Title: Next-Generation Implications of Open Access

Date: Sep 09, 2008 11:00 AM

URL: http://pirsa.org/07090080

Abstract: True open access to scientific publications not only gives readers the possibility to read articles without paying subscription, but also makes the material available for automated ingestion and harvesting by 3rd parties. Once articles and associated data become universally treatable as computable objects, openly available to 3rd party aggregators and value-added services, what new services can we expect, and how will they change the way that researchers interact with their scholarly communications infrastructure? I will discuss straightforward applications of existing ideas and services, including citation analysis, collaborative filtering, external database linkages, interoperability, and other forms of automated markup, and speculate on the sociology of the next generation of users.

Next-Generation Implications of Open Access

Paul Ginsparg

CIS and Physics, Cornell University

Abstract: True open access to scientific publications not only gives readers the possibility to read articles without paying subscription, but also makes the material available for automated ingestion and harvesting by 3rd parties. Once articles and associated data become universally treatable as computable objects, openly available to 3rd party aggregators and value-added services, what new services can we expect, and how will they change the way that researchers interact with their scholarly communications infrastructure? I will discuss straightforward applications of existing ideas and services, including citation analysis, collaborative filtering, external database linkages, interoperability, and other forms of automated markup, and speculate on the sociology of the next generation of users.

Next-Generation Implications of Open Access

Paul Ginsparg

CIS and Physics, Cornell University

Abstract: True open access to scientific publications not only gives readers the possibility to read articles without paying subscription, but also makes the material available for automated ingestion and harvesting by 3rd parties. Once articles and associated data become universally treatable as computable objects, openly available to 3rd party aggregators and value-added services, what new services can we expect, and how will they change the way that researchers interact with their scholarly communications infrastructure? I will discuss straightforward applications of existing ideas and services, including citation analysis, collaborative filtering, external database linkages, interoperability, and other forms of automated markup, and speculate on the sociology of the next generation of users.

Next-Generation Implications of Open Access

Paul Ginsparg
CIS and Physics, Cornell University

Abstract: True open access to scientific publications not only gives readers the possibility to read articles without paying subscription, but also makes the material available for automated ingestion and harvesting by 3rd parties. Once articles and associated data become universally treatable as computable objects, openly available to 3rd party aggregators and value-added services, what new services can we expect, and how will they change the way that researchers interact with their scholarly communications infrastructure? I will discuss straightforward applications of existing ideas and services, including citation analysis, collaborative filtering, external database linkages, interoperability, and other forms of automated markup, and speculate on the sociology of the next generation of users.

Fantasy

Reflect for a moment

Current practice:

- **free access articles, background material from authors, slide presentations, video, related software, on-line animations, blog discussions, 3rd party notes**
- **course websites, e-mail, course blogs, wiki for notes**

New expectations (harvest all related, activity maps, concept browse).

Collapse internet resources to subset of unique ideas, authenticated.

Marketplace for preresearch barter of tools, resources, capabilities.

Authoring tools.

New generation of users.

Next-Generation Implications of Open Access

Paul Ginsparg

CIS and Physics, Cornell University

Abstract: True open access to scientific publications not only gives readers the possibility to read articles without paying subscription, but also makes the material available for automated ingestion and harvesting by 3rd parties. Once articles and associated data become universally treatable as computable objects, openly available to 3rd party aggregators and value-added services, what new services can we expect, and how will they change the way that researchers interact with their scholarly communications infrastructure? I will discuss straightforward applications of existing ideas and services, including citation analysis, collaborative filtering, external database linkages, interoperability, and other forms of automated markup, and speculate on the sociology of the next generation of users.

Fantasy

Reflect for a moment

Current practice:

- **free access articles, background material from authors, slide presentations, video, related software, on-line animations, blog discussions, 3rd party notes**
- **course websites, e-mail, course blogs, wiki for notes**

New expectations (harvest all related, activity maps, concept browse).

Collapse internet resources to subset of unique ideas, authenticated.

Marketplace for preresearch barter of tools, resources, capabilities.

Authoring tools.

New generation of users.

Fantasy

Reflect for a moment

Current practice:

- free access articles, background material from authors, slide presentations, video, related software, on-line animations, blog discussions, 3rd party notes
- course websites, e-mail, course blogs, wiki for notes

New expectations (harvest all related, activity maps, concept browse).

Collapse internet resources to subset of unique ideas, authenticated.

Marketplace for preresearch barter of tools, resources, capabilities.

Authoring tools.

New generation of users.

Collective Action

Large groups work collectively to synthesize knowledge (feedback mechanisms, Wikipedia, open source software, social-networking sites, bloggers feed news coverage, ...)

Frameworks for analyzing online information systems still based on 1990s models for studying collections of hyperlinked documents. Need graph and network algorithms, data mining, natural-language processing, machine learning, probabilistic modeling, human-computer interaction, algorithmic game theory (individuals seek out advantageous positions within the network).

Maps make social network visible to the group itself: feedback loops increase participant awareness of the state of relationships among participants as a whole to increase participation.

Temporal dynamics of aggregate user populations (news cycle, flash crowds, ...) follow quantiable laws, leads to design principles.

Fantasy

Reflect for a moment

Current practice:

- free access articles, background material from authors, slide presentations, video, related software, on-line animations, blog discussions, 3rd party notes
- course websites, e-mail, course blogs, wiki for notes

New expectations (harvest all related, activity maps, concept browse).

Collapse internet resources to subset of unique ideas, authenticated.

Marketplace for preresearch barter of tools, resources, capabilities.

Authoring tools.

New generation of users.

Fantasy

Reflect for a moment

Current practice:

- free access articles, background material from authors, slide presentations, video, related software, on-line animations, blog discussions, 3rd party notes
- course websites, e-mail, course blogs, wiki for notes

New expectations (harvest all related, activity maps, concept browse).

Collapse internet resources to subset of unique ideas, authenticated.

Marketplace for preresearch barter of tools, resources, capabilities.

Authoring tools.

New generation of users.

Collective Action

Large groups work collectively to synthesize knowledge (feedback mechanisms, Wikipedia, open source software, social-networking sites, bloggers feed news coverage, ...)

Frameworks for analyzing online information systems still based on 1990s models for studying collections of hyperlinked documents. Need graph and network algorithms, data mining, natural-language processing, machine learning, probabilistic modeling, human-computer interaction, algorithmic game theory (individuals seek out advantageous positions within the network).

Maps make social network visible to the group itself: feedback loops increase participant awareness of the state of relationships among participants as a whole to increase participation.

Temporal dynamics of aggregate user populations (news cycle, flash crowds, ...) follow quantiable laws, leads to design principles.

NIH Public Access Policy Becomes Mandate in 2007

http://info-libraries.mit.edu/scholarly/open-access-initiatives/

On December 26, 2007, President Bush signed a spending bill that requires the US National Institutes of Health (NIH) to mandate open online access to all research it funds.

This is the first mandate for a major public funding agency in the US that requires research to be openly available; it changes the 2005 NIH Public Access Policy, which requested, but did not require, open access to NIH-funded research.

The new language stipulates that investigators funded by the NIH submit their peer-reviewed manuscripts to the National Library of Medicines open access repository PubMed Central when the manuscript is accepted for publication [additional > 70k/year]. The manuscript would then become openly available via PubMed Central within 12 months of publication in a journal. The policy will be implemented "in a manner consistent with copyright law."

Harvard To Collect, Disseminate Scholarly Articles For Faculty

http://www.fas.harvard.edu/home/news_and_events/releases/scholarly_02122008.html

Cambridge, Mass. - February 12, 2008 - In a move to disseminate faculty research and scholarship more broadly, the Harvard University Faculty of Arts and Sciences voted today to give the University a worldwide license to make each faculty member's scholarly articles available and to exercise the copyright in the articles, provided that the articles are not sold for a profit.

In proposing the legislation, Professor Stuart M. Shieber said, ". . . scholarly journals have historically allowed scholars to distribute their research to audiences around the world. But, the scholarly publishing system has become far more restrictive than it need be. Many publishers will not even allow scholars to use and distribute their own work. And, the cost of journals has risen to such astronomical levels that many institutions and individuals have cancelled subscriptions, further reducing the circulation of scholars' works."

Why necessary?

Congress Passes Law Requiring Users to Post to Youtube, Flickr, ...

arXiv.org

- e-mail interface started August 1991
  - download data available from start
  - WWW usage logs starting from 1993
- 456,000 full text documents (with full graphics), as of end 2007
  - physics, mathematics, computer science, non-linear science
  - growing at 60,000 new submissions per year (est. 2008 $\Rightarrow$ $>$ 516,000 at end of year)
  - 20 references per article (over 9 million total)
- over 50 million full text downloads during calendar year '07
  - over 600 downloads per article from '96-'07 ($>$250M total)
- overall: 12.2k ingested links to 8k articles (1.6% of 500k)

  '08 (so far): 2.1k ingested links to 1.5k articles (4% of 40k)
- Now managed by CU library (starting roughly 2001)

arXiv.org

- e-mail interface started August 1991
  - •download data available from start
  - •WWW usage logs starting from 1993
- 456,000 full text documents (with full graphics), as of end 2007
  - •physics, mathematics, computer science, non-linear science
  - •growing at 60,000 new submissions per year (est. 2008 $\Rightarrow$ $>$ 516,000 at end of year)
  - •20 references per article (over 9 million total)
- over 50 million full text downloads during calendar year '07
  - •over 600 downloads per article from '96-'07 ($>$250M total)
- overall: 12.2k ingested links to 8k articles (1.6% of 500k)

  '08 (so far): 2.1k ingested links to 1.5k articles (4% of 40k)
- Now managed by CU library (starting roughly 2001)

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

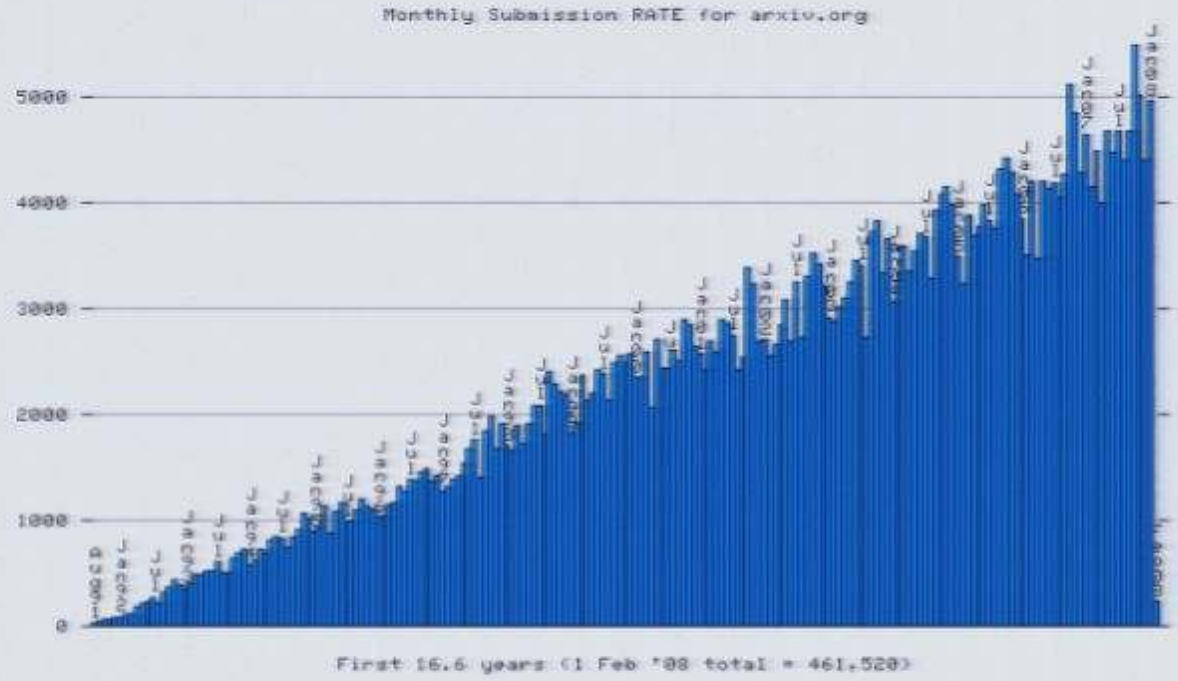
Cornell University Library

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).

