At a Physics/InfoSci Intersection

Paul Ginsparg

Physics and InfoSci, Cornell University

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.

Colloquium Paco Ynduráin, Universidad Autónoma de Madrid - 25 Mar 2015

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

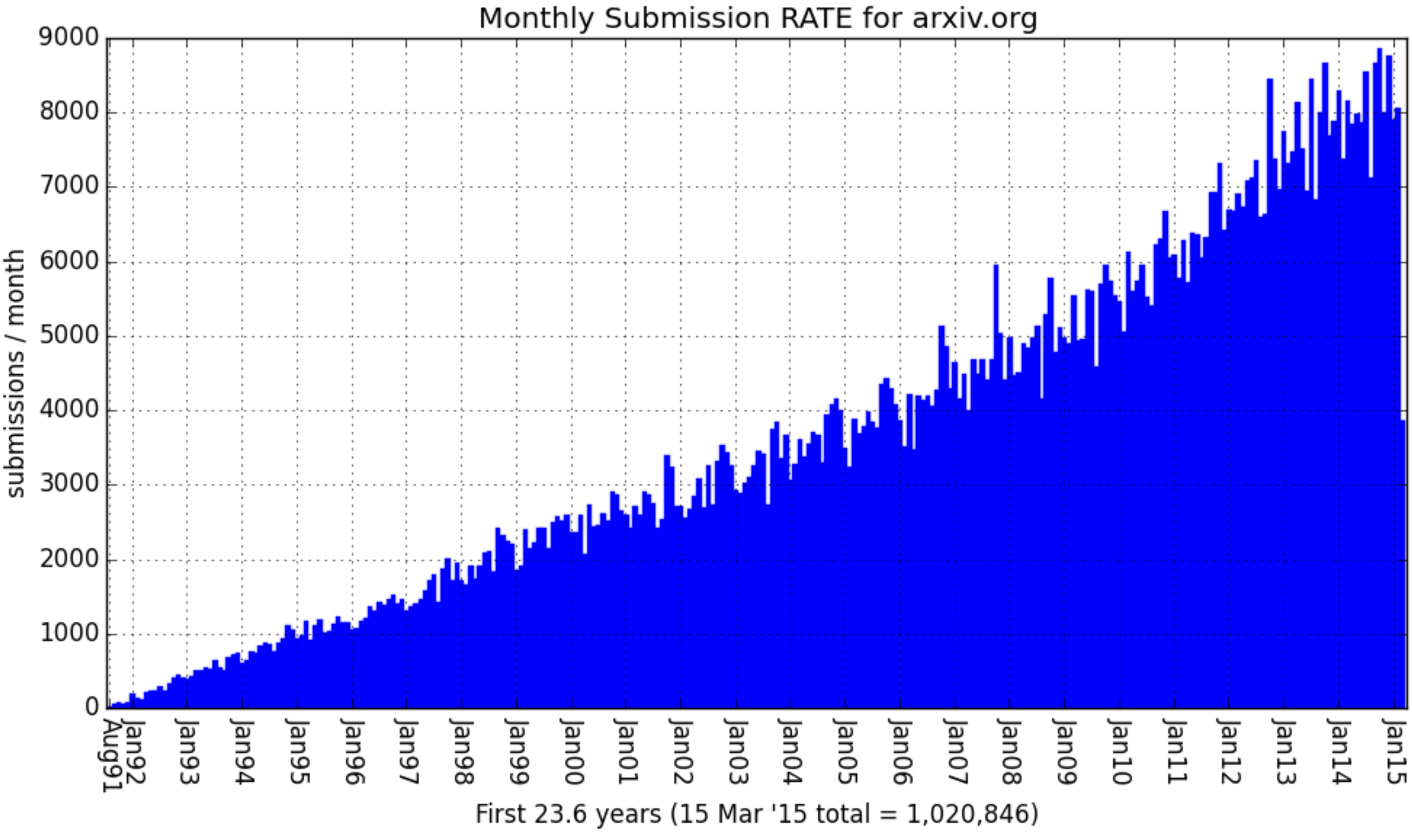
Cornell University Library

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 - ’15

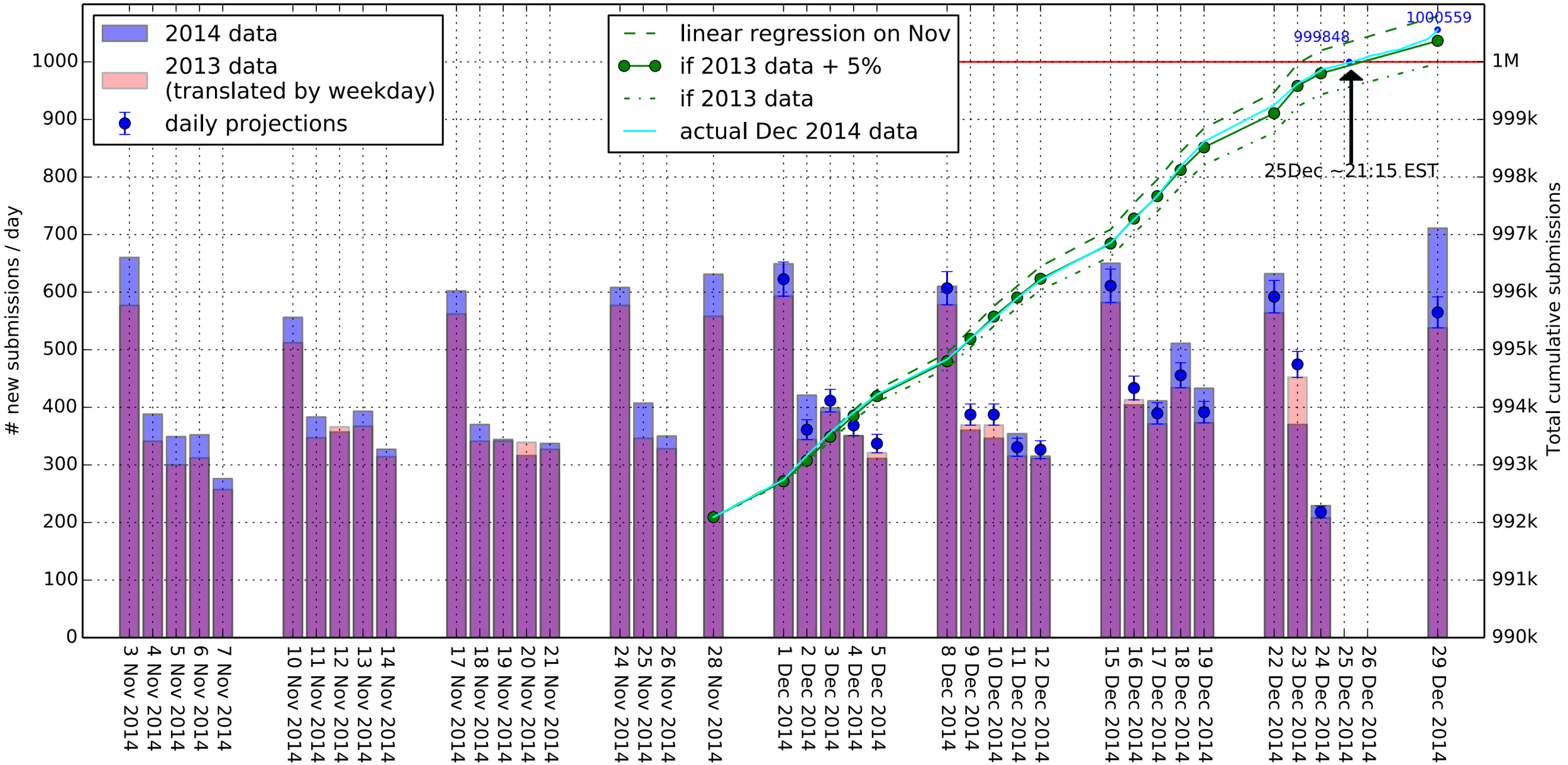


arXiv.org

- **e-mail interface started August 1991**
  - **download data available from start**
  - **WWW usage logs starting from 1993**
- **1,020,000 full text documents (with full graphics), 15 Mar 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)



nature
International weekly journal of science

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.

Miljoenste artikel toegevoegd aan arXiv

31 DECEMBER 2014 DOOR ARIE NOUWEN ▪ REAGEER

Op de populaire preprint server arXiv.org, waar wetenschappers hu
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 volg

30 DICEMBRE 2014

I milioni di arXiv

*Ci sono voluti 17 anni per accumulare mezzo milione di manc
'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 dell
californiani del fondatore Paul Ginsparg

TheScientist
EXPLORING LIFE, INSPIRING INNOVATION

Q&A: One Million Preprints and Counting

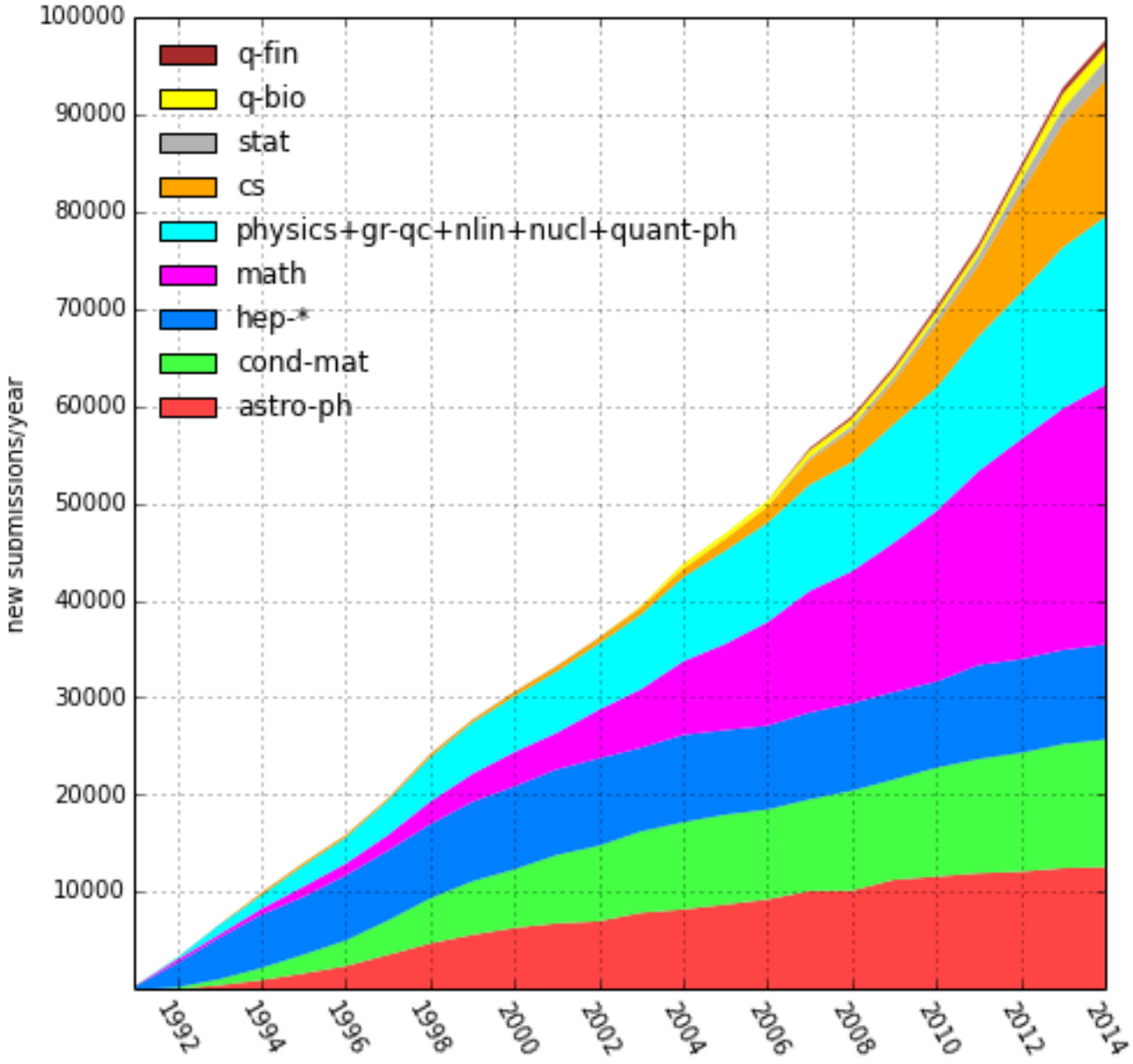
A conversation with ArXiv founder Paul Ginsparg

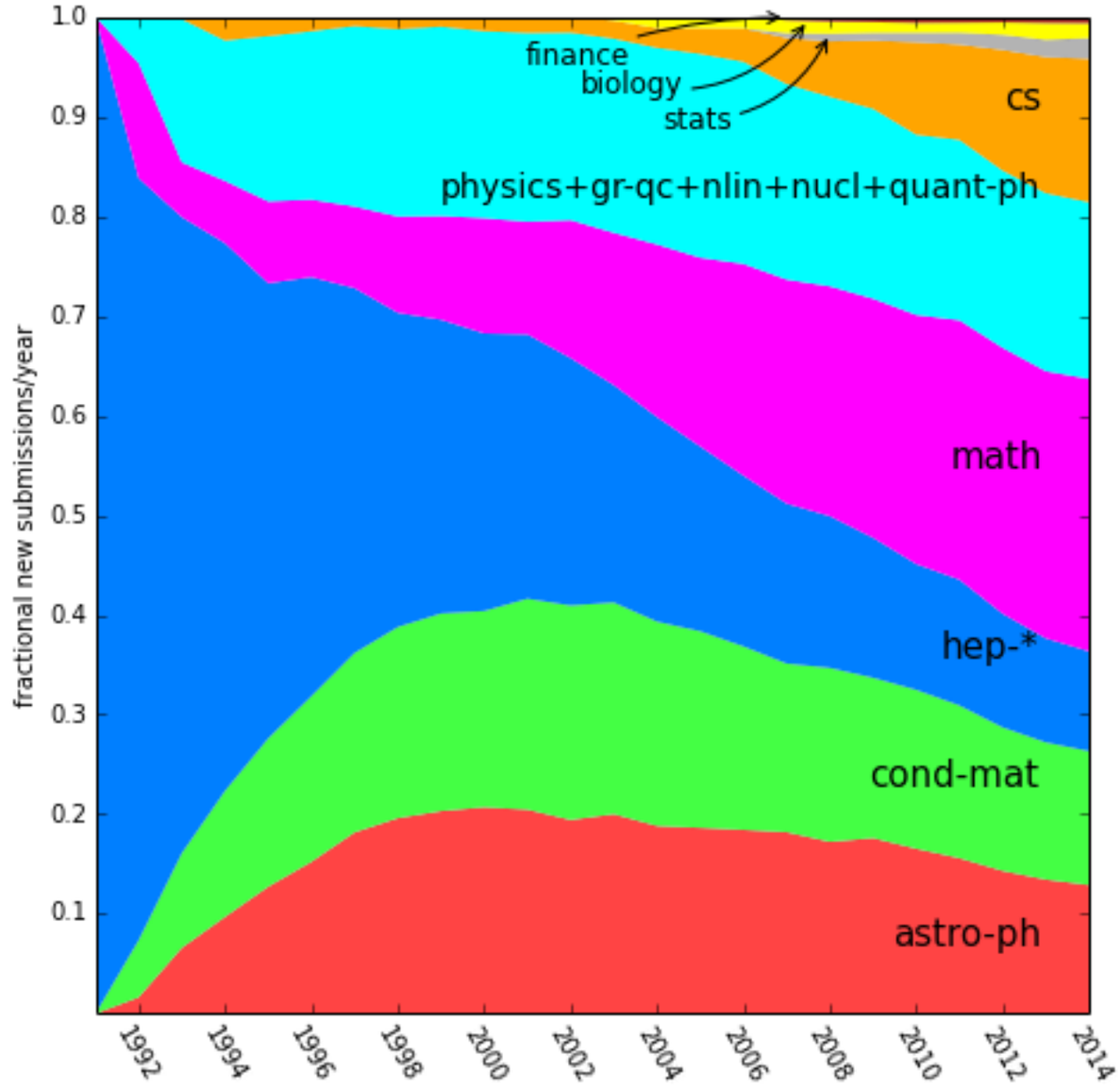
By Tracy Vence | December 29, 2014

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





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.

Full Text Databases

- **Text as computable object: literature-based discovery via centralized web-based platform, open repository with pre-parsed ontological properties and statistically based relationships, available for analysis by user-contributed algorithms.**
- **More powerful when centralized and critical mass user base**
- **Goal: semi-supervised, self-incentivized, self-maintaining knowledge structure, navigated via synthesized concepts, w/o redundancy/ambiguity, sourced, authenticated, highlighted for novelty**
- **Neo-Minsky: "Can you imagine they used to have an internet in which authors, databases, articles, and readers didn't talk to each other?"**
- **arXiv.org: has already dedicated user community, we've done a variety of text datamining and usage log experiments, but just skimming the surface, open to a broader community (modulo privacy concerns)**

