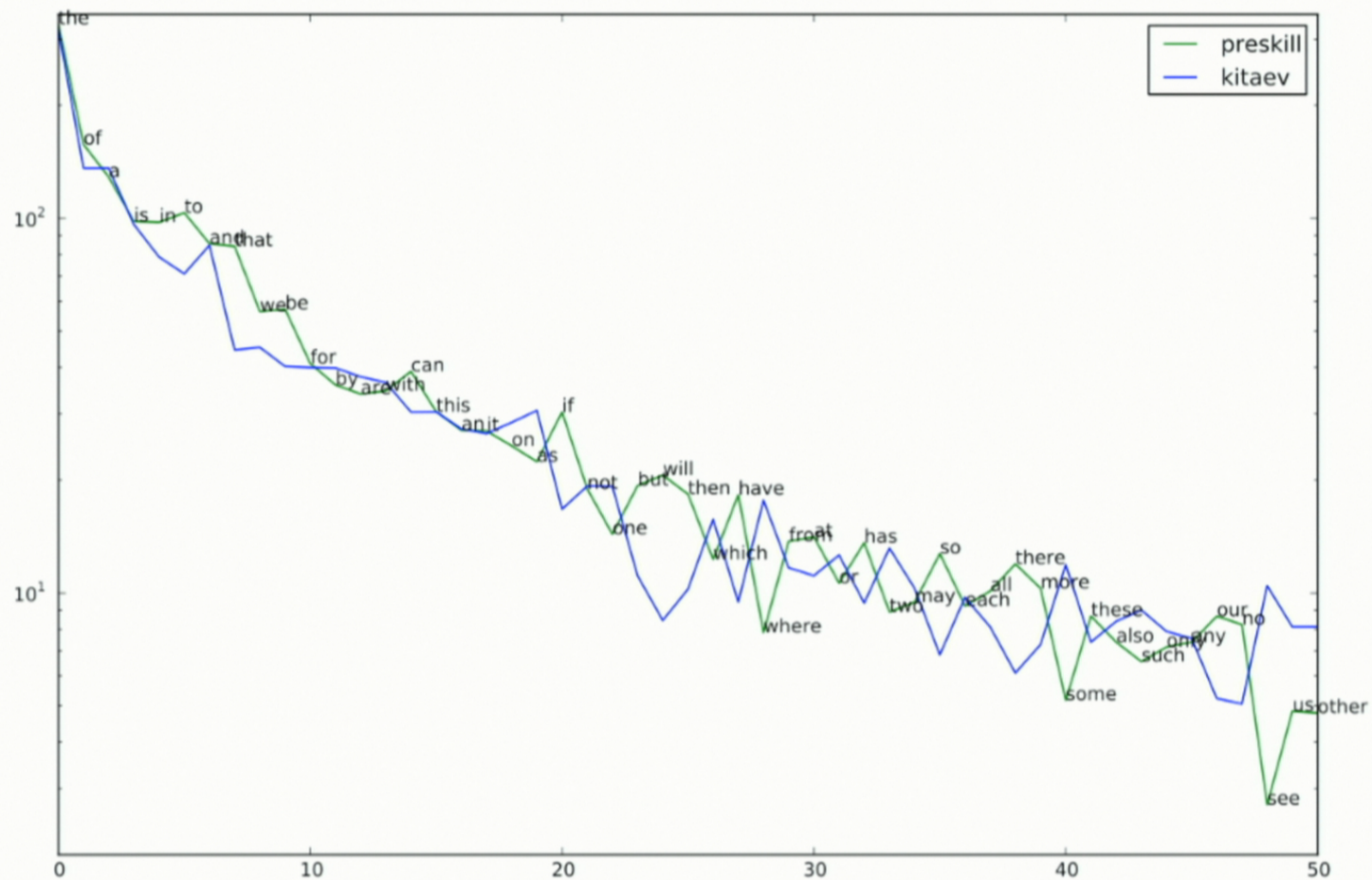
# Singular Value Decomposition

$$M = U\Sigma V^T$$

(generalizes $M = O\Lambda O^T$)

- weather data

- document word (LSA)

- stock data

- genomic data

- apple itunes genius

- microarray data

- netflix challenge (500k $\times$ 17k)

- ...

# a.k.a. Schmidt decomposition

$$M = U\Sigma V^\dagger$$

(generalizes $M = U\Lambda U^\dagger$)

**Familiar to physicists as the Schmidt decomposition**

$$|\psi\rangle = M_{ij}|\phi_A^i\rangle \otimes |\phi_B^j\rangle = \sum_i \sigma_i |\psi_A^i\rangle \otimes |\psi_B^i\rangle$$

**where orthonormal bases:** $\langle\psi_A^i|\psi_A^j\rangle = \langle\psi_B^i|\psi_B^j\rangle = \delta_{ij}$
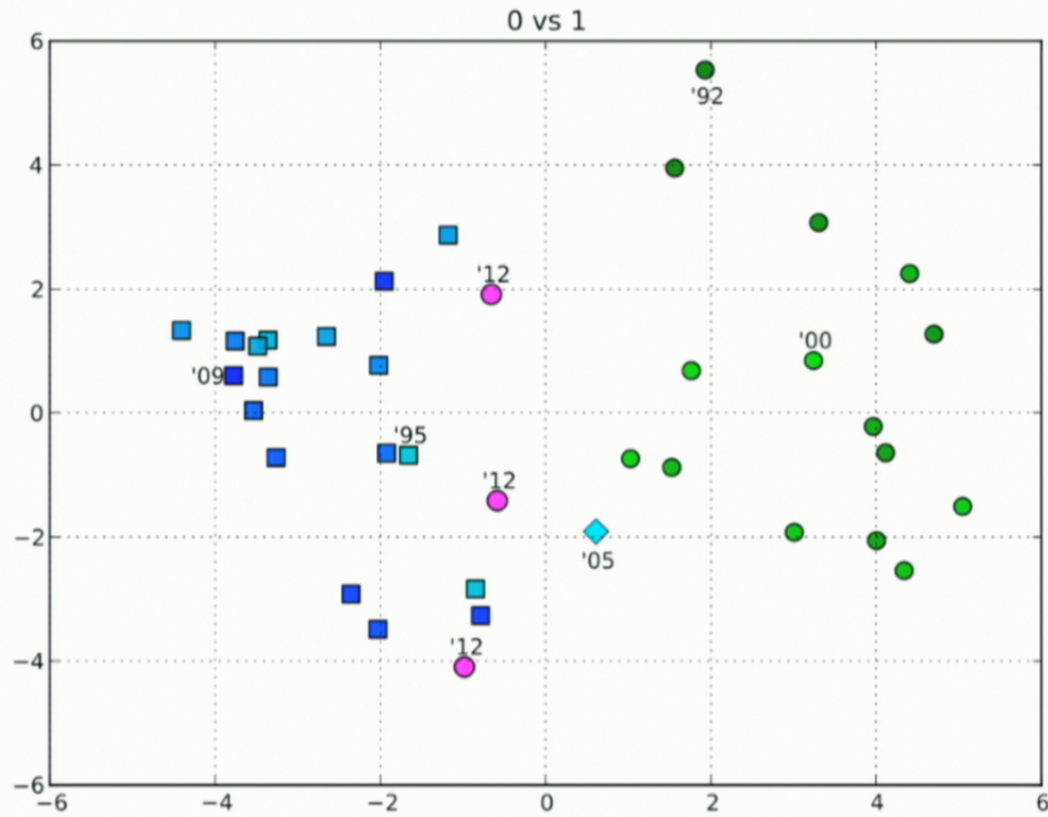**(components correspond to columns of $U$ and $V$).**

**With** $\sigma_i = \exp(-\xi_i/2)$, **entanglement spectrum "energy levels"** $\xi_i$ **give more info than entanglement entropy** $S = \sum_i \xi_i \exp(-\xi_i)$ **(a single number, thermodynamic entropy at $T = 1$) , and probe topological order of ground state** (Li/Haldane, arXiv:0805.0332)
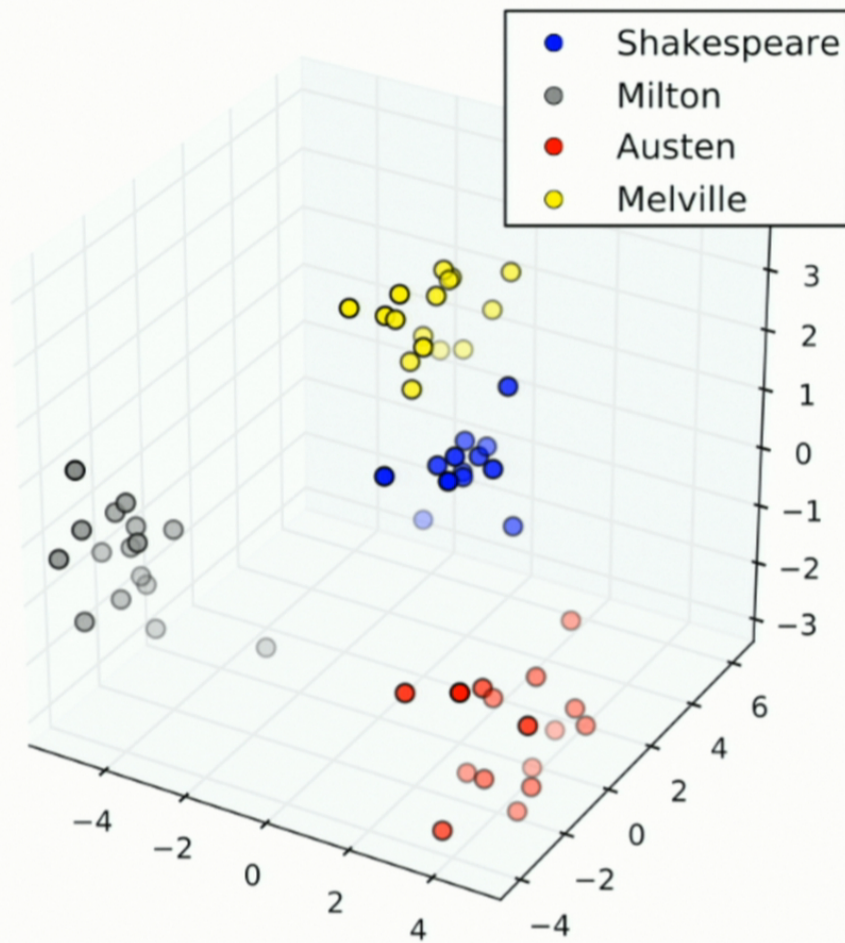
# Correspondence

## ArXiv screens spot fake papers

Unlike the computer-generated nonsense papers in some peer-reviewed subscription services (see *Nature* http://doi.org/r3n; 2014), the 500 or so preprints received daily by the automated repository arXiv are not pre-screened by humans. But sometimes automated assessment can be better than human diligence at enforcing standards.

The automated screens for outliers in arXiv include analysis of the probability distributions of words and their combinations, ensuring that they fall into patterns that are consistent with existing subject classes. This serves as a check of the subject categorizations provided by submitters, and helps to detect non-research content.

Fake papers generated by SCIgen software, for example, have a 'native dialect' that can be picked up by simple stylometric analysis (see J. N. G. Binongo *Chance* **16**, 9–17; 2003). The



**COUNTERFEIT CLUSTERS**

Nonsense papers generated by software such as SCIgen and Mathgen cluster separately from human-authored arXiv papers when analysed for stylistic word features.

- SCIgen
- ▼ Mathgen
- ■ SCIgen-physics
- Ike Antkare (SCIgen)
- ★ arXiv 14 March 2014
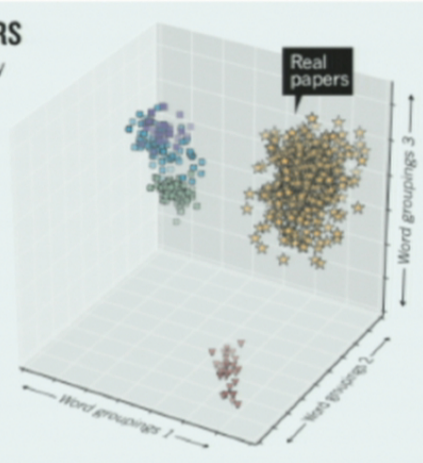
however, science advisers may encounter a conflict of interest if they are involved in administering public research funding.

Gluckman is the New Zealand Prime Minister's chief science adviser and chaired the panel that last year selected the National Science Challenges. He has been instrumental in publicizing and defending the new funding mechanism for meeting these goals (see go.nature.com/cmgkx1), which the government

## Journals must boost data sharing

The journal ecosystem is a powerful filter of scientific literature, promoting the best work into the best journals. Why not use a similar mechanism to encourage more comprehensive data sharing?

Several journals have introduced policies mandating that data be shared on a public archive at publication (see,

## Projects powered by free computing grid

Herman Tse describes the scientific output of IBM's World Community Grid as "lacklustre" (*Nature* **507**, 431; 2014). This is not the case: the 22 projects we have supported so far have generated more than 35 peer-reviewed papers in prominent journals. Our donated computing power has resulted in several important practical scientific advances.

For example, Japan's Chiba Cancer Center used our free computing power to screen three million drug candidates for treating neuroblastoma, a common childhood cancer. This yielded seven promising compounds that have no apparent side effects (Y. Nakamura *et al. Cancer Med.* **3**, 25–35; 2014).