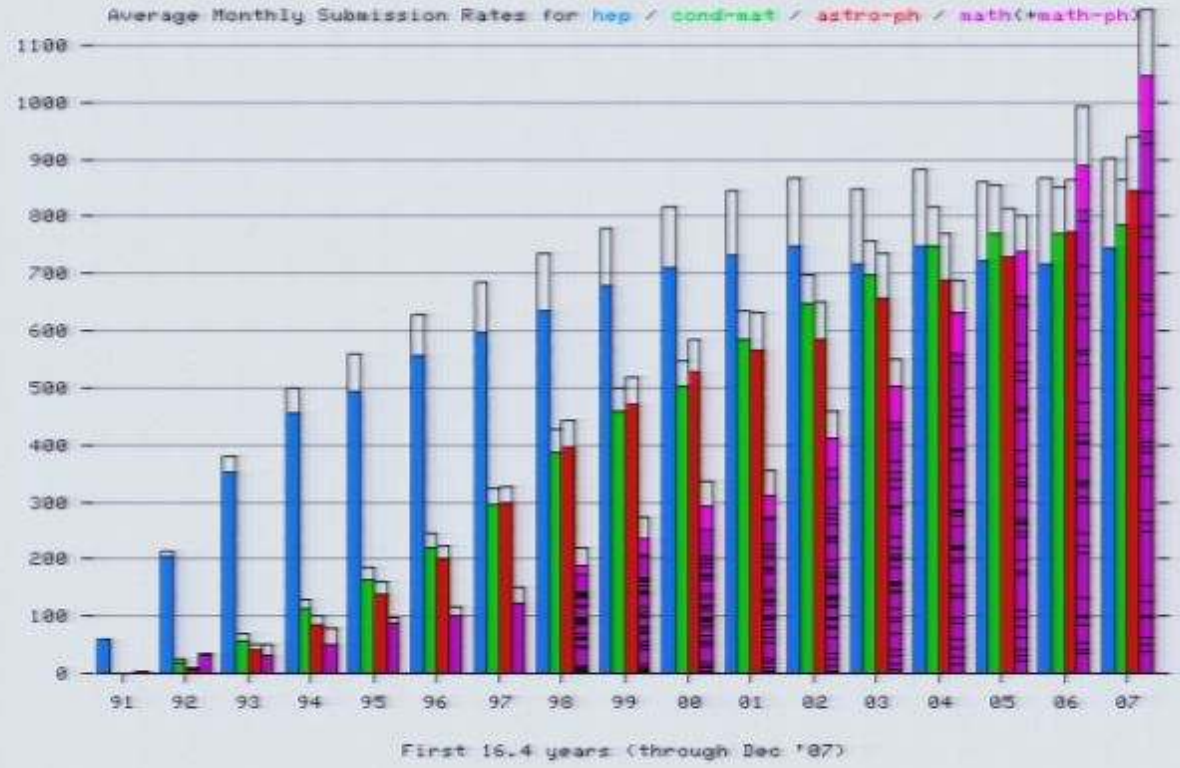
www-admin@arxiv.org

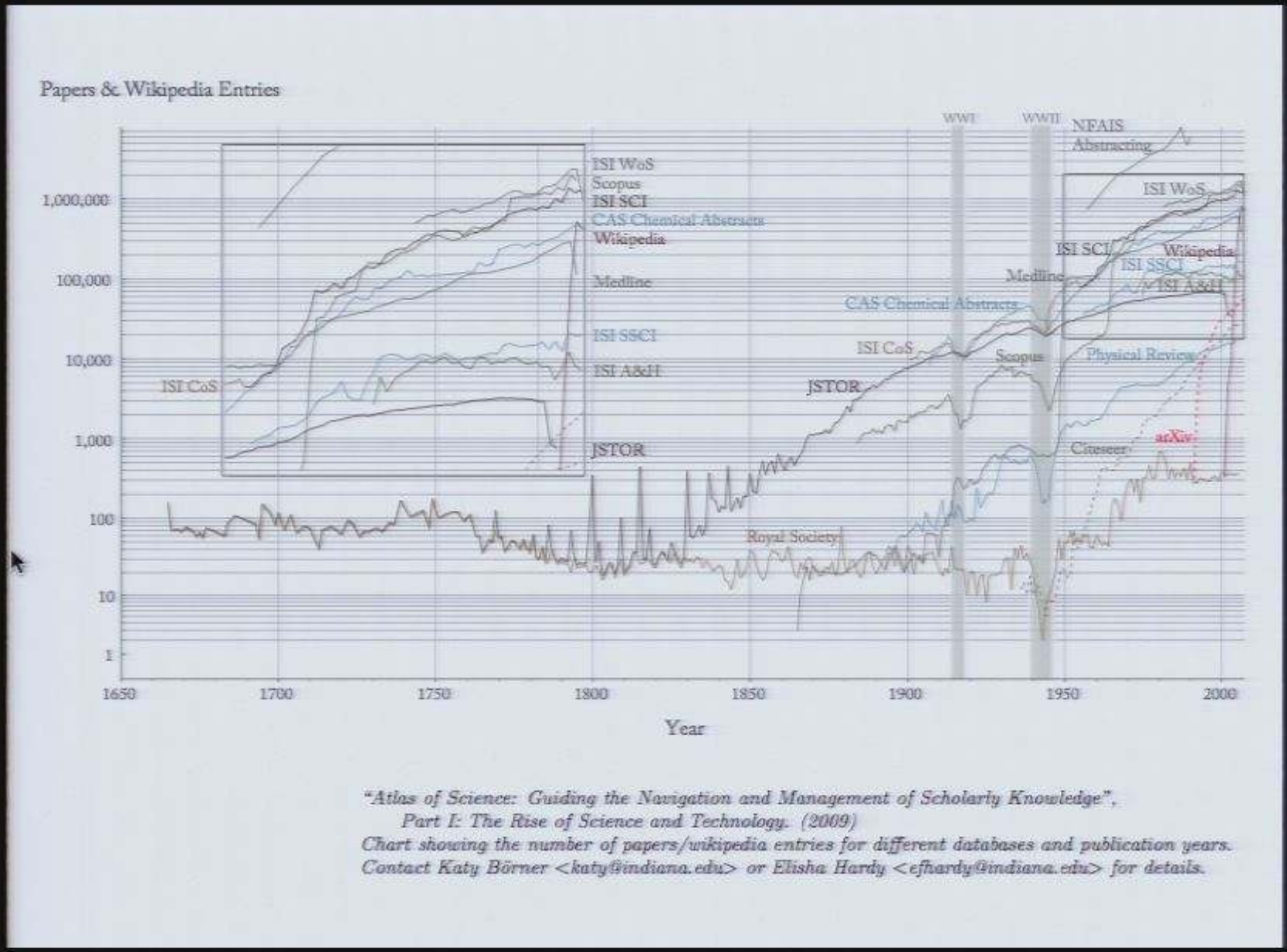
Submissions per month, '91 – '08



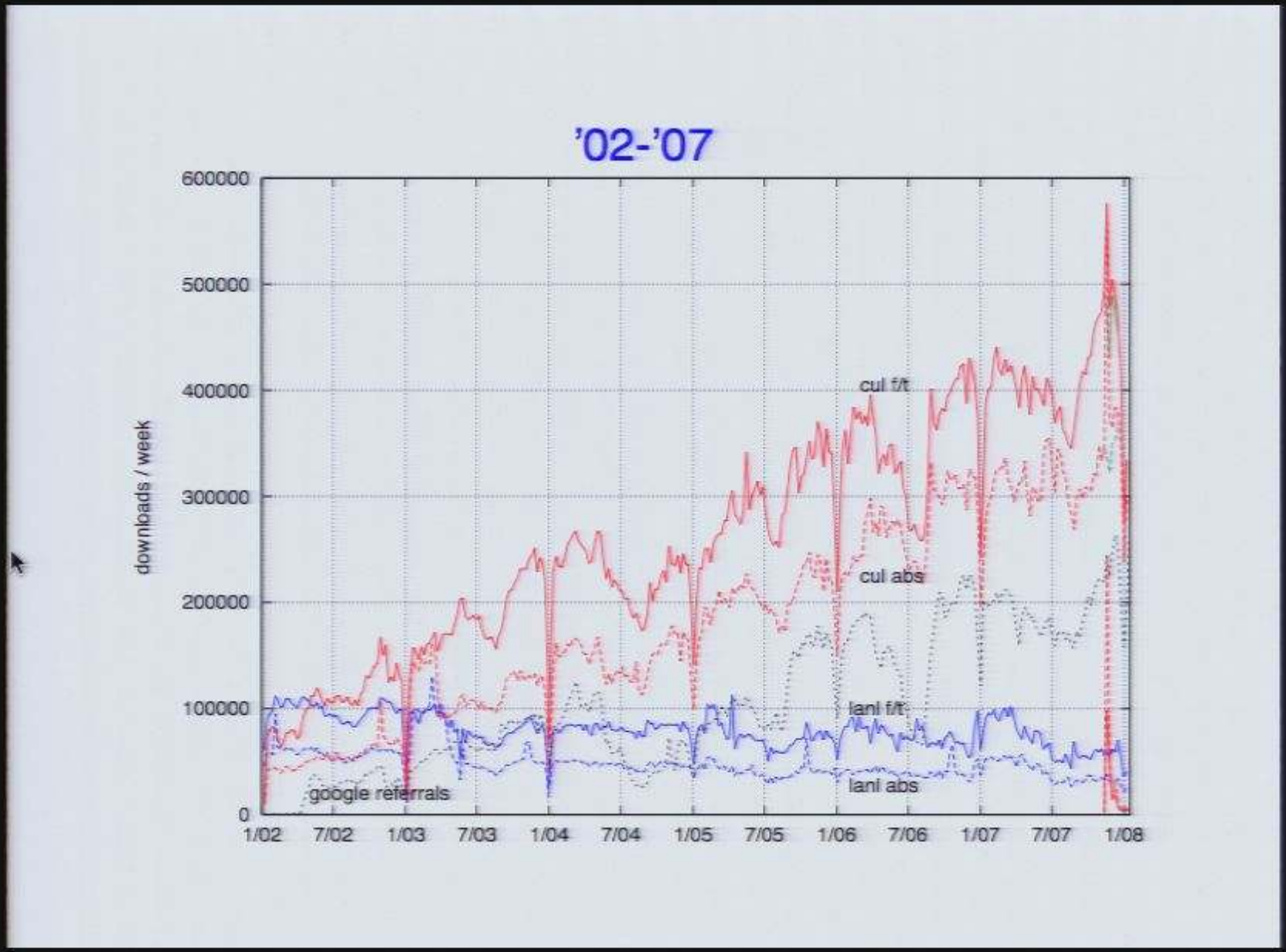
Total ≈ 495,000 (Aug 2008)

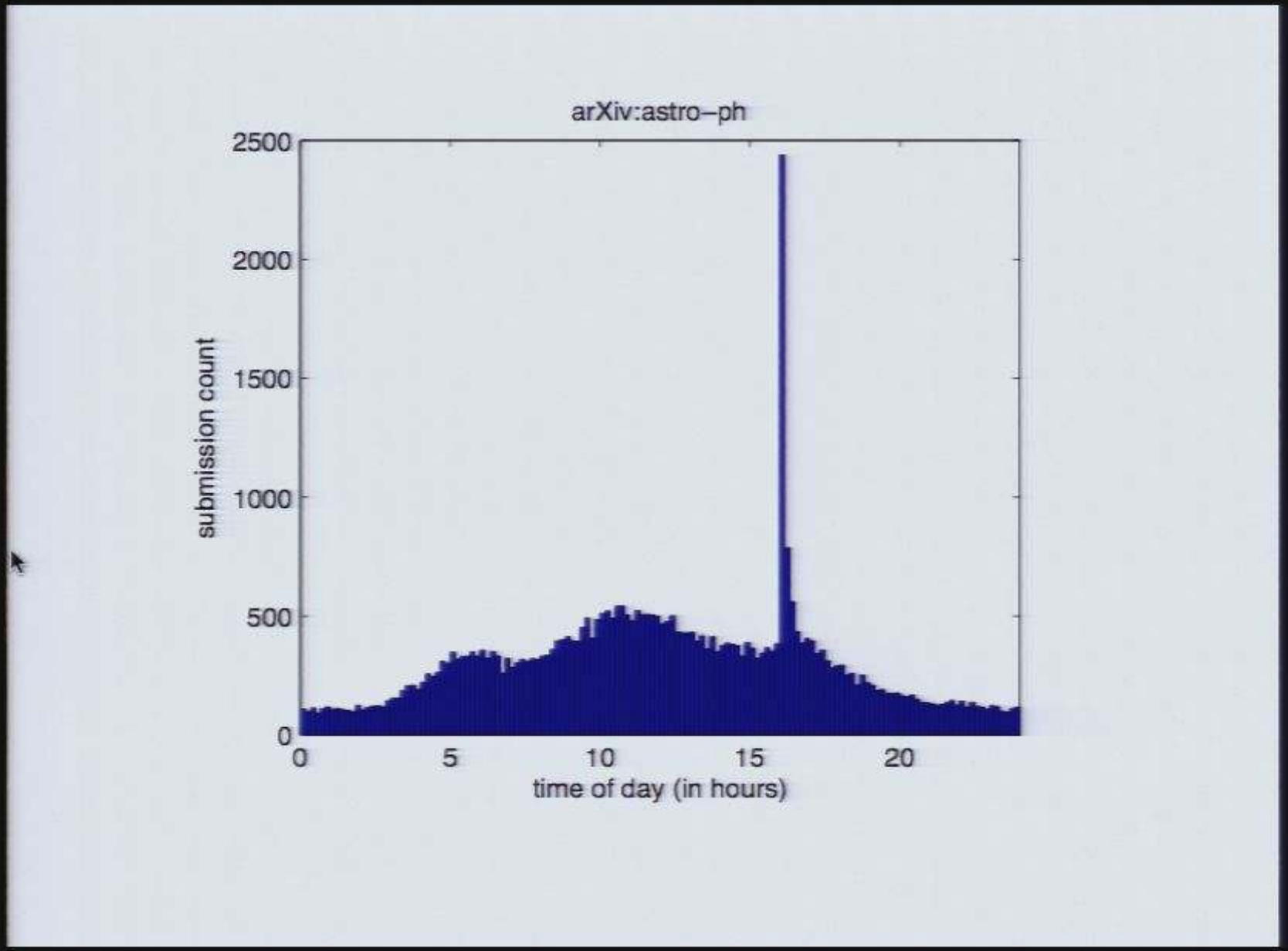
Top four subject areas





"Atlas of Science: Guiding the Navigation and Management of Scholarly Knowledge",
Part I: The Rise of Science and Technology. (2009)
Chart showing the number of papers/wikipedia entries for different databases and publication years.
Contact Katy Börner <katy@indiana.edu> or Elisha Hardy <efhardy@indiana.edu> for details.