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

astro-ph.∗

cond-mat.∗

cs.∗

gr-qc
hep-(ex/lat/th/ph)
math-ph

math.∗

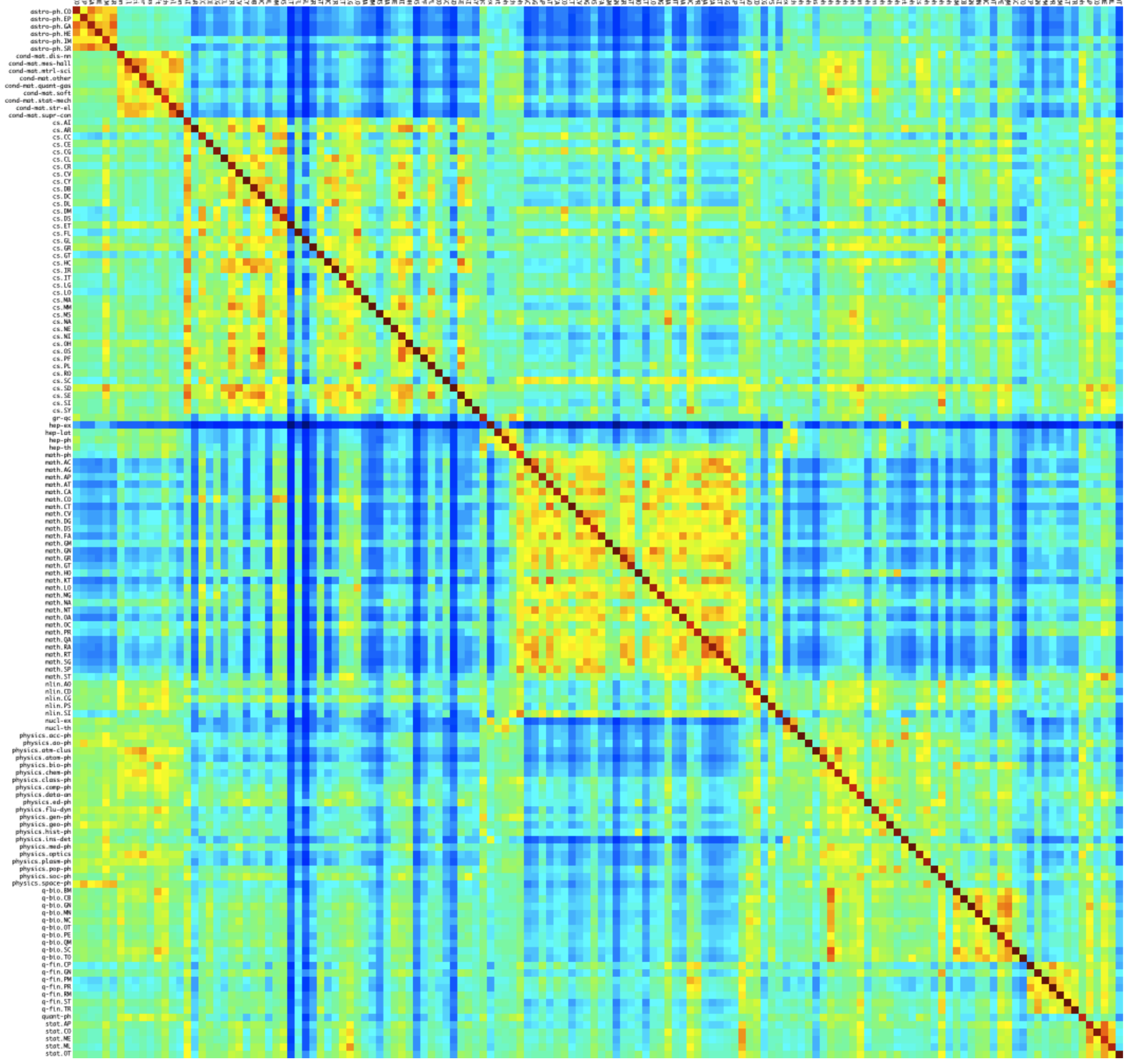
nlin.∗
nucl-∗

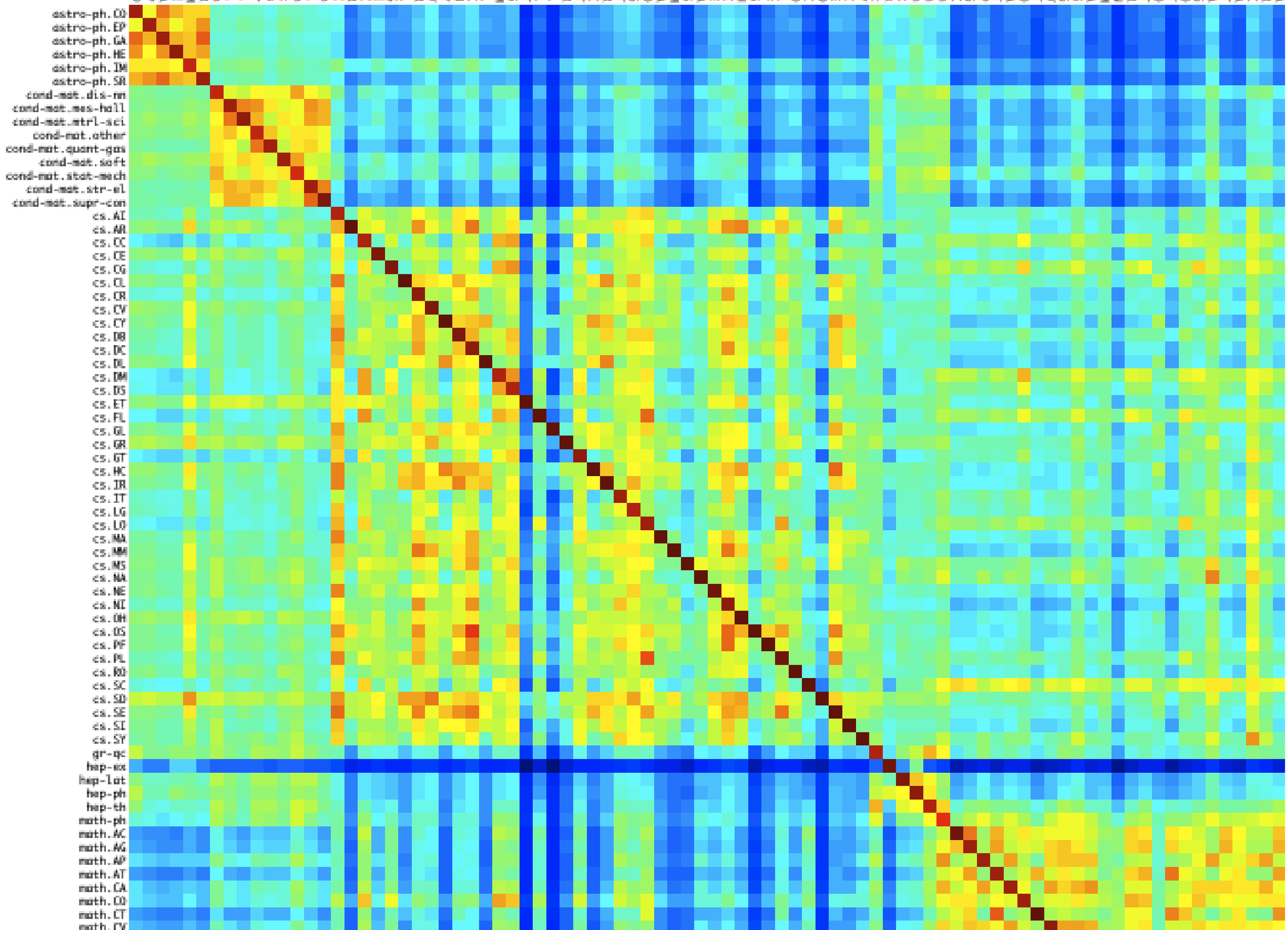
physics.∗

q-bio.∗

q-fin.∗

quant-ph
stat.∗




astro-ph.CO
astro-ph.EP
astro-ph.GA
astro-ph.HE
astro-ph.IM
astro-ph.SR
cond-mat.dis-nn
cond-mat.mes-hall
cond-mat.mtrl-sci
cond-mat.other
cond-mat.quant-gas
cond-mat.soft
cond-mat.stat-mech
cond-mat.str-el
cond-mat.supr-con
cs.AI
cs.AR
cs.CC
cs.CE
cs.CG
cs.CL
cs.CR
cs.CV
cs.CY
cs.DB
cs.DC
cs.DL
cs.DM
cs.DS
cs.ET
cs.FL
cs.GL
cs.GR
cs.GT
cs.HC
cs.IR
cs.IT
cs.LG
cs.LO
cs.MA
cs.MM
cs.MS
cs.NA
cs.NE
cs.NI
cs.OH
cs.OS
cs.PF
cs.PL
cs.RO
cs.SC
cs.SD
cs.SE
cs.SI
cs.SY
gr-qc
hep-ex
hep-lat
hep-ph
hep-th
math-ph
math.AC
math.AG
math.AP
math.AT
math.CA
math.CO
math.CT
math.CV
math.DG
math.DS
math.FA
math.GM
math.GN
math.GR
math.GT
math.HO
math.IT
math.LO
math.MG
math.MP
math.NA
math.NT
math.OA
math.OC
math.PR
math.QA

nature International weekly journal of science

NATURE | NEWS

Impact of UK research revealed in 7,000 case studies

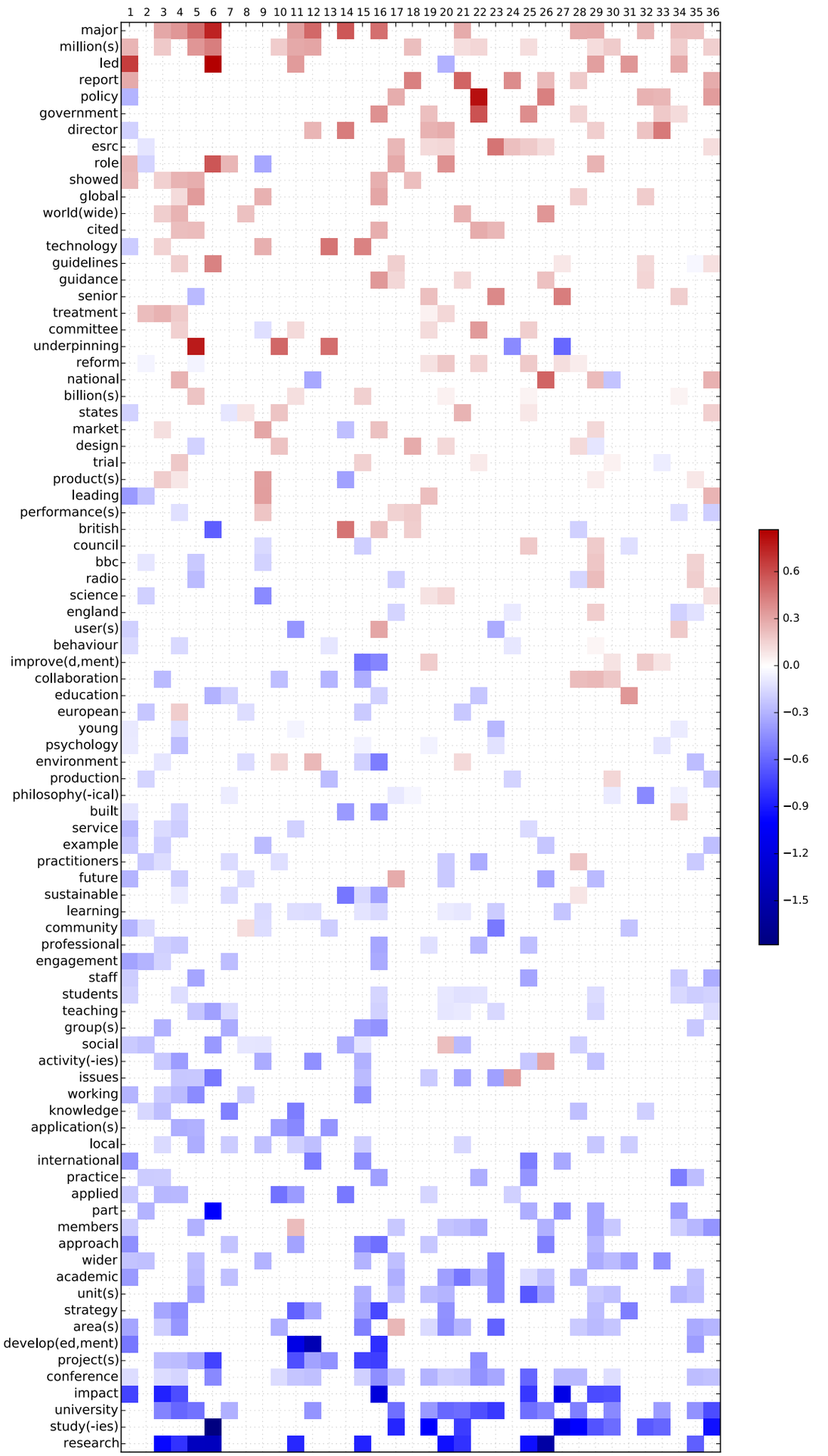
Language analysis reflects how projects succeeded in unique assessment.

Richard Van Noorden

11 February 2015

Science benefits society in myriad ways — but how to identify and encourage work with high impact is an obsession of funding agencies the world over. Last month, the United Kingdom brought new data to bear on the problem: almost 7,000 case studies chronicling the economic, cultural and social benefits of the nation's scholarship, which were solicited as part of a unique assessment exercise. As policy-makers pore over the documents, *Nature* has commissioned its own analysis, revealing how researchers described the worth of their work to their paymasters, and hinting at buzzwords, including 'million' and 'market', that garnered high marks.

Many funding bodies ask academics to plan for the broader impacts of their work when they apply for grants. But the United Kingdom wanted to reward impact that had already been achieved, says Steven Hill, head of research policy at the Higher Education Funding Council for England (HEFCE). The country already has an audit culture: it grades the quality of university research every few years, and hands out £2 billion (US$3 billion) annually on the basis of that assessment. For the 2014 audit, known as the Research Excellence Framework, or REF, HEFCE tweaked the rules. It added a requirement that universities send in case studies detailing their work's wider impact during 2008–13, and announced that 20% of an institution's final grade would be based on these contributions (see *Nature* http://doi.org/zx8; 2014).



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

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: M. B. Gavela
Author 2: Luis E. Ibáñez
Author 3: Paco Ynduráin
Author 4:
Author 5:

Generate Reset

Arnicin: Visualization of Vacuum Tubes

M. B. Gavela, Luis E. Ibáñez, Paco Ynduráin

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 important when 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-

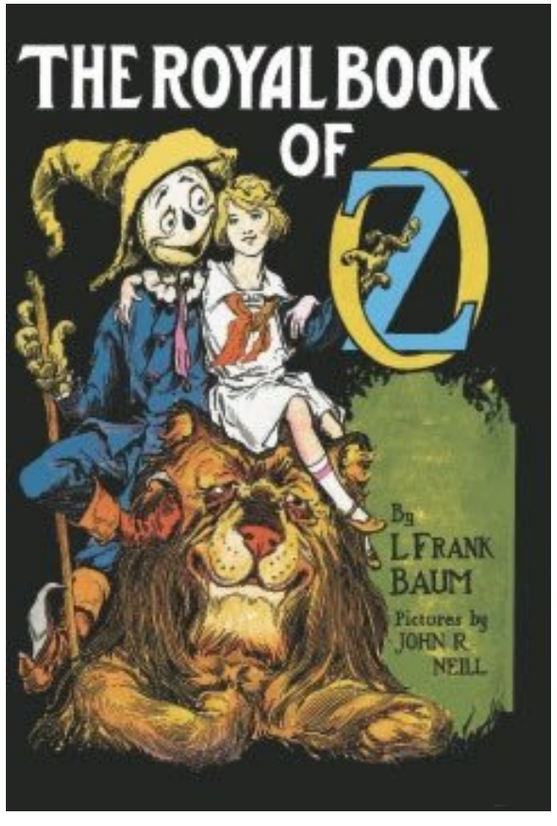
"Ike Antkare, One of the Great Stars in the Scientific Firmament"

(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:

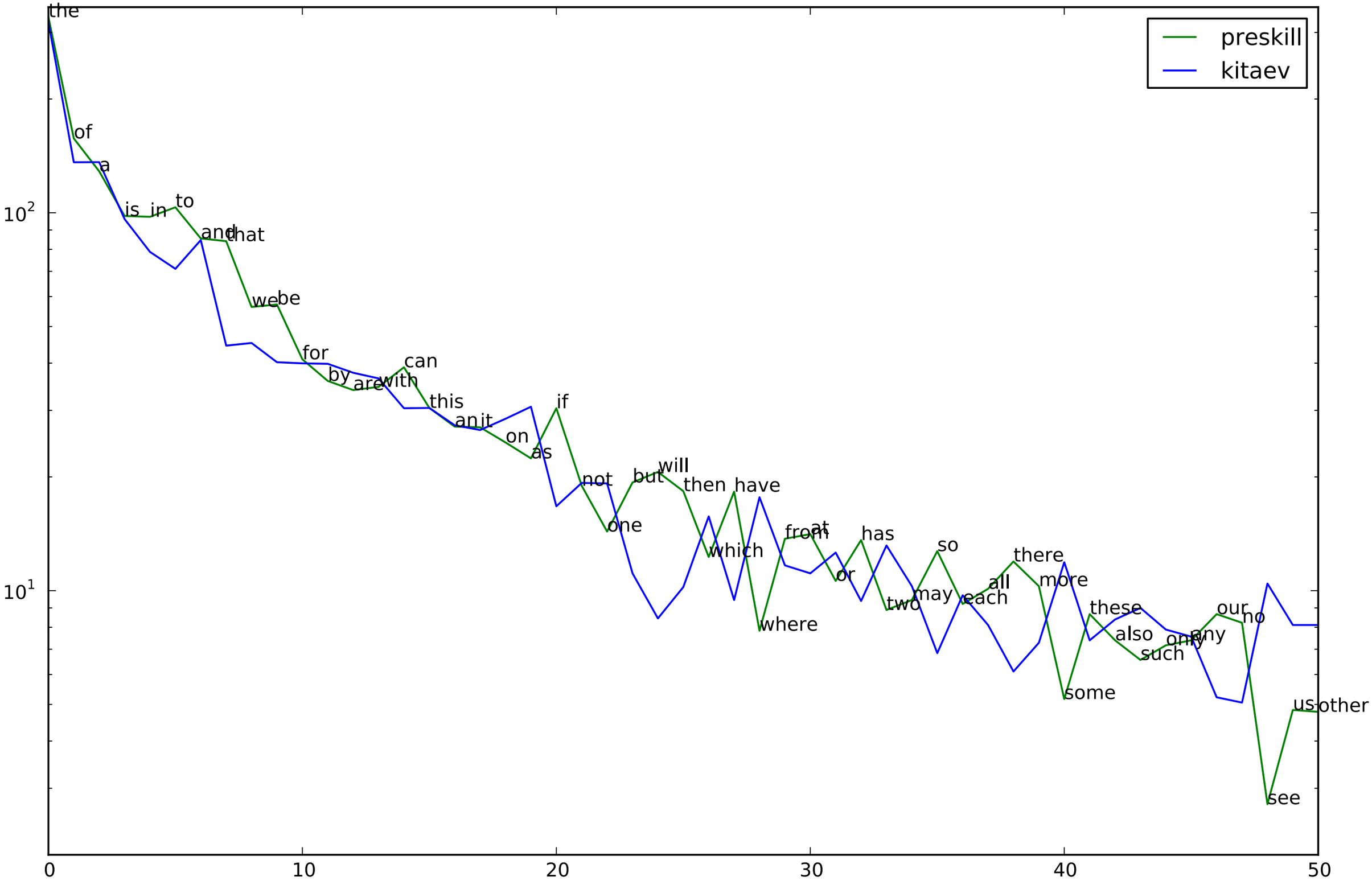
#13. The Magic of Oz (1919)

#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

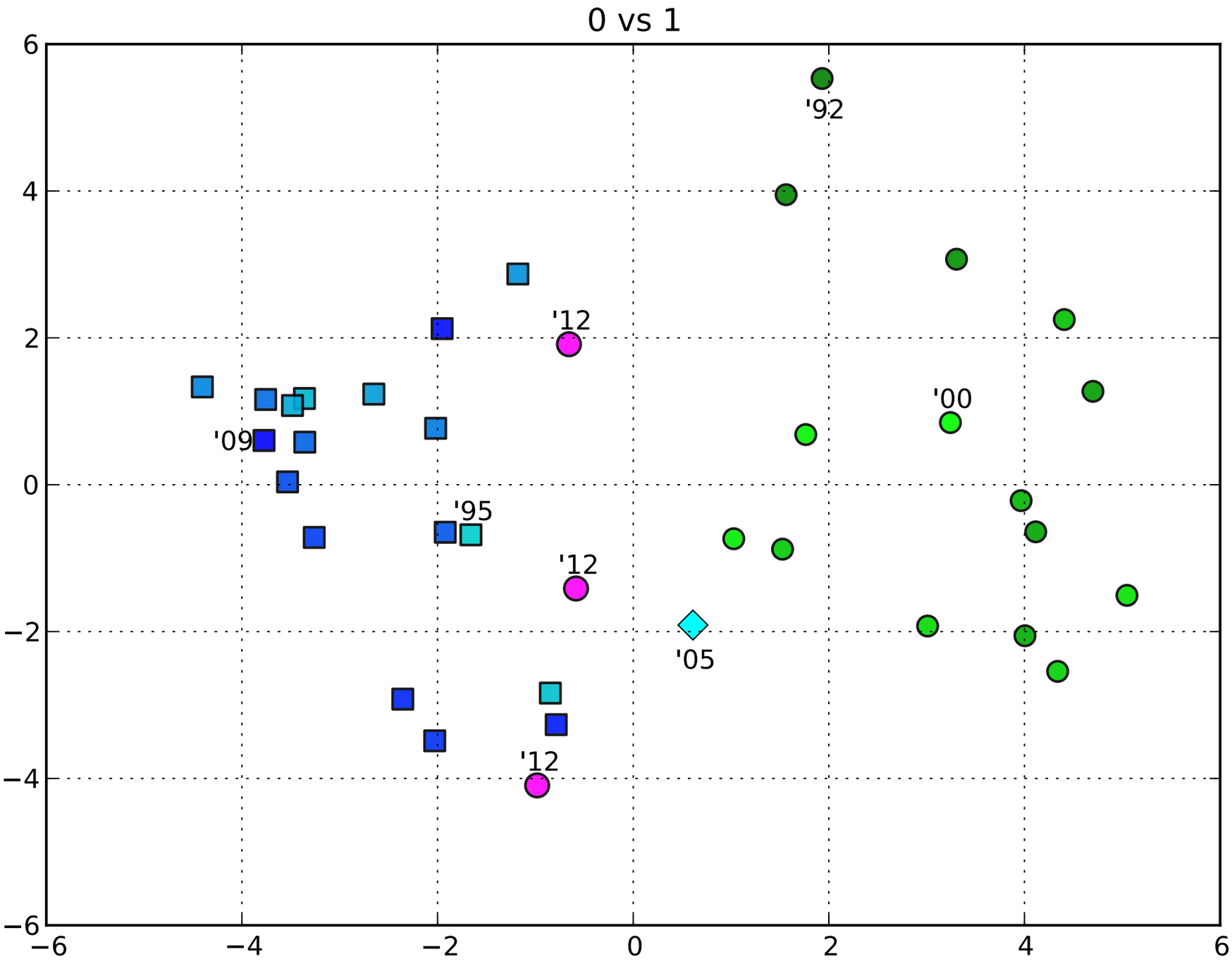
$$|\psi\rangle = M_{ij}|\phi_A^i\rangle \otimes |\phi_B^j\rangle = \sum_i \sigma_i |\psi_A^i\rangle \otimes |\psi_B^i\rangle$$