Last June, Harvard University's Clean Energy Project announced some 35,000 organic materials that could double the efficiency of carbon-based solar cells, after using our grid to scan more than

# PCA on the Stopword Distributions
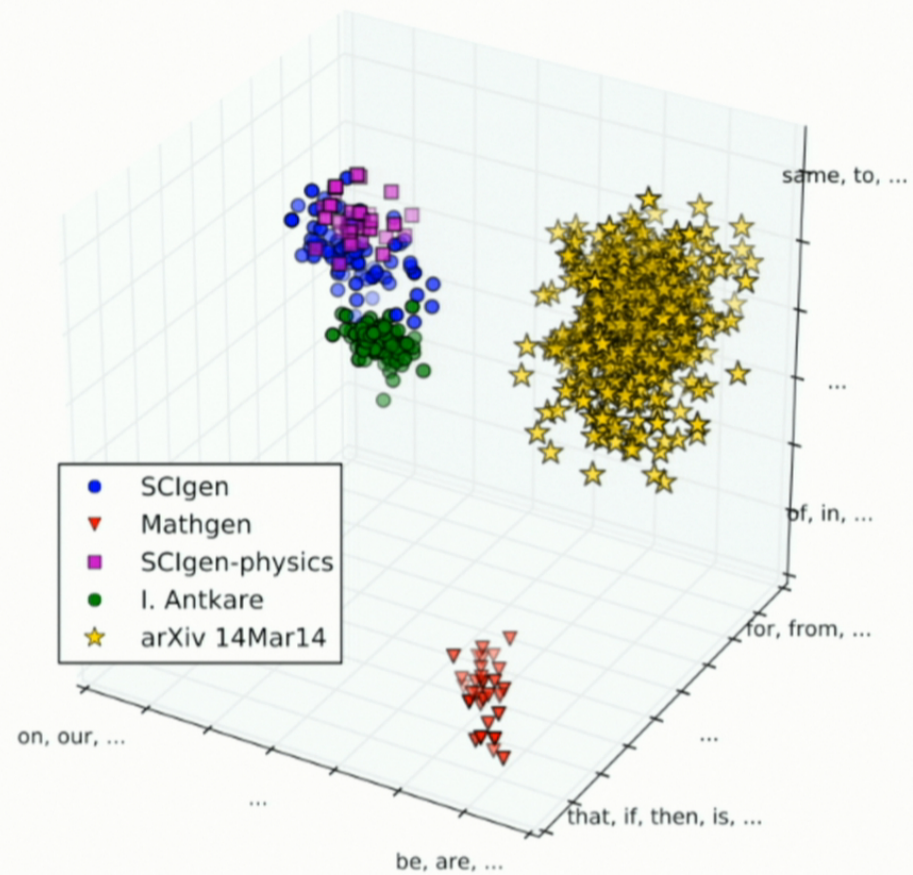
# Springer and Université Joseph Fourier release SciDetect to discover fake scientific papers

The new, open source software is publically available for free to the scientific and publishing communities

Grenoble | Heidelberg | New York, 23 March 2015

**Université Joseph Fourier** GRENOBLE

After intensive collaboration with Dr. Cyril Labbé from Université Joseph Fourier in Grenoble, France, Springer announces the release of *SciDetect*, a new software program that automatically checks for fake scientific papers. The open source software discovers text that has been generated with the SCIgen computer program and other fake-paper generators like Mathgen and Physgen. Springer uses the software in its production workflow to provide additional, fail-safe checking. Springer and the University are releasing the software under the GNU General Public License, Version 3.0 (GPLv3) so others in the scientific and publishing communities can benefit.

*SciDetect* scans Extensible Markup Language (XML) and Adobe Portable Document Format (PDF) files and compares them against a corpus of fake scientific papers. *SciDetect* indicates whether an entire document or its parts are genuine or not. The software reports suspicious activity by relying on sensitivity thresholds that can be easily adjusted. SciDetect is highly flexible and can be quickly customized to cope with new methods of automatically generating fake or random text.

syllogism                    a:b :: c:d

Paris - France + Italy = ?

syllogism             a:b :: c:d

Paris - France + Italy =

Rome

## arxplor.lassp.cornell.edu

(20 slides from A.Alemi presentation at March Meeting '14 Denver)

After filtering:

- 7 years: Apr 2007 - Feb 2014
- 488,072 articles.
- 422,704 authors.
- 1,285,320 unique "words".

Example "words":

- "singular_value_decomposition",
- "black_hole",
- "aps_march_meeting"

| word | cosine | word | cosine | word | cosine |
|---|---|---|---|---|---|
| electron | 1.0 | physics | 1.0 | blue | 1.0 |
| electrons | 0.83 | theoretical_physics | 0.76 | red | 0.91 |
| positron | 0.67 | particle_physics | 0.72 | orange | 0.87 |
| conduction_electron | 0.65 | nuclear_physics | 0.7 | cyan | 0.87 |
| carriers | 0.64 | astrophysics | 0.68 | purple | 0.86 |
| unpaired_electron | 0.63 | astronomy | 0.68 | magenta | 0.86 |
| electron_gas | 0.63 | materials_science | 0.67 | yellow | 0.85 |
| electronhole | 0.63 | phyiscs | 0.66 | blue_red | 0.85 |
| impurity | 0.63 | astronomy_louisiana_state_univ... | 0.66 | violet | 0.85 |
| oneelectron | 0.62 | bern_bern_switzerland_18 | 0.66 | blue_green | 0.83 |
| mobile_electrons | 0.62 | tennessee_knoxville_tennessee_... | 0.66 | light_blue | 0.82 |

| word | cosine | word | cosine | word | cosine |
|---|---|---|---|---|---|
| svd | 1.0 | graphene | 1.0 | python | 1.0 |
| truncated_svd | 0.7 | single_layer_graphene | 0.9 | source_code | 0.83 |
| svds | 0.69 | monolayer_graphene | 0.9 | python_program... | 0.83 |
| qr_decomposition | 0.66 | bilayer_graphene | 0.88 | java | 0.83 |
| svd_decomposition | 0.65 | multilayer_graphene | 0.88 | scripting_language | 0.82 |
| eigendecomposition | 0.62 | graphene_monolayer | 0.88 | command_line | 0.82 |
| music_algorithm | 0.61 | graphene_sheets | 0.87 | scipy | 0.82 |
| omp_algorithm | 0.61 | suspended_graphene | 0.85 | package | 0.82 |
| cholesky_factorization | 0.61 | graphene_monolayers | 0.85 | scripts | 0.82 |
| alternating_mini... | 0.6 | graphene_sheet | 0.85 | libraries | 0.82 |

| word | cosine | word | cosine | word | cosine | word | cosine | word | cosine |
|---|---|---|---|---|---|---|---|---|---|
| john | ● 1.0 | dmitri | ● 1.0 | wang | ● 1.0 | stefano | ● 1.0 | pierre | ● 1.0 |
| william | ♥ 0.86 | dmitry | ♥ 0.78 | chen | ♥ 0.95 | paolo | ♥ 0.9 | alain | ♥ 0.82 |
| michael | ♥ 0.84 | mikhail | ♥ 0.78 | zhang | ♥ 0.94 | francesco | ♥ 0.89 | olivier | ♥ 0.8 |
| edward | ♥ 0.84 | oleg | ♥ 0.77 | liu | ♥ 0.94 | matteo | ♥ 0.88 | philippe | ♥ 0.79 |
| david | ♥ 0.84 | sergey | ♥ 0.76 | zhou | ♥ 0.93 | michele | ♥ 0.88 | frederic | ♥ 0.79 |
| robert | ♥ 0.84 | konstantin | ♥ 0.76 | zhao | ♥ 0.93 | giuseppe | ♥ 0.87 | stephane | ♥ 0.78 |
| andrew | ♥ 0.84 | igor | ♥ 0.76 | zhu | ♥ 0.92 | alessandro | ♥ 0.87 | sylvain | ♥ 0.78 |
| james | ♥ 0.83 | alexey | ♥ 0.75 | huang | ♥ 0.91 | davide | ♥ 0.87 | jean_francois | ♥ 0.78 |
| peter | ♥ 0.83 | anatoly | ♥ 0.75 | fang | ♥ 0.9 | giorgio | ♥ 0.87 | sebastien | ♥ 0.78 |
| stephen | ♥ 0.82 | ilya | ♥ 0.74 | wei | ♥ 0.9 | riccardo | ♥ 0.87 | benoit | ♥ 0.78 |
| brian | ♥ 0.82 | ivan | ♥ 0.73 | guo | ♥ 0.9 | enrico | ♥ 0.87 | thierry | ♥ 0.77 |
| philip | ♥ 0.82 | nikolay | ♥ 0.73 | ding | ♥ 0.89 | francesca | ♥ 0.87 | guillaume | ♥ 0.77 |

| word | cosine | word | cosine | word | cosine | word | cosine | word | cosine |
|---|---|---|---|---|---|---|---|---|---|
| john | 1.0 | dmitri | 1.0 | wang | 1.0 | stefano | 1.0 | pierre | 1.0 |
| william | 0.86 | dmitry | 0.78 | chen | 0.95 | paolo | 0.9 | alain | 0.82 |
| michael | 0.84 | mikhail | 0.78 | zhang | 0.94 | francesco | 0.89 | olivier | 0.8 |
| edward | 0.84 | oleg | 0.77 | liu | 0.94 | matteo | 0.88 | philippe | 0.79 |
| david | 0.84 | sergey | 0.76 | zhou | 0.93 | michele | 0.88 | frederic | 0.79 |
| robert | 0.84 | konstantin | 0.76 | zhao | 0.93 | giuseppe | 0.87 | stephane | 0.78 |
| andrew | 0.84 | igor | 0.76 | zhu | 0.92 | alessandro | 0.87 | sylvain | 0.78 |
| james | 0.83 | alexey | 0.75 | huang | 0.91 | davide | 0.87 | jean_francois | 0.78 |
| peter | 0.83 | anatoly | 0.75 | fang | 0.9 | giorgio | 0.87 | sebastien | 0.78 |
| stephen | 0.82 | ilya | 0.74 | wei | 0.9 | riccardo | 0.87 | benoit | 0.78 |
| brian | 0.82 | ivan | 0.73 | guo | 0.9 | enrico | 0.87 | thierry | 0.77 |
| philip | 0.82 | nikolay | 0.73 | ding | 0.89 | francesca | 0.87 | guillaume | 0.77 |

torque - force + momentum =

orbital_angular_momentum

newtonian_mechanics - isaac_newton + albert_einstein =

newtonian_mechanics - isaac_newton + albert_einstein =

special_relativity

gravity - Newton + Hawking =

gravity - Newton + Hawking =

black_hole_evaporation

smuon - muon + higgs =

**gravity-newton+hawking**

hawking ❤0.8

an_evaporating_black_hole ❤0.73

black_hole_evaporation ❤0.7

hawking_effect ❤0.7

cosmic_censorship ❤0.69

gravity ❤0.66

**torque-force+momentum=**

momentum ❤0.72

angular_momentum_conservation ❤0.59

longitudinal_component ❤0.56

orbital_angular_momentum ❤0.55

angular_momentum_lz ❤0.55

helicity ❤0.54

**newtonian_mechanics-isaac_newton+albert_einstein=**

newtonian_mechanics ❤0.78

special_relativity ❤0.57

newtonian_dynamics ❤0.55

planetary_motions ❤0.54

material_bodies ❤0.53

aristotelian ❤0.51

**smuon-muon+higgs**

higgsino ❤0.81

sfermions ❤0.79

higgsinos ❤0.78

heavy_higgs ❤0.78

lightest_higgs_boson ❤0.78

smuon ❤0.78

# Using the same projections, different parts of vector space.

tau

higgs

stau

slepton   higgsino gluino squark

lepton

sneutrino

muon meutrino

smuon

quark gluon

selectron

electron
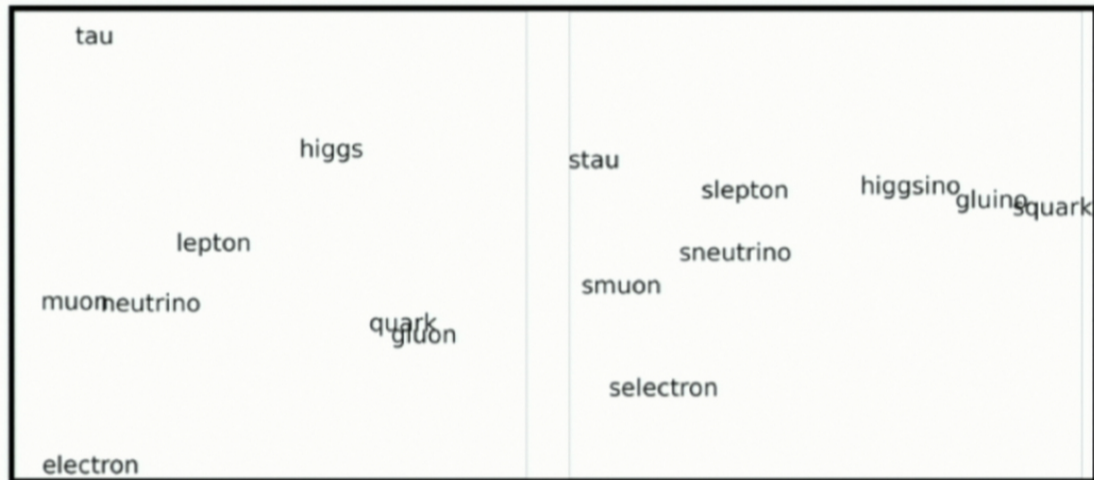
## word2vec knows supersymmetry

Full text of the higgs article, colored by K-means clustering of words

- 🔴 astro
- 🔵 cond-mat
- 🟢 cs
- 🟣 hep
- 🟠 math
- 🟡 places
- 🟤 references
- 🩷 names
- ⚪ equations
- ⚪ english

# James Clerk Maxwell (Feb 1856)

===================== 88 =====================

ESSAY FOR THE APOSTLES ON
'ANALOGIES IN NATURE'

FEBRUARY 1856

From Campbell and Garnett, *Life of Maxwell*[1]

## ARE THERE REAL ANALOGIES IN NATURE?[1]

In the ancient and religious foundation of Peterhouse there is observed this rule, that whoso makes a pun shall be counted the author of it, but that whoso pretends to find it out shall be counted the publisher of it, and that both shall be fined. Now, as in a pun two truths lie hid under one expression, so in an analogy one truth is discovered under two expressions. Every question concerning analogies is therefore the reciprocal of a question concerning puns, and the solutions can be transposed by reciprocation. But since we are still in doubt as to the legitimacy of reasoning by analogy, and as reasoning even by paradox has been pronounced less heinous than reasoning by puns, we must adopt the direct method with respect to analogy, and then, if necessary, deduce by reciprocation the theory of puns.

That analogies appear to exist is plain in the face of things, for all parables, fables, similes, metaphors, tropes, and figures of speech are analogies, natural or revealed, artificial or concealed. The question is entirely of their reality. Now, no question exists as to the possibility of an analogy without a mind to recognise it – that is rank nonsense. You might as well talk of a demonstration or refutation existing unconditionally. Neither is there any question as to the occurrence of analogies to our minds. They are as plenty as reasons, not to say blackberries. For, not to mention all the things in external nature which men have seen as the projections of things in their own minds, the whole framework

# Basic Intuition    (From GloVE)

For semantic applications like the analogy task, the vector space embedding should respect the ratios of conditional probabilities. For example, the ratio

$$\frac{p(k|\text{ice})}{p(k|\text{steam})}$$

is high for $k$ = solid,
intermediate for $k$ = water, fashion
and low for $k$ = gas.

So if we're interested in thermodynamic phase, we learn that solid and gas are relevant to the distinction between ice and steam, and water and fashion are not.

$$\left[ p(k|i) = X_{ik}/X_i \right. = \text{probability word } k \text{ appears in context of word } i,$$
$$X_{ik} = \text{co-occurrence count}, X_i = \text{total \# occurrences of word } i. \left.\right]$$

# The analogy task: solve a:b :: c:?

Want word vectors $v_a, v_b, v_c, v_d$ to satisfy linear relation

$$d = \operatorname{argmin}_d |(v_b - v_a) - (v_d - v_c)|^2$$

How to find vectors that do this?

Example: man:woman :: king:?