'02-'07
600000
500000
400000
300000
200000
100000
0
downloads / week
cul f/t
cul abs
lanl f/t
lanl abs
google referrals
1/02
7/02
1/03
7/03
1/04
7/04
1/05
7/05
1/06
7/06
1/07
7/07
1/08


arXiv:astro–ph
2500
2000
1500
1000
500
0
submission count
0
5
10
15
20
time of day (in hours)

Game Theory

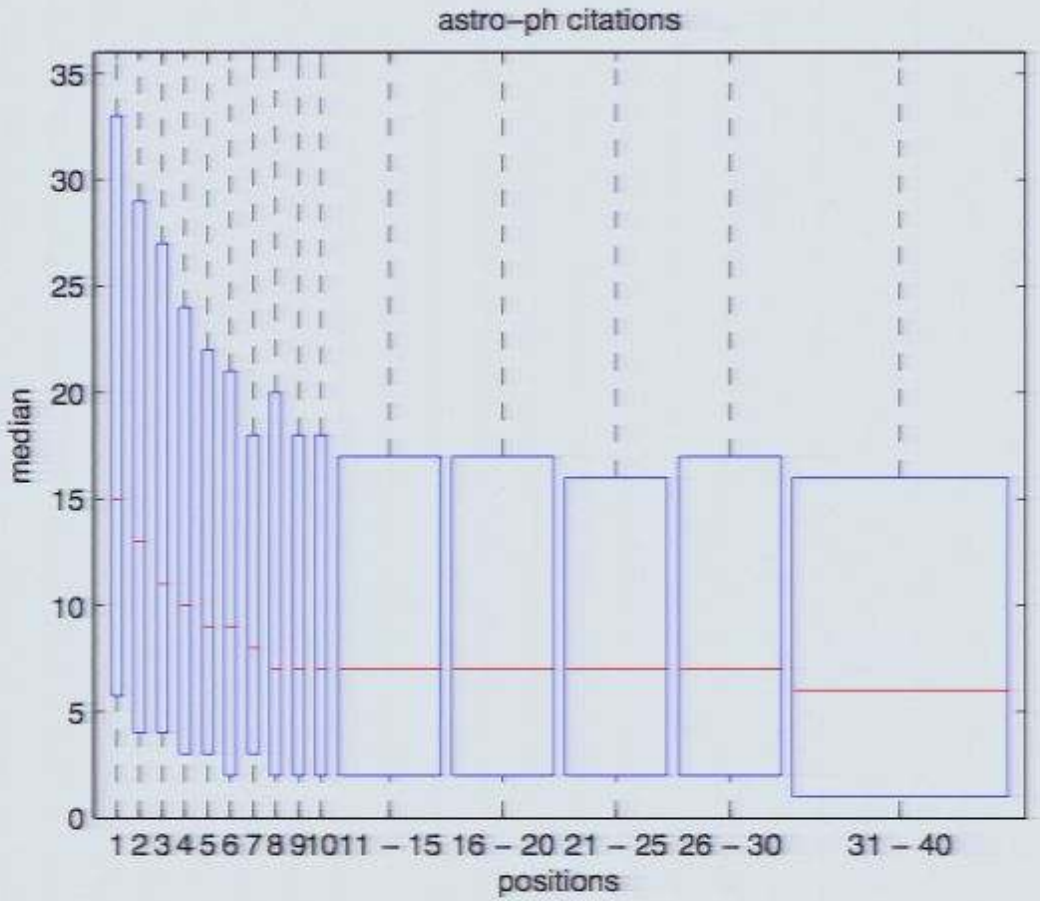
A few percent submitted in first 60 seconds (increasing with time...).

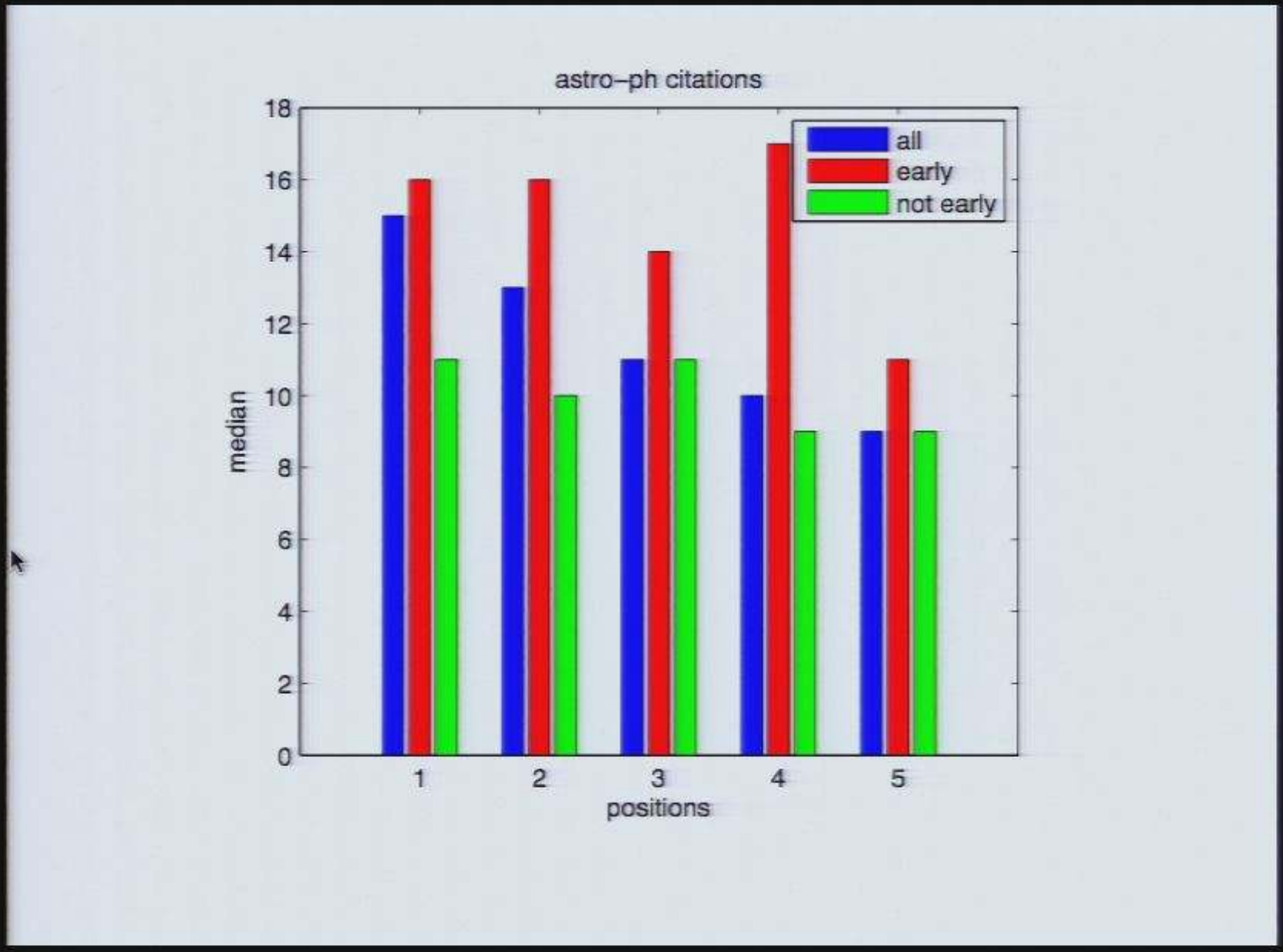
Dietrich (2008): At or near the top receive more citations by a factor of 2 (first six: 90 +/- 9, 10-40: 45+/-.9)