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"** ξ_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)

Kitaev

Preskill



Cornell Stylometric Connection:

"Literary Data Processing Conference"

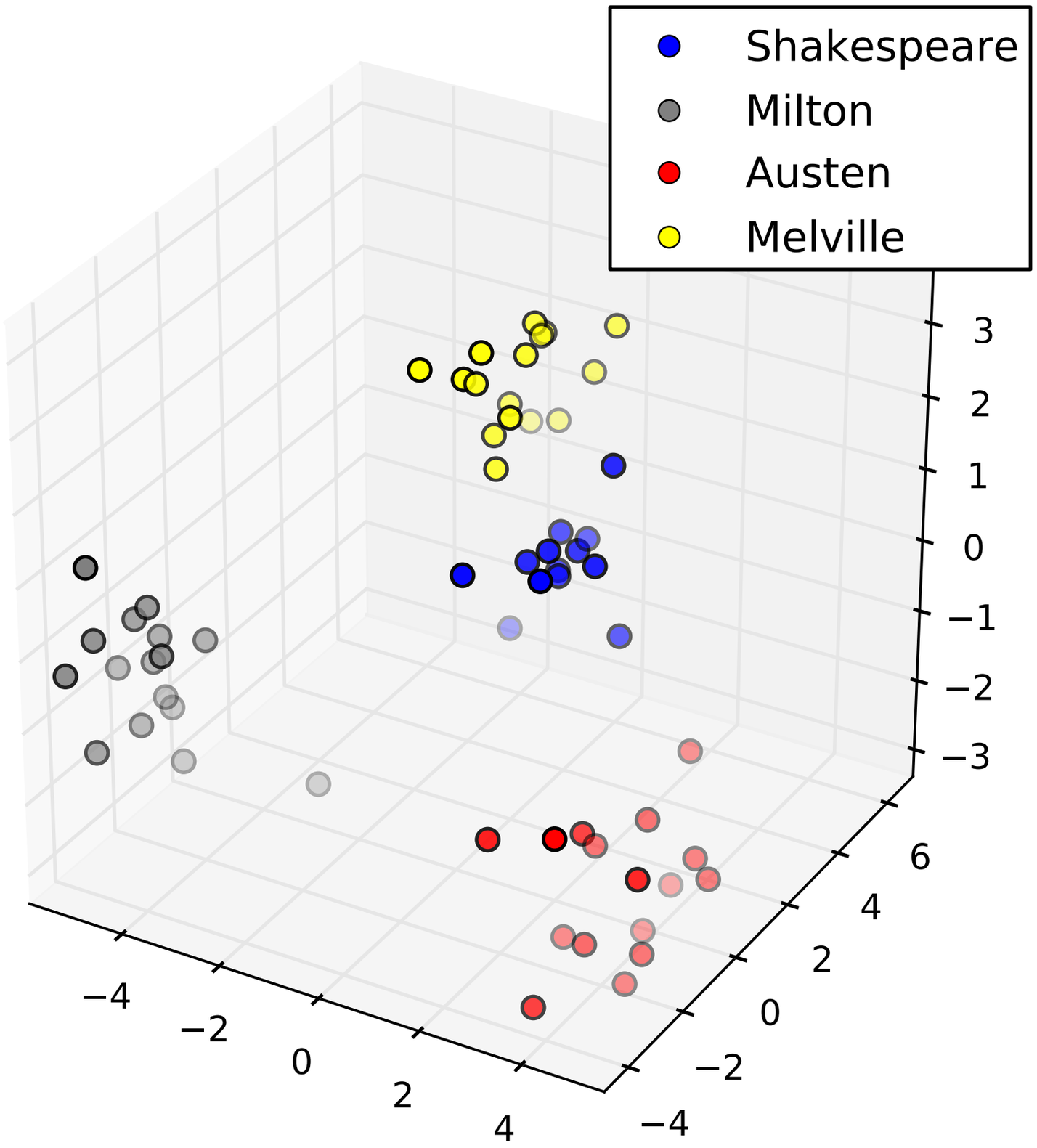
(Sep 1964, first conference on computers and humanities research?).
co-chaired by Stephen M. Parrish, Cornell, English Dept)

included "plea to the audience not to abandon their punch cards and magnetic tapes after their concordances were printed and (hopefully) published."

In Parrish's conference summary:

"when all the libraries or at least all pertinent bibliographical references are readily available on tape or in core memory, there will be no excuse for ignorance."

"...the perfection of attribution study or source study or influence study by computer techniques will make obsolete the studies that rely on the judgment and the memory of one poor fallible human scholar"


Shakespeare
Milton
Austen
Melville
3
2
1
0
−1
−2
−3
−4
−2
0
2
4
6
4
2
0
−2
−4

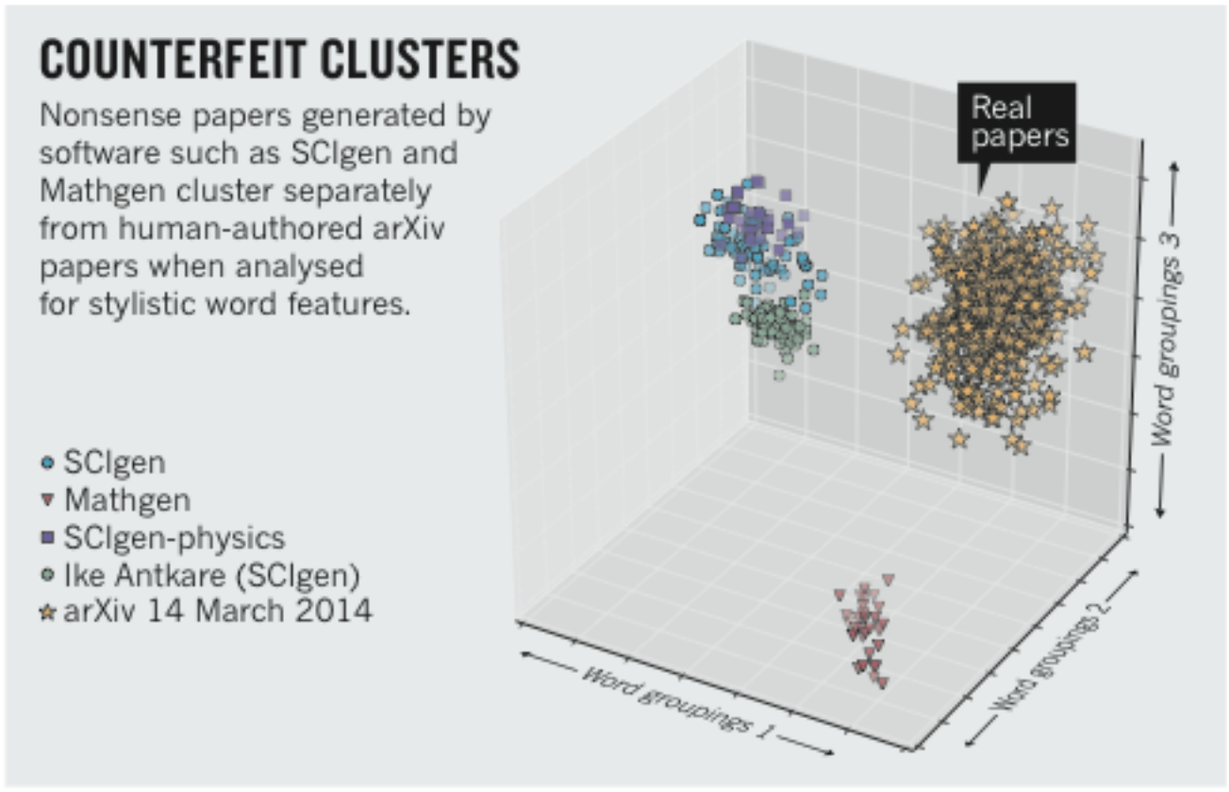
Correspondence

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



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,

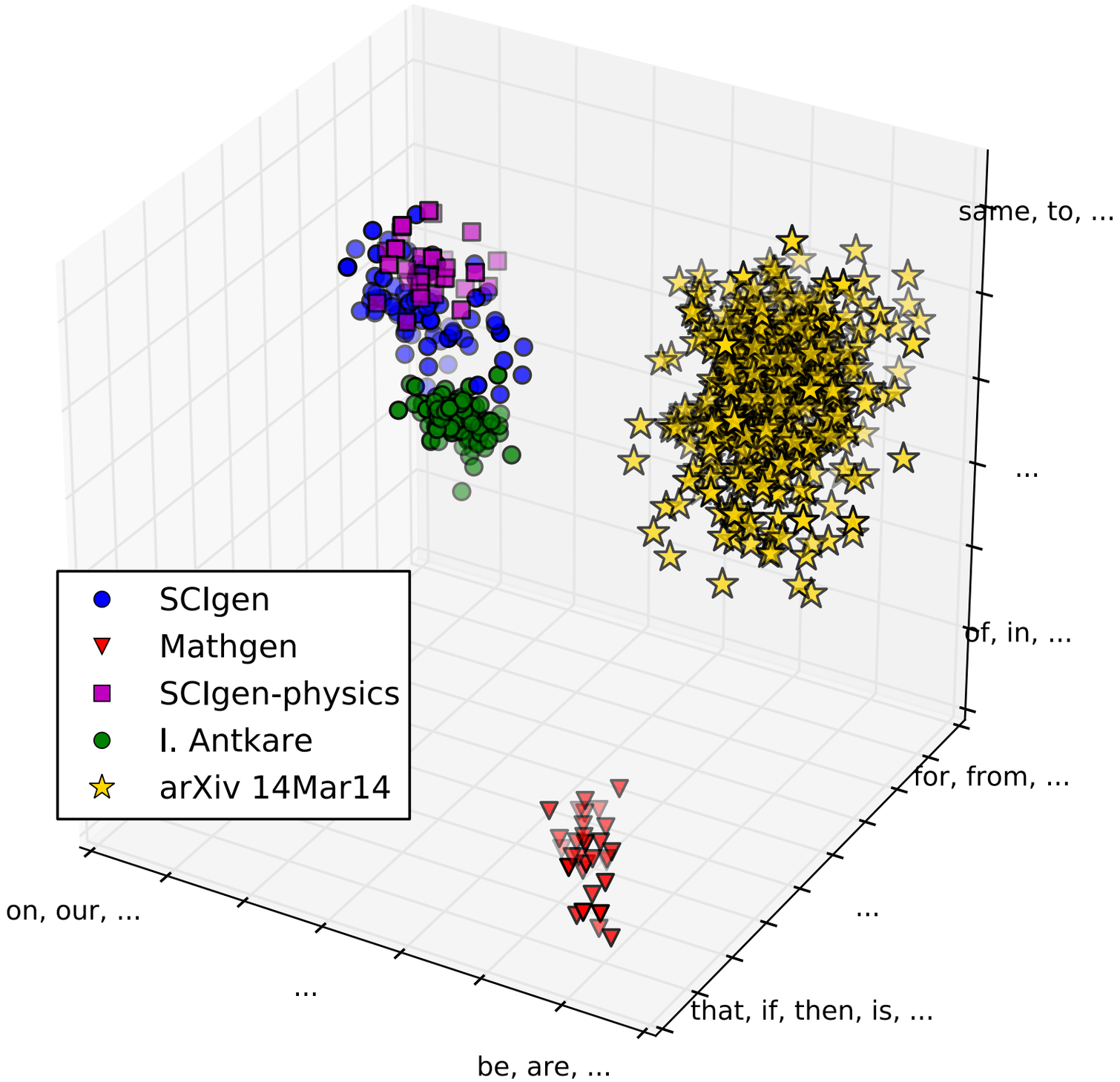
Projects powered by free computing grid

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

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.

word2vec

code.google.com/p/word2vec

- arxiv:1301.3781
- arxiv:1310.4546
- arxiv:1309.4168

Words generated by combining common tokens together.

$$\frac{\text{count}(w_1, w_2) - \delta}{\text{count}(w_1)\text{count}(w_2)} > \theta$$

Four passes with decreasing threshold.

$$\delta = 30, \theta \in \{400, 300, 200, 100\}$$

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":

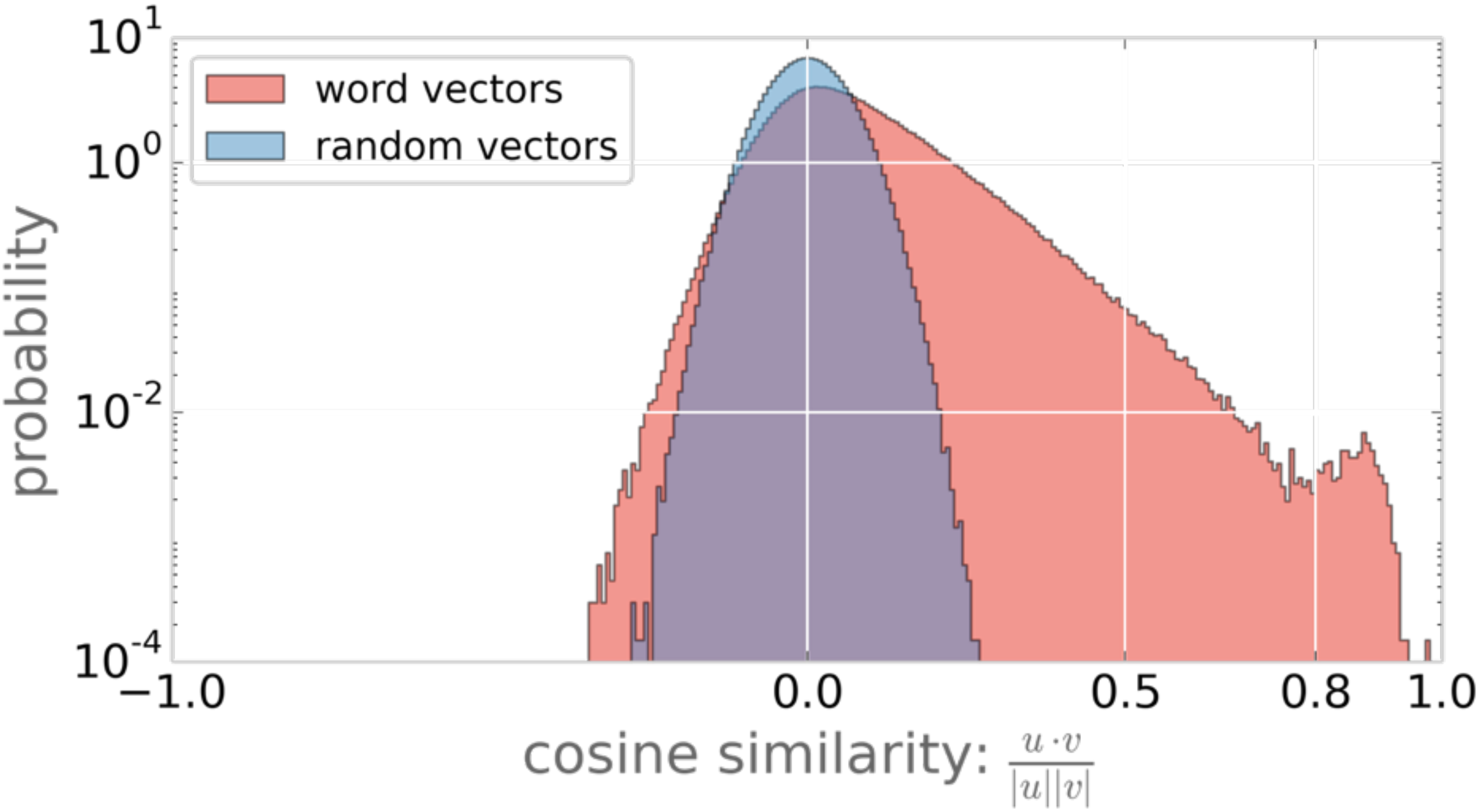
- "singular_value_decomposition",
- "black_hole",
- "aps_march_meeting"

Continuous skip-gram model. Single layer neural network.
Mapping of words to vectors: w_i.
Maximize the log probability of nearby words.

$$\sum_i \sum_{-n \le j \le n, j \neq 0} \log p(w_j | w_i)$$

where

$$p(w_j | w_i) = \frac{1}{Z} \exp\left(w_j \cdot w_i\right)$$


word vectors
random vectors
probability
10^{1}
10^{0}
10^{-2}
10^{-4}
−1.0
0.0
0.5
0.8
1.0
cosine similarity: $\frac{u \cdot v}{|u||v|}$

word	cosine
electron	1.0
electrons	0.83
positron	0.67
conduction_electron	0.65
carriers	0.64
unpaired_electron	0.63
electron_gas	0.63
electronhole	0.63
impurity	0.63
oneelectron	0.62
mobile_electrons	0.62

word	cosine
physics	1.0
theoretical_physics	0.76
particle_physics	0.72
nuclear_physics	0.7
astrophysics	0.68
astronomy	0.68
materials_science	0.67
phyiscs	0.66
astronomy_louisiana_state_univ...	0.66
bern_bern_switzerland_18	0.66
tennessee_knoxville_tennessee_...	0.66

word	cosine
blue	1.0
red	0.91
orange	0.87
cyan	0.87
purple	0.86
magenta	0.86
yellow	0.85
blue_red	0.85
violet	0.85
blue_green	0.83
light_blue	0.82

word	cosine
expects	1.0
expecting	0.63
would_allow	0.61
might_argue	0.6
expect	0.6
prefers	0.58
still_expects	0.57
one_could_imagine	0.57
can_afford	0.55
anticipate	0.54

word	cosine
almost	1.0
approximately	0.57
roughly	0.57
remarkably	0.55
fairly	0.53
still	0.51
however	0.51
nearly_constant	0.5
exponentially	0.49
although	0.49

word	cosine
want	1.0
wish	0.86
intend	0.74
do_not_need	0.71
do_not_know_how	0.67
don_t	0.66
trying	0.65
one_needs	0.63
going	0.62
come_back	0.62

word	cosine
svd	1.0
truncated_svd	0.7
svds	0.69
qr_decomposition	0.66
svd_decomposition	0.65
eigendecomposition	0.62
music_algorithm	0.61
omp_algorithm	0.61
cholesky_factorization	0.61
alternating_mini...	0.6

word	cosine
graphene	1.0
single_layer_graphene	0.9
monolayer_graphene	0.9
bilayer_graphene	0.88
multilayer_graphene	0.88
graphene_monolayer	0.88
graphene_sheets	0.87
suspended_graphene	0.85
graphene_monolayers	0.85
graphene_sheet	0.85

word	cosine
python	1.0
source_code	0.83
python_program...	0.83
java	0.83
scripting_language	0.82
command_line	0.82
scipy	0.82
package	0.82
scripts	0.82
libraries	0.82

word	cosine
supreme_court	1.0
politics	0.76
lawyers	0.74
justices	0.73
senate	0.73
congressional	0.72
politicians	0.72
public_opinion	0.71
republican_party	0.71
political	0.7

word	cosine
tfidf	1.0
tf_idf	0.8
cooccurrence	0.7
training_corpus	0.68
cosine_similarity	0.68
bm25	0.68
tf_idf_weighting	0.68
stop_words	0.67
unigrams	0.67
cosine_similarity_between	0.66

word	cosine
scientific_impact	1.0
interdisciplinarity	0.79
citation_impact	0.77
bibliometrics	0.76
citation_analysis	0.75
citation_patterns	0.74
bibliometric	0.73
journals	0.73
scholarly	0.73
peer_review	0.72

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

syllogism

a:b :: c:d

Paris - France + Italy = ?