First need analytic representation of task. Note that most contexts $\chi$ have

$$\frac{p(\chi|\text{king})}{p(\chi|\text{queen})} \approx \frac{p(\chi|\text{man})}{p(\chi|\text{woman})}$$

since the ratios will be roughly one for most words (contexts not sensitive). But different for gendered words: $\chi$ = dress, he, she, Elizabeth, Harry, . . .

# Computational Objective

So average over contexts, and to solve analogy find word w that minimizes:

$$\sum_{\chi}\left(\log\frac{p(\chi|\mathrm{man})}{p(\chi|\mathrm{woman})} - \log\frac{p(\chi|\mathrm{king})}{p(\chi|\mathrm{w})}\right)^2 \qquad (\star)$$

But this analytic approach to analogy task is not computationally efficient: expensive to assay for all words w in large vocabulary.

Much easier: simple vector addition on low dimensional space?

Digression: what does it mean to factorize a low-rank matrix?
Consider familiar case of symmetric $N \times N$ matrix, diagonalize:

$$M = O\Lambda O^T$$

$$\Lambda = \begin{pmatrix} \lambda_1 & & & & & & \\ & \ddots & & & & & \\ & & \lambda_n & & & & \\ & & & \varepsilon_1 & & & \\ & & & & \varepsilon_2 & & \\ & & & & & \ddots & \\ & & & & & & \varepsilon_{N-n} \end{pmatrix}$$

Ignore all the small eigenvalues: $M_{ij} \approx \sum_{a=1}^{n} \lambda_a O_{ia} O_{ja} = \vec{v}_i \cdot \vec{v}_j$
where $(v_i)_a = \sqrt{\lambda_a} O_{ia}$ are the components of vectors living in a much
smaller space for $n \ll N$

# Suppositions

Recall mutual information

$$I(X, Y) = \sum_{x,y} p(x, y) \log \frac{p(x,y)}{p(x)p(y)} = \sum_{x,y} p(x, y) \text{PMI}(x, y)$$

is a sum of "pointwise mutual information"s $\text{PMI}(x, y) = \log \frac{p(x,y)}{p(x)p(y)}$

Suppose:

1) $\text{PMI}(w, w') = \log \frac{p(w,w')}{p(w)p(w')}$ is low rank ($n \sim \log N$ rather than $N$),
   so it factorizes:

$$\text{PMI}(w, w') \approx v_w \cdot v_{w'}$$

2) word vectors are 'isotropic': $\sum_w |w\rangle\langle w| \approx I$ (eigenvalues $1 + \delta$)

[In a generative model, can show $\log p(w) \approx |v_w|^2/2d - \log Z$,
$\log p(w, w') \approx |v_w + v'_w|^2/2d - 2\log Z$,
$\implies$ 1) $\text{PMI}(w, w') \propto v_w \cdot v'_w$ (factorization)]

# Key Relation: $(1,2) \Longrightarrow (\star)$

By factorization (1),

$$v \cdot w \sim \text{PMI}(v, w) = \log \frac{p(w, v)}{p(w)p(v)} = \log \frac{p(w|v)}{p(w)}$$

By isotropy (2), $|v|^2 \approx \sum_w \langle v|w \rangle \langle w|v \rangle = \sum_w (v \cdot w)^2$, and thus

$$\min_d \left| (v_b - v_a) - (v_d - v_c) \right|^2$$

$$\propto \min_d \sum_w \left| (v_b \cdot w - v_a \cdot w) - (v_d \cdot w - v_c \cdot w) \right|^2$$

$$\propto \min_d \sum_w \left( \log \frac{p(w|a)}{p(w|b)} - \log \frac{p(w|c)}{p(w|d)} \right)^2 \qquad (\star)$$

# Summary

The word **d** that minimizes

$$\min_{d} \sum_{w} \left( \log \frac{p(w|\text{a})}{p(w|\text{b})} - \log \frac{p(w|\text{c})}{p(w|\text{d})} \right)^2$$

is a likely solution to the analogy task a:b::c:?

Under assumptions (1,2), if we can find isotropic vectors such that $v_w \cdot v_{w'} \propto \text{PMI}(w, w')$, then the vector $v_d$ that minimizes

$$\min_{d} |(v_b - v_a) - (v_d - v_c)|^2$$

solves the same task.

The (low dimensional) word vectors $v_w$ are precomputed once and for all, so finding the above minimum is computationally efficient.

# Readers and Authors live in the same vector space

Extend article context to readers:

Reader vectors => Recommender System

Extend article context to authors:

Author vectors => Referee Selection

# Text Overlap

Text "reuse" by global researchers in a scholarly corpus

Simple n-gram analysis of the texts in arXiv covering over 20 years

Everything from

- dozens of pages verbatim from 3rd party lecture notes for PhD theses

- large sections of Wikipedia entries for introductory material in articles

- series of articles by overlapping authors each greater than 50% overlap with preceding

- articles assembled in whole or part from one or more other articles by different authors, with or without attribution

Majority have found way undetected into conventional publication venues

Shed light on sociology, mentality, methodology, and demography of perps?
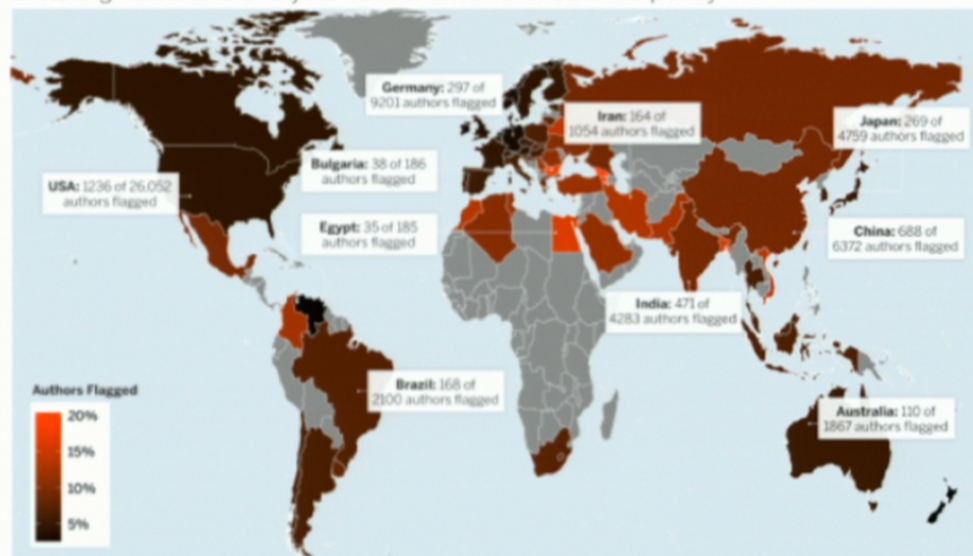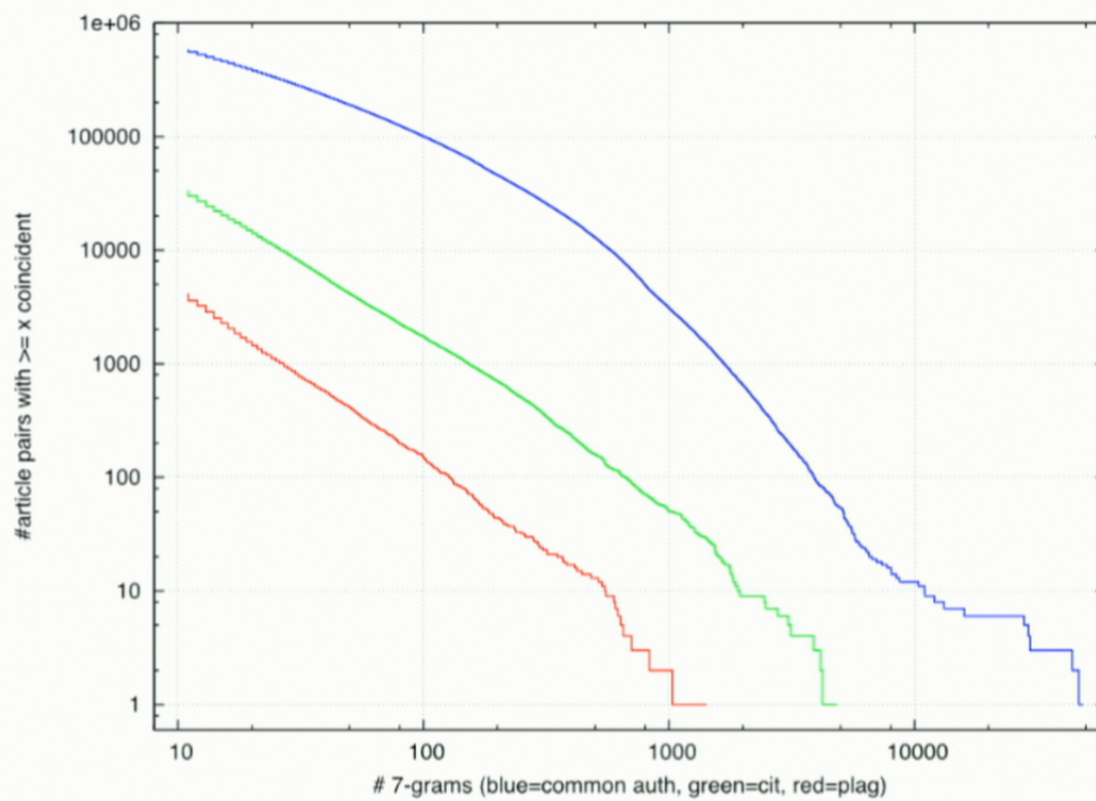
Number of article pairs with at least the number of overlapping 7 grams given on horizontal, log–log scale, red signifies without attribution, green with attribution, and blue with at least one common author.

# Caveats

- Not "plagiarism" in its most general form — i.e., unattributed use of ideas (whether or not text is copied).

- no attempt to detect text copied from sources outside of arXiv

- simple factual statement regarding textual overlap of materials only within arXiv (not Wikipedia, print literature, web search etc.)

- watch out for: famous quotes, experimental articles (author lists), review articles, conf proceedings [but note cs/info ?], other benign (refs not stripped), math (?), explicit quotes, hidden pdf text, ...

# high threshold, cont'd

high threshold

Articles flagged as having text overlap with articles "by other authors" must have at least multiple consecutive sentences in common. Overlap between articles having at least one coauthor in common is permitted an even higher threshold: typically at least roughly 1/3 of the content of

Threshold for appearance of the admin note is set quite high – many articles with smaller amounts of detected overlap are not noted.

"The appearance of an arXiv admin note does not suggest misconduct on the part of the author, or that an article does not contain original work. Sometimes it simply serves to suggest a related article, or can serve as a quality flag. (There is a statistically significant correlation between the amount of reused content in an article and a smaller number of citations received years later.)"

(in practice also use size of contiguous blocks)

Additional exceptions for articles having a coauthor in common: articles marked by authors in the "Comments:" as review articles, or theses, conference proceedings, book contributions, and so on, are not noted, because such overlaps, whether or not desirable, appear to be common practice.

1109.4897: Measurement of the neutrino velocity with the OPERA detect...

Legend:
- total
- None
- arxiv.org
- bbc.co.uk
- facebook.com
- press.web.cern.ch
- guardian.co.uk
- public.web.cern.ch
- google
- nature.com
- nytimes.com

accesses / week (total 412450+133712=546162)

From: "Albert E." <a.einstein@patents.bern.ch>
To: opera@grandsasso.it
Date: Tue, Sep 26, 1905 at 12:40 PM
Subject: Re: Überlichtgeschwindigkeit

Sehr geehrte Opera,

Bitte überprüfen Sie die Glasfaser-Anschlüsse!

Mit freundlichen Grüßen,
AE
(der wackere Schwabe)

1503.01104 You Can Run, You Can Hide: The Epidemiology and Statistical ...

**arXiv.org > q-bio > arXiv:1503.01104**

Search or Article–id (Help | Advanced search)
[          ] All papers [◇] Go!

Quantitative Biology > Populations and Evolution

# You Can Run, You Can Hide: The Epidemiology and Statistical Mechanics of Zombies

Alexander A. Alemi, Matthew Bierbaum, Christopher R. Myers, James P. Sethna

(Submitted on 4 Mar 2015 (v1), last revised 5 Mar 2015 (this version, v2))

We use a popular fictional disease, zombies, in order to introduce techniques used in modern epidemiology modelling, and ideas and techniques used in the numerical study of critical phenomena. We consider variants of zombie models, from fully connected continuous time dynamics to a full scale exact stochastic dynamic simulation of a zombie outbreak on the continental United States. Along the way, we offer a closed form analytical expression for the fully connected differential equation, and demonstrate that the single person per site two dimensional square lattice version of zombies lies in the percolation universality class. We end with a quantitative study of the full scale US outbreak, including the average susceptibility of different geographical regions.

Comments: 12 pages, 13 figures
Subjects: Populations and Evolution (q-bio.PE); Popular Physics (physics.pop-ph)
Cite as: arXiv:1503.01104 [q-bio.PE]
    (or arXiv:1503.01104v2 [q-bio.PE] for this version)

## Submission history

From: Alexander Alemi [view email]
[v1] Wed, 4 Mar 2015 00:36:09 GMT (3710kb,D)
[v2] Thu, 5 Mar 2015 03:24:37 GMT (3710kb,D)

**Download:**
- PDF
- Other formats

Current browse context:
q-bio.PE
< prev | next >
new | recent | 1503

Change to browse by:
physics
    physics.pop-ph
q-bio

References & Citations
- NASA ADS

10 blog links (what is this?)

Bookmark (what is this?)

# In the news, cont'd

- earth and planetary (comets, meteors, moon)

- cosmology (wmap, cc, ...)