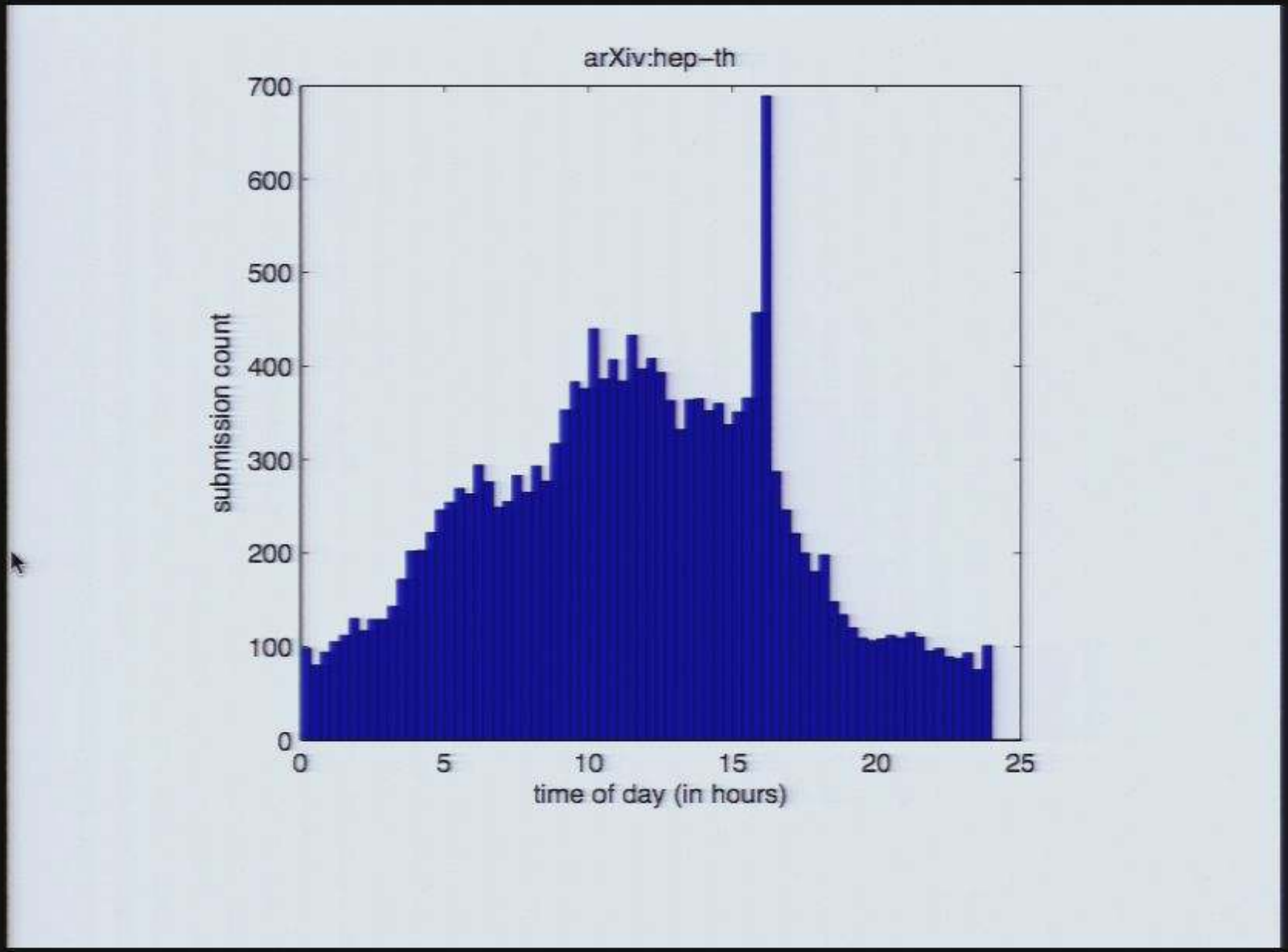
- Visibility Bias
- Self-promotion Bias

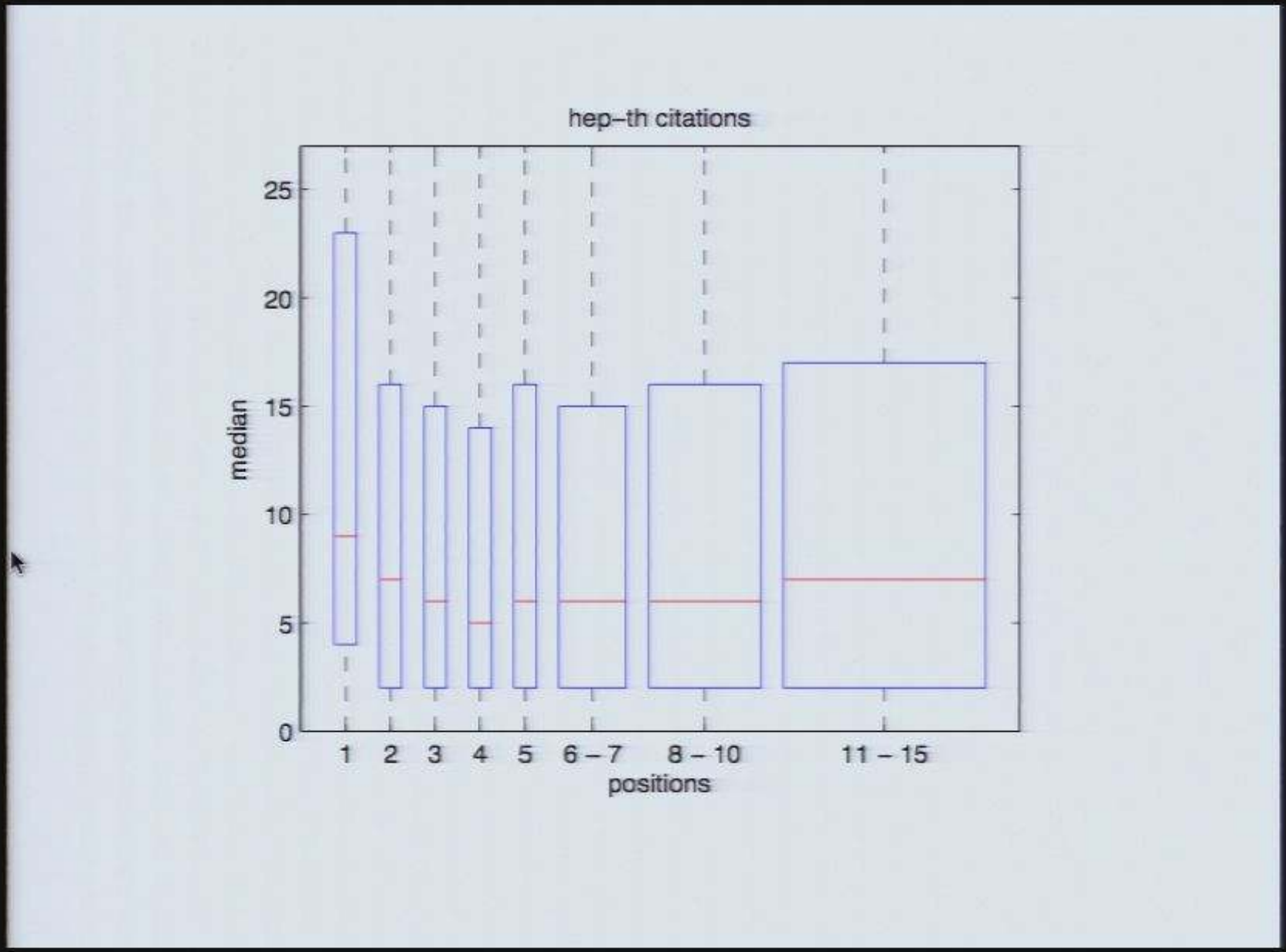
SP highest, but also difference definite VB effect (cited not necessarily due to inherent quality)

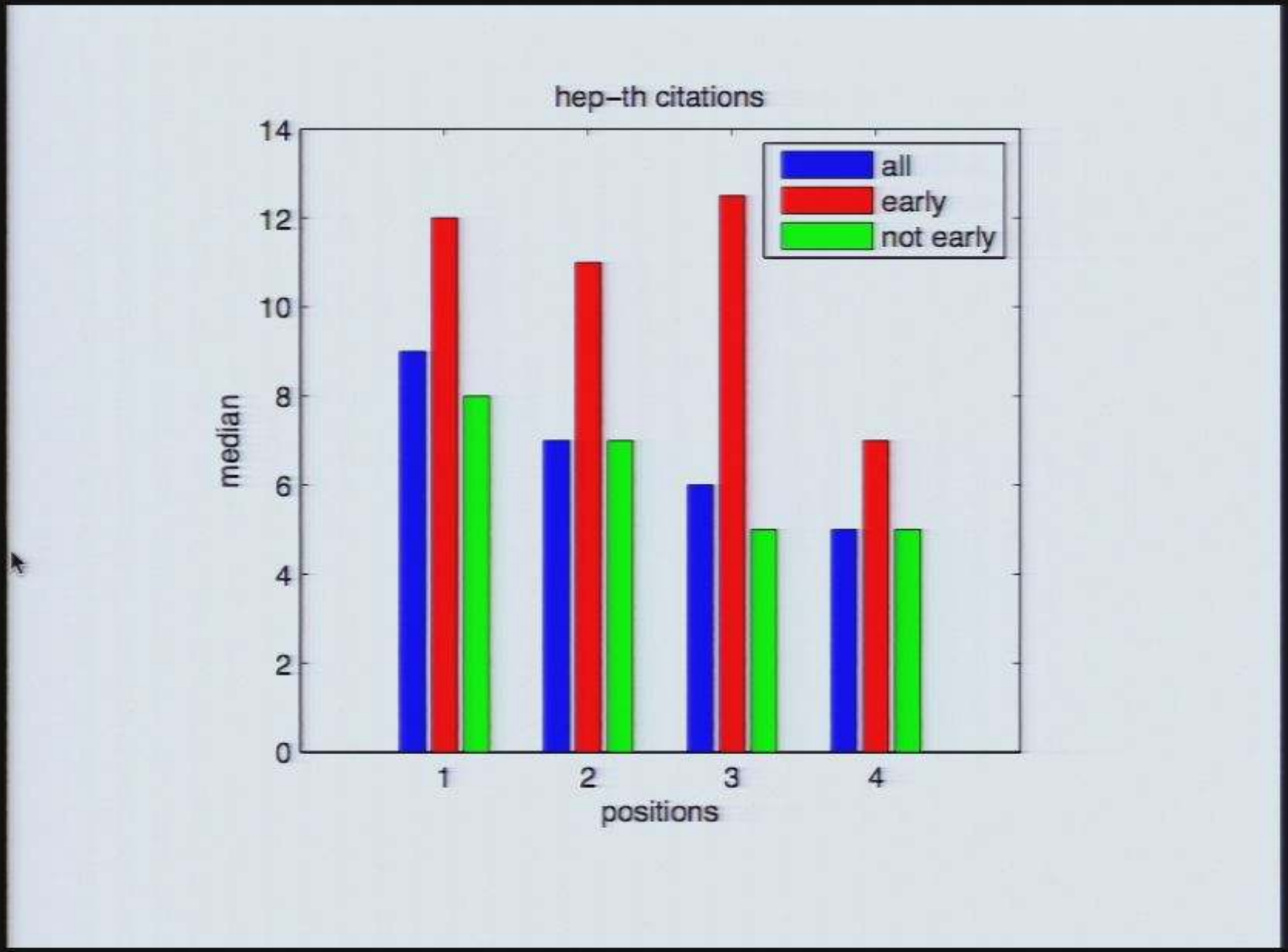
Information overload so some overlooked?
Need more tools to sort by relevance.

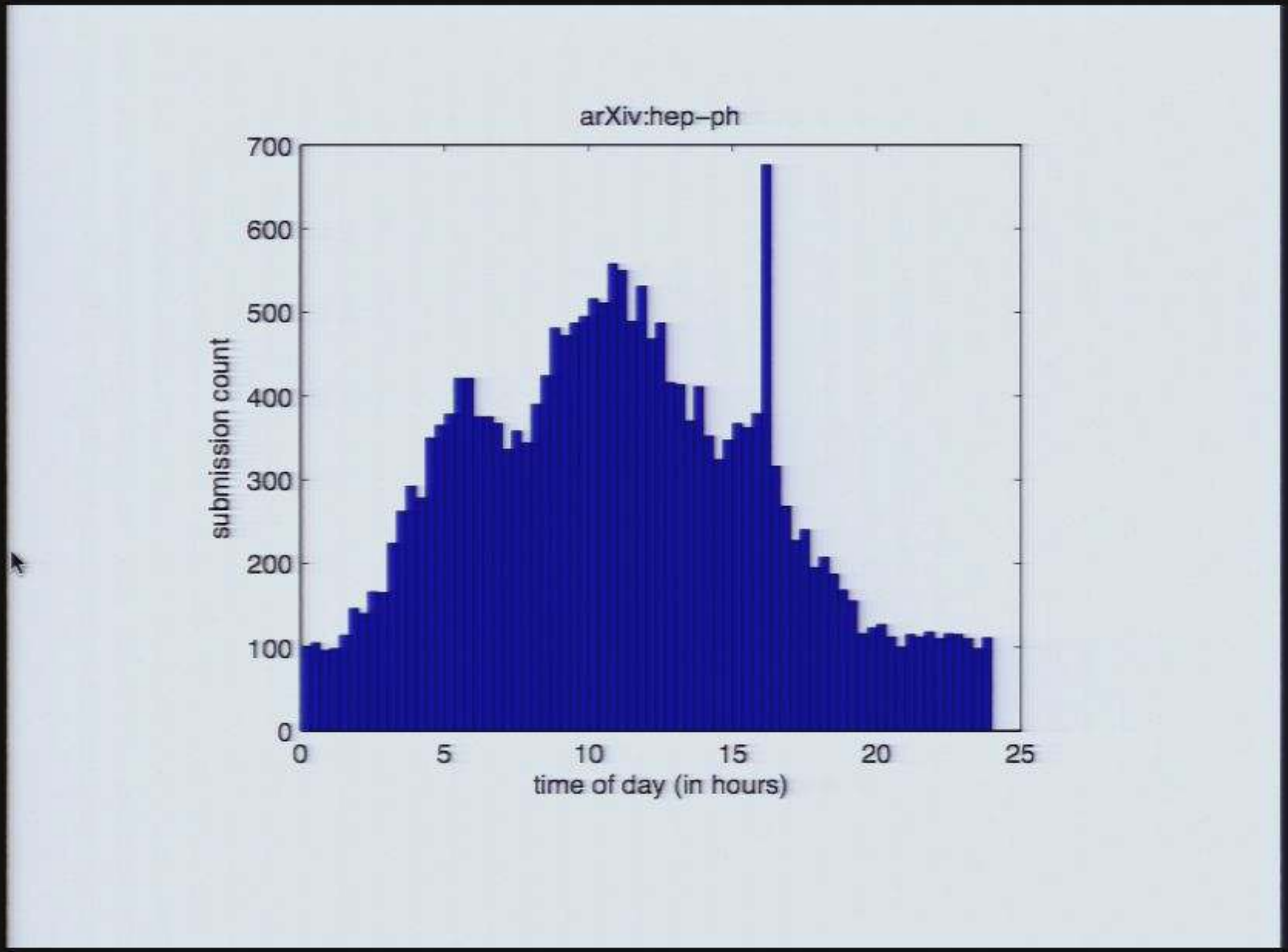

astro−ph citations
median
0
5
10
15
20
25
30
35
1 2 3 4 5 6 7 8 9 10 11 – 15 16 – 20 21 – 25 26 – 30 31 – 40
positions

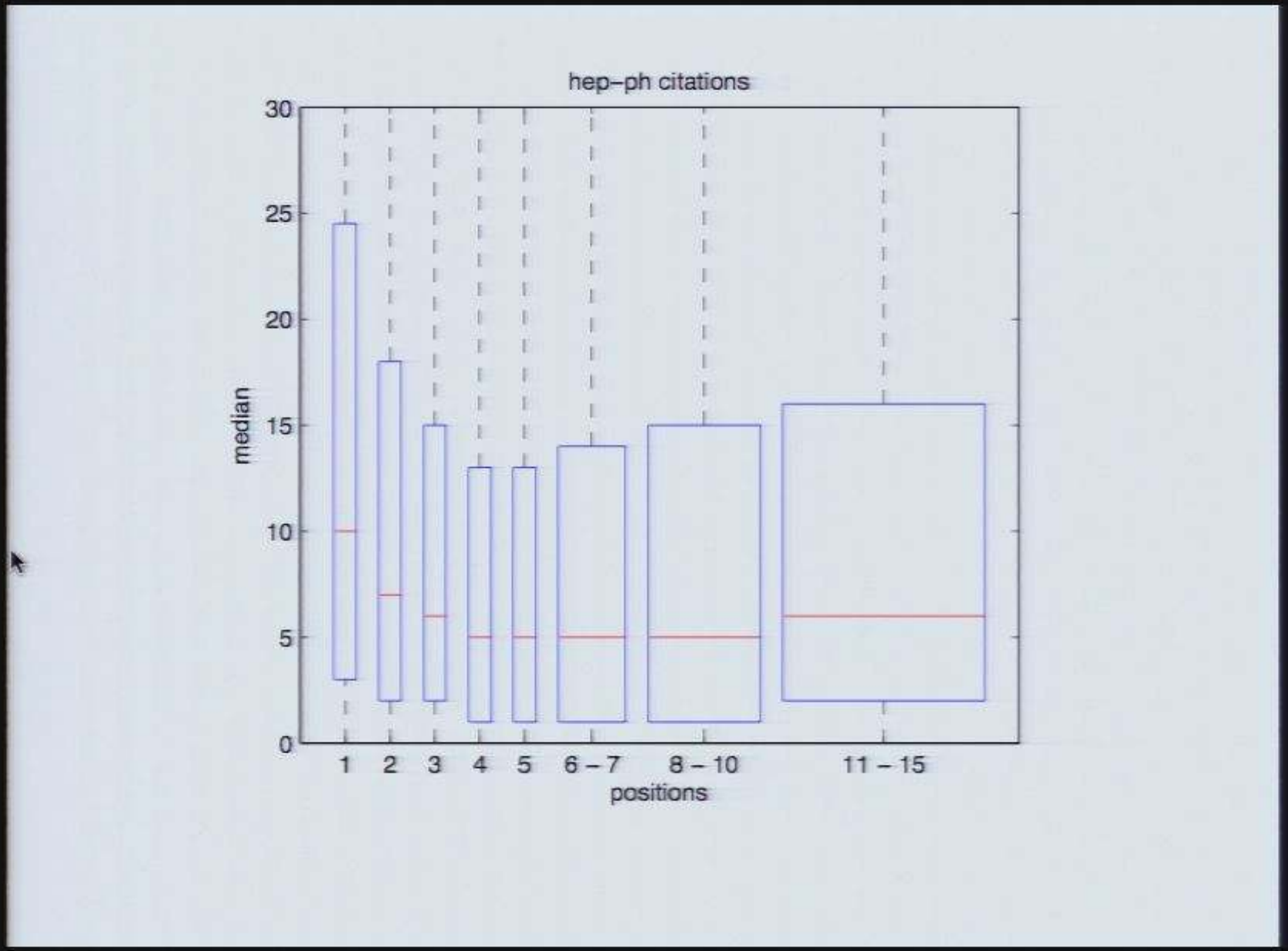

astro–ph citations
all
early
not early
median
positions
0
2
4
6
8
10
12
14
16
18
1
2
3
4
5