torque - force + momentum =

orbital_angular_momentum

newtonian_mechanics - isaac_newton +
albert_einstein =

special_relativity

gravity - Newton + Hawking =

black_hole_evaporation

smuon - muon + higgs =

higgsino

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

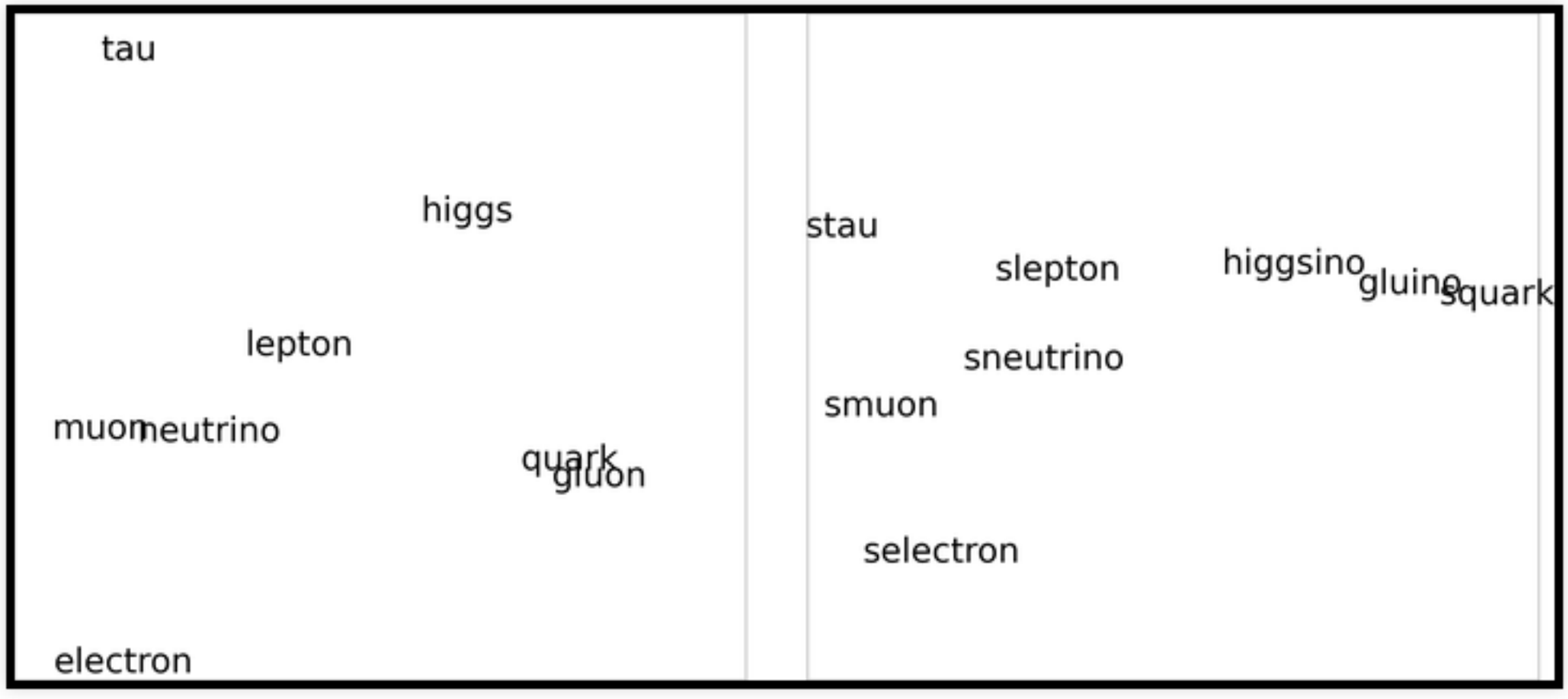
newtonian_mechanics-isaac_newton+albert_einstein=

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.



word2vec knows supersymmetry

physics+buzzword=

buzzword	0.79
physics	0.77
life_science	0.7
philosophy	0.68
interdisciplinary_research	0.67
biology	0.66

chemistry+physics=

chemistry	0.92
physics	0.9
theoretical_chemistry	0.73
solid_state_physics	0.72
chemical_engineering	0.7
theoretical_physics	0.69

biology-buzzword=

physiology	0.44
ecology	0.42
biological_macromolecules	0.41
cell_motility	0.4
flegg	0.39
phenotypes_nat_rev_genet	0.39

sm_higgs_boson the search for the higgs_boson the only elementary_particle in the sm that has_not_yet been_observed is one of the highlights of the large_hadron_collider 11 lhc physics_programme indirect limits on the sm_higgs_boson mass of mh 158 gev at_95_confidence_level_cl have_been set using global_fits to precision_electroweak results 12 direct_searches at lep 13 the tevatron 14 16 and the lhc 17 18 have previously excluded at_95_cl a sm_higgs_boson with mass below 600_gev apart_from some mass regions between 116 gev and 127 gev both the atlas and cms_collaborations reported excesses of events in their 2011 datasets of protonproton pp_collisions_at centre of mass energy s 7 tev at the lhc which were compatible_with sm_higgs_boson_production and decay in the mass region 124 126 gev with significances of 2 9 and 3 1 standard_deviations s respectively 17 18 the cdf and do experiments at the tevatron have also recently_reported a broad excess in the mass region 120 135 gev using the existing lhc constraints the observed local significances for mh_125_gev are 2 7 s for cdf 14 1 1 s for do 15 and 2 8 s for their combination 16 the previous atlas_searches in 4 6 4 8 fb 1 of data at s 7 tev are combined here with new searches for h zz 41 h gg and h ww

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

astro
cond-mat
cs
hep
math
places
references
names
equations
english

art	title	cosine
1207.7214	Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC	1.0
1307.1427	Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC	0.98
1207.7235	Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC	0.98
1303.4571	Observation of a new boson with mass near 125 GeV in pp collisions at sqrt(s) = 7 and 8 TeV	0.97
1106.2748	Limits on the production of the Standard Model Higgs Boson in pp collisions at sqrt(s) =7 TeV with the ATLAS detector	0.97
1206.0756	Search for the Standard Model Higgs boson in the H -> WW(*) -> l nu l nu decay mode with 4.7 /fb of ATLAS data at sqrt(s) = 7 TeV	0.97
1211.6956	Search for the neutral Higgs bosons of the Minimal Supersymmetric Standard Model in pp collisions at sqrt(s)=7 TeV with the ATLAS detector	0.97

James Clerk Maxwell (Feb 1856)

ESSAY FOR THE APOSTLES ON 'ANALOGIES IN NATURE'

FEBRUARY 1856

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

ARE THERE REAL ANALOGIES IN NATURE?[2]

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.

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

From 1502.03520 (S. Arora et al.)

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

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

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

How to find vectors that do this?

Example: man:woman :: king:?

First need analytic representation of task. Note that most contexts χ 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:

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

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

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

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

Suppose:

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

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

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) $\mathrm{PMI}(w,w') \propto v_w \cdot v'_w$ (factorization)]

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

By factorization (1),

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

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

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

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

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

Summary

The word d that minimizes

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

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

Under assumptions (1,2), if we can find isotropic vectors such that $v_w \cdot v_{w'} \propto \mathrm{PMI}(w, w')$, 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.

Text Segmentation based on Semantic Word Embeddings

Alexander A Alemi, Paul Ginsparg

(Submitted on 18 Mar 2015)

We explore the use of semantic word embeddings in text segmentation algorithms, including the C99 segmentation algorithm and new algorithms inspired by the distributed word vector representation. By developing a general framework for discussing a class of segmentation objectives, we study the effectiveness of greedy versus exact optimization approaches and suggest a new iterative refinement technique for improving the performance of greedy strategies. We compare our results to known benchmarks, using known metrics. We demonstrate state-of-the-art performance for an untrained method with our Content Vector Segmentation (CVS) on the Choi test set. Finally, we apply the segmentation procedure to an in-the-wild dataset consisting of text extracted from scholarly articles in the arXiv.org database.

Comments: 10 pages, 4 figures. KDD2015 submission
Subjects: **Computation and Language (cs.CL)**; Information Retrieval (cs.IR)
Cite as: arXiv:1503.05543 [**cs.CL**]
(or arXiv:1503.05543v1 [**cs.CL**] for this version)


Text Segmentation based on Semantic Word Embeddings

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

Recommendation redux

Complaints about information overload date back 2 millennia

coreadership (proxy for svd or more sophisticated)

- **netflix prize**
- **itunes genius**

Evaluation metric? (free bagels and cream cheese for duration)

The information layer vs. the social layer (Google vs. Facebook): optimal referral mechanism?

Dangers of recommendation systems: local vs. global diversity

- **imperfect filter worse than none?**

Personalization: readers inherit topics from articles

Recommender Systems

Example: NASA ADS (Astrophysical Data System) uses (anonymized) arXiv usage data

(a) infer topics from readership data and keyword assignment

- **classify articles and users (based on past activity) according to topics**
- **measures of proximity of articles to people, and articles to themselves.**
- **reader can be presented with a menu of recent papers on subjects of interest (ordered according to closeness of match, or by importance as measured by readership or citation, ...)**

(b) find the 40 most similar articles (augments data for sparse readership) to make article-based recommendations, via a few algorithms

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?

Full text analysis

winnowed 7-grams (w/ J. Gehrke, D.Sorokina , S. Warner 2007), after Schleimer et al. (2003)

Detection software now more feasible than ever, computation of fingerprint in memory with 96Gb machine, hundreds of lookups per second

practical implications for running arXiv site: problem authors are inconvenience to readers, but screening was haphazard, no systematic baseline to identify outliers

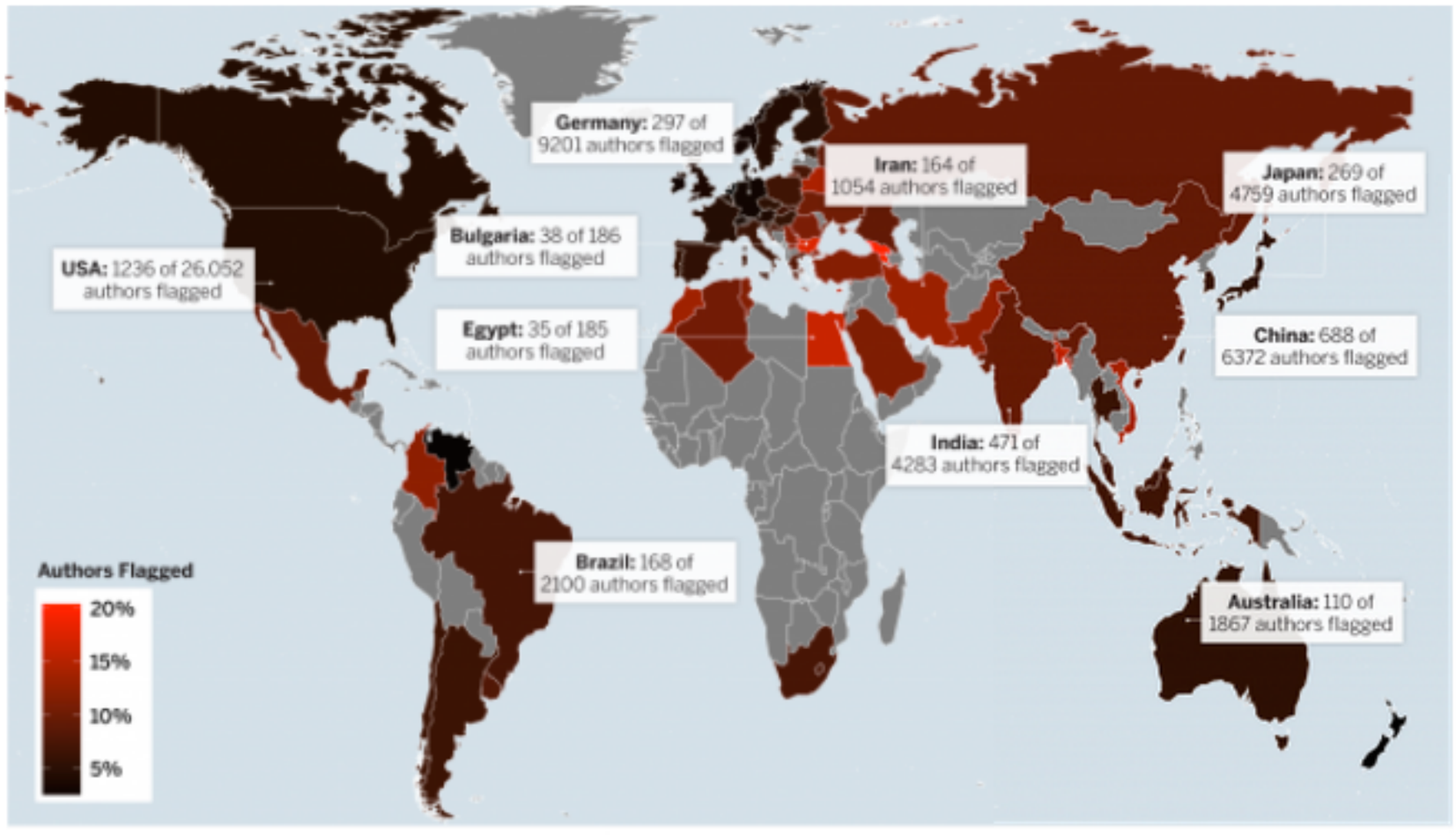
cs Meng project Scott Rogoff (in conjunction with S. Warner), spring 2011

more systematically summer 2012 w/ Daniel Citron*

*D.Citron and PG, "Patterns of text reuse in a scientific corpus," PNAS 2015 112 (1) 6-7, arXiv:1412.2716

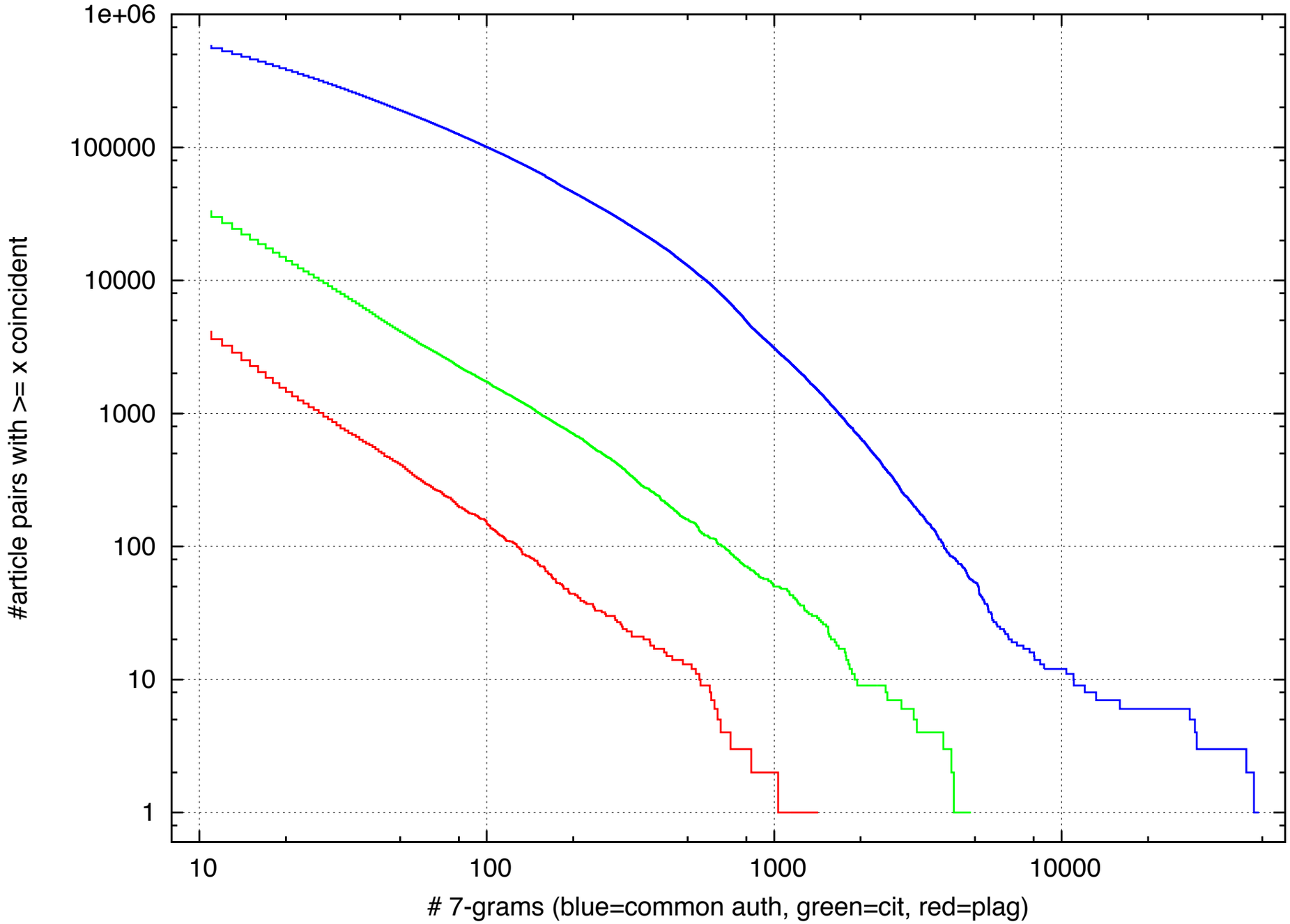
SCIENCEINSIDER

Breaking news and analysis from the world of science policy



Study of massive preprint archive hints at the geography of plagiarism

By John Bohannon | 11 December 2014 3:00 am | 69 Comments



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.

http://arxiv.org/help/overlap

Starting in Jun 2011, some submissions to arXiv marked with an "admin note", indicating text overlap with other arXiv submissions (200-250 / month currently flagged).

"Such notes are intended as informational to readers, and as well to authors from different educational backgrounds. Readers frequently find it useful to know when an article draws heavily from another, or supersedes an earlier article. Some authors, by contrast, are not aware that importing large sections of text either from their earlier articles, or from articles by others, is not common practice."

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

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.)”

high threshold, cont'd

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 the newer text must be taken verbatim from the earlier article in order to be noted.

(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.

author reactions

a) none (replace w/o changing, do they even notice?)

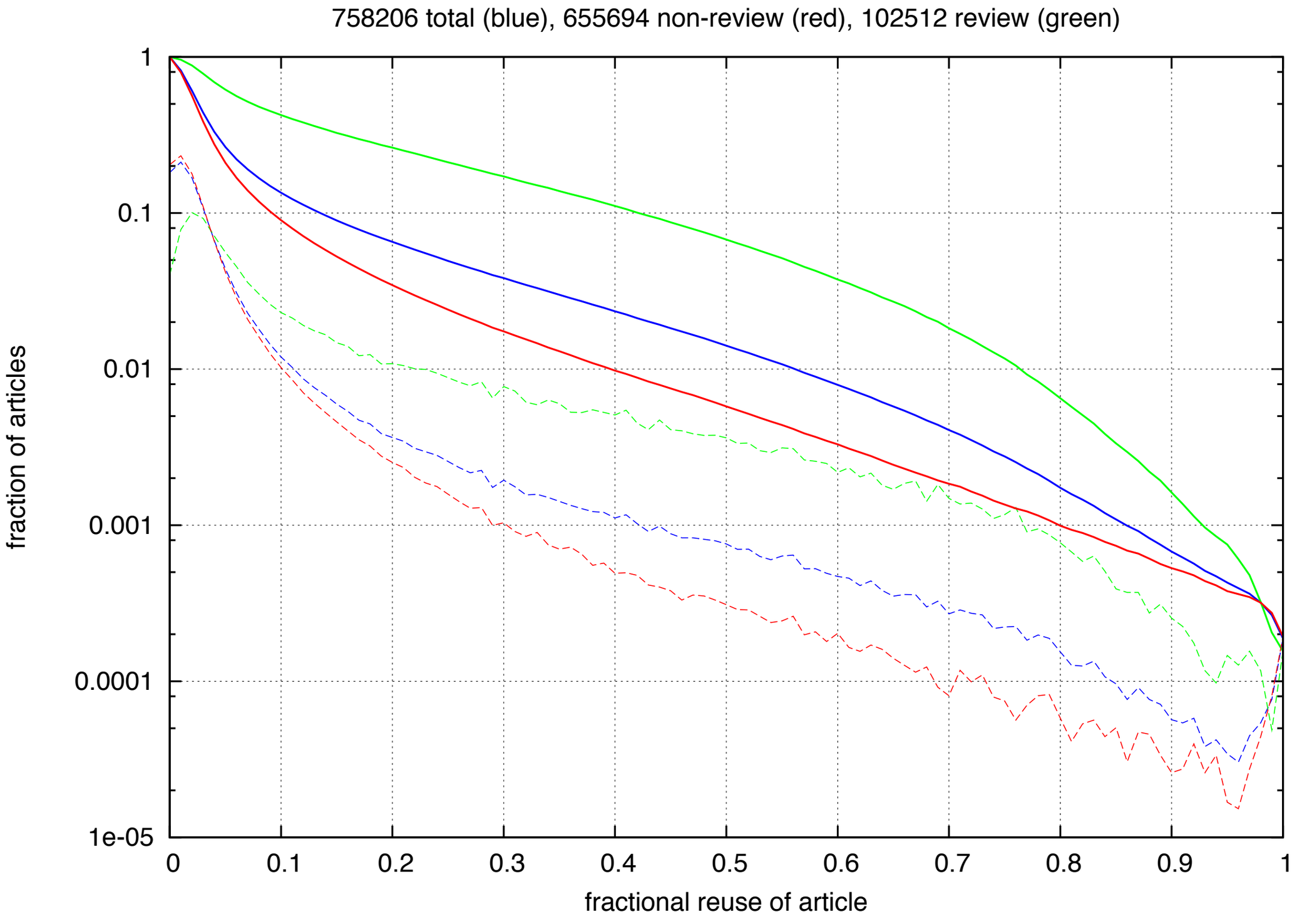
b) try to remove overlaps, not always successful (can't even find ?!?)