- HEP Exp (Higgs, new particles)

- HEP Th+Ph (theoretical developments)

- Computer Sci (networks, games)

- April Fools'

- Neo-Einstein / The Fringe

# Amateur Fascination

- **Math: Riemann Hypothesis, Goldbach Conjecture, Fermat's Last theorem, 4-color theorem**

- **Computer Science: P v NP**

- **Physics: Special Relativity is Wrong, General Relativity is Wrong, Quantum Mechanics is Wrong, Theories of Everything**

- **More generally: progress (real or imagined) on famous problems**

'15 so far
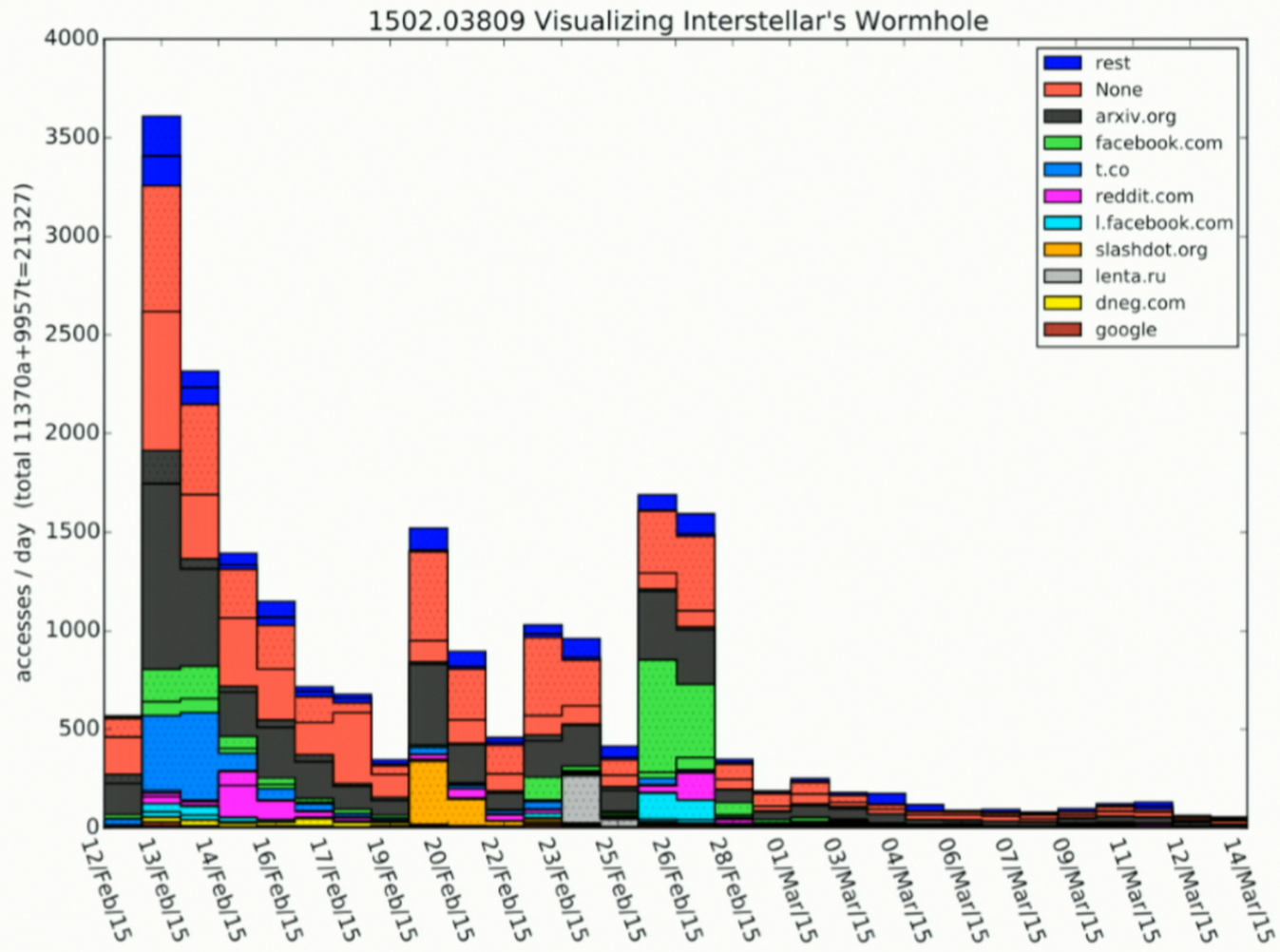
1502.03809 Visualizing Interstellar's Wormhole

**General Relativity and Quantum Cosmology**

# Visualizing Interstellar's Wormhole

Oliver James (1), Eugenie von Tunzelmann (1), Paul Franklin (1), Kip S. Thorne (2)
((1) Double Negative Ltd (2) California Institute of Technology)

(Submitted on 12 Feb 2015 (v1), last revised 16 Feb 2015 (this version, v2))

Christopher Nolan's science fiction movie Interstellar offers a variety of opportunities for students in elementary courses on general relativity theory. This paper describes such opportunities, including: (i) At the motivational level, the manner in which elementary relativity concepts underlie the wormhole visualizations seen in the movie. (ii) At the briefest computational level, instructive calculations with simple but intriguing wormhole metrics, including, e.g., constructing embedding diagrams for the three-parameter wormhole that was used by our visual effects team and Christopher Nolan in scoping out possible wormhole geometries for the movie. (iii) Combining the proper reference frame of a camera with solutions of the geodesic equation, to construct a light-ray-tracing map backward in time from a camera's local sky to a wormhole's two celestial spheres. (iv) Implementing this map, for example in Mathematica, Maple or Matlab, and using that implementation to construct images of what a camera sees when near or inside a wormhole. (v) With the student's implementation, exploring how the wormhole's three parameters influence what the camera sees---which is precisely how Christopher Nolan, using our implementation, chose the parameters for \emph{Interstellar}'s wormhole. (vi) Using the student's implementation, exploring the wormhole's Einstein ring, and particularly the peculiar motions of star images near the ring; and exploring what it looks like to travel through a wormhole.

**Download:**
- PDF
- Other formats

Current browse context:
gr-qc
< prev | next >
new | recent | 1502

Change to browse by:
physics
  physics.pop-ph

References & Citations
- INSPIRE HEP
  (refers to | cited by )
- NASA ADS

1 blog link (what is this?)

Bookmark (what is this?)

1502.03182 PowerSpy: Location Tracking using Mobile Device Power Analys...

arXiv.org > gr-qc > arXiv:1502.03809

**General Relativity and Quantum Cosmology**

# Visualizing Interstellar's Wormhole

Oliver James (1), Eugenie von Tunzelmann (1), Paul Franklin (1), Kip S. Thorne (2)
((1) Double Negative Ltd (2) California Institute of Technology)

Christopher Nolan's science fiction movie Interstellar offers a variety of opportunities for students in elementary courses on general relativity theory. This paper describes such opportunities, including: (i) At the motivational level, the manner in which elementary relativity concepts underlie the wormhole visualizations seen in the movie. (ii) At the briefest computational level, instructive calculations with simple but intriguing wormhole metrics, including, e.g., constructing embedding diagrams for the three-parameter wormhole that was used by our visual effects team and Christopher Nolan in scoping out possible wormhole geometries for the movie. (iii) Combining the proper reference frame of a camera with solutions of the geodesic equation, to construct a light-ray-tracing map backward in time from a camera's local sky to a wormhole's two celestial spheres. (iv) Implementing this map, for example in Mathematica, Maple or Matlab, and using that implementation to construct images of what a camera sees when near or inside a wormhole. (v) With the student's implementation, exploring how the wormhole's three parameters influence what the camera sees---which is precisely how Christopher Nolan, using our implementation, chose the parameters for \emph{Interstellar}'s wormhole. (vi) Using the student's implementation, exploring the wormhole's Einstein ring, and particularly the peculiar motions of star images near the ring; and exploring what it looks like to travel through a wormhole.

**Download:**

- PDF
- Other formats

Current browse context:

gr-qc

< prev | next >

new | recent | 1502

Change to browse by:

physics
    physics.pop-ph
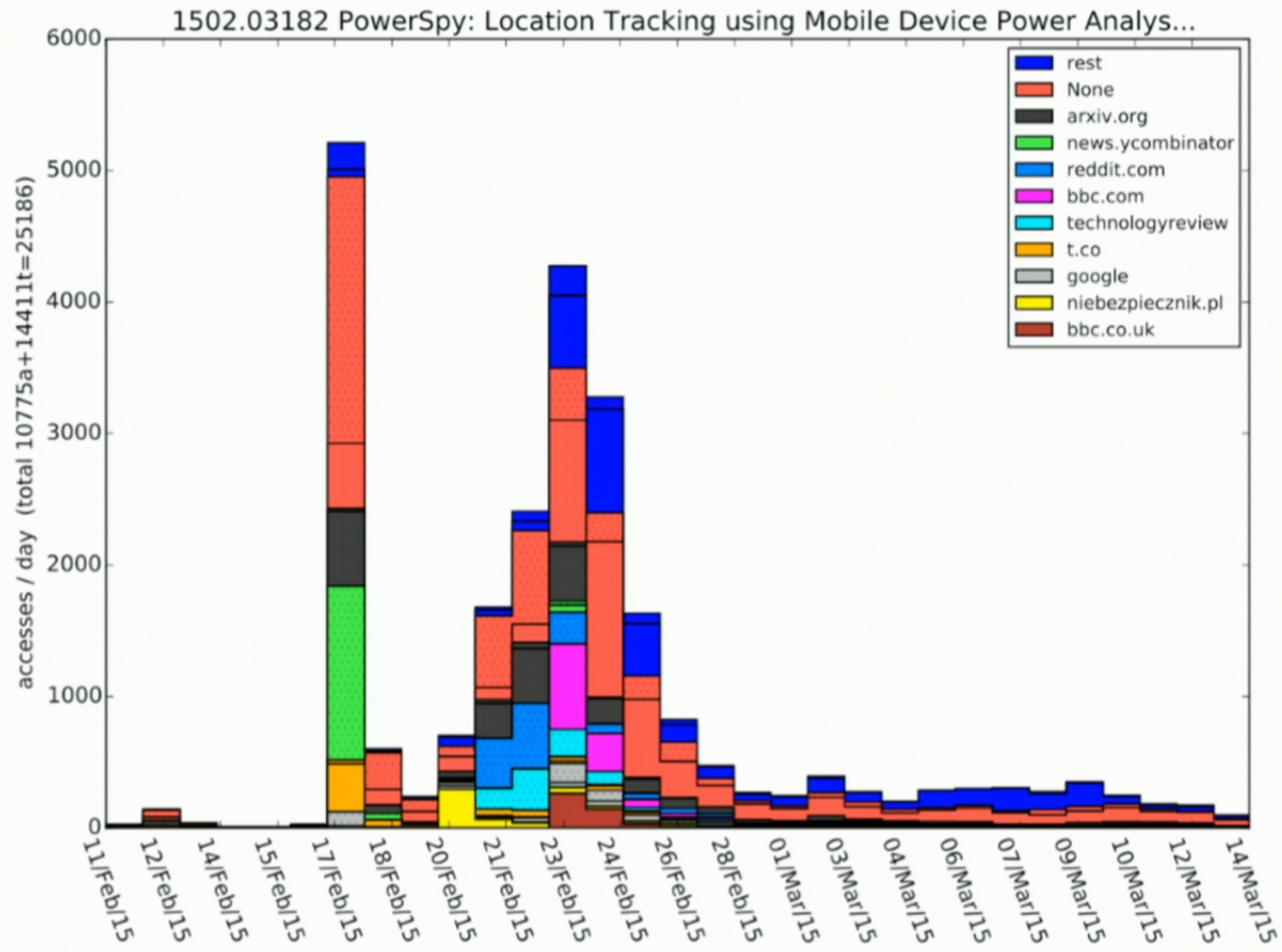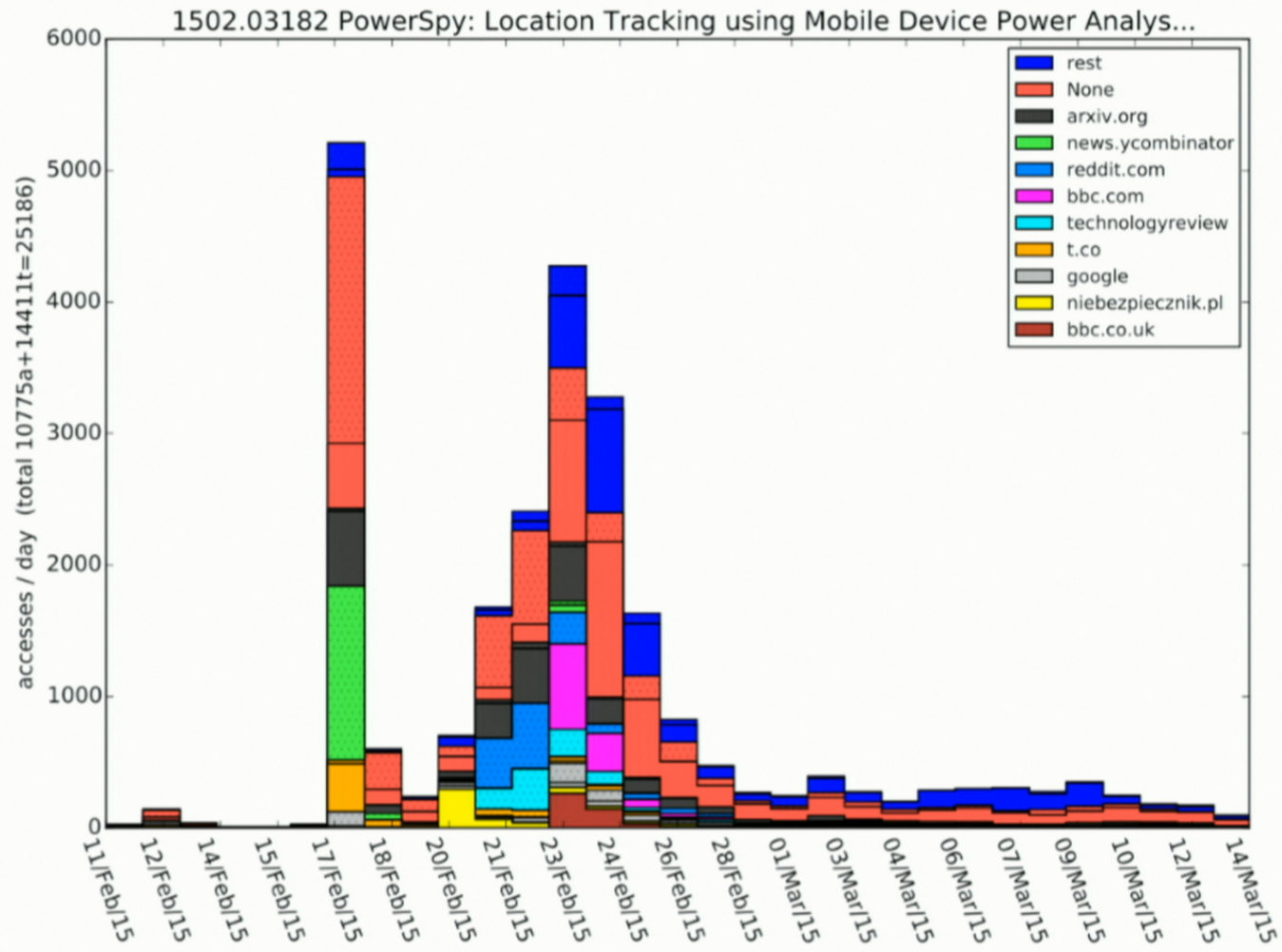
References & Citations

- INSPIRE HEP
    (refers to | cited by )
- NASA ADS

1 blog link (what is this?)

Bookmark (what is this?)

1502.03182 PowerSpy: Location Tracking using Mobile Device Power Analys...

1502.01852 Delving Deep into Rectifiers: Surpassing Human-Level Perform...

arXiv.org > cs > arXiv:1502.01852

Search or Article-id (Help | Advanced search)
All papers  Go!