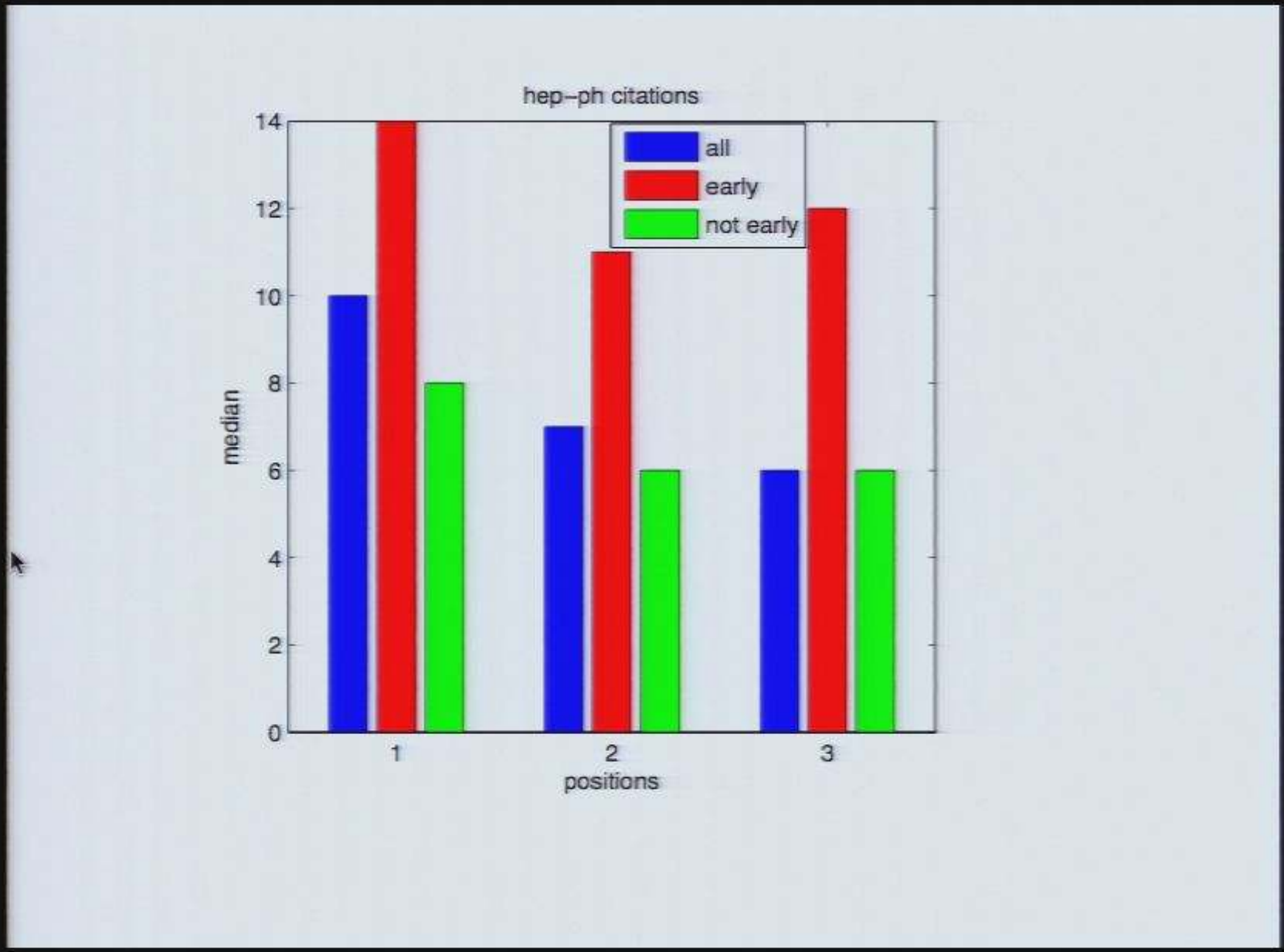

arXiv:hep–th
700
600
500
400
300
200
100
0
submission count
0
5
10
15
20
25
time of day (in hours)

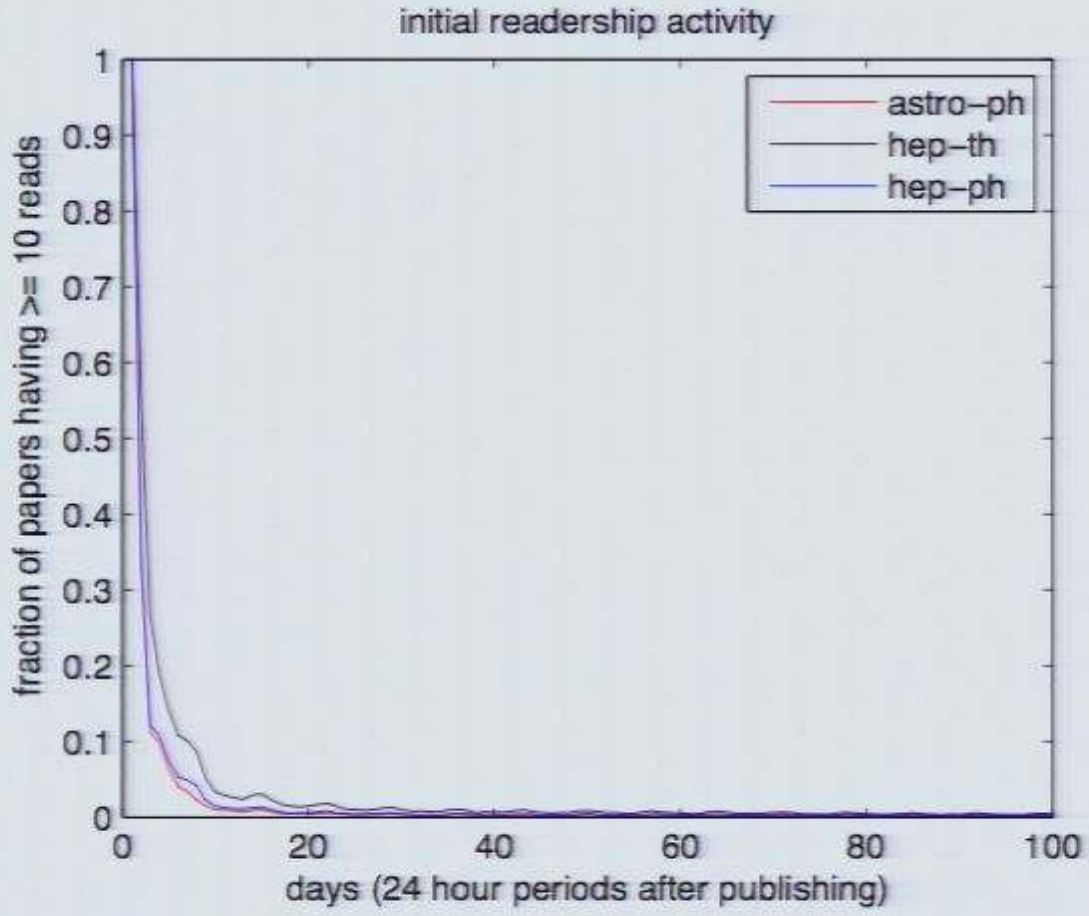

hep–th citations
median
0
5
10
15
20
25
1
2
3
4
5
6 – 7
8 – 10
11 – 15
positions


hep–th citations
14
12
10
8
6
4
2
0
median
1
2
3
4
positions
all
early
not early

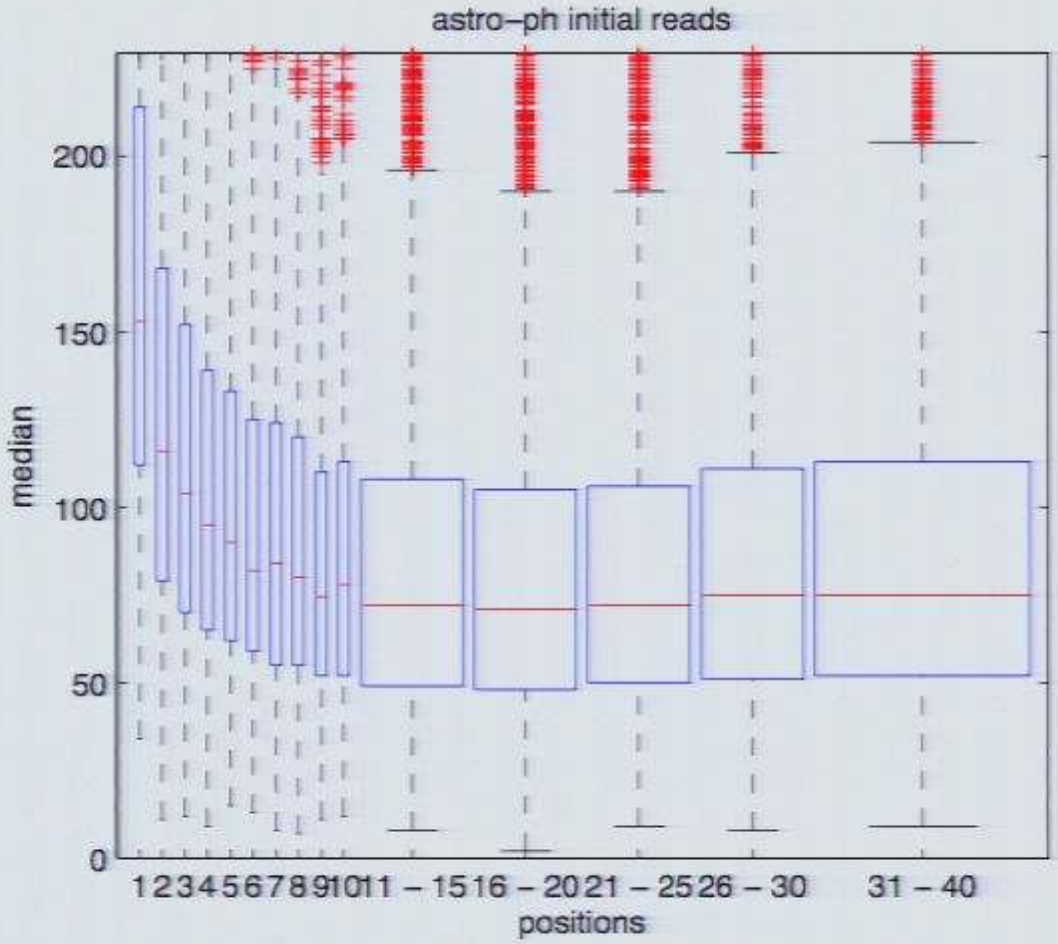

arXiv:hep-ph
700
600
500
400
300
200
100
0
submission count
0
5
10
15
20
25
time of day (in hours)

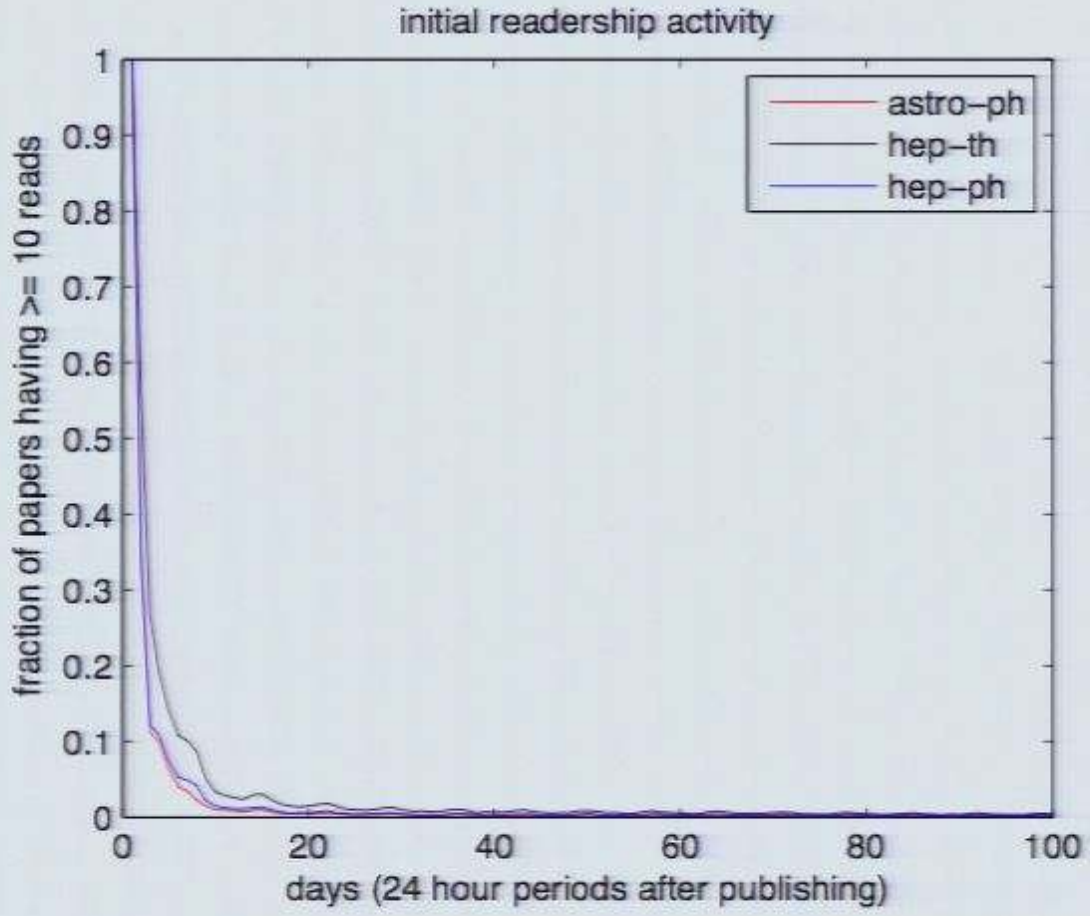

hep–ph citations
30
25
20
15
10
5
0
median
1
2
3
4
5
6 – 7
8 – 10
11 – 15
positions


