

Global and Local Knowledge in Citation Network Formation

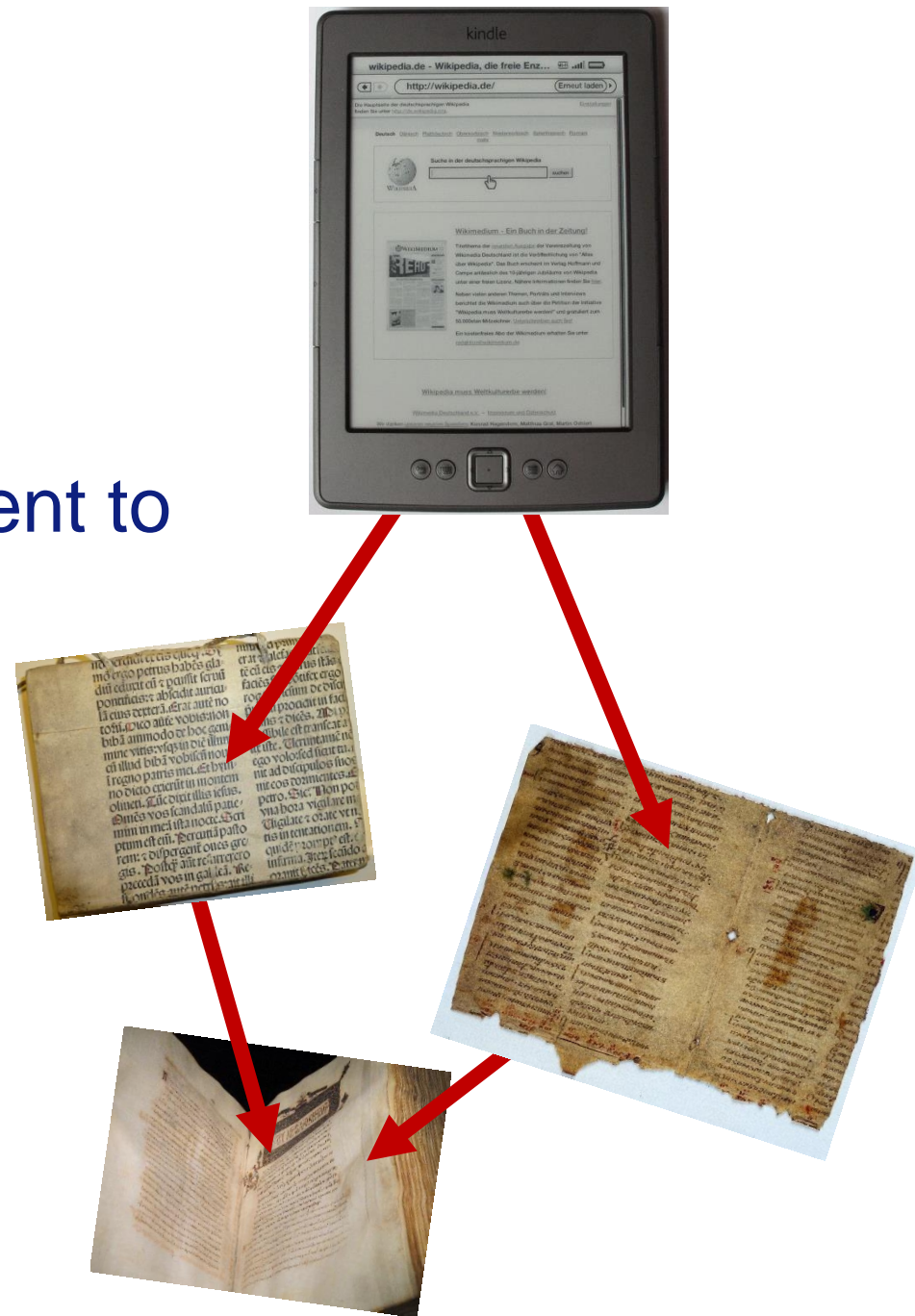
Tim Evans

Work done with:
Sophia Goldberg &
Hannah Anthony



Citation Networks

- Nodes = Documents
- Directed Edges from bibliography of document to older works
 - Arrow of Time see James Clough's talk
- Examples:
Academic papers,
Patents,
Court Documents