Computer Science > Computer Vision and Pattern Recognition

# Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification

Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun

(Submitted on 6 Feb 2015)

Rectified activation units (rectifiers) are essential for state-of-the-art neural networks. In this work, we study rectifier neural networks for image classification from two aspects. First, we propose a Parametric Rectified Linear Unit (PReLU) that generalizes the traditional rectified unit. PReLU improves model fitting with nearly zero extra computational cost and little overfitting risk. Second, we derive a robust initialization method that particularly considers the rectifier nonlinearities. This method enables us to train extremely deep rectified models directly from scratch and to investigate deeper or wider network architectures. Based on our PReLU networks (PReLU-nets), we achieve 4.94% top-5 test error on the ImageNet 2012 classification dataset. This is a 26% relative improvement over the ILSVRC 2014 winner (GoogLeNet, 6.66%). To our knowledge, our result is the first to surpass human-level performance (5.1%, Russakovsky et al.) on this visual recognition challenge.

Subjects: Computer Vision and Pattern Recognition (cs.CV); Artificial Intelligence (cs.AI); Learning (cs.LG)
Cite as: arXiv:1502.01852 [cs.CV]
(or arXiv:1502.01852v1 [cs.CV] for this version)

## Submission history

From: Kaiming He [view email]

**Download:**
- PDF
- Other formats

Current browse context:
cs.CV
< prev | next >
new | recent | 1502

Change to browse by:
cs
    cs.AI
    cs.LG

References & Citations
- NASA ADS

Bookmark (what is this?)

1502.01710 Text Understanding from Scratch

Legend:
- rest
- None
- arxiv.org
- news.ycombinator
- weibo.com
- t.co
- reddit.com
- google
- datatau.com
- facebook.com
- feedly.com

y-axis: accesses / day (total 18526a+17587t=36113)

**Computer Science > Learning**
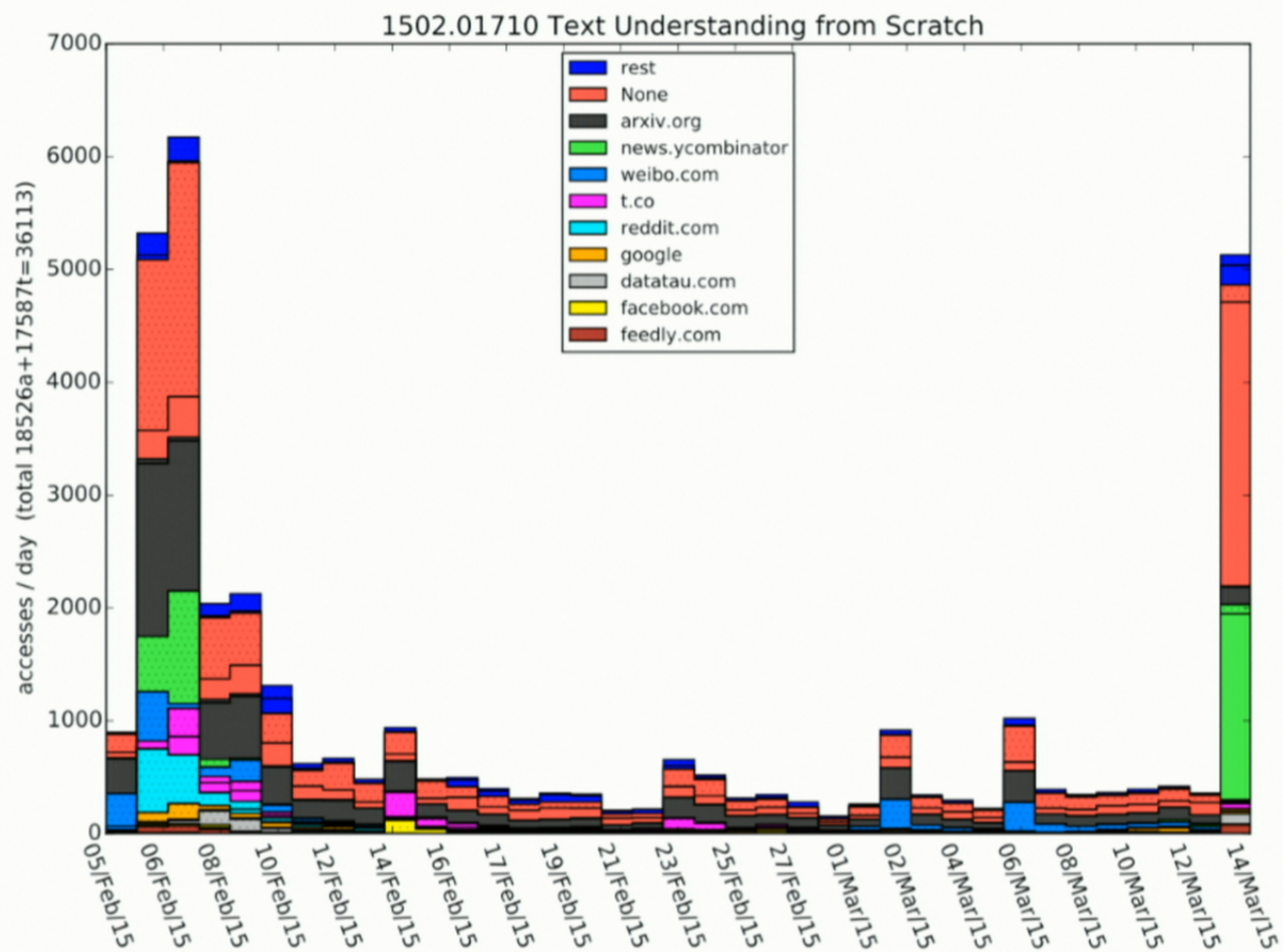
# Text Understanding from Scratch

Xiang Zhang, Yann LeCun

(Submitted on 5 Feb 2015)

This article demontrates that we can apply deep learning to text understanding from character-level inputs all the way up to abstract text concepts, using temporal convolutional networks (ConvNets). We apply ConvNets to various large-scale datasets, including ontology classification, sentiment analysis, and text categorization. We show that temporal ConvNets can achieve astonishing performance without the knowledge of words, phrases, sentences and any other syntactic or semantic structures with regards to a human language. Evidence shows that our models can work for both English and Chinese.

Subjects: **Learning (cs.LG)**; Computation and Language (cs.CL)
Cite as:    **arXiv:1502.01710 [cs.LG]**
            (or **arXiv:1502.01710v1 [cs.LG]** for this version)

**Submission history**

From: Xiang Zhang [view email]
[v1] Thu, 5 Feb 2015 20:45:19 GMT (85kb,D)

*Which authors of this paper are endorsers? | Disable MathJax (What is MathJax?)*

Link back to: arXiv, form interface, contact.

**Download:**

- PDF
- Other formats

Current browse context:
**cs.LG**
< prev | next >
new | recent | 1502

Change to browse by:

cs
    cs.CL

**References & Citations**

- NASA ADS

**Bookmark** (what is this?)

1501.03798 A nearby M star with three transiting super-Earths discovere...

arXiv.org > astro-ph > arXiv:1501.03798

Search or Article-id     (Help | Advanced search)

All papers     Go!

Astrophysics > Earth and Planetary Astrophysics
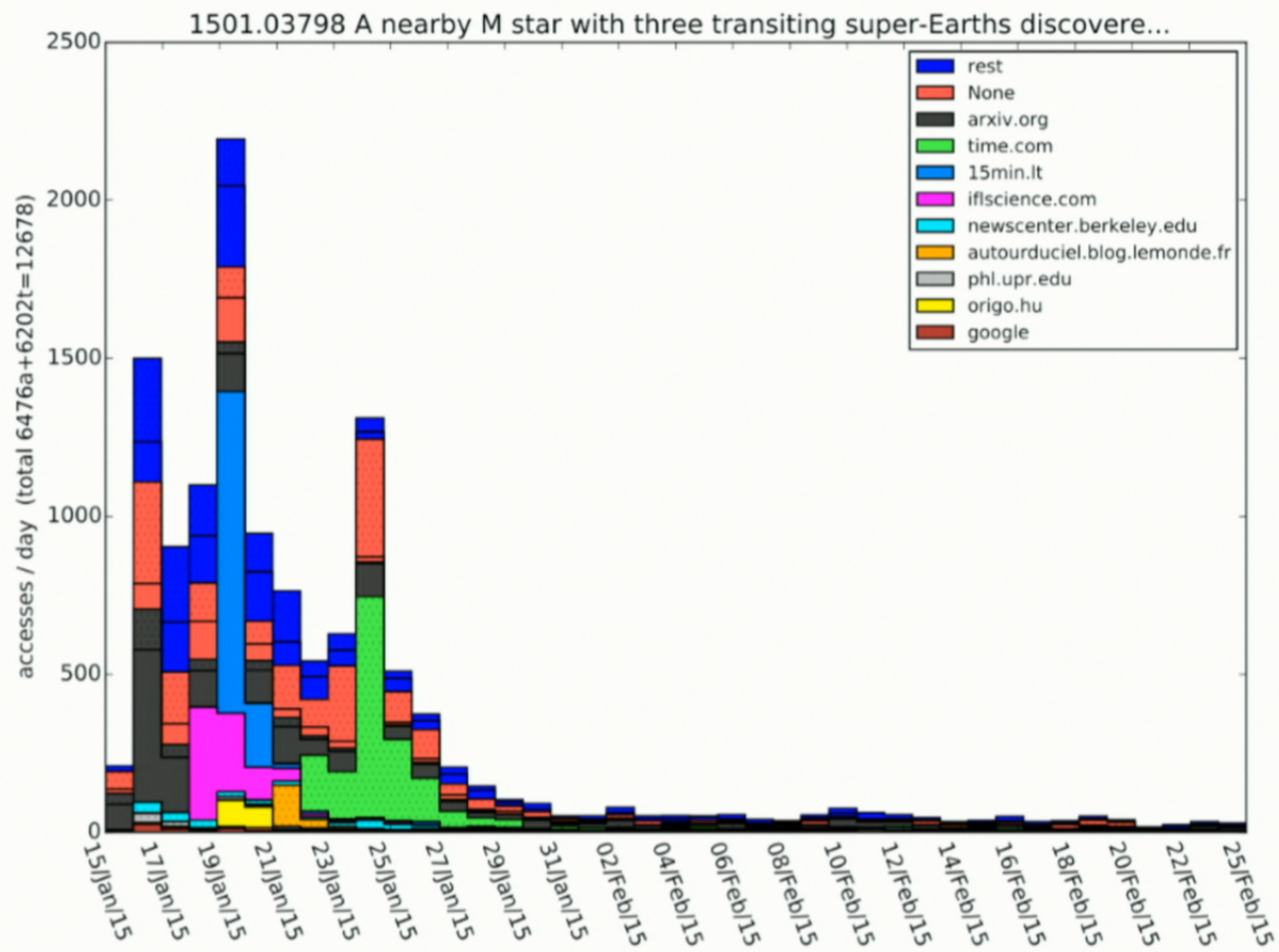
# A nearby M star with three transiting super–Earths discovered by K2

Ian J. M. Crossfield, Erik Petigura, Joshua Schlieder, Andrew W. Howard, B.J. Fulton, Kimberly M. Aller, David R. Ciardi, Sebastien Lepine, Thomas Barclay, Imke de Pater, Katherine de Kleer, Elisa V. Quintana, Jessie L. Christiansen, Eddie Schlafly, Lisa Kaltenegger, Justin R. Crepp, Thomas Henning, Christian Obermeier, Niall Deacon, Lauren M. Weiss, Howard T. Isaacson, Brad M. S. Hansen, Michael C. Liu, Tom Greene, Steve B. Howell, Travis Barman, Christoph Mordasini

Small, cool planets represent the typical end–products of planetary formation. Studying the archi– tectures of these systems, measuring planet masses and radii, and observing these planets' atmospheres during transit directly informs theories of planet assembly, migration, and evolution. Here we report the discovery of three small planets orbiting a bright (Ks = 8.6 mag) M0 dwarf using data collected as part of K2, the new transit survey using the re–purposed Kepler spacecraft. Stellar spectroscopy and K2 photometry indicate that the system hosts three transiting planets with radii 1.5–2.1 R_Earth, straddling the transition region between rocky and increasingly volatile–dominated compositions. With orbital periods of 10–45 days the planets receive just 1.5–10x the flux incident on Earth, making these some of the coolest small planets known orbiting a nearby star; planet d is located near the inner edge of the system's habitable zone. The bright, low–mass star makes this system an excellent laboratory to determine the planets' masses via Doppler spectroscopy and to constrain their atmospheric compositions via transit spectroscopy. This discovery

**Download:**
- PDF
- Other formats

Current browse context:
astro-ph.EP
< prev | next >
new | recent | 1501

Change to browse by:
astro-ph
    astro-ph.SR

References & Citations
- NASA ADS

Bookmark (what is this?)