hep–ph citations
all
early
not early
median
positions
0
2
4
6
8
10
12
14
1
2
3

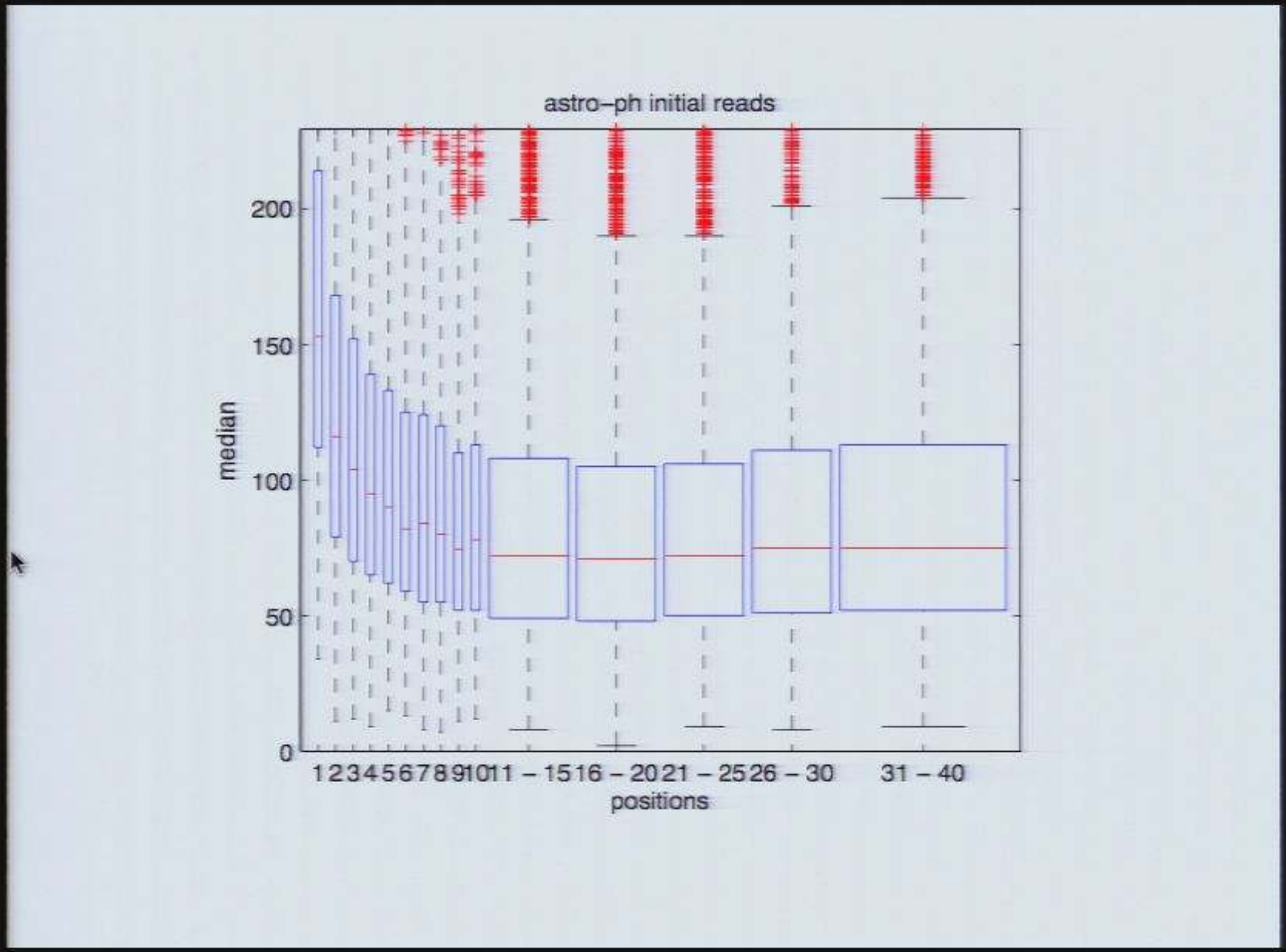


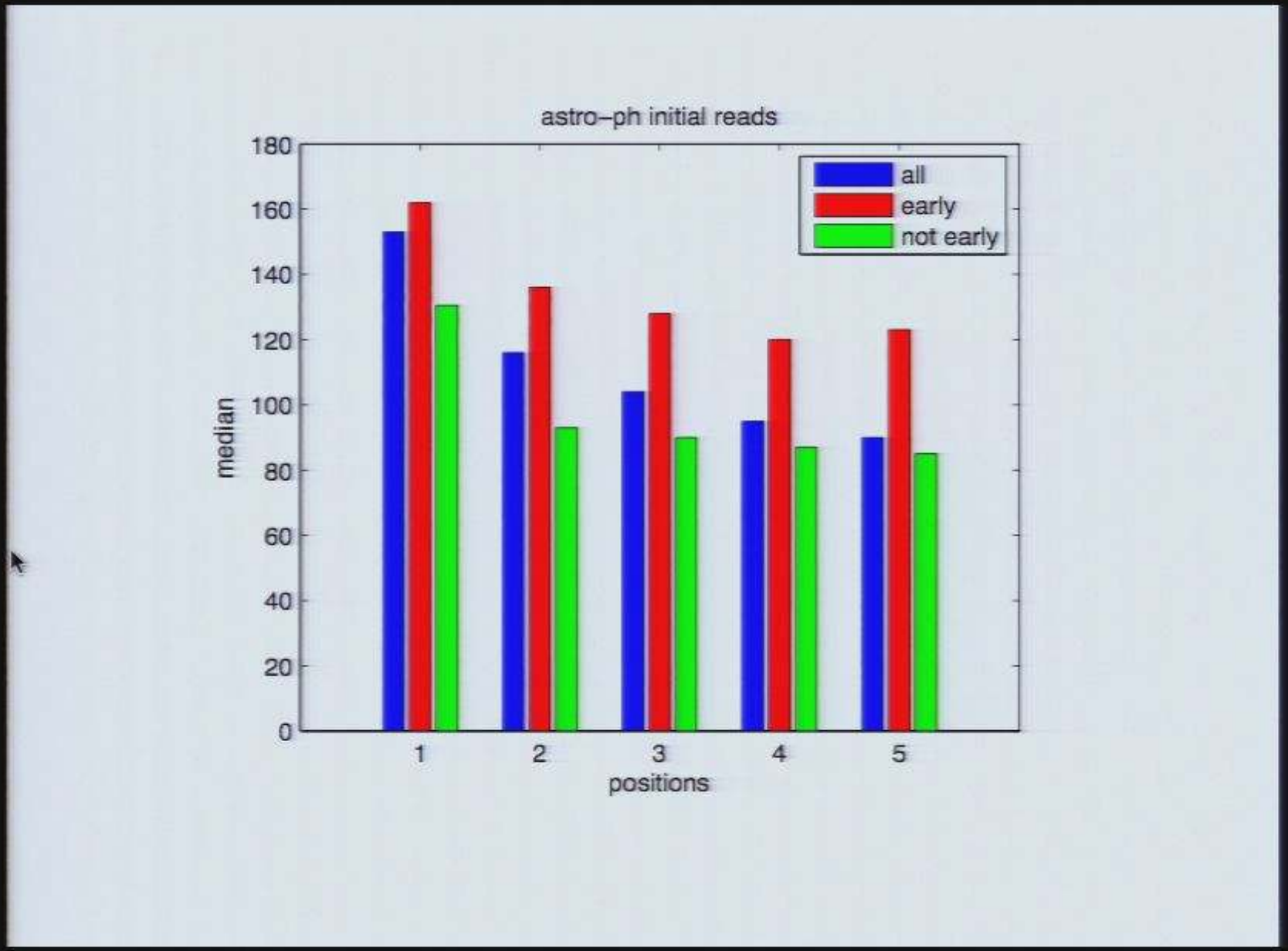
"Active" periods: 25, 15, 10 days to $< .01$

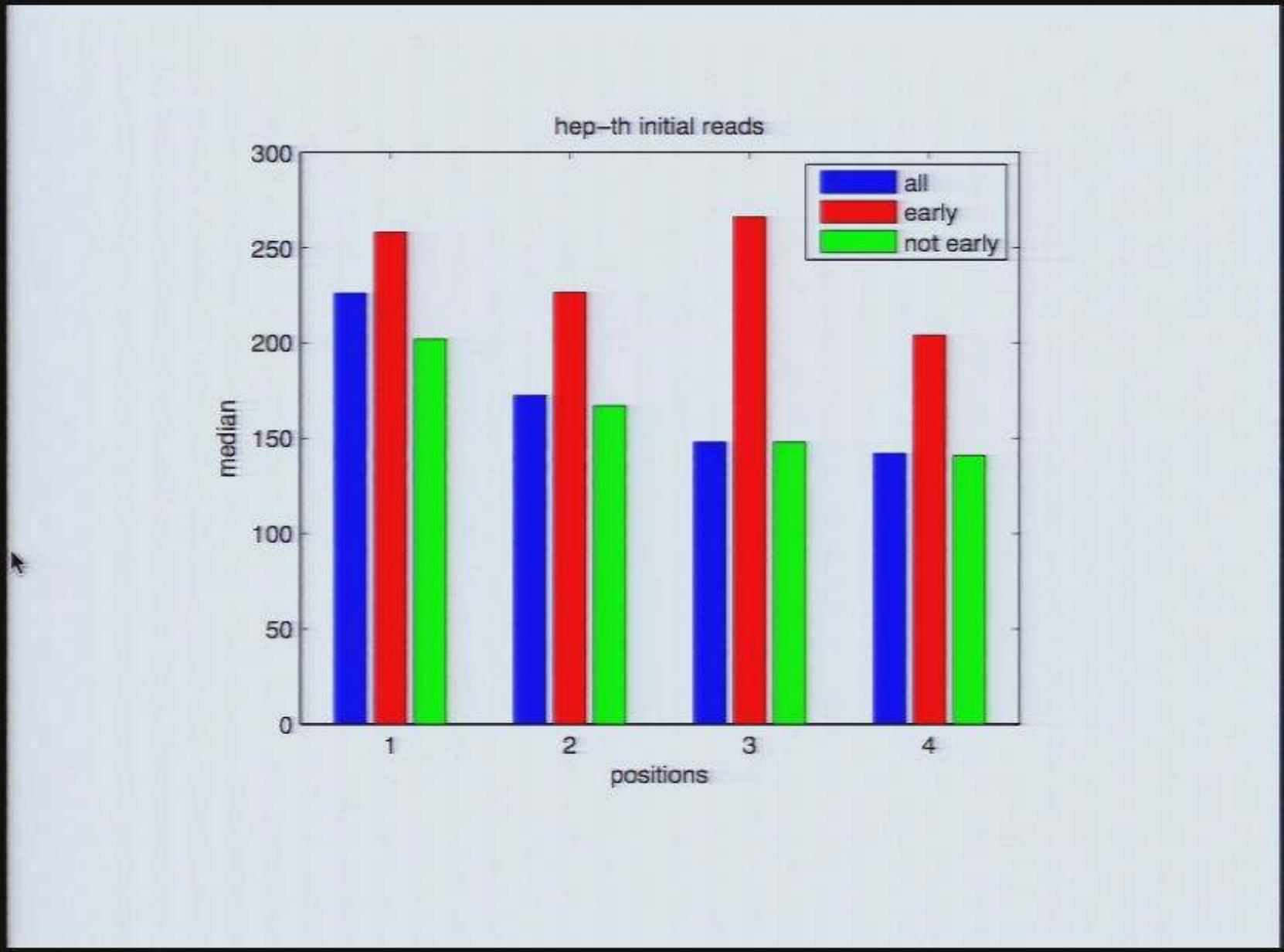

astro–ph initial reads
median
0
50
100
150
200
1 2 3 4 5 6 7 8 9 10 11 – 15 16 – 20 21 – 25 26 – 30 31 – 40
positions