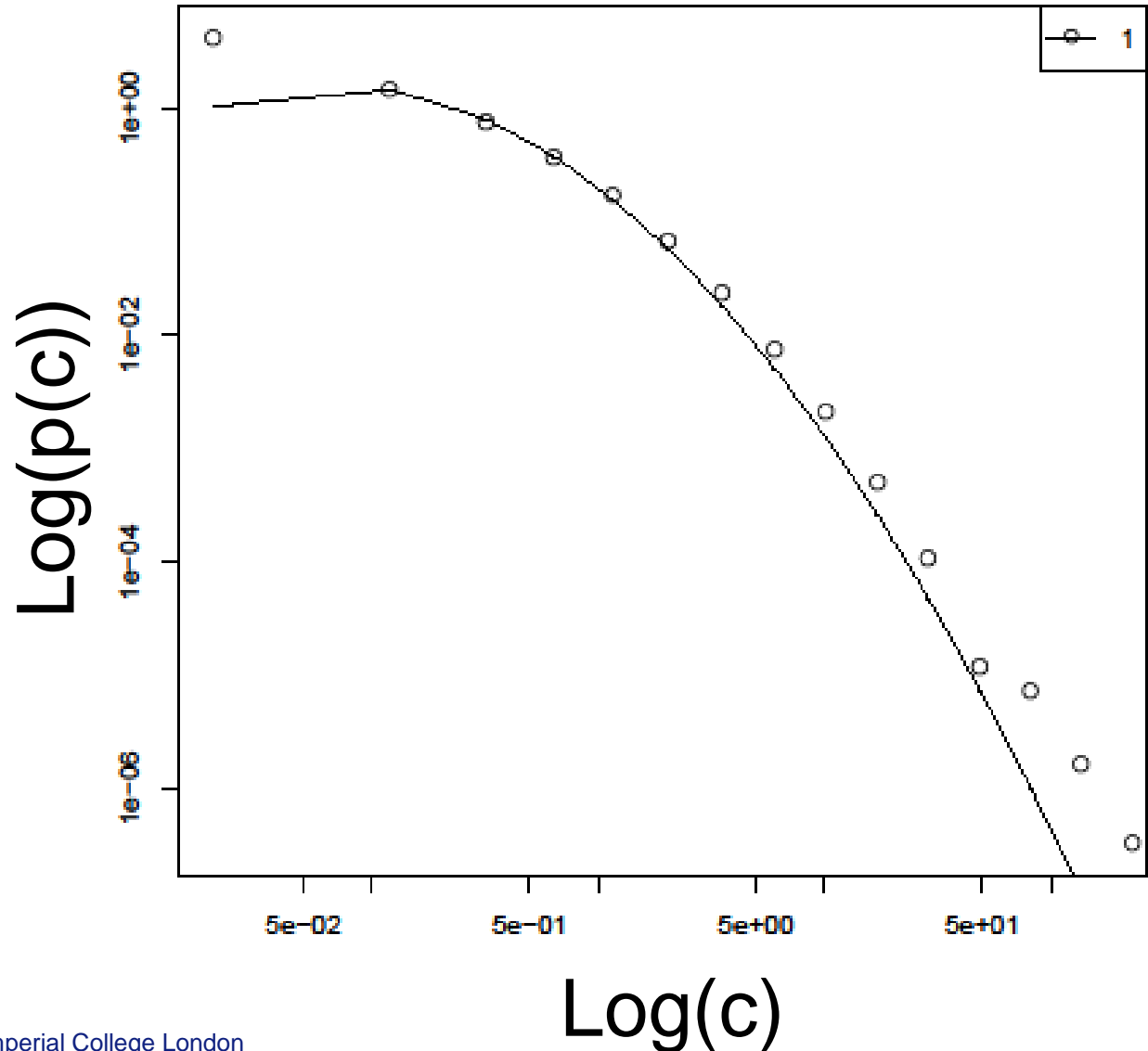


Citation Distribution

Fat-Tail
(clearer on
log-log plot)