1412.6213 Measurements on the reality of the wavefunction

Legend:
- rest
- arxiv.org
- nytimes.com
- None
- mobile.nytimes.com
- google
- theregister.co.uk
- news.ycombinator
- t.co
- facebook.com
- phys.org

Search or Article-id

(Help | Advanced search)

All papers    Go!
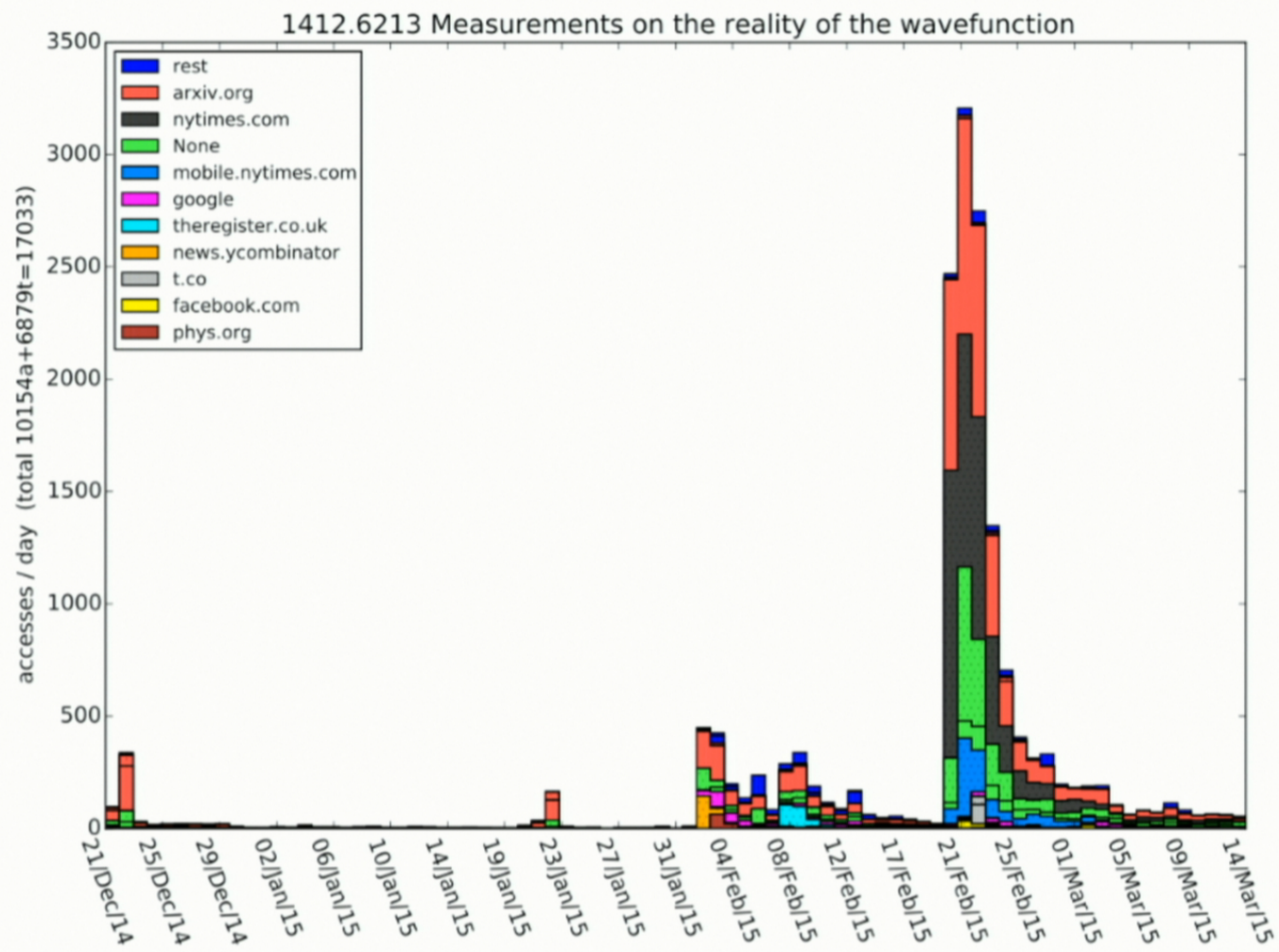
**Quantum Physics**

# Measurements on the reality of the wavefunction

Martin Ringbauer, Ben Duffus, Cyril Branciard, Eric G. Cavalcanti, Andrew G. White, Alessandro Fedrizzi

(Submitted on 19 Dec 2014 (v1), last revised 20 Jan 2015 (this version, v2))

Quantum mechanics is an outstandingly successful description of nature, underpinning fields from biology through chemistry to physics. At its heart is the quantum wavefunction, the central tool for describing quantum systems. Yet it is still unclear what the wavefunction actually is: does it merely represent our limited knowledge of a system, or is it an element of reality? Recent no-go theorems argued that if there was any underlying reality to start with, the wavefunction must be real. However, that conclusion relied on debatable assumptions, without which a partial knowledge interpretation can be maintained to some extent. A different approach is to impose bounds on the degree to which knowledge interpretations can explain quantum phenomena, such as why we cannot perfectly distinguish non-orthogonal quantum states. Here we experimentally test this approach with single photons. We find that no knowledge interpretation can fully explain the indistinguishability of non-orthogonal quantum states in three and four dimensions. Assuming that some underlying reality exists, our results strengthen the view that the entire wavefunction should be real. The only alternative is to adopt more unorthodox concepts such as backwards-in-time causation, or to completely abandon any notion of objective reality.

Comments: 7 pages, 4 figures
Subjects:     **Quantum Physics (quant-ph)**

**Download:**
- PDF
- Other formats

Current browse context:
quant-ph
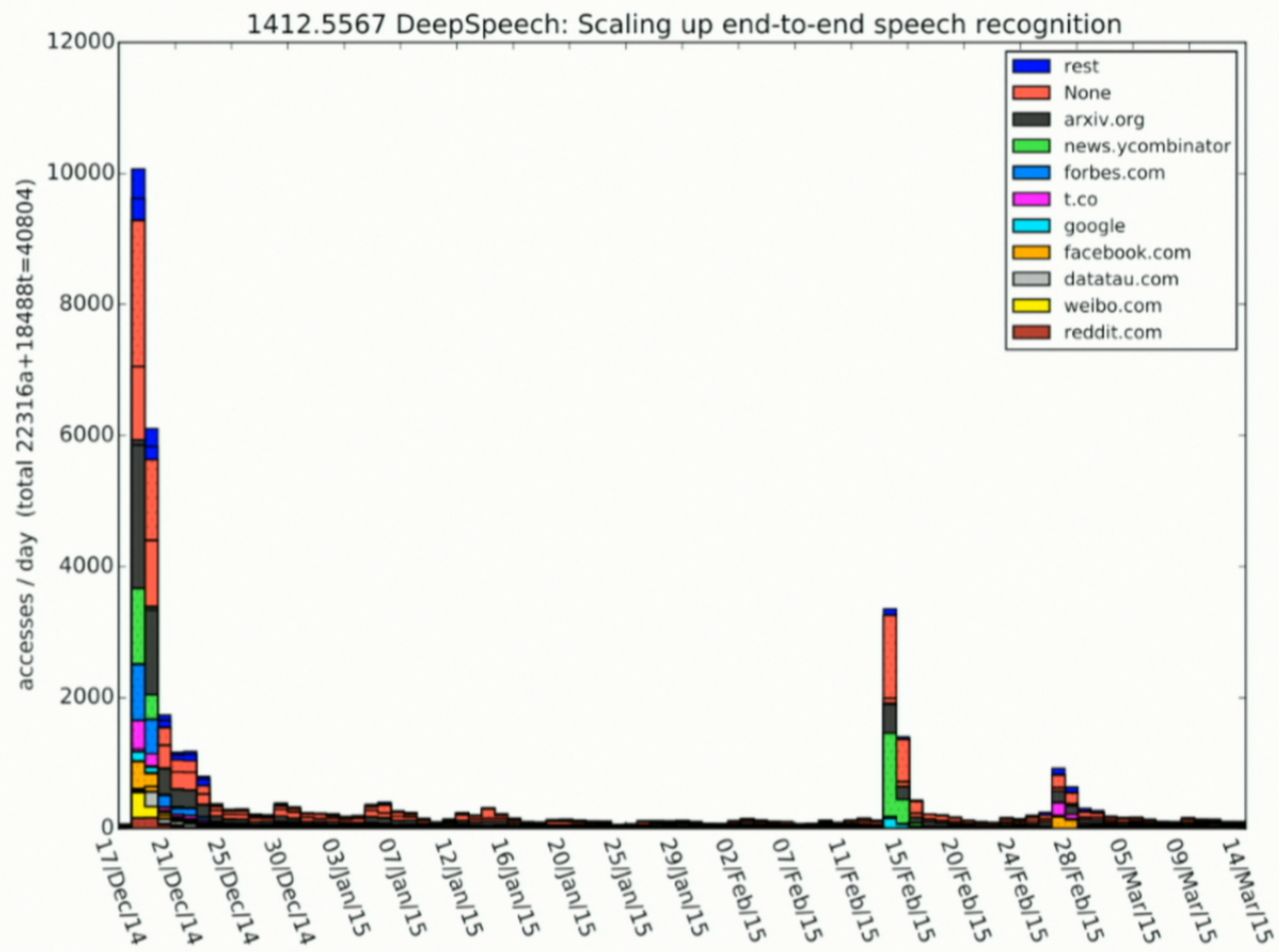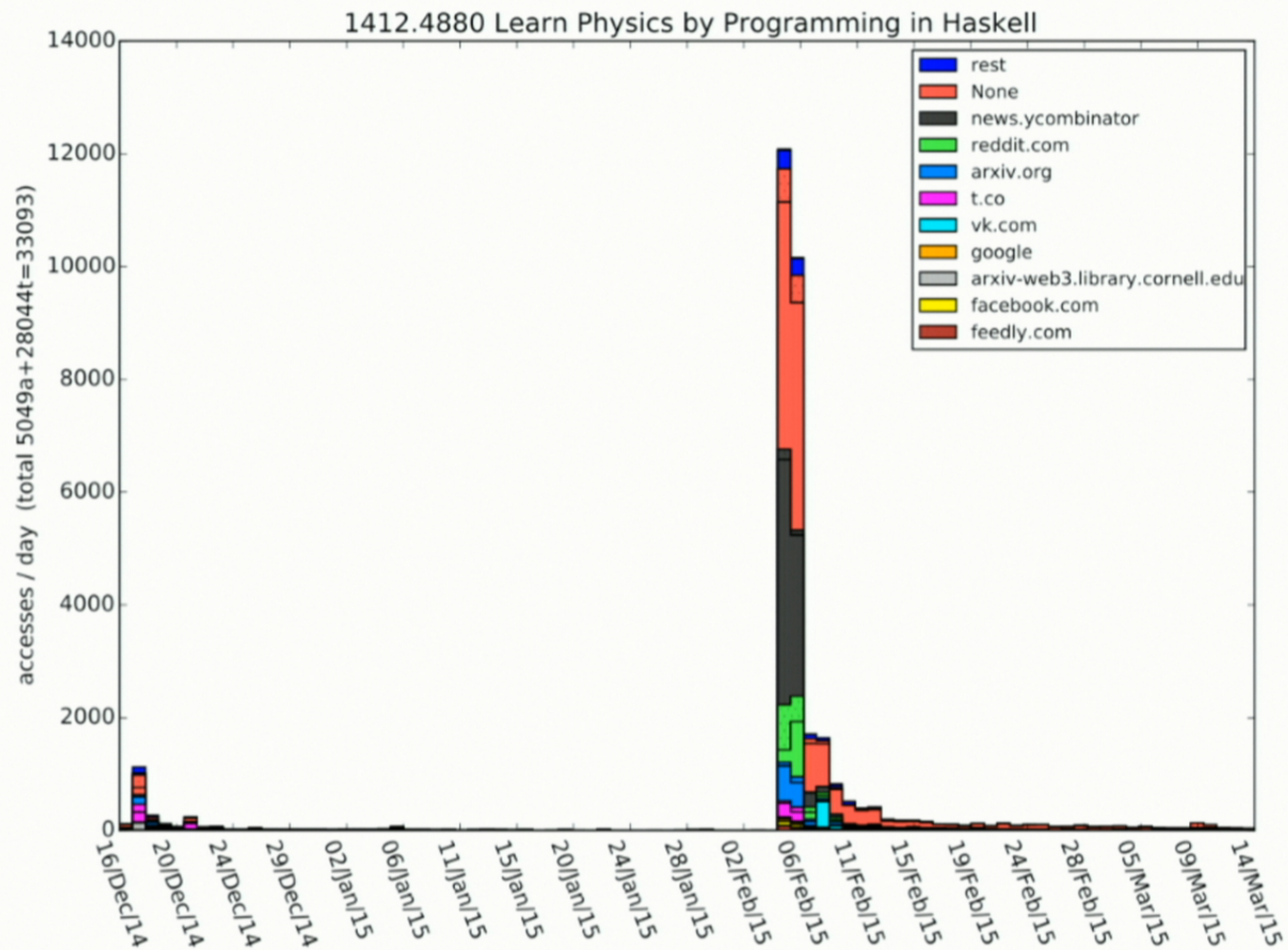< prev | next >
new | recent | 1412

References & Citations
- INSPIRE HEP
  (refers to | cited by )
- NASA ADS

Bookmark (what is this?)

1412.5567 DeepSpeech: Scaling up end-to-end speech recognition

1412.4880 Learn Physics by Programming in Haskell

arXiv.org > cs > arXiv:1412.4880

Search or Article-id       (Help | Advanced search)
[                    ]  All papers [ ]  Go!

**Computer Science > Computers and Society**

# Learn Physics by Programming in Haskell

Scott N. Walck (Lebanon Valley College, Annville, Pennsylvania, USA)

(Submitted on 16 Dec 2014)

We describe a method for deepening a student's understanding of basic physics by asking the student to express physical ideas in a functional programming language. The method is implemented in a second-year course in computational physics at Lebanon Valley College. We argue that the structure of Newtonian mechanics is clarified by its expression in a language (Haskell) that supports higher-order functions, types, and type classes. In electromagnetic theory, the type signatures of functions that calculate electric and magnetic fields clearly express the functional dependency on the charge and current distributions that produce the fields. Many of the ideas in basic physics are well-captured by a type or a function.

**Download:**

- PDF
- PostScript
- Other formats

Current browse context:
cs.CY
< prev | next >
new | recent | 1412

Change to browse by:
cs
  cs.PL
physics
  physics.ed-ph

References & Citations

- NASA ADS
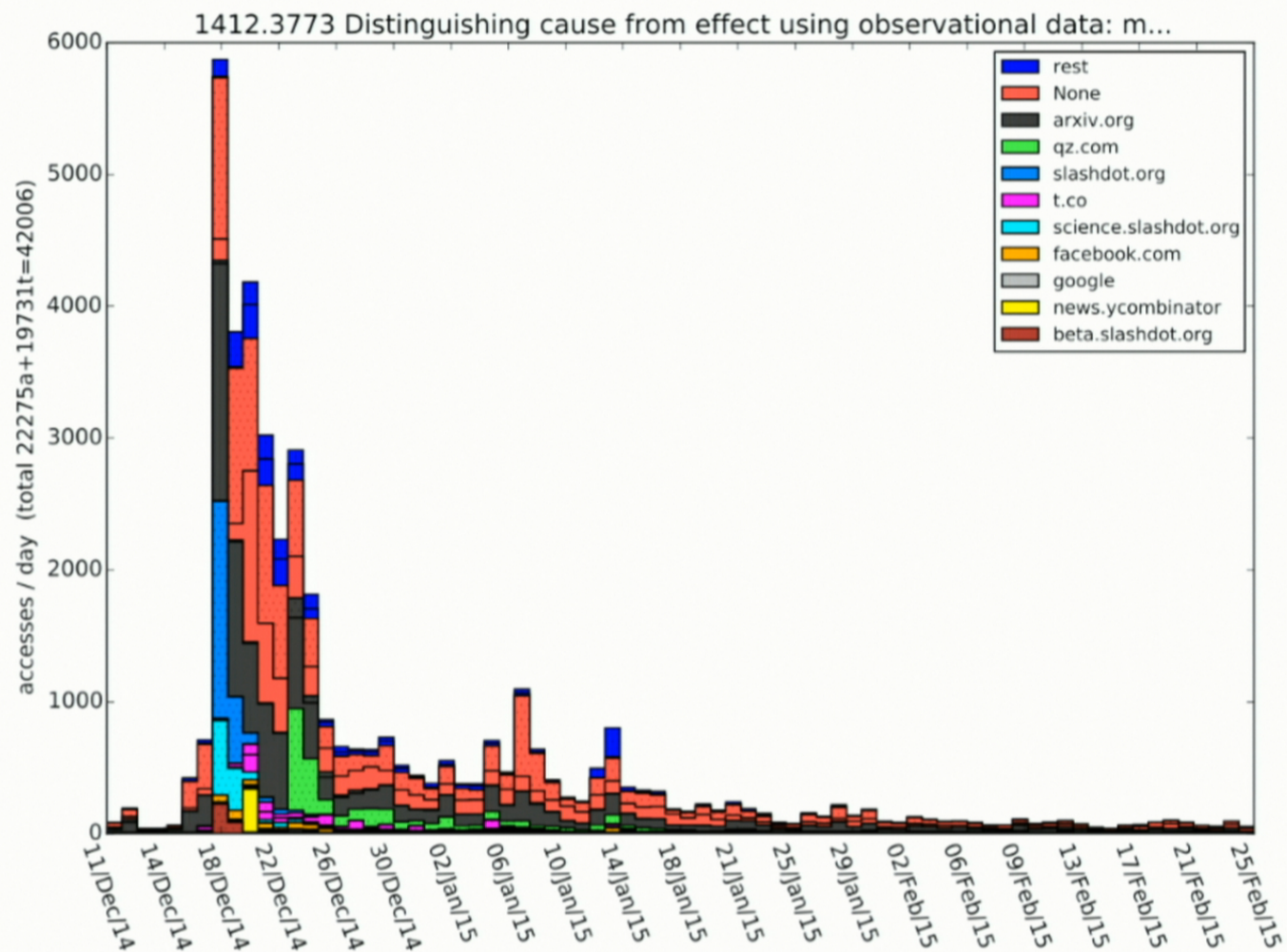
DBLP – CS Bibliography
  listing | bibtex
  Scott N. Walck

Bookmark (what is this?)