c) virulently object (crowdsourced quality control of methodology, though usual complaints misguided, exposing confusion about what is standard practice, as statistically confirmed by arXiv corpus)

[but note Kiesler et al, 2010, "Regulating Behavior in on-line communities":
Design Claim 15: Publicly displaying many examples of inappropriate behavior on the site will lead members to believe this is common and expected.]

start to distinguish

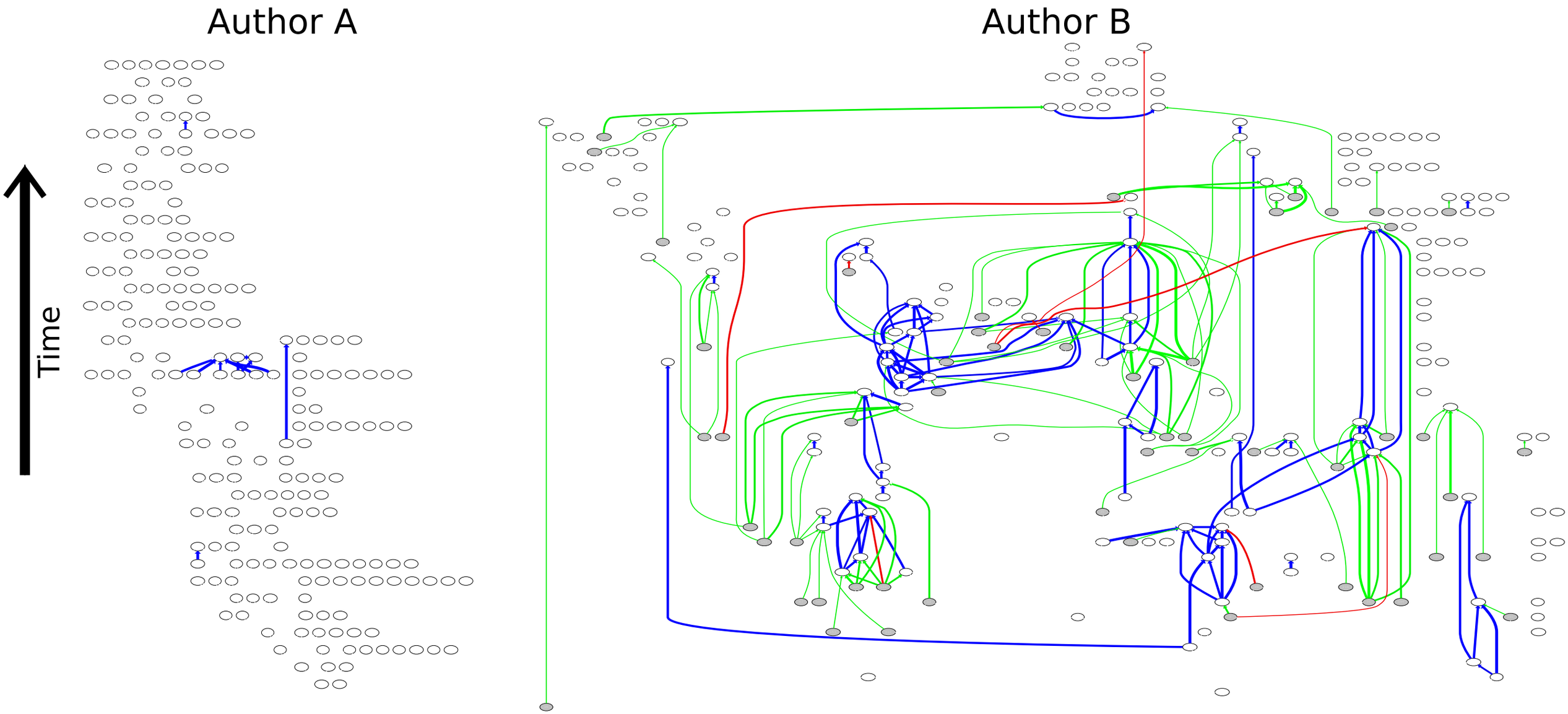
previous graph looked bad, but is some of it "acceptable" recycling?



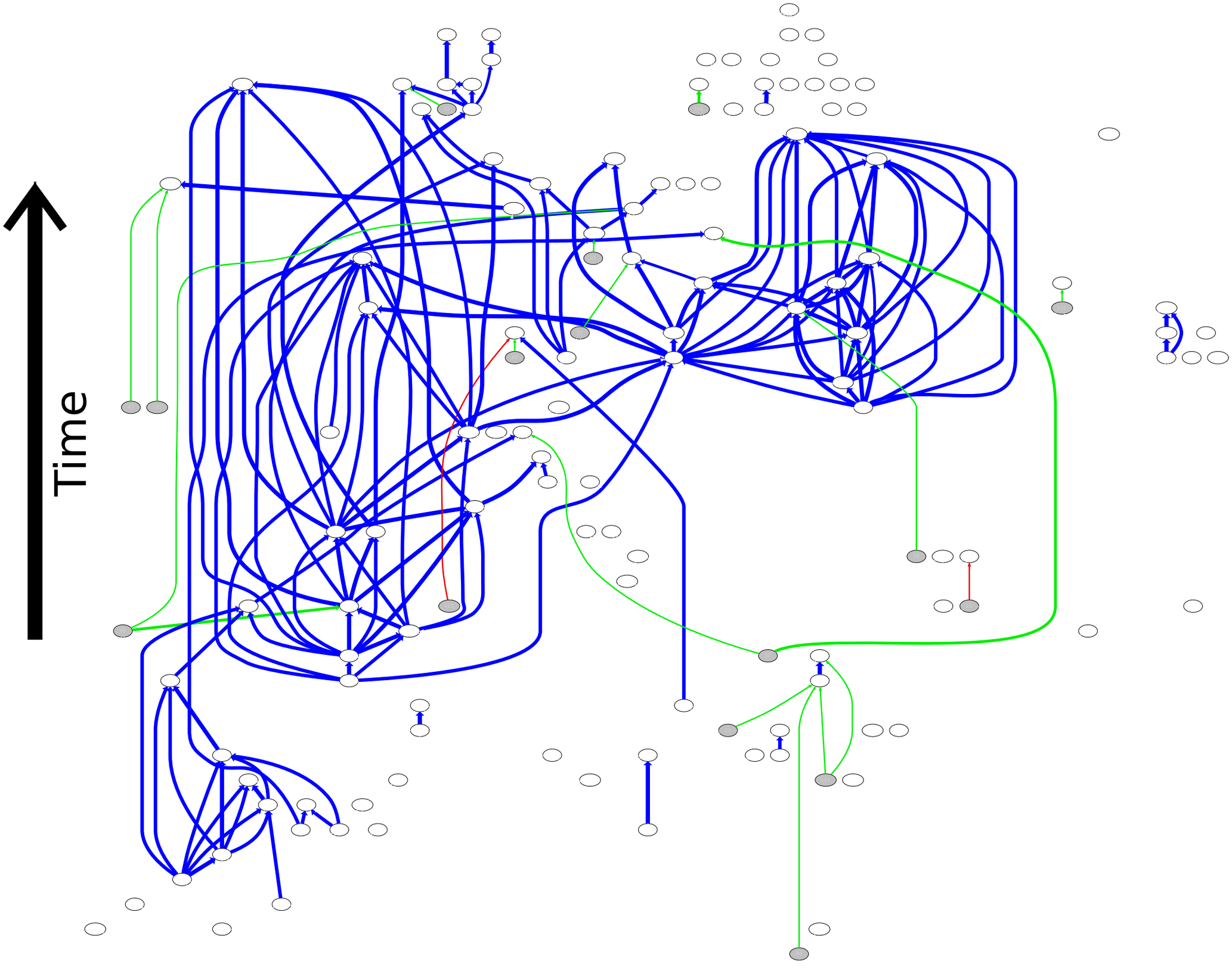
Fraction of articles on vertical with at least the indicated fraction of reused 7-grams on horizontal. green = “review”, red = non-“review”, blue= all

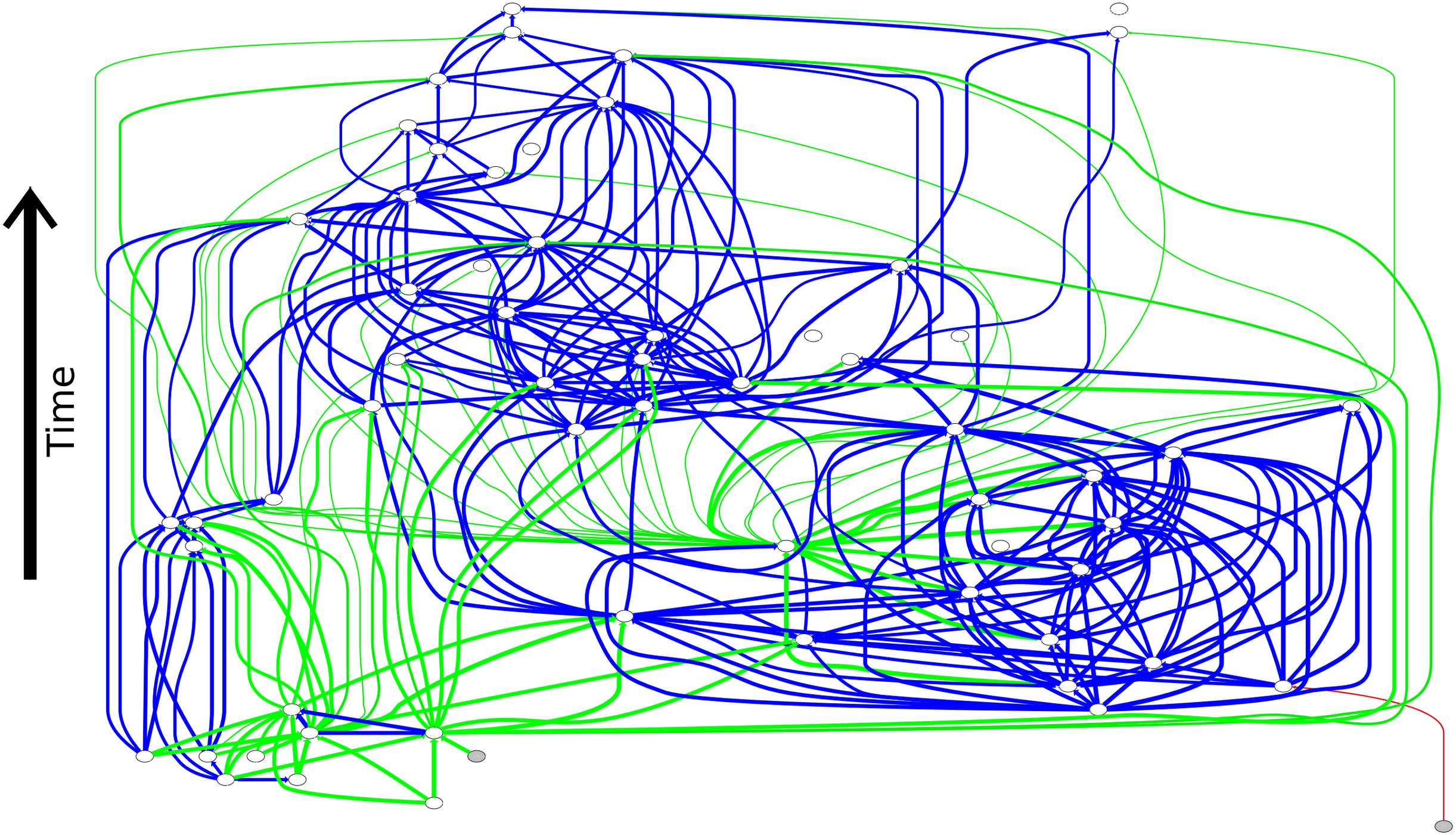
but how distributed?

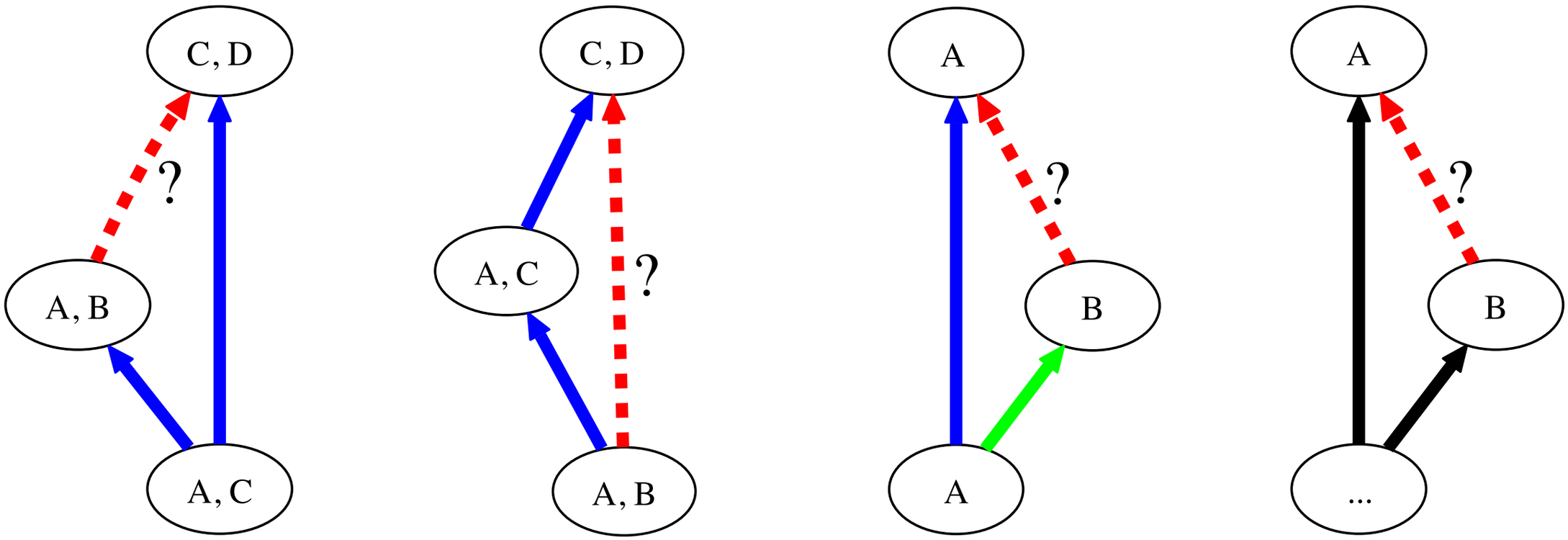
previous graph looked bad, but is it all of the authors some of the time, or some of the authors all of the time?



Tale of two authors: edges representing self-copying, cited material, and material recopied without citation are colored **blue**, **green**, and **red**, respectively. The edge thickness increases with the amount of overlap between the two articles. Nodes colored grey are attributed to other authors.


Time


Time

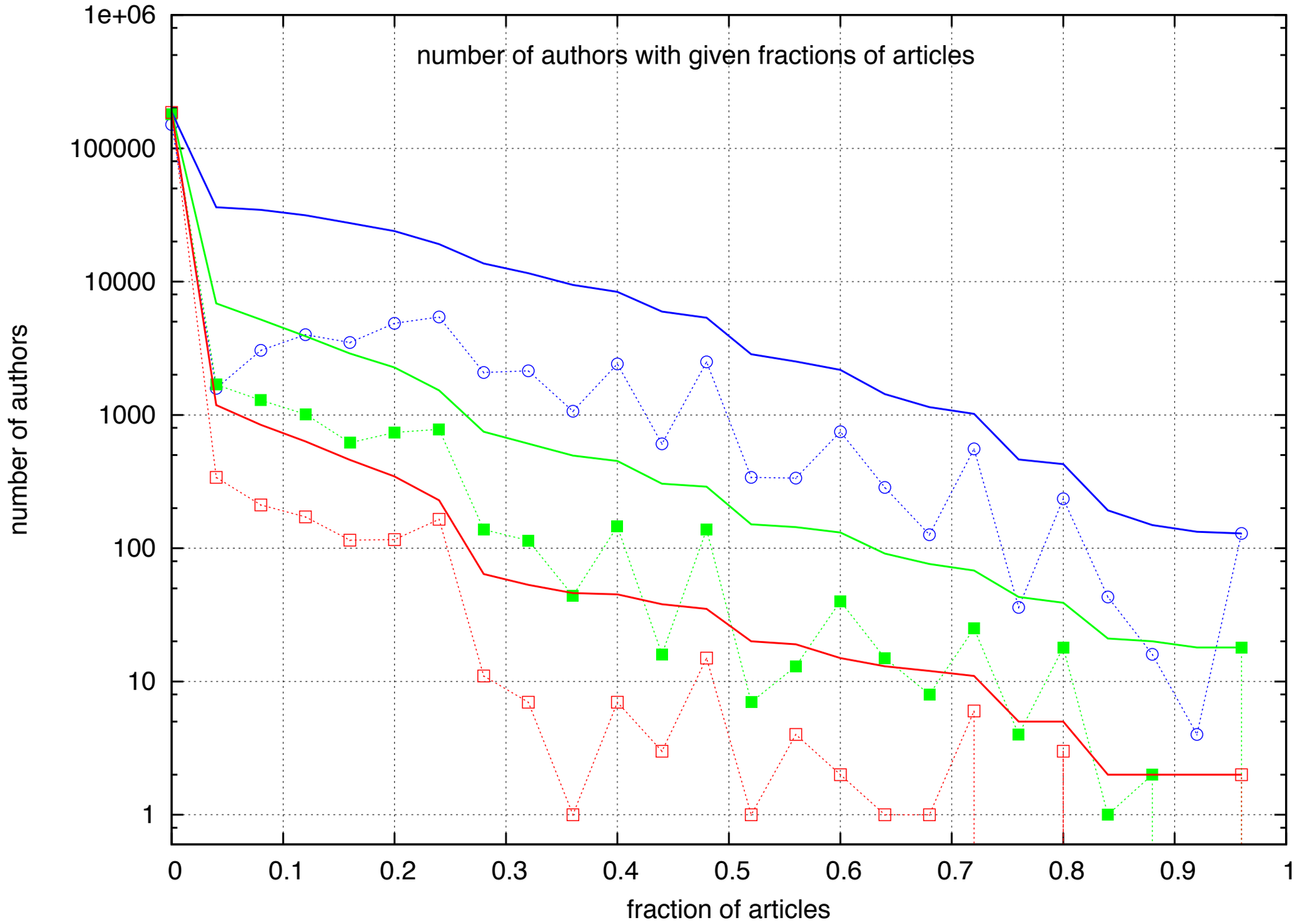


prior co-authored, but “ok”

demonstrably not by C,D but ?