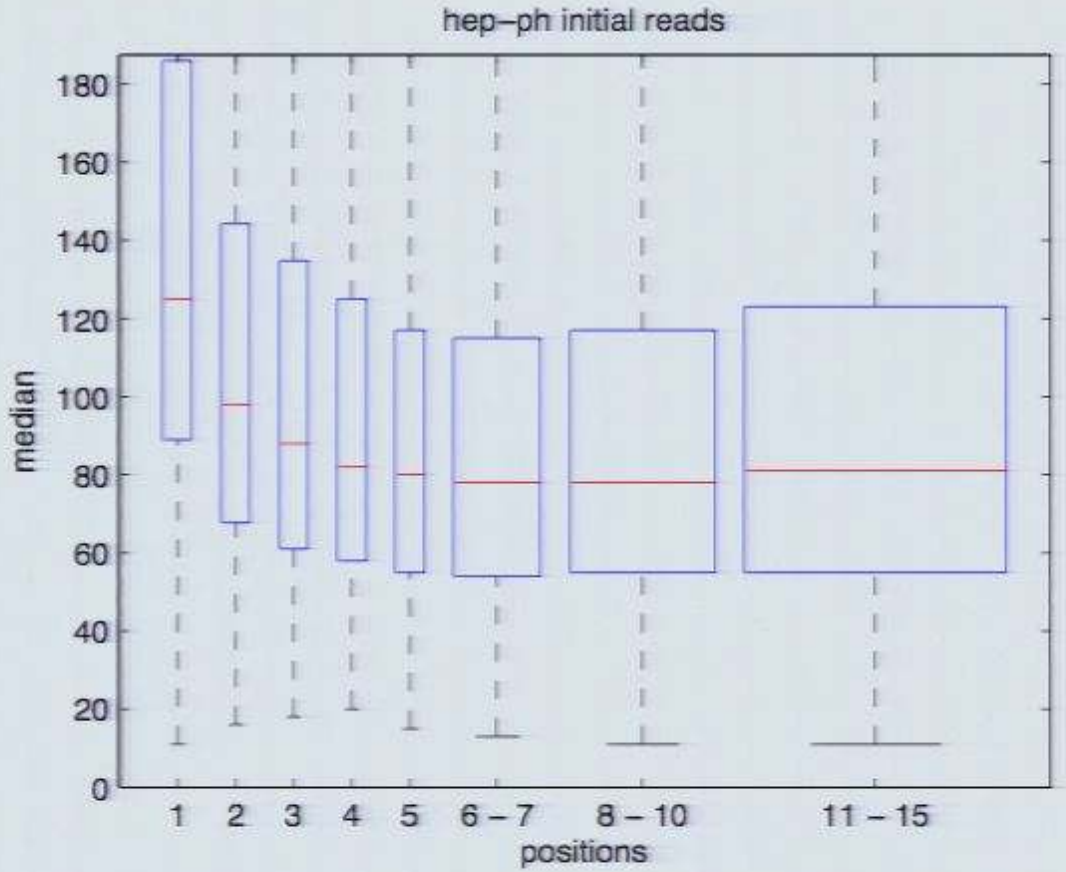


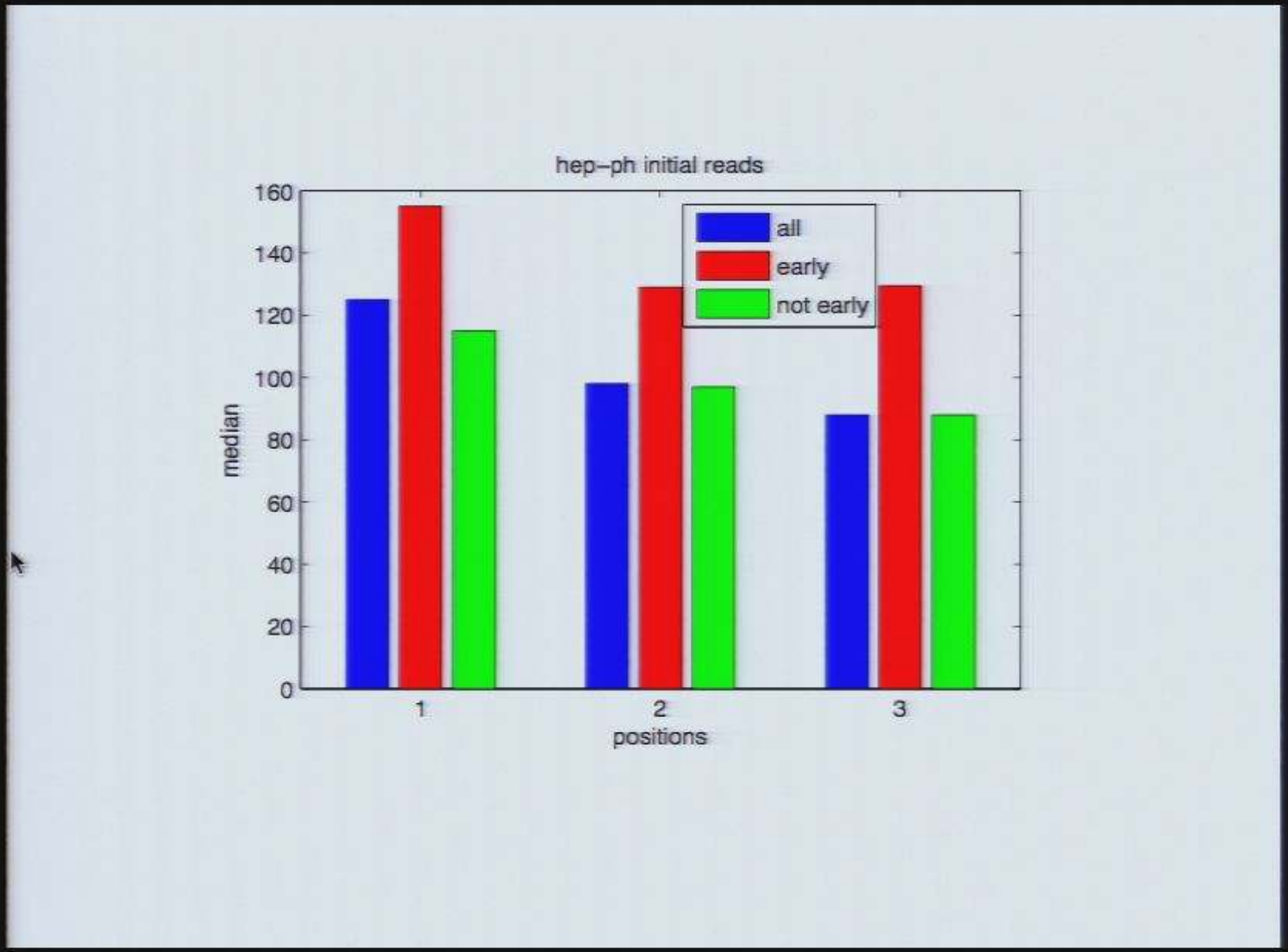
"Active" periods: 25, 15, 10 days to $< .01$


astro–ph initial reads
median
0
50
100
150
200
1 2 3 4 5 6 7 8 9 10 11 – 15 16 – 20 21 – 25 26 – 30 31 – 40
positions


astro−ph initial reads
all
early
not early
median
positions
0
20
40
60
80
100
120
140
160
180
1
2
3
4
5


hep–th initial reads
300
250
200
150
100
50
0
median
1
2
3
4
positions
all
early
not early


hep–ph initial reads
median
0
20
40
60
80
100
120
140
160
180
1
2
3
4
5
6 – 7
8 – 10
11 – 15
positions


hep–ph initial reads
160
140
120
100
80
60
40
20
0
median
1
2
3
positions
all
early
not early

Learning from Logs (Joachims/Radlinski)

Learn ranking function from implicit feedback (which clicked on) — if ranked above, but not clicked on, less relevant.

Eliminate presentation bias by flipping adjacent pairs.

Query chains: users perform sequence, or chain, of queries with similar information need. New types of preference judgments from search engine logs, taking advantage of user intelligence in reformulating queries (learned rankings outperform static ranking function)

Co-access recommender better than co-citation (and available sooner...)