1412.3773 Distinguishing cause from effect using observational data: m...

arXiv.org > cs > arXiv:1412.3773

**Computer Science > Learning**

# Distinguishing cause from effect using observational data: methods and benchmarks

Joris M. Mooij, Jonas Peters, Dominik Janzing, Jakob Zscheischler, Bernhard Schölkopf

(Submitted on 11 Dec 2014)

The discovery of causal relationships from purely observational data is a fundamental problem in science. The most elementary form of such a causal discovery problem is to decide whether X causes Y or, alternatively, Y causes X, given joint observations of two variables X, Y . This was often considered to be impossible. Nevertheless, several approaches for addressing this bivariate causal discovery problem were proposed recently. In this paper, we present the benchmark data set CauseEffectPairs that consists of 88 different "cause-effect pairs" selected from 31 datasets from various domains. We evaluated the performance of several bivariate causal discovery methods on these real-world benchmark data and on artificially simulated data. Our empirical results provide evidence that additive-noise methods are indeed able to distinguish cause from effect using only purely observational data. In addition, we prove consistency of the additive-noise method proposed by Hoyer et al. (2009).

**Download:**

- PDF
- Other formats

Current browse context:
cs.LG
< prev | next >
new | recent | 1412

Change to browse by:
cs
    cs.AI
stat
    stat.ML
    stat.OT

**References & Citations**

- NASA ADS

2 blog links (what is this?)

**DBLP – CS Bibliography**
listing | bibtex
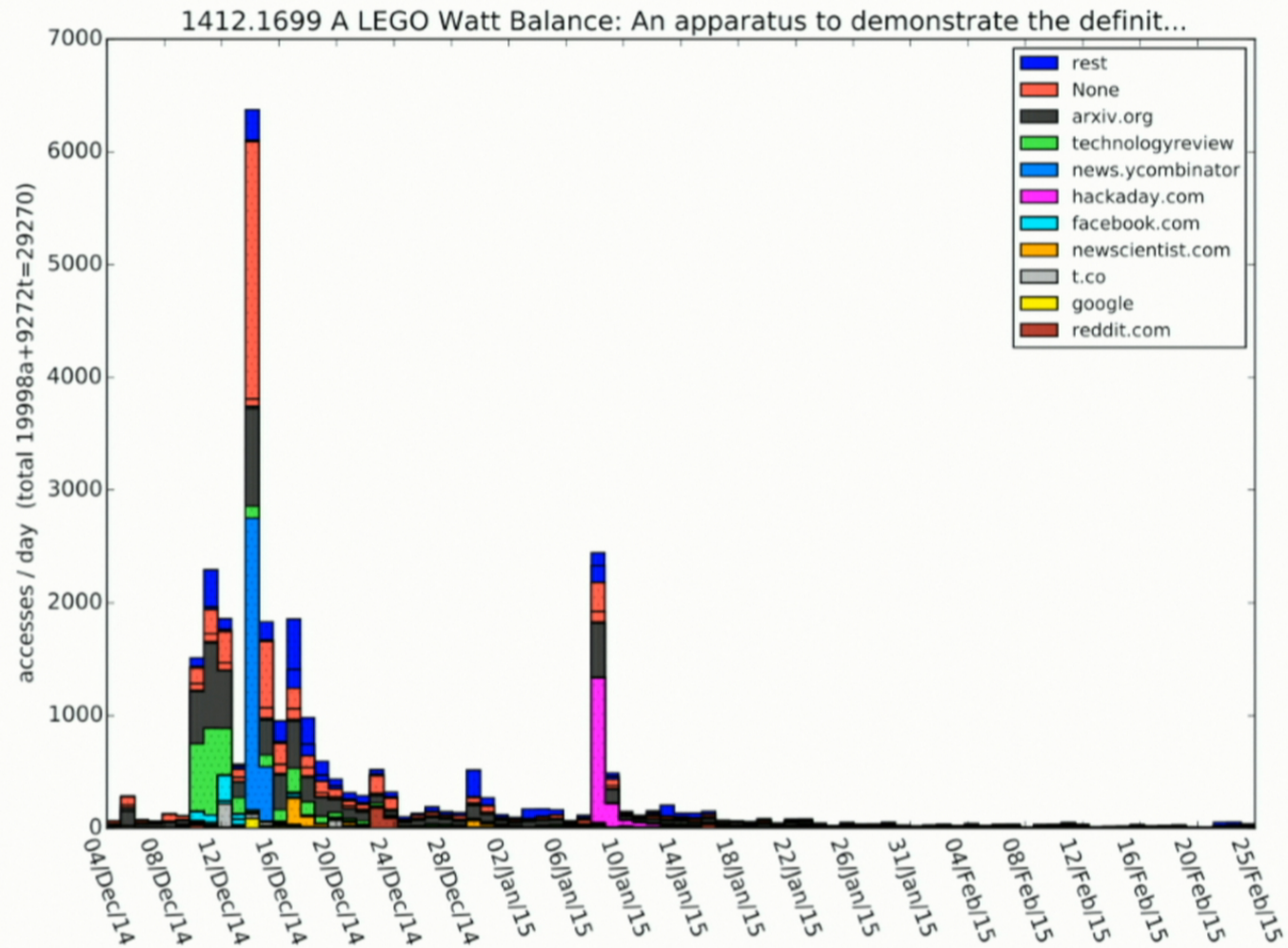
Joris M. Mooij
Jonas Peters
Dominik Janzing
Jakob Zscheischler
Bernhard Schölkopf

**Bookmark** (what is this?)

1412.1699 A LEGO Watt Balance: An apparatus to demonstrate the definit...

Legend:
- rest
- None
- arxiv.org
- technologyreview
- news.ycombinator
- hackaday.com
- facebook.com
- newscientist.com
- t.co
- google
- reddit.com

y-axis: accesses / day (total 19998a+9272t=29270)

Search or Article-id    (Help | Advanced search)

[            ]    All papers    Go!

**Physics > Instrumentation and Detectors**

# A LEGO Watt Balance: An apparatus to demonstrate the definition of mass based on the new SI

L.S. Chao, S. Schlamminger, D.B. Newell, J.R. Pratt, G. Sineriz, F. Seifert, A. Cao, D. Haddad, X. Zhang

(Submitted on 4 Dec 2014 (v1), last revised 12 Dec 2014 (this version, v2))

A global effort to redefine our International System of Units (SI) is underway and the change to the new system is expected to occur in 2018. Within the newly redefined SI, the present base units will still exist but be derived from fixed numerical values of seven reference constants. More specifically, the unit of mass, the kilogram, will be realized through a fixed value of the Planck constant $h$. For instance, a watt balance can be used to realize the kilogram unit of mass within a few parts in $10^8$. Such a balance has been designed and constructed at the National Institute of Standards and Technology. For educational outreach and to demonstrate the principle, we have constructed a LEGO tabletop watt balance capable of measuring a gram size mass to 1 % relative uncertainty. This article presents the design, construction, and performance of the LEGO watt balance and its ability to determine $h$

Comments:    11 pages, 6 figures. Also included is a partlist to build your own LEGO watt balance. The article has been submitted to American Journal of Physics. After it is published, it will be found at this http URL

Subjects:    **Instrumentation and Detectors (physics.ins-det)**; Classical Physics (physics.class-ph); Popular Physics (physics.pop-ph); Physics and Society (physics.soc-ph)

Cite as:    arXiv:1412.1699 [physics.ins-det]
         (or arXiv:1412.1699v2 [physics.ins-det] for this version)

**Download:**
- PDF
- Other formats

Current browse context:
**physics.ins-det**
< prev | next >
new | recent | 1412

Change to browse by:
physics
    physics.class-ph
    physics.pop-ph
    physics.soc-ph

**References & Citations**
- INSPIRE HEP
  (refers to | cited by )
- NASA ADS

1 blog link (what is this?)

Bookmark (what is this?)

Search or Article-id    (Help | Advanced search)
All papers    Go!

Astrophysics > Earth and Planetary Astrophysics

# On the abundance of extraterrestrial life after the Kepler mission

Amri Wandel

(Submitted on 3 Dec 2014)

The data recently accumulated by the Kepler mission have demonstrated that small planets are quite common and that a significant fraction of all stars may have an Earth-like planet within their Habitable Zone. These results are combined with a Drake-equation formalism to derive the space density of biotic planets as a function of the relatively modest uncertainty in the astronomical data and of the (yet unknown) probability for the evolution of biotic life, Fb. I suggest that Fb may be estimated by future spectral observations of exoplanet biomarkers. If Fb is in the range 0.001 -- 1 then a biotic planet may be expected within 10 -- 100 light years from Earth. Extending the biotic results to advanced life I derive expressions for the distance to putative civilizations in terms of two additional Drake parameters – the probability for evolution of a civilization, Fc, and its average longevity. For instance, assuming optimistic probability values (Fb Fc 1) and a broadcasting longevity of a few thousand years, the likely distance to the nearest civilizations detectable by SETI is of the order of a few thousand light years. The probability of detecting intelligent signals with present and future radio telescopes is calculated as a function of the Drake parameters. Finally, I describe how the detection of intelligent signals would constrain the Drake parameters.

**Download:**
- PDF only

Current browse context:
astro-ph.EP
< prev | next >
new | recent | 1412

Change to browse by:
astro-ph

References & Citations
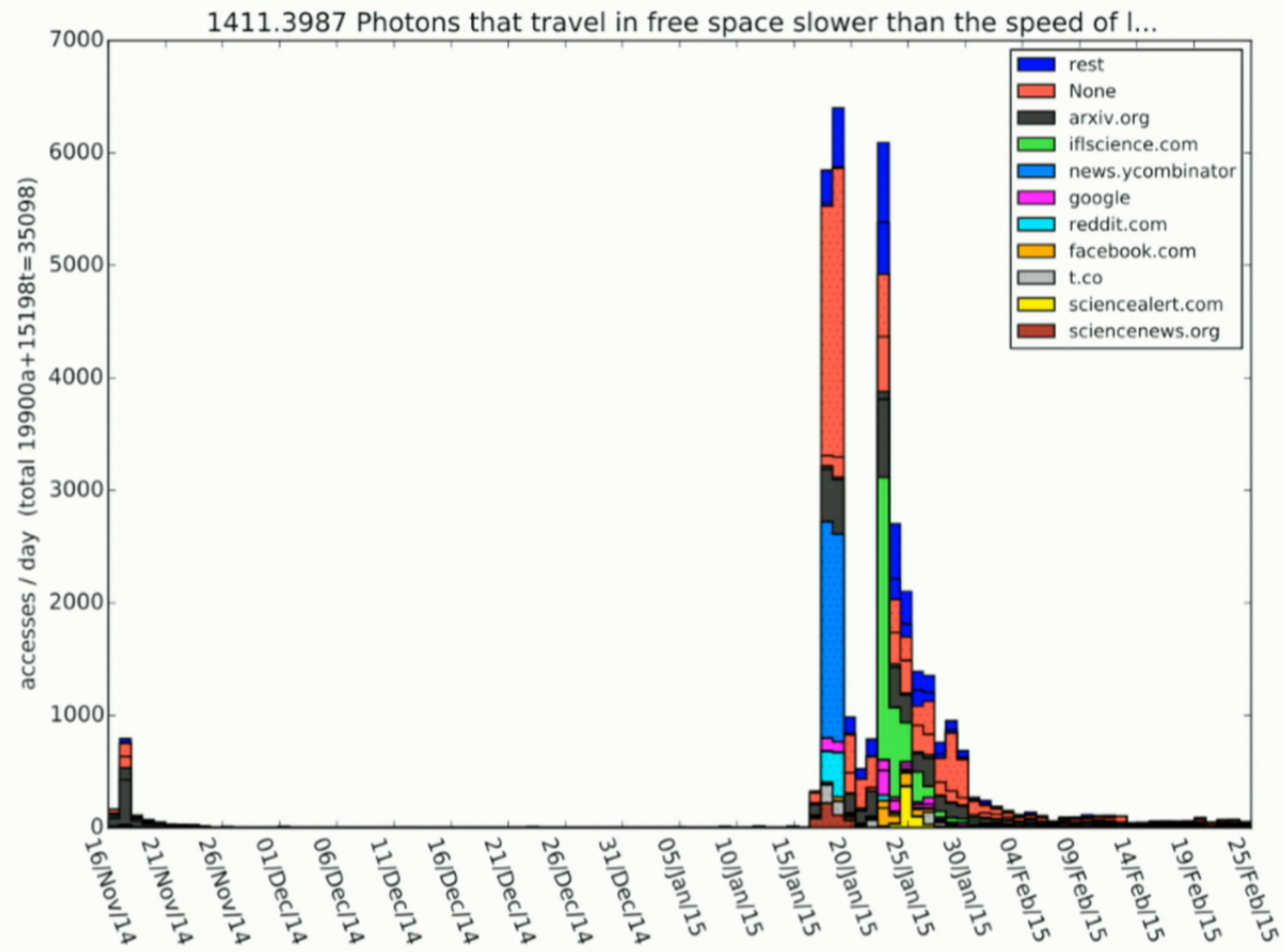- NASA ADS

1 blog link (what is this?)

Bookmark (what is this?)