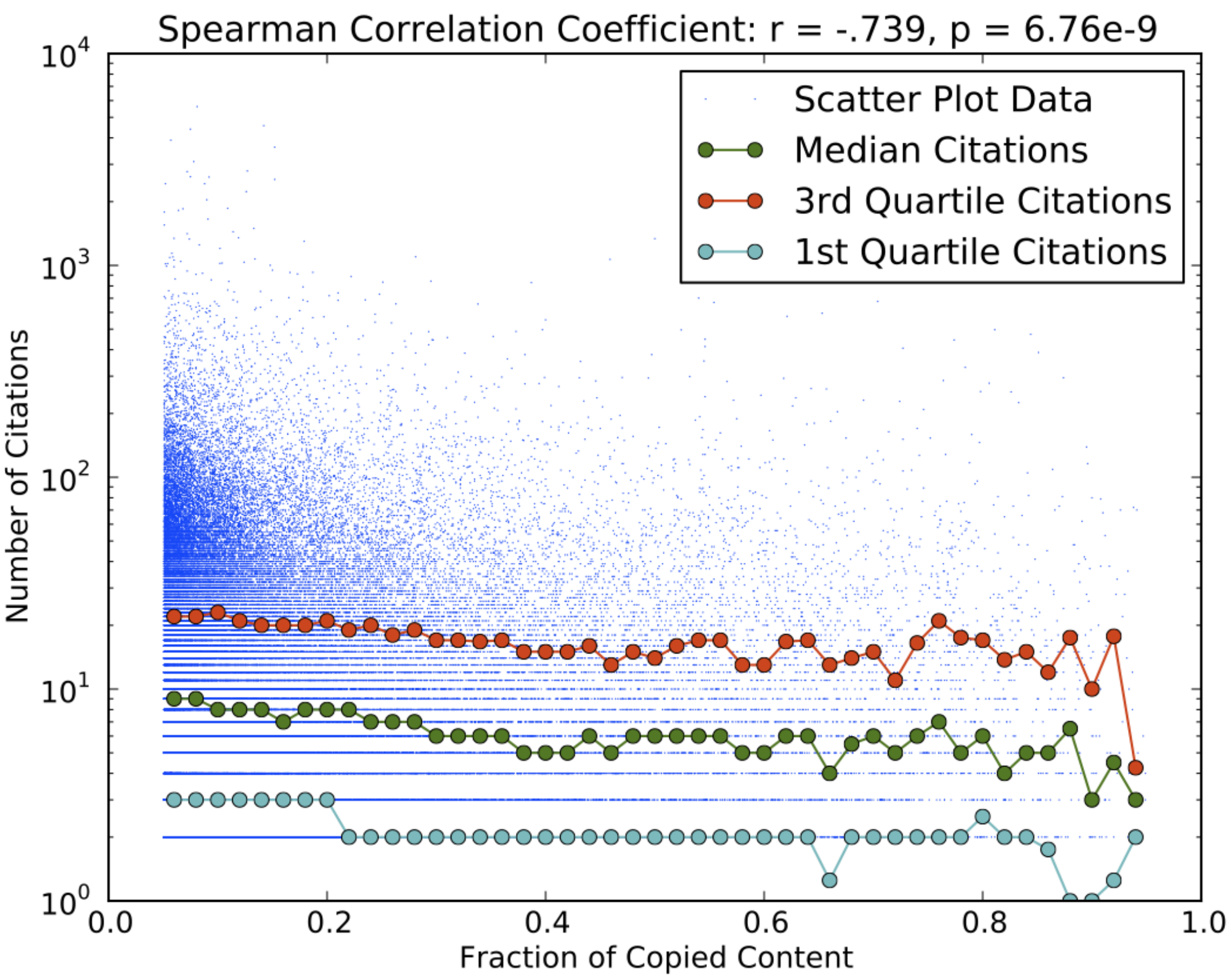
actually self

common source

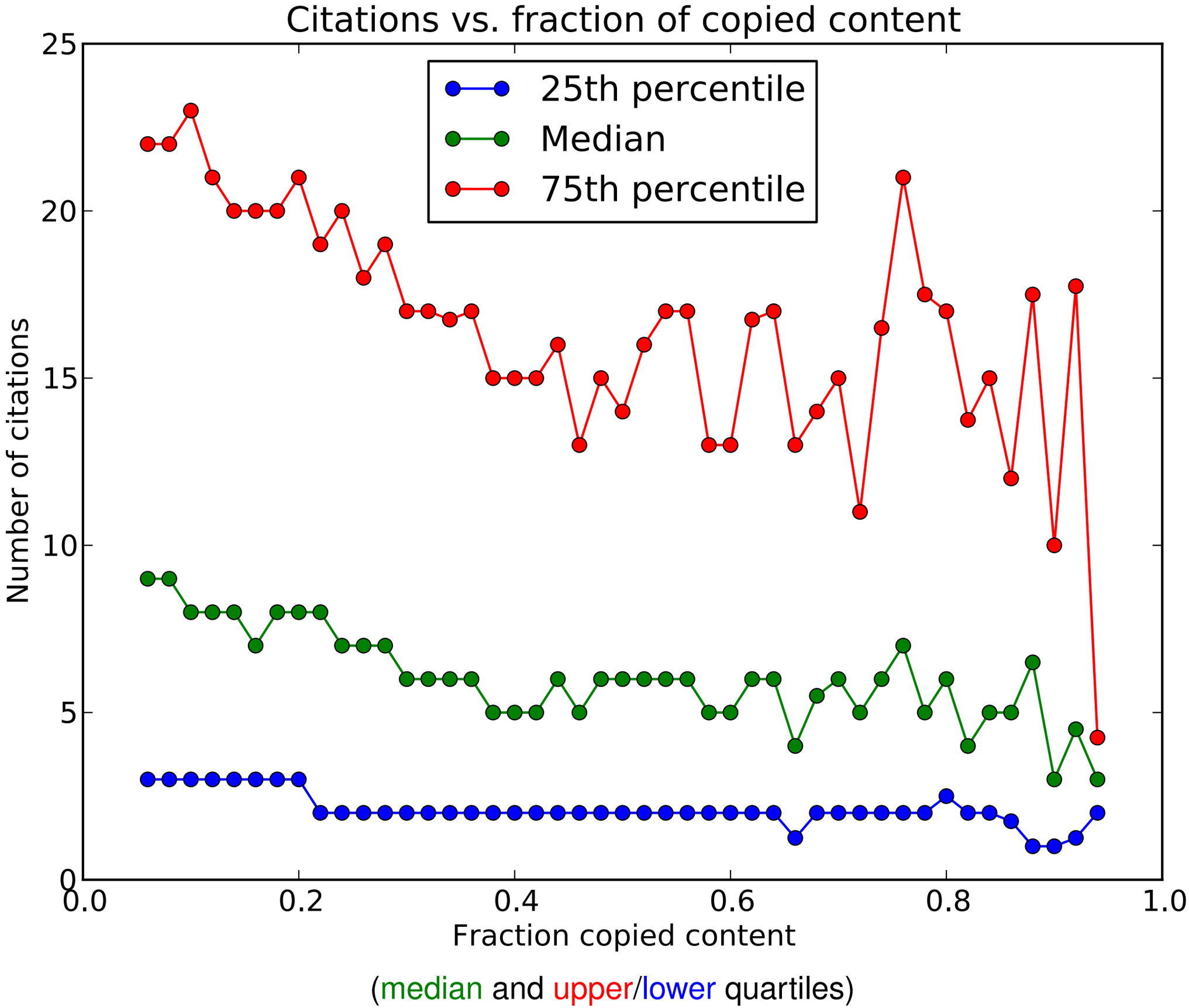


Authors vs. fraction of articles that include text overlaps. $\sim$ 1720 authors have at least 50% of articles with significant CA text overlap. Of 392,850 authors, only 49,830 have at least 1% CA, only 8990 with CI, and only 1630 contain PI (vast majority OK). Moreover only 10,550, 1130, and 130 authors have at least 25%, resp.

a quality flag?



citations vs. fraction of copied content (blue). median citations vs. fraction of copied content in green, negative correlation (116,490 pre 2011 articles, self-citations removed)


Citations vs. fraction of copied content
25th percentile
Median
75th percentile
Number of citations
0
5
10
15
20
25
0.0
0.2
0.4
0.6
0.8
1.0
Fraction copied content
(median and upper/lower quartiles)

Underlying sociology?

non-native speakers of English

(exacerbated by the ease of text reuse in the electronic format? but also easier to detect)

perhaps not willful fraudulence but different (deficient?) educational systems

an act of magic to produce a new idea? of course articles are produced by weaving together texts from existing sources (as was done by mentors)

In summary

- **it's easy, out there, no one has really looked (except for turnitin, moss.stanford.edu , ...)**
- **text overlap not plagiarism (though there are a few instances of duplicate articles by different authors, thoroughly inexplicable)**
- **not the most creative authors**
- **an educational issue re common practice (systematic reuse ok for review articles, as opposed to lecture notes or conf proc?)**
- **(or perhaps that's changing, e.g. wikipedia comment "not necessary to cite dynamically produced content")**
- **uncited reuse rare (question of training, cite but include blocks of text)**
- **lessons for how we train undergrads in modern networked world?**
- **still aren't many comprehensive OA corpora available, but can be done on some? (nature, science, phys rev, pubmedcentral)**

1101.5456

I take this opportunity to express my deep sense of gratitude to my supervisor, Dr. Sanjay Kumar, for his constant encouragement, cooperation and invaluable guidance in the successful accomplishment of this dissertation. I also express my gratitude to Prof. B. K. Dass, Head, Department of Mathematics, University of Delhi for providing necessary facilities and constant encouragement during the course of this study.

I also wish to extend my thanks to all the faculty members of the Department of Mathematics, University of Delhi for their help, guidance and motivation for the work presented in the dissertation. They have always been there for me whenever I needed support from them, providing me critical research insights and answering my questions with their valuable time. Their academic excellence has also been a great value to my dissertation.

I am also thankful to my friends and fellow research scholars (specially Sumit Nagpal, Kuldeep Prakash, Sarika Goyal and Rani Kumari) for their help and discussion during the course of my study. I am also thankful to M.M.Mishra, Assistant Professor, Hansraj college, for his valuable guidance in Latex.

I also wish to express my gratitude to the C.S.I.R for granting me the fellowship which was a great financial assistance in the completion of my M. Phil programme. I am sincerely thankful to my parents for motivating me to do higher studies. I would also like to extend my gratitude to my brothers and sisters for helping me in every possible way and encouraging me to achieve my long cherished goal.

Above all, I thank, The Almighty, for all his blessings bestowed upon me in completing this work successfully.

1407.8478

I take this opportunity to express my deep sense of gratitude to my supervisor, Dr. Sanjay Kumar, for his constant encouragement, cooperation and invaluable guidance in the successful accomplishment of this dissertation. I also express my gratitude to Prof. Ajay Kumar, Head, Department of Mathematics, University of Delhi for providing necessary facilities and constant encouragement during the course of this study.

I also wish to extend my thanks to all the faculty members of the Department of Mathematics, University of Delhi for their help, guidance and motivation for the work presented in the dissertation. They have always been there for me whenever I needed support from them, providing me critical research insights and answering my questions with their valuable time. Their academic excellence has also been a great value to my dissertation. I am also thankful to the organizers of ATM schools of geometry and topology, which I attended in CEMS Almora, NEHU Shillong and HRI Allahabad, which helped me to learn many facts related to this field.

I am also thankful to Prof. Ravi S. Kulkarni, who gave the idea of this work and discussed the problem, and Prof. Anant R. Shastri, for his guidance in better understanding of the subject. I am also thankful to my friends and fellow research scholars (specially Dinesh Kumar and Gopal Datt) for their help and discussion during the course of my study.

I also wish to express my gratitude to the U.G.C. for granting me the fellowship which was a great financial assistance in the completion of my M. Phil program.

I am sincerely thankful to my parents for motivating me to do higher studies and encouraging me to achieve my long cherished goal.

Above all, I thank, The Almighty, for all his blessings bestowed upon me in completing this work successfully.

1306.3408

First and foremost, I wish to extend my gratitude to my supervisors, Professors W. David McComb and Arjun Berera. Without their continued support I would never have completed this thesis. I wish to thank Prof. McComb for his patient guidance and motivation towards research. I thank Prof. Berera for sharing his knowledge and enthusiasm with me, as well as for being approachable with any problems I had. I have learnt a lot from working with both of them.
Particular thanks are due to Dr. Matthew Salewski, for his friendship and many stimulating discussions on the topic of turbulence.
I cannot describe how indebted I am to my wonderful girlfriend, Amanda, whose love and encouragement will always motivate me to achieve all that I can. I could not have written this thesis without her support; in particular, my peculiar working hours and erratic behaviour towards the end could not have been easy to deal with!
Of course, I would never have made it this far without the love and support of my family, particularly my mum and brother, Joe. Their interest (a facade though it may have been!) in my work and pride at my achievements has always been an inspiration.
I could also have not made it through without the many friends I have made along the way. I particularly wish to thank my colleagues and flatmates Gavin and Liam, as living and working with them was a privilege. I also thank Eoin for the many jamming sessions and encouraging the creation of the physics dept. football team, the Feynmen.
When I joined the particle theory group, I was instantly made to feel welcome and included, for which I owe additional thanks to Erik, Claudia, Simone, Thomas and Brian. I extend my thanks and best wishes to the entire PPT corridor and the students and post-docs I got to share lunch, coffee and/or (several) pints with.
I would like to thank Jane Patterson for her kindness and ensuring my PhD career ran smoothly.
I gratefully acknowledge the generosity and support of the Edinburgh Compute and Data Facility. My funding was provided by the STFC, to whom I am eternally grateful for this opportunity.

1408.4411

First and foremost, I wish to extend my gratitude to my supervisors, Prof. Dino Anthony Jaroszynski and Dr. Adam Noble, whom I also find to be a close friend. Without their continued support I would never have completed this thesis. I wish to thank Dr. Noble for his patient guidance and motivation towards research. I thank Prof. Jaroszynski for sharing his knowledge and enthusiasm with me, as well as for being approachable with any problems I had. I have learnt a lot from working with both of them.
Particular thanks are due to Dr. Samuel Yoffe, for his friendship and many stimulating discussions on the topic of radiation reaction and quantum corrections.
I cannot describe how indebted I am to my wonderful wife, Renata, whose love and encouragement will always motivate me to achieve all that I can. I could not have written this thesis without her support; in particular, my peculiar working hours and erratic behaviour towards the end could not have been easy to deal with!
Of course, I would never have made it this far without the love and support of my beloved grandparents. Their interest in my work and pride at my achievements has always been an inspiration.
I could also have not made it through without the many friends I have made along the way. I particularly wish to thank my colleagues at the University of Strathclyde and senior researchers at Lancaster University as working with them was a privilege. I also thank my cat Fluffy for keeping me smiling at the downfalls of my project.
When I joined the SILIS group, I was instantly made to feel welcome and included, for which I owe additional thanks to Bernhard, Gaurav, Enrico, Silvia and Gregory. I extend my thanks and best wishes to all the students and post-docs I got to share lunch, coffee and/or (several) pints with.
I would like to thank Kirsten Munro, Catherine Cheshire and Lynn Gilmour for their kind approach in dealing with all the administrative matters and ensuring my PhD ran smoothly.
I would like to thank Jane Patterson for her kindness and ensuring my PhD career ran smoothly.
I gratefully acknowledge the generosity and support of the Scottish Universities Physics Alliance (SUPA) and University of Strathclyde, who provided me with a Prize Studentship enabling me to undertake this PhD. I am eternally grateful for this opportunity.

"Signal is a physical quantity that vacillates with time, space or any other alienated variable. ..."

"Signal is a physical quantity that vacillates with time, space or any other alienated variable."

("Signals are physical quantities that change as a function of time, space, or some other independent variable.")

Spectrum conflict management,..., and the (thus far incomplete) Search for Extraterrestrial Intelligence (SETI) all alleviate on ferreting the propinquity of radio signals of concealed frequency, power, and modulation.

... all rely on detecting the presence of radio signals of unknown frequency, power, and modulation."

or

"escalates the rate at which sampled signals can purl through the processor"

"increases the rate at which sampled signals can flow through the processor".

(later two taken from intro of 2001 thesis at Monterey Naval Postgraduate School:
Charles T. Dorcey, "FFT-based spectrum analysis using a Digital Signal Processor.")

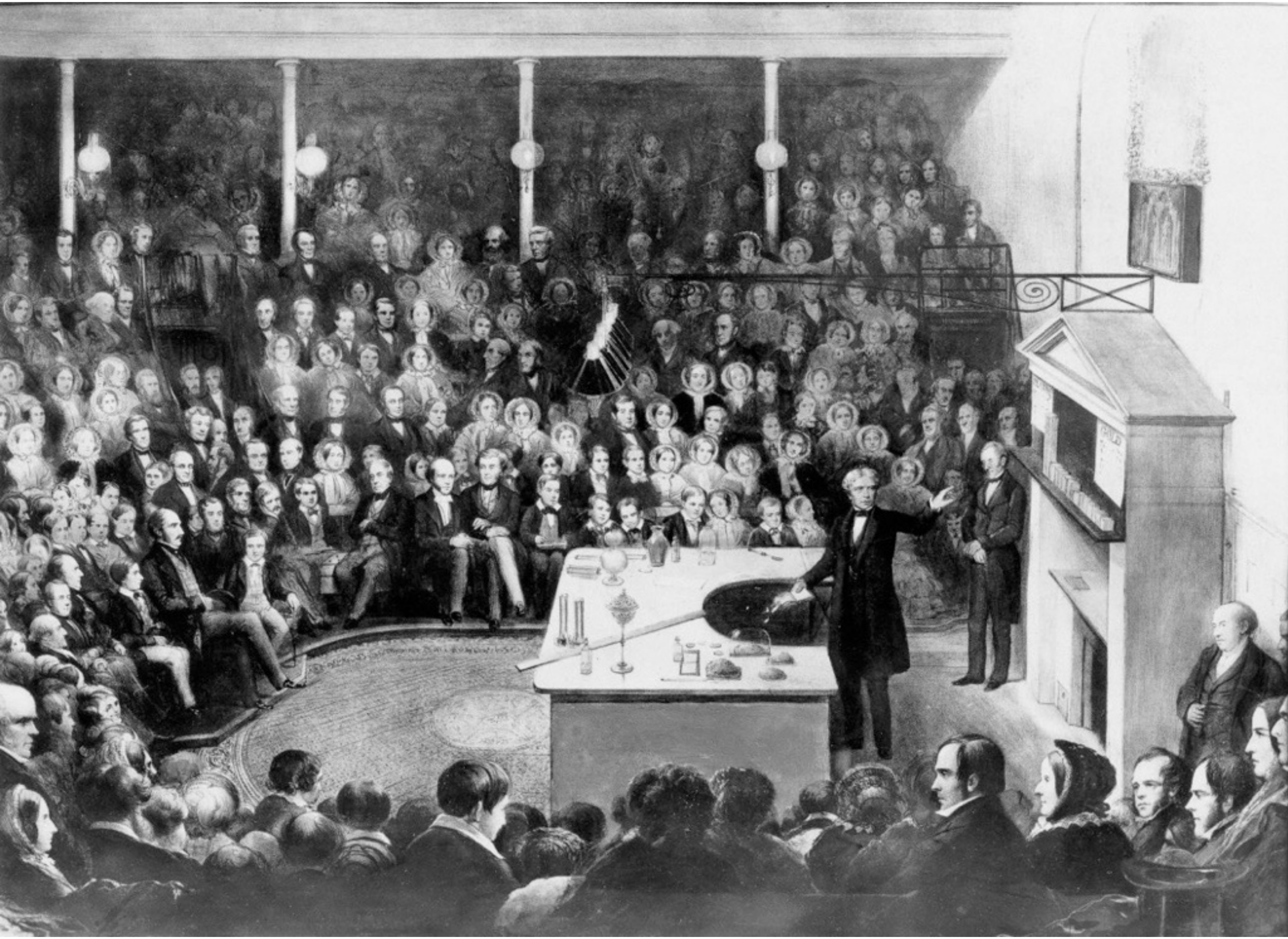
or from http://www.tutorialsweb.com/rf-measurements/spectrum-analyzer.htm :

Spectrum analyzer is a device used to [examine -> anatomize] the spectral composition of electric, acoustic or optical waveform [7]. It is a wideband and [very -> eminent] sensitive receiver. It works on the [principle -> ethic] of super heterodyne receiver which [converts -> transmogrifies] higher frequencies to measurable quantities. Received frequency spectrum is [slowly -> apathetically] swept through a range of preselected frequencies converting the selected frequency to a measurable and [displaying -> unveiling] on the CRT. These are [capable -> adept] in measuring the frequency response of power levels as low as -120dbm.

or finally:

“The proposed work in this thesis is having a lot of potential for further research in the area of [edge detection] using different paradigm making the work more versatile and flexible."

Enter the Public



General Relativity and Quantum Cosmology

A `warp drive' with more reasonable total energy requirements

Chris Van Den Broeck

(Submitted on 21 May 1999 (v1), last revised 21 Sep 1999 (this version, v5))

I show how a minor modification of the Alcubierre geometry can dramatically improve the total energy requirements for a `warp bubble' that can be used to transport macroscopic objects. A spacetime is presented for which the total negative mass needed is of the order of a few solar masses, accompanied by a comparable amount of positive energy. This puts the warp drive in the mass scale of large traversable wormholes. The new geometry satisfies the quantum inequality concerning WEC violations and has the same advantages as the original Alcubierre spacetime.