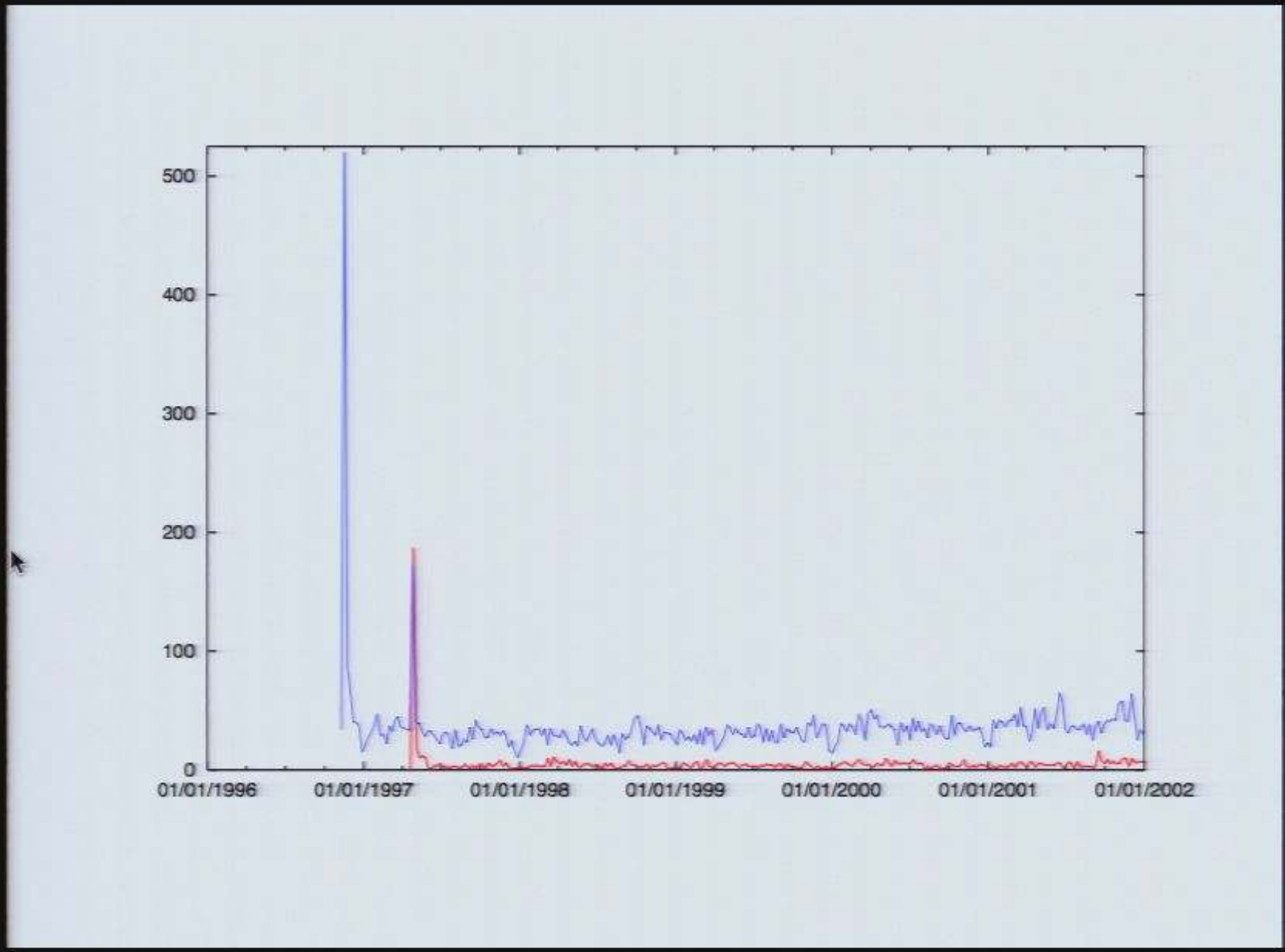
No firehose, focused community

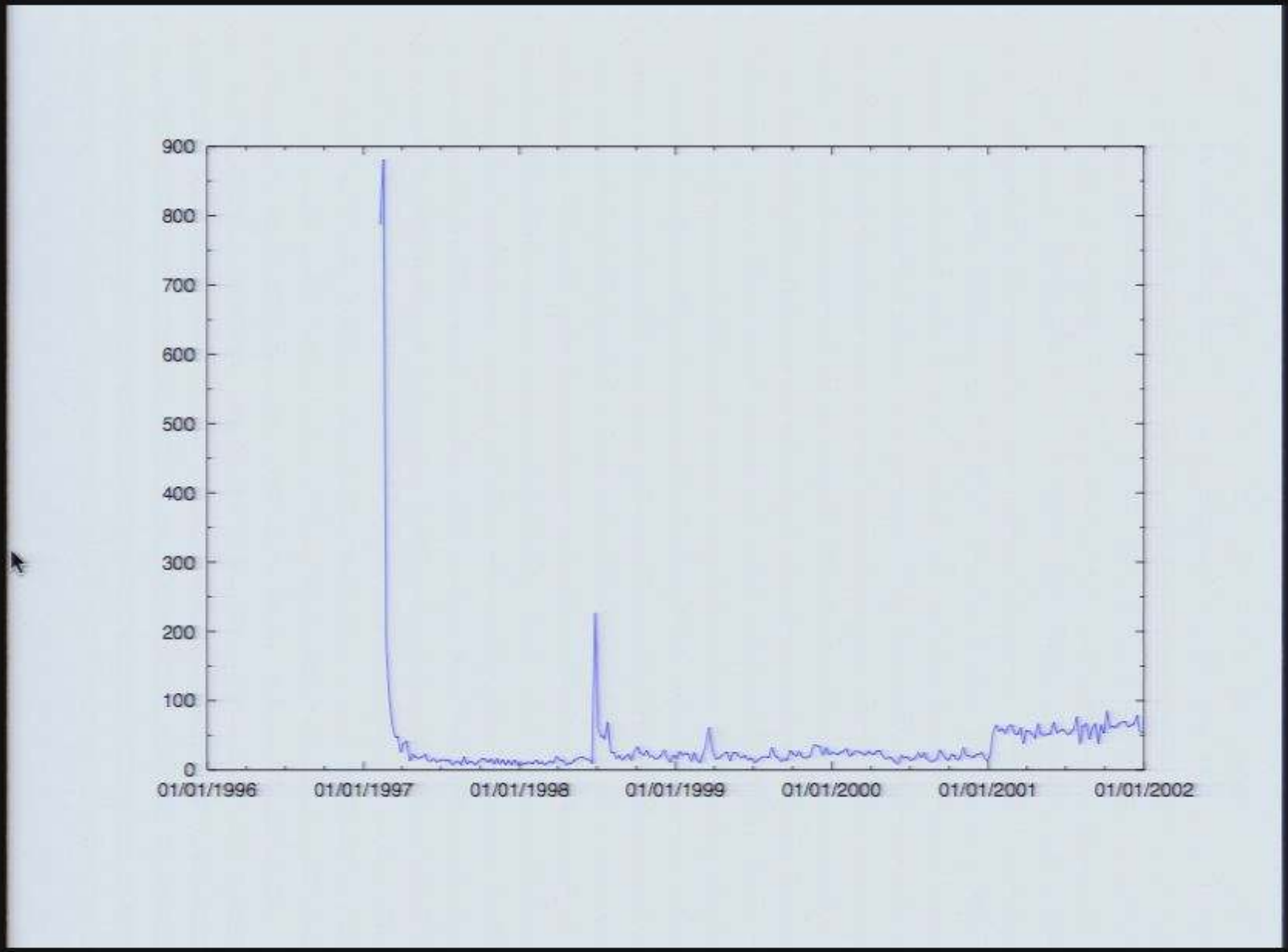
More on readership

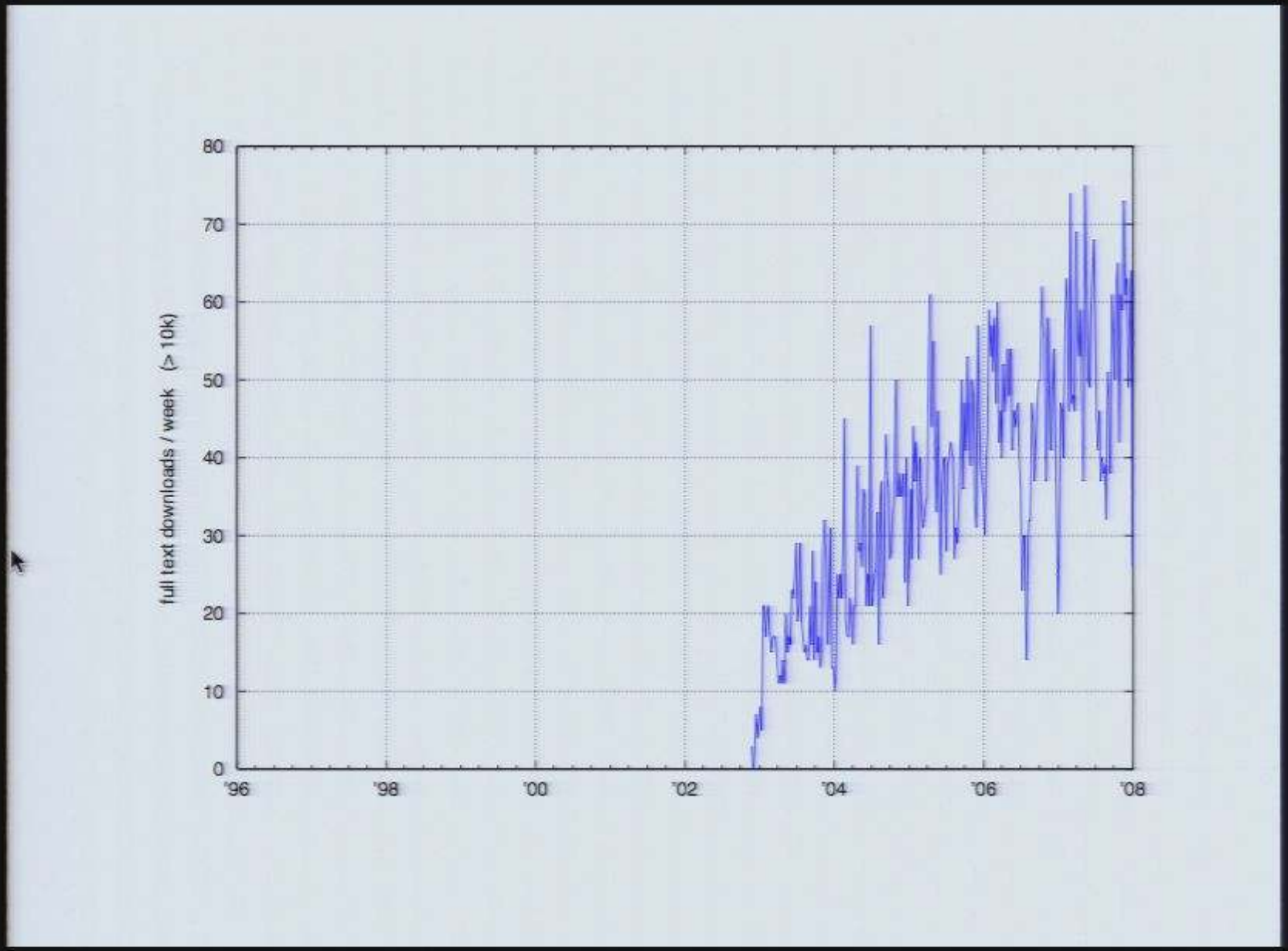
Different from cites

Almost deterministic

More network effects


500
400
300
200
100
0
01/01/1996
01/01/1997
01/01/1998
01/01/1999
01/01/2000
01/01/2001
01/01/2002


900
800
700
600
500
400
300
200
100
0
01/01/1996
01/01/1997
01/01/1998
01/01/1999
01/01/2000
01/01/2001
01/01/2002


full text downloads / week (> 10k)
80
70
60
50
40
30
20
10
0
'96
'98
'00
'02
'04
'06
'08

Mining full text

- Auto-classify
- Cluster
- Ontological Markup (via API interoperate with PMC)
- Plagiarism

Search Results: Display

- Summary
- Brief
- XML
- Taxonomy Tree
- Cited in Books
- CancerChrom Links
- Conserved Domain Links
- 3D Domain Links
- GEO DataSet Links
- Gene Links
- Genome Links
- Genome Project Links
- GENSAT Links
- GEO Profile Links
- HomoloGene Links
- Nucleotide Links
- OMIM Links
- Compound Links
- Substance Links
- PopSet Links
- Protein Links
- PubMed Links
- Cited Articles
- SNP Links
- Structure Links
- Taxonomy Links
- UniSTS Links

Article: Related Material

- PubMed record
- PubMed related arts
- PubMed LinkOut
- Gene
- HomoloGene
- Nucleotide
- Omim
- GEO Profiles
- Protein
- PubChem Compound
- PubChem Substance
- Taxonomy
- Taxonomy tree

Backdoor Route to Open Access

- More than one-third of the high-impact journal articles in a sample of biological/medical journals published in 2003 were found at nonjournal Web sites (Wren, 2005).
- Unsystematic (Ginsparg, 2006, "As We May Read"): 75% of publications from 2000 or later posted at web site of incoming president of Society for Neuroscience available without subscription (preprints, open-access journal sites, copies at nonjournal web sites).
  Perhaps already farther along than most realize?
- Expectations of next generation independent of outcome of government mandate debate

What will Open Access Mean?

First survey current generic functionality:

- PMC, ADS, Citeseer, PLoS, scholar.google, SLAC-Spires, . . .
- APS, ISI, Highwire, ScienceDirect, IoP, . . .
- nytimes, youtube, video.google, amazon, . . .

Scholarship $\Longleftrightarrow$ Shopping $\Longleftrightarrow$ Entertainment

(sniff: we're no longer the bleeding edge)

Note importance of community building / social networking

Avoid emulating abacus

Watch out for interactions with blogspace/media

clearly no citation advantage ...

Item specific

- standard metadata (title, author(s), submitter)
- browse related items, related keywords
  - ▷ local, in 3rd party (e.g., pubmed, ISI, scholar.google . . .)
- add tags, labels ("flowering of the commons")
- more from this user
- rate this item
- save to favorites
- add to groups
- share, e-mail to friend
- blog this item
- post to 3rd party site (e.g., myspace)
- flag as inappropriate
- comments, responses, eletters (read, add)

Item specific, cont'd

- full text
- supplemental data
- show references, citations
- addenda, corrigenda
- related web pages
- export citation; cite or link using DOI
- alert when cited
- same object in 3rd party (e.g., pubmed citation)
- search 3rd party database (e.g., by same authors in scholar.google, h-index)
- flavors of relatedness by text, co-citation, co-reference, co-usage (also read)

Site specific

- subscribe
- alert to new issues
- upload
- personalization
  - ▷ my articles (view collection)
  - ▷ add/subtract from private library

(Note: enhancement for other users, but privacy issues?)

Browse

- groups, categories, subject area
- most recent
- recently featured
- most viewed
- top rated
- most discussed
- top favorites
- most linked
- most honored
- most shared
- most blogged
- most searched

Present

More than a new means of distribution?

Crippled by document format? (TeX, Word → PDF, 70's methodology)

Implications of next generation open XML document format [.docx, . . .] not yet appreciated.

(Commercial tools for authoring in NLM/NCBI DTD?
Article authoring add-in for MS Word 2007)

Paradox of physics: some well-established areas could fit into a semantic web context, amenable to a "commons" approach via open ontologies and sets of relationships

(more generally, tie semantic content in existing centralized literature databases to distributed network databases using relevant ontologies and machine-readable document standards)

Past Confusion

Still no wysiwig?

Metastable state?

Efficacy of search engines?

Other fields? (not just information processing...)

Wikipedia?

Caution: new developments no longer academic-centric

Future

Challenge from Word developers to Scientists:

Suggest 20 functions to provide optimal environment for scientific authorship (handshakes to networked databases, etc.)

Active + Passive user participation in bottom-up approach to QC

- actively add tags, links; contribute to ontologies, correct wiki entries
- passively ingest readership, bookmarking, annotation behavior

Incentive Question: expertise-intensive efforts beyond conventional journal publication (annotation, linkage, . . .) = scholarly achievement?

articles + blog commentary → more modular objects

glue databases together into knowledge structure

Goal: semi-supervised, self-incentivized, self-maintaining knowledge structure, navigated via synthesized concepts, w/o redundancy/ambiguity, sourced, authenticated, highlighted for novelty

Obvious

- Automated Markup (genomes, proteins, organisms, glossary items)
- Missing refs? Superfluous refs?
  Example: AdS Collaborative filtering
  - ▷ **Recent:** Keyword search gives recent articles on a subject
  - ▷ **Popular:** Usage data gives most read
  - ▷ **Useful:** Citation data gives most referred
  - ▷ **Pedagogical:** Citation data gives most referring
- user tagging (e.g., flagged review articles)
- interoperability with [blog/news/wiki]-space
- systematically mine and bookmark interiors (relatedness tree), modular components
- NVO, Mathworld, etc.

Embarrassments: a) plagiarism, b) not yet public

Network benefits to readers and authors

algorithms with access to personal and collective user behaviors ⇒ more comprehensive browsing

linkages to explanatory and complementary resources tied to words, equations, figures, and data ⇒ more incisive reading

Network-aware authoring tools will analyze draft document content in progress, suggesting links to related external text and data resources, including semantic linkages.

Take advantage of the continued growth in distributed network databases, new interoperability protocols, machine-readable document standards, and relevant ontologies.

Neo-Minsky: "Can you imagine they used to have an internet in which authors, databases, articles, and readers didn't talk to each other?"

Essential question

How will the analog of NCBI/PubMedCentral be provided for other communities? (Who? With whose money?)

Common web service protocols, common languages (e.g., for manipulating, visualizing data), data interchange standards

Distributed version for other fields

networked resources ⇒ new nonlinear reading strategies
ubiquitous mobile devices ⇒ new usage of short-, long-term memory

Qualitatively new research and cognitive methodologies, transformation in the way we process scientific information, with academic community as role model for the creation and dissemination of knowledge to the public

Network benefits to readers and authors

algorithms with access to personal and collective user behaviors ⇒ more comprehensive browsing

linkages to explanatory and complementary resources tied to words, equations, figures, and data ⇒ more incisive reading

Network-aware authoring tools will analyze draft document content in progress, suggesting links to related external text and data resources, including semantic linkages.

Take advantage of the continued growth in distributed network databases, new interoperability protocols, machine-readable document standards, and relevant ontologies.

Neo-Minsky: "Can you imagine they used to have an internet in which authors, databases, articles, and readers didn't talk to each other?"

Essential question

How will the analog of NCBI/PubMedCentral be provided for other communities? (Who? With whose money?)

Common web service protocols, common languages (e.g., for manipulating, visualizing data), data interchange standards

Distributed version for other fields

networked resources $\Rightarrow$ new nonlinear reading strategies

ubiquitous mobile devices $\Rightarrow$ new usage of short-, long-term memory

Qualitatively new research and cognitive methodologies, transformation in the way we process scientific information, with academic community as role model for the creation and dissemination of knowledge to the public