1411.3987 Photons that travel in free space slower than the speed of l...

Physics > Optics

# Photons that travel in free space slower than the speed of light

Daniel Giovannini, Jacquiline Romero, Vaclav Potocek, Gergely Ferenczi, Fiona Speirits, Stephen M. Barnett, Daniele Faccio, Miles J. Padgett

(Submitted on 14 Nov 2014)

That the speed of light in free space is constant is a cornerstone of modern physics. However, light beams have finite transverse size, which leads to a modification of their wavevectors resulting in a change to their phase and group velocities. We study the group velocity of single photons by measuring a change in their arrival time that results from changing the beam's transverse spatial structure. Using time-correlated photon pairs we show a reduction of the group velocity of photons in both a Bessel beam and photons in a focused Gaussian beam. In both cases, the delay is several microns over a propagation distance of the order of 1 m. Our work highlights that, even in free space, the invariance of the speed of light only applies to plane waves. Introducing spatial structure to an optical beam, even for a single photon, reduces the group velocity of the light by a readily measurable amount.

Submission history

**Download:**

- PDF only

Current browse context:
physics.optics
< prev | next >
new | recent | 1411

Change to browse by:
physics
quant-ph

References & Citations

- NASA ADS

Bookmark (what is this?)

Search or Article-id    (Help | Advanced search)
All papers    Go!

**Astrophysics > Cosmology and Nongalactic Astrophysics**

# Multiple Images of a Highly Magnified Supernova Formed by an Early-Type Cluster Galaxy Lens

Patrick L. Kelly (UCB), Steven A. Rodney (JHU), Tommaso Treu (UCLA), Ryan J. Foley (Illinois), Gabriel Brammer (STScI), Kasper B. Schmidt (UCSB), Adi Zitrin (Caltech), Alessandro Sonnenfeld (UCLA), Louis-Gregory Strolger (STScI), Or Graur (NYU/AMNH), Alexei V. Filippenko (UCB), Saurabh W. Jha (Rutgers), Adam G. Riess (JHU/STScI), Marusa Bradac (UCD), Benjamin J. Weiner (Arizona), Daniel Scolnic (Chicago), Matthew A. Malkan (UCLA), Anja von der Linden (DARK/Stanford), Michele Trenti (Melbourne), Jens Hjorth (DARK), Raphael Gavazzi (IAP), Adriano Fontana (INAF-OAR), Julian Merten (Caltech), Curtis McCully (LCOGT/UC Santa Barbara), Tucker Jones (UCLA), Marc Postman (STScI), Alan Dressler (Carnegie Obs.), Brandon Patel (Rutgers), S. Bradley Cenko (GSFC/UMD), Melissa L. Graham (UCB), Bradley E. Tucker (UCB/ANU)

(Submitted on 21 Nov 2014 (v1), last revised 5 Mar 2015 (this version, v3))

In 1964, Refsdal hypothesized that a supernova whose light traversed multiple paths around a strong gravitational lens could be used to measure the rate of cosmic expansion. We report the discovery of such a system. In Hubble Space Telescope imaging, we have found four images of a single supernova forming an Einstein cross configuration around a redshift $z=0.54$ elliptical galaxy in the MACS J1149.6+2223 cluster. The cluster's gravitational potential also creates multiple images of the $z=1.49$ spiral supernova host galaxy, and a future appearance of the supernova elsewhere in the cluster field is expected. The magnifications and staggered arrivals of the supernova images probe the cosmic expansion rate, as well as the distribution of matter in the galaxy and cluster lenses.

**Download:**
- PDF
- Other formats

Current browse context:
astro-ph.CO
< prev | next >
new | recent | 1411

Change to browse by:
astro-ph
    astro-ph.GA

References & Citations
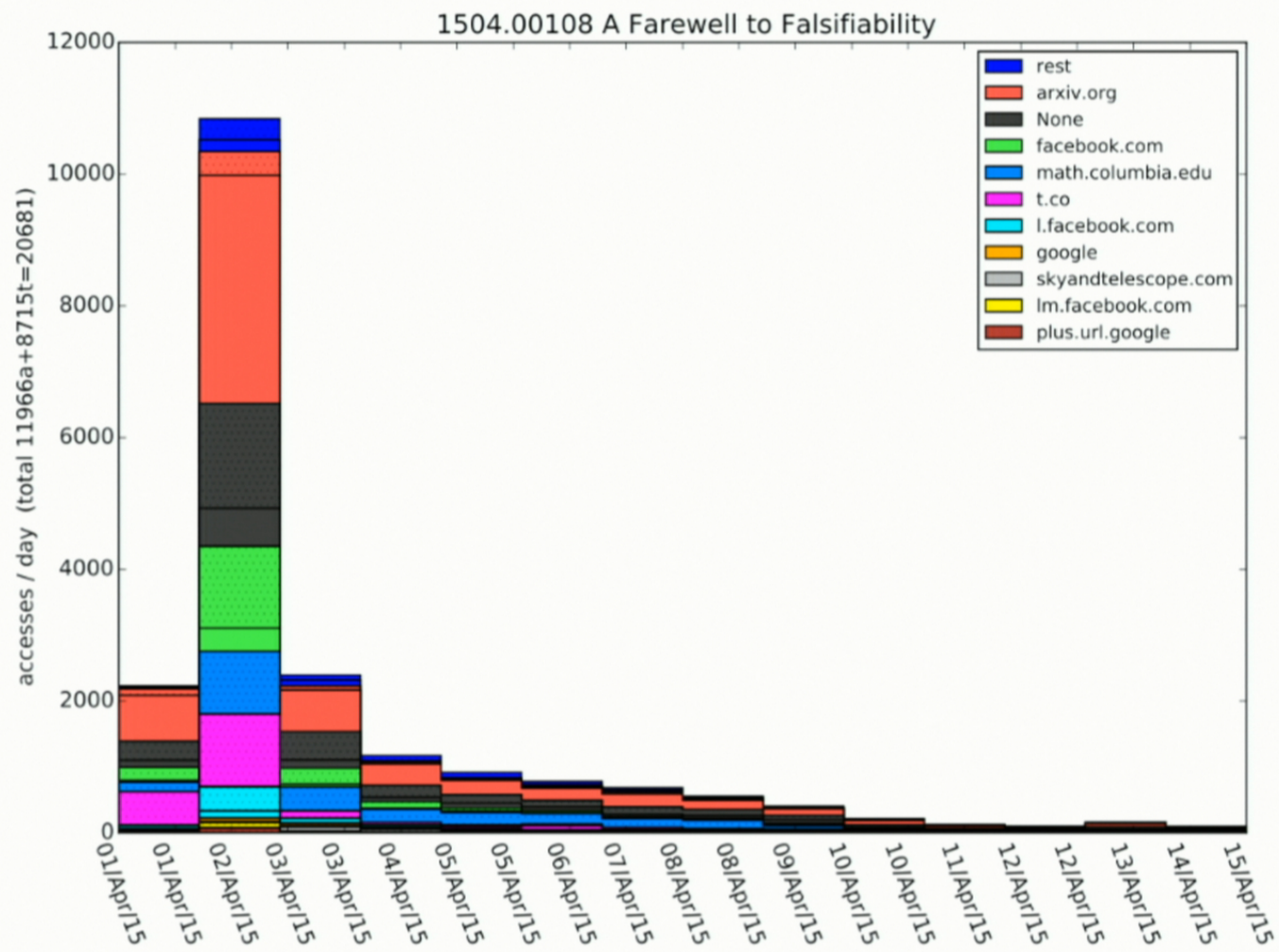- INSPIRE HEP
    (refers to | cited by )
- NASA ADS

2 blog links (what is this?)

Bookmark (what is this?)

1504.00108 A Farewell to Falsifiability

arXiv.org > astro-ph > arXiv:1504.00108

Search or Ar

Astrophysics > Cosmology and Nongalactic Astrophysics

# A Farewell to Falsifiability

Douglas Scott, Ali Frolop, Ali Narimani, Andrei Frolov

(Submitted on 1 Apr 2015)

Some of the most obviously correct physical theories – namely string theory and the multiverse – make no testable predictions, leading many to question whether we should accept something as scientific even if it makes no testable predictions and hence is not refutable. However, some far-thinking physicists have proposed instead that we should give up on the notion of Falsifiability itself. We endorse this suggestion but think it does not go nearly far enough. We believe that we should also dispense with other outdated ideas, such as Fidelity, Frugality, Factuality and other "F" words. And we quote a lot of famous people to support this view.

**Submission history**

-------------------------------------------------------------------------------
\\
Paper: 9203227
From: harvey@witten.uchicago.edu (J. B. Harvey)
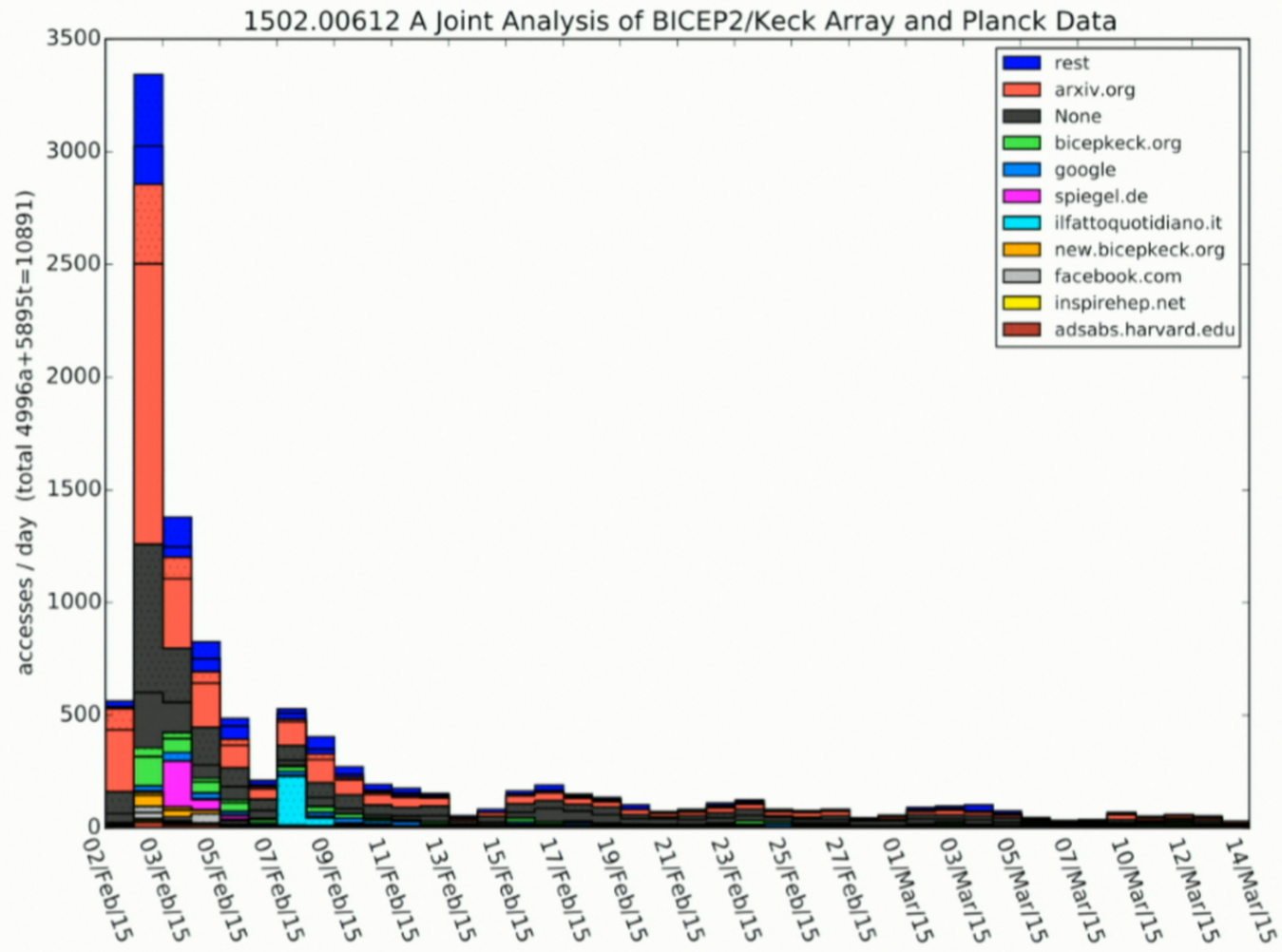Date: Wed 1 Apr 1992 00:25 CST 1992

A solvable string theory in four dimensions,
by J. Harvey, G. Moore, N. Seiberg, and A. Strominger, 30 pp
\\

We construct a new class of exactly solvable string theories by generalizing
the heterotic construction to connect a left-moving non-compact Lorentzian
coset algebra with a right-moving supersymmetric Euclidean coset algebra. These
theories have no spacetime supersymmetry, and a generalized set of anomaly
constraints allows only a model with four spacetime dimensions, low energy
gauge groups SU(3) and spontaneously broken SU(2)xU(1), and three families
of quarks and leptons. The model has a complex dilaton whose radial mode
is automatically eaten in a Higgs-like solution to the cosmological
constant problem, while its angular mode survives to solve the strong CP
problem at low energy. By adroit use of the theory of parabolic cylinder
functions, we calculate the mass spectrum of this model to all orders in
the string loop expansion. The results are within 5% of measured values,
with the discrepancy attributable to experimental error. We predict a top
quark mass of $175 \pm 2$ GeV, and no physical Higgs particle in the spectrum.
\\

1502.00612 A Joint Analysis of BICEP2/Keck Array and Planck Data

Legend:
- rest
- arxiv.org
- None
- bicepkeck.org
- google
- spiegel.de
- ilfattoquotidiano.it
- new.bicepkeck.org
- facebook.com
- inspirehep.net
- adsabs.harvard.edu

**Computer Science > Databases**

# Knowledge–Based Trust: Estimating the Trustworthiness of Web Sources

Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shaohua Sun, Wei Zhang

(Submitted on 12 Feb 2015)

The quality of web sources has been traditionally evaluated using exogenous signals such as the hyperlink structure of the graph. We propose a new approach that relies on endogenous signals, namely, the correctness of factual information provided by the source. A source that has few false facts is considered to be trustworthy. The facts are automatically extracted from each source by information extraction methods commonly used to construct knowledge bases. We propose a way to distinguish errors made in the extraction process from factual errors in the web source per se, by using joint inference in a novel multi-layer probabilistic model. We call the trustworthiness score we computed Knowledge-Based Trust (KBT). On synthetic data, we show that our method can reliably compute the true trustworthiness levels of the sources. We then apply it to a database of 2.8B facts extracted from the web, and thereby estimate the trustworthiness of 119M webpages. Manual evaluation of a subset of the results confirms the effectiveness of the method.

Subjects:   **Databases (cs.DB)**; Information Retrieval (cs.IR)
Cite as:    **arXiv:1502.03519 [cs.DB]**
            (or **arXiv:1502.03519v1 [cs.DB]** for this version)

**Submission history**

Download:
- PDF
- Other formats

Current browse context:
cs.DB
< prev | next >
new | recent | 1502

Change to browse by:
cs
cs.IR

References & Citations
- NASA ADS

Bookmark (what is this?)