Comments:	9 pages, 1 figure; error in calculation corrected
Subjects:	**General Relativity and Quantum Cosmology (gr-qc)**
Journal reference:	Class.Quant.Grav. 16 (1999) 3973-3979
DOI:	10.1088/0264-9381/16/12/314
Report number:	KUL-TF-99/18
Cite as:	**arXiv:gr-qc/9905084**
	(or **arXiv:gr-qc/9905084v5** for this version)

Submission history

From: Chris Van Den Broeck [view email]
[v1] Fri, 21 May 1999 12:50:59 GMT (8kb)
[v2] Tue, 1 Jun 1999 14:25:26 GMT (9kb)
[v3] Wed, 16 Jun 1999 12:07:25 GMT (9kb)

Astrophysics

The Effects of Moore's Law and Slacking on Large Computations

C Gottbrath, J Bailin, C Meakin, T Thompson, J.J. Charfman

(Submitted on 9 Dec 1999)

We show that, in the context of Moore's Law, overall productivity can be increased for large enough computations by `slacking' or waiting for some period of time before purchasing a computer and beginning the calculation.

Subjects: **Astrophysics (astro-ph)**
Cite as: **arXiv:astro-ph/9912202**
(or **arXiv:astro-ph/9912202v1** for this version)

Submission history

From: Christopher Gottbrath [view email]
[v1] Thu, 9 Dec 1999 20:06:55 GMT (16kb)

Mathematics > Differential Geometry

The entropy formula for the Ricci flow and its geometric applications

Grisha Perelman

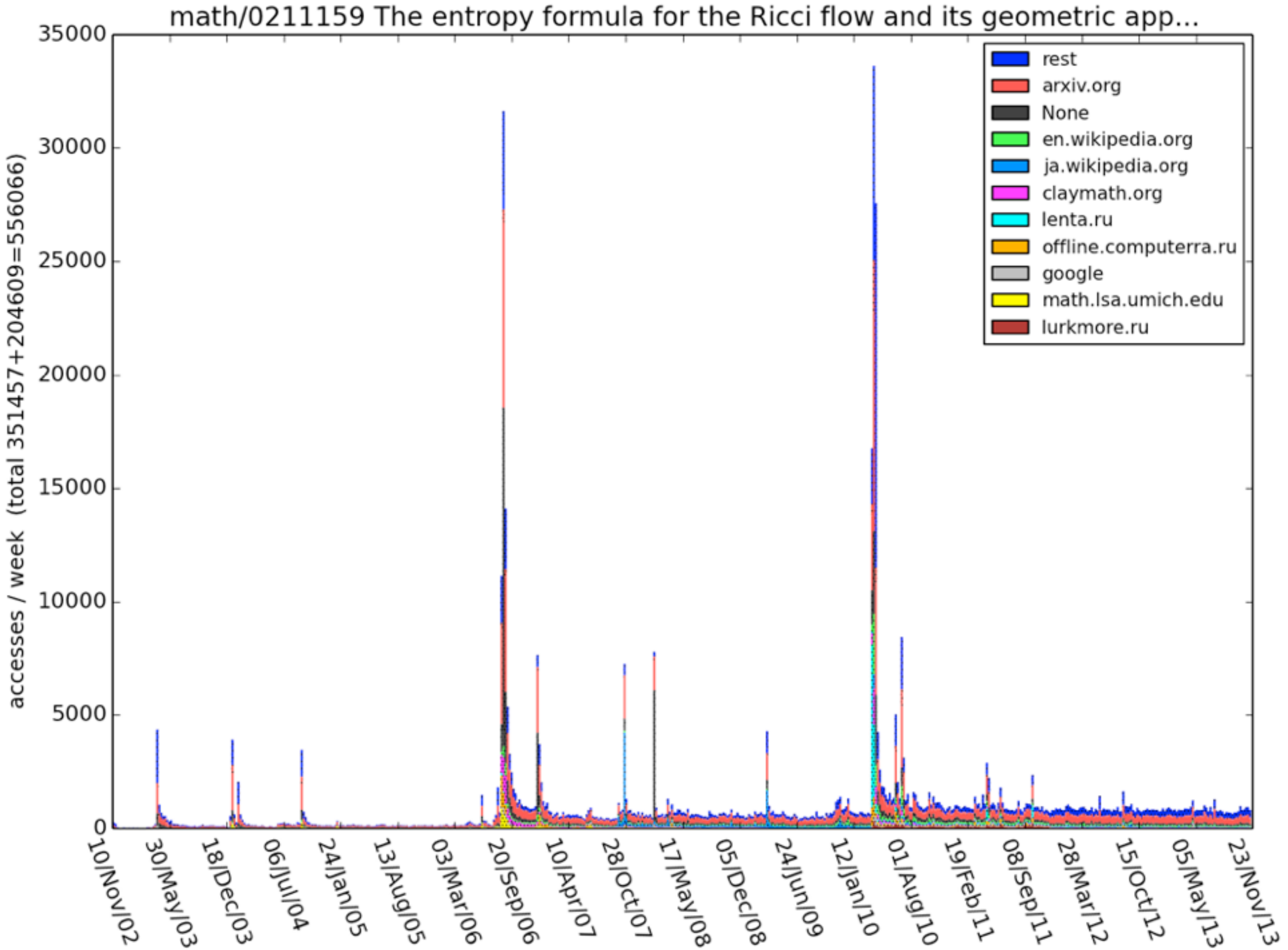
(Submitted on 11 Nov 2002)

We present a monotonic expression for the Ricci flow, valid in all dimensions and without curvature assumptions. It is interpreted as an entropy for a certain canonical ensemble. Several geometric applications are given. In particular, (1) Ricci flow, considered on the space of riemannian metrics modulo diffeomorphism and scaling, has no nontrivial periodic orbits (that is, other than fixed points); (2) In a region, where singularity is forming in finite time, the injectivity radius is controlled by the curvature; (3) Ricci flow can not quickly turn an almost euclidean region into a very curved one, no matter what happens far away. We also verify several assertions related to Richard Hamilton's program for the proof of Thurston geometrization conjecture for closed three-manifolds, and give a sketch of an eclectic proof of this conjecture, making use of earlier results on collapsing with local lower curvature bound.

Comments: 39 pages
Subjects: **Differential Geometry (math.DG)**
MSC classes: 53C
Cite as: **arXiv:math/0211159 [math.DG]**
(or **arXiv:math/0211159v1 [math.DG]** for this version)

Submission history

From: Grisha Perelman [view email]
[v1] Mon, 11 Nov 2002 16:11:49 GMT (33kb)


math/0211159 The entropy formula for the Ricci flow and its geometric app...
accesses / week (total 351457+204609=556066)
35000
30000
25000
20000
15000
10000
5000
0
rest
arxiv.org
None
en.wikipedia.org
ja.wikipedia.org
claymath.org
lenta.ru
offline.computerra.ru
google
math.lsa.umich.edu
lurkmore.ru
10/Nov/02
30/May/03
18/Dec/03
06/Jul/04
24/Jan/05
13/Aug/05
03/Mar/06
20/Sep/06
10/Apr/07
28/Oct/07
17/May/08
05/Dec/08
24/Jun/09
12/Jan/10
01/Aug/10
19/Feb/11
08/Sep/11
28/Mar/12
15/Oct/12
05/May/13
23/Nov/13

Meme (from Wikipedia)

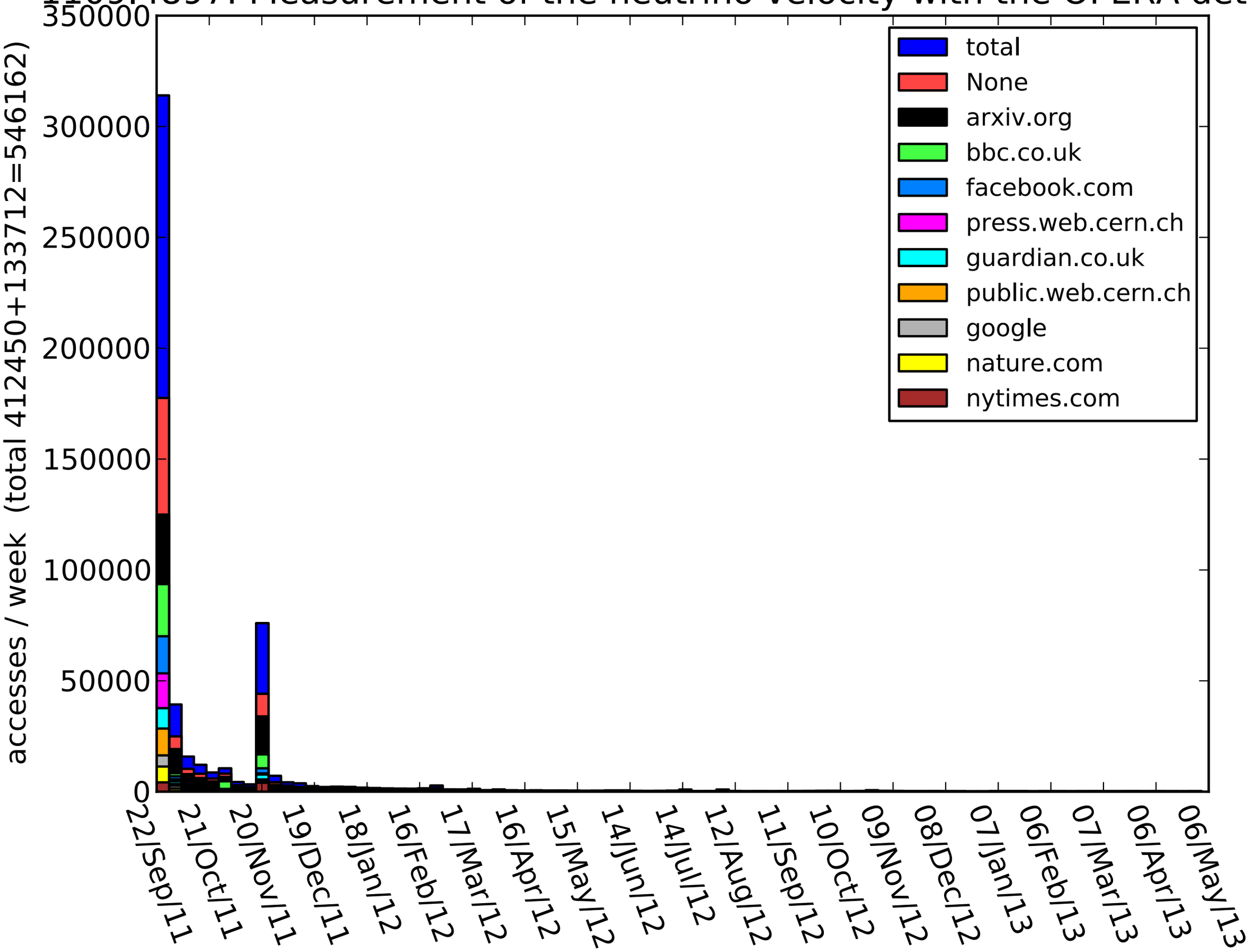
A meme is "an idea, behavior, or style that spreads from person to person within a culture." A meme acts as a unit for carrying cultural ideas, symbols, or practices that can be transmitted from one mind to another through writing, speech, gestures, rituals, or other imitable phenomena . . . cultural analogues to genes in that they self-replicate, mutate, and respond to selective pressures.

. . . coined by the British evolutionary biologist Richard Dawkins in The Selfish Gene (1976) as a concept for discussion of evolutionary principles in explaining the spread of ideas and cultural phenomena. Examples of memes given in the book included melodies, catch-phrases, fashion, and the technology of building arches.

. . .may evolve by natural selection analogous to biological evolution . . . through variation, mutation, competition, and inheritance. . . A field of study called memetics arose in the 1990s to explore the concepts and transmission of memes in terms of an evolutionary model.)

Example: Einstein is wrong

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



High Energy Physics - Experiment

Measurement of the neutrino velocity with the OPERA detector in the CNGS beam

The OPERA Collaboration: T. Adam, N. Agafonova, A. Aleksandrov, O. Altinok, P. Alvarez Sanchez, A. Anokhina, S. Aoki, A. Ariga, T. Ariga, D. Autiero, A. Badertscher, A. Ben Dhahbi, A. Bertolin, C. Bozza, T. Brugiere, R. Brugnera, F. Brunet, G. Brunetti, S. Buontempo, B. Carlus, F. Cavanna, A. Cazes, L. Chaussard, M. Chernyavsky, V. Chiarella, A. Chukanov, G. Colosimo, M. Crespi, N. D'Ambrosio, G. De Lellis, M. De Serio, Y. Declais, P. del Amo Sanchez, F. Di Capua, A. Di Crescenzo, D. Di Ferdinando, N. Di Marco, S. Dmitrievsky, M. Dracos, D. Duchesneau, S. Dusini, T. Dzhatdoev, J. Ebert, I. Efthymiopoulos, O. Egorov, A. Ereditato, L. S. Esposito, J. Favier, T. Ferber, R. A. Fini, T. Fukuda, A. Garfagnini, G. Giacomelli, M. Giorgini, M. Giovannozzi, C. Girerd, J. Goldberg, C. Gollnitz, et al. (132 additional authors not shown)

(Submitted on 22 Sep 2011 (v1), last revised 12 Jul 2012 (this version, v4))

The OPERA neutrino experiment at the underground Gran Sasso Laboratory has measured the velocity of neutrinos from the CERN CNGS beam over a baseline of about 730 km. The measurement is based on data taken by OPERA in the years 2009, 2010 and 2011. Dedicated upgrades of the CNGS timing system and of the OPERA detector, as well as a high precision geodesy campaign for the measurement of the neutrino baseline, allowed reaching comparable systematic and statistical accuracies. An arrival time of CNGS muon neutrinos with respect to the one computed assuming the speed of light in vacuum of (6.5 +/- 7.4(stat.)((+8.3)(-8.0)sys.))ns was measured corresponding

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)

more memes

- star trek (warp drive, tractor beam, teleportation)
- harry potter (invisibility cloak)
- star wars (tatooine)

COMPUTERWORLD

SECURITY IS SEXY
By Darlene Storm | Follow

NEWS ANALYSIS

Physics researchers map where to run and hide during a zombie apocalypse

Credit: Steve Baker

Cornell University researchers presented "You Can Run, You Can Hide: The Epidemiology and Statistical Mechanics of Zombies" and created a zombie susceptibility map, Zombie-town USA, which simulates a zombie infestation based on how diseases spread in real

les inRocKs actu musiques concerts cinémas séries médias livres arts scène
politique • société • monde • médias • blogs • buzzodrome •

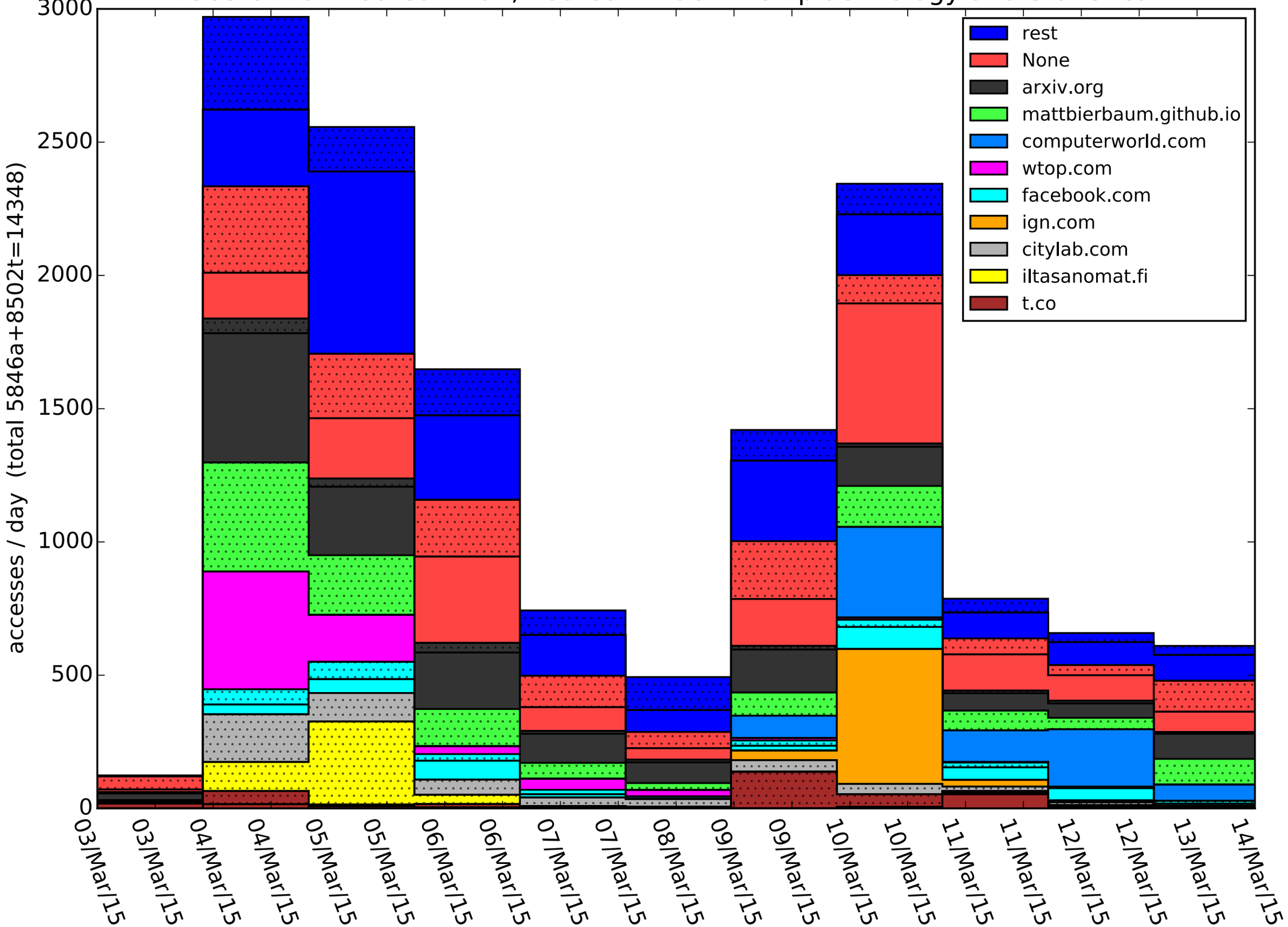
Quel est le meilleur endroit pour se cacher en cas d'invasion de zombies ?

05/03/2015 | 13h42

The Walking Dead. Source: Allociné

Une étude mathématique très sérieuse intitulée "The Statistical Mechanics of Zombies" et présentée aujourd'hui à l'American Physical Society par un groupe d'universitaires s'est posée la question de l'avenir des Etats-Unis en cas d'attaque de zombies. Etude de cas.

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



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)

From *The Atlantic* CITYLAB

NAVIGATOR | CITYFIXER | MAPS | PHOTOS

COMMUTE WORK HOUSING WEATHER

Scientists Agree: In Case of Zombie Outbreak, Leave the City

Researchers at Cornell University modeled what would actually happen if zombies attacked. Spoiler: The news is not good for city-dwellers.

AARIAN MARSHALL | @AarianMarshall | Mar 3, 2015 | 6 Comments

2.1k Shares | Share on Facebook | Tweet

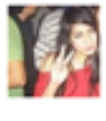
S. Kuelcue / Shutterstock.com

BuzzFeed

Scientists Figured Out What Would Really Happen During A Zombie Outbreak

Here's how the infection would spread in the United States. New York City is so screwed.

posted on March 9, 2015, at 12:54 p.m.



Natasha Umer
BuzzFeed Staff

A team of Cornell researchers figured out how fast a zombie outbreak would spread across the United States. You know, just in case...

9 MAR 2015

SCIENTISTS DETERMINE SAFEST PLACES IN U.S. TO SURVIVE A ZOMBIE APOCALYPSE

349 How fast until the zombies hit your city?

BY RACHEL HAAS→ A group of Cornell University researchers have conducted a study to determine the safest places in the U.S. to hide out during a zombie outbreak.

Nyt tätäkin on tutkittu: Näin nopeasti zombit valtaisivat Amerikan

Kuinka ihmiskunta pärjäisi zombien invaasiota vastaan? Ei kovin hyvin, väittävät Cornellin yliopiston jatko-opiskelijat. Kuvituskuva. *Colourbox*

Julkaistu: 5.3.2015 0:57

Yhdysvaltalaistutkijat halusivat tietää, kuinka elokuvista tutussa zombihyökkäyksessä kävisi oikeasti.

Texasin San Antoniossa pidetyssä tutkijatilaisuudessa esiteltiin keskiviikkona varsin erikoinen tutkimus. *You Can Run, You Can Hide: The Epidemiology and Statistical Mechanics of Zombies* -artikkeli on teoreettisen fysiikan jatko-opiskelijoiden luomus, jossa nämä spekuloivat zombihyökkäyksen etenemistä manner-Yhdysvalloissa.

In the news

- Condensed Matter (new technologies, ...)
- Games (sudoku, rubik's, nintendo,)
- popular physics (slinky, ...)
- quantum weirdness
- physics and society
- math (famous theorems)
- astro ("goldilocks" planets)

In the news, cont'd

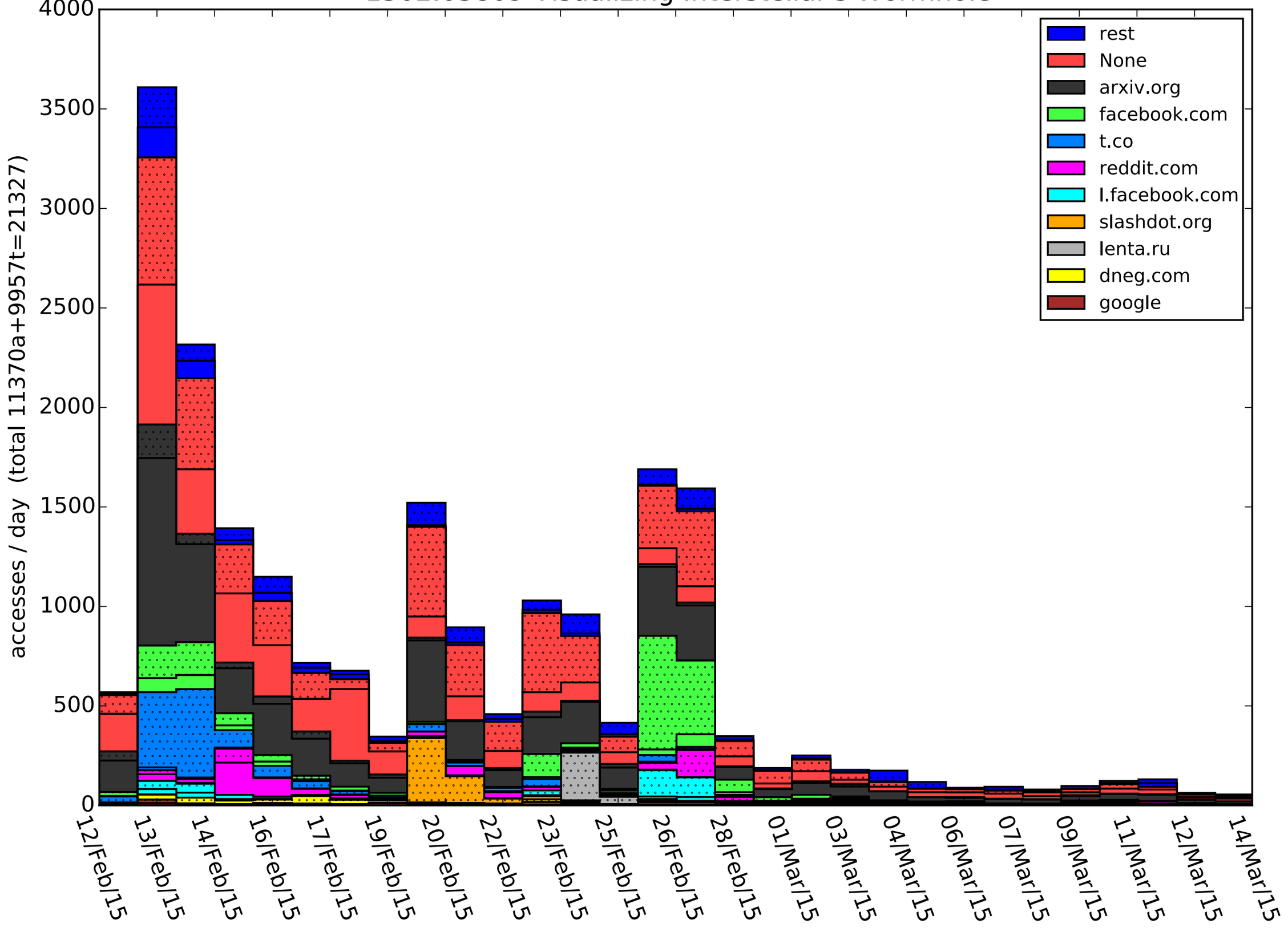
- earth and planetary (comets, meteors, moon)
- cosmology (wmap, cc, ...)
- HEP Exp (Higgs, new particles)
- HEP Th+Ph (theoretical developments)
- Computer Sci (networks, games)
- April Fools'
- Neo-Einstein / The Fringe

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**

a few from ’15 so far

1502.03809 Visualizing Interstellar's Wormhole



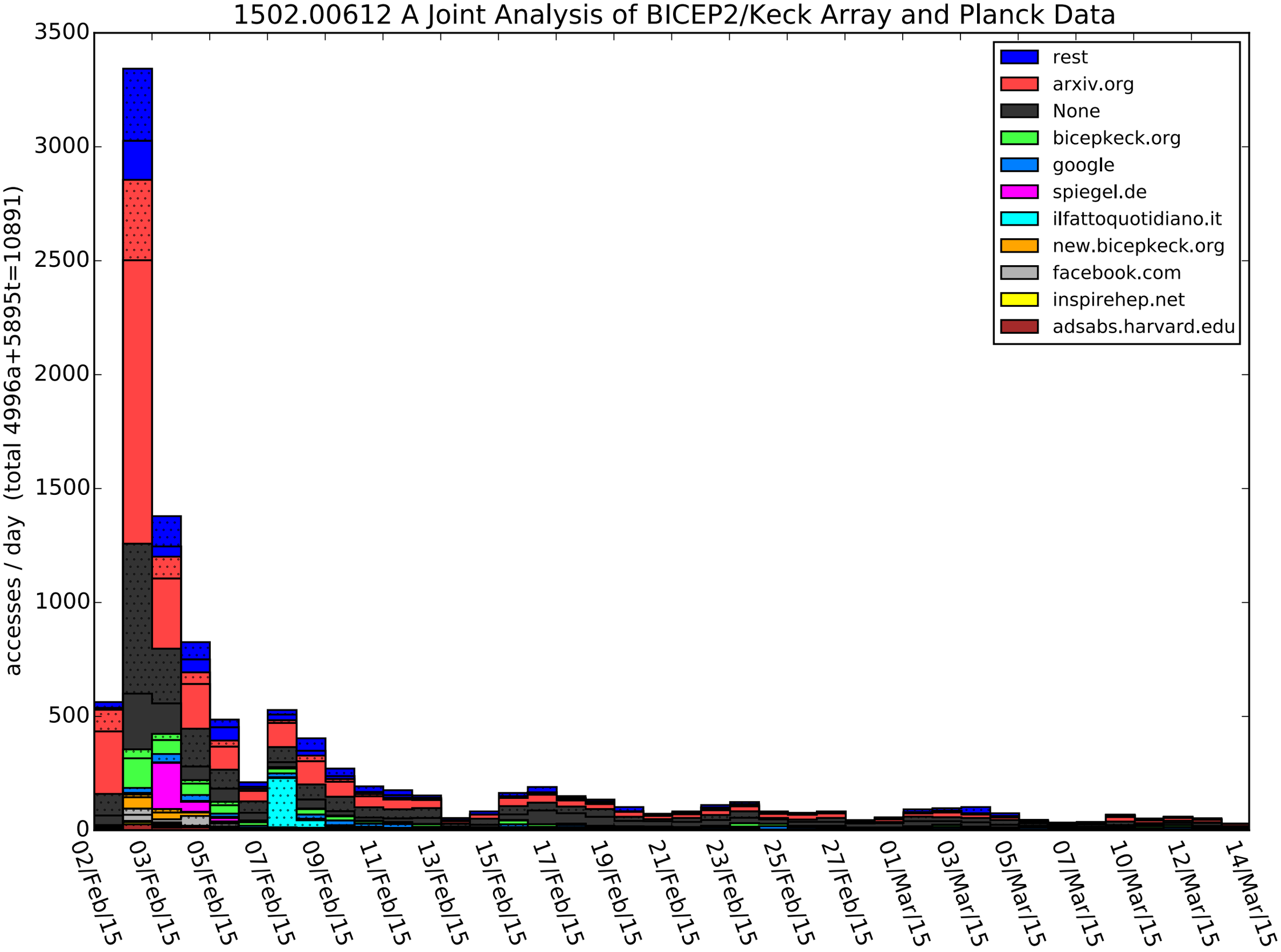
General Relativity and Quantum Cosmology

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.


1502.00612 A Joint Analysis of BICEP2/Keck Array and Planck Data
accesses / day (total 4996a+5895t=10891)
3500
3000
2500
2000
1500
1000
500
0
02/Feb/15
03/Feb/15
05/Feb/15
07/Feb/15
09/Feb/15
11/Feb/15
13/Feb/15
15/Feb/15
17/Feb/15
19/Feb/15
21/Feb/15
23/Feb/15
25/Feb/15
27/Feb/15
01/Mar/15
03/Mar/15
05/Mar/15
07/Mar/15
10/Mar/15
12/Mar/15
14/Mar/15
rest
arxiv.org
None
bicepkeck.org
google
spiegel.de
ilfattoquotidiano.it
new.bicepkeck.org
facebook.com
inspirehep.net
adsabs.harvard.edu

A Joint Analysis of BICEP2/Keck Array and Planck Data

BICEP2/Keck, Planck Collaborations: P. A. R. Ade, N. Aghanim, Z. Ahmed, R. W. Aikin, K. D. Alexander, M. Arnaud, J. Aumont, C. Baccigalupi, A. J. Banday, D. Barkats, R. B. Barreiro, J. G. Bartlett, N. Bartolo, E. Battaner, K. Benabed, A. Benoit-Lévy, S. J. Benton, J.-P. Bernard, M. Bersanelli, P. Bielewicz, C. A. Bischoff, J. J. Bock, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, J. A. Brevik, M. Bucher, I. Buder, E. Bullock, C. Burigana, R. C. Butler, V. Buza, E. Calabrese, J.-F. Cardoso, A. Catalano, A. Challinor, R.-R. Chary, H. C. Chiang, P. R. Christensen, L. P. L. Colombo, C. Combet, J. Connors, F. Couchot, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, et al. (216 additional authors not shown)

(Submitted on 2 Feb 2015)

We report the results of a joint analysis of data from BICEP2/Keck Array and Planck. BICEP2 and Keck Array have observed the same approximately 400 deg^2 patch of sky centered on RA 0h, Dec. -57.5 deg. The combined maps reach a depth of 57 nK deg in Stokes Q and U in a band centered at 150 GHz. Planck has observed the full sky in polarization at seven frequencies from 30 to 353 GHz, but much less deeply in any given region (1.2 μK deg in Q and U at 143 GHz). We detect 150×353 cross-correlation in B-modes at high significance. We fit the single- and cross-frequency power spectra at frequencies above 150 GHz to a lensed-ΛCDM model that includes dust and a possible contribution from inflationary gravitational waves (as parameterized by the tensor-to-scalar ratio r). We probe various model variations and extensions, including adding a synchrotron component in combination with lower frequency data, and find that these make little difference to the r constraint. Finally we present an alternative analysis which is similar to a map-based cleaning of the dust contribution, and show that this gives similar constraints. The final result is expressed as a likelihood curve for r, and yields an upper limit $r_{0.05} < 0.12$ at 95% confidence. Marginalizing over dust and r, lensing B-modes are detected at $7.0\,\sigma$ significance.

nature International weekly journal of science

NATURE | TRENDWATCH

Trail of dust and gravitational waves tracked in arXiv papers

Manuscripts posted to preprint website tell a tale of increasing scepticism.

Richard Van Noorden

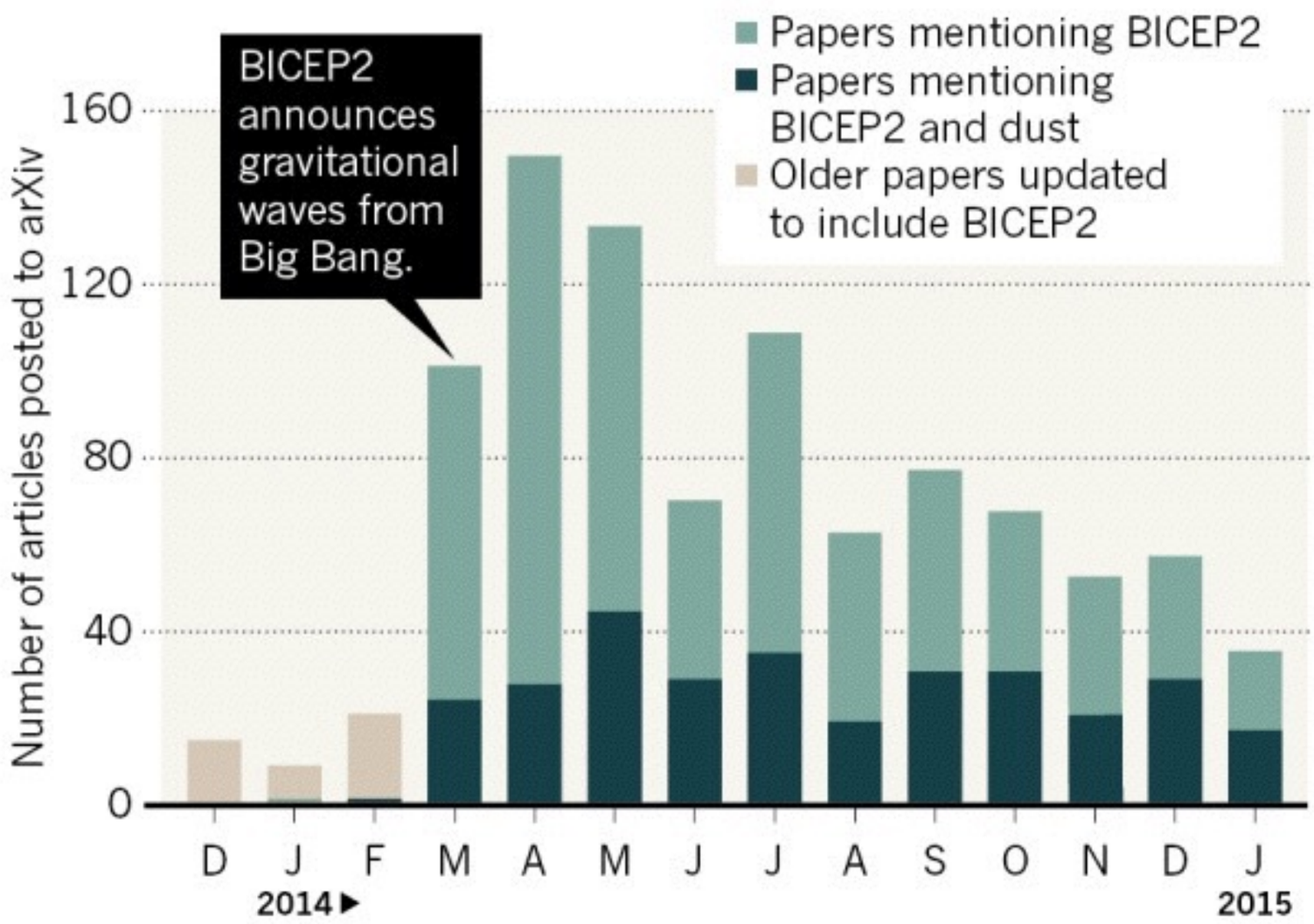
04 February 2015

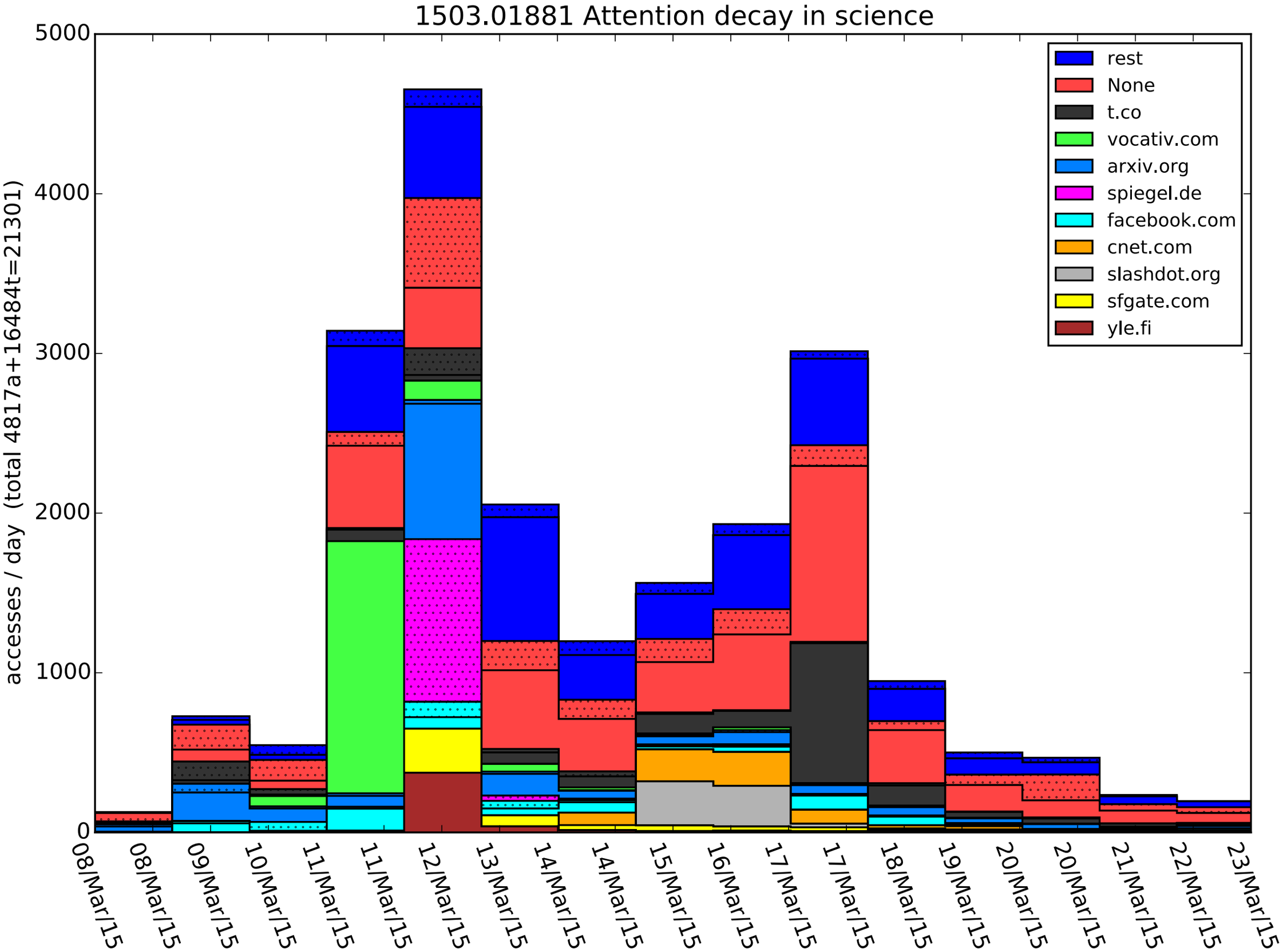
The trail of manuscripts posted to the preprint server arXiv s
embraced, then doubted and gradually lost interest in one of
announcements: the discovery of gravitational waves genera

Soon after the announcement last March, papers questionin
final nail in the coffin came last week, when researchers offi
Way was responsible for the signal seen by the South Pole-

THE TRAIL OF DUST

Interest waned in the BICEP2 experiment as it became clearer that dust had been mistaken for a signal of gravitational waves.




1503.01881 Attention decay in science
accesses / day (total 4817a+16484t=21301)
5000
4000
3000
2000
1000
0
rest
None
t.co
vocativ.com
arxiv.org
spiegel.de
facebook.com
cnet.com
slashdot.org
sfgate.com
yle.fi
08/Mar/15
08/Mar/15
09/Mar/15
10/Mar/15
11/Mar/15
11/Mar/15
12/Mar/15
13/Mar/15
14/Mar/15
14/Mar/15
15/Mar/15
16/Mar/15
17/Mar/15
17/Mar/15
18/Mar/15
19/Mar/15
20/Mar/15
20/Mar/15
21/Mar/15
22/Mar/15
23/Mar/15