Data from **hep-th**
section of arXiv

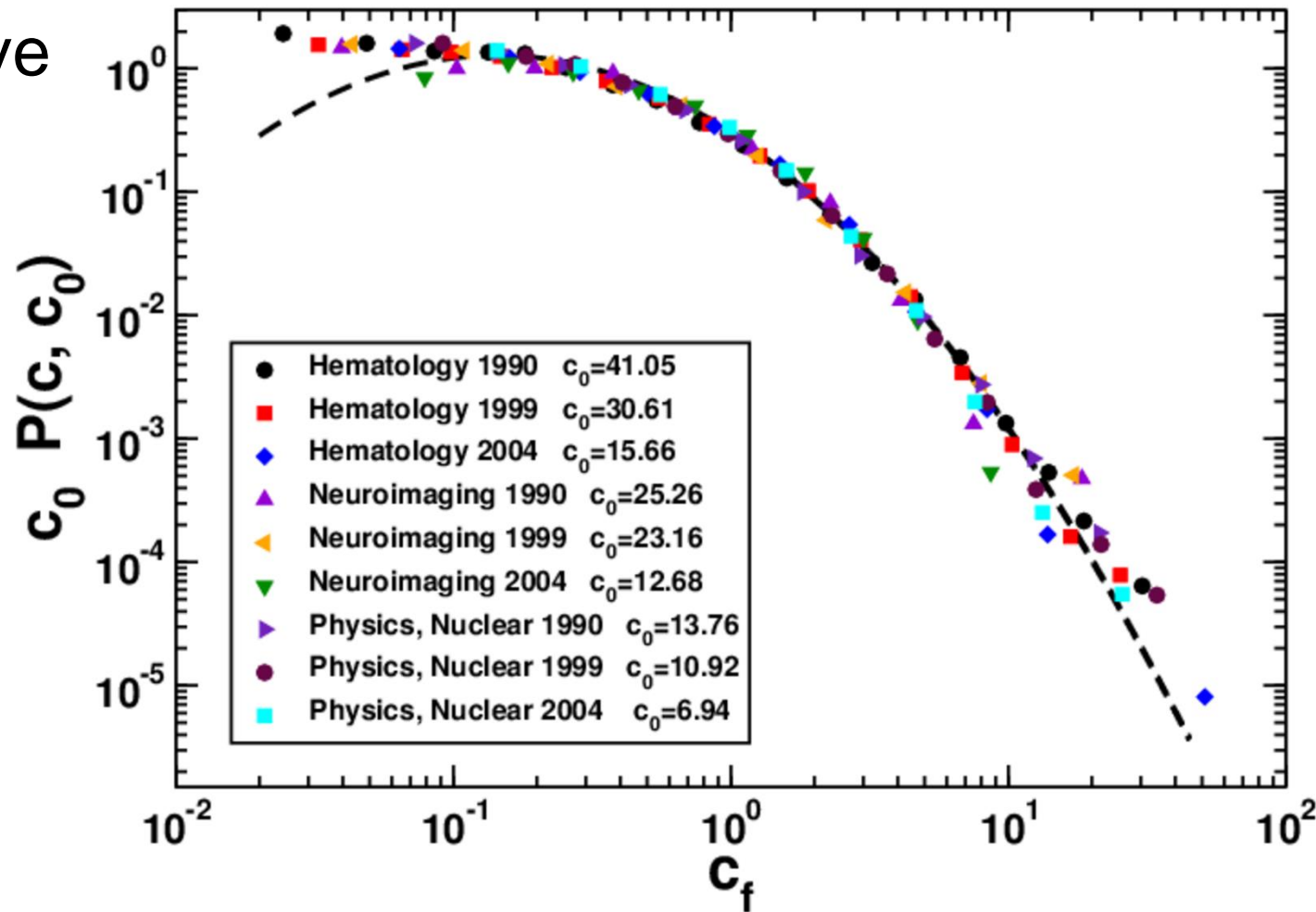
Raw Data 1992–2002



Citations to Publications published in One Year in one Field

Rescale relative to average number of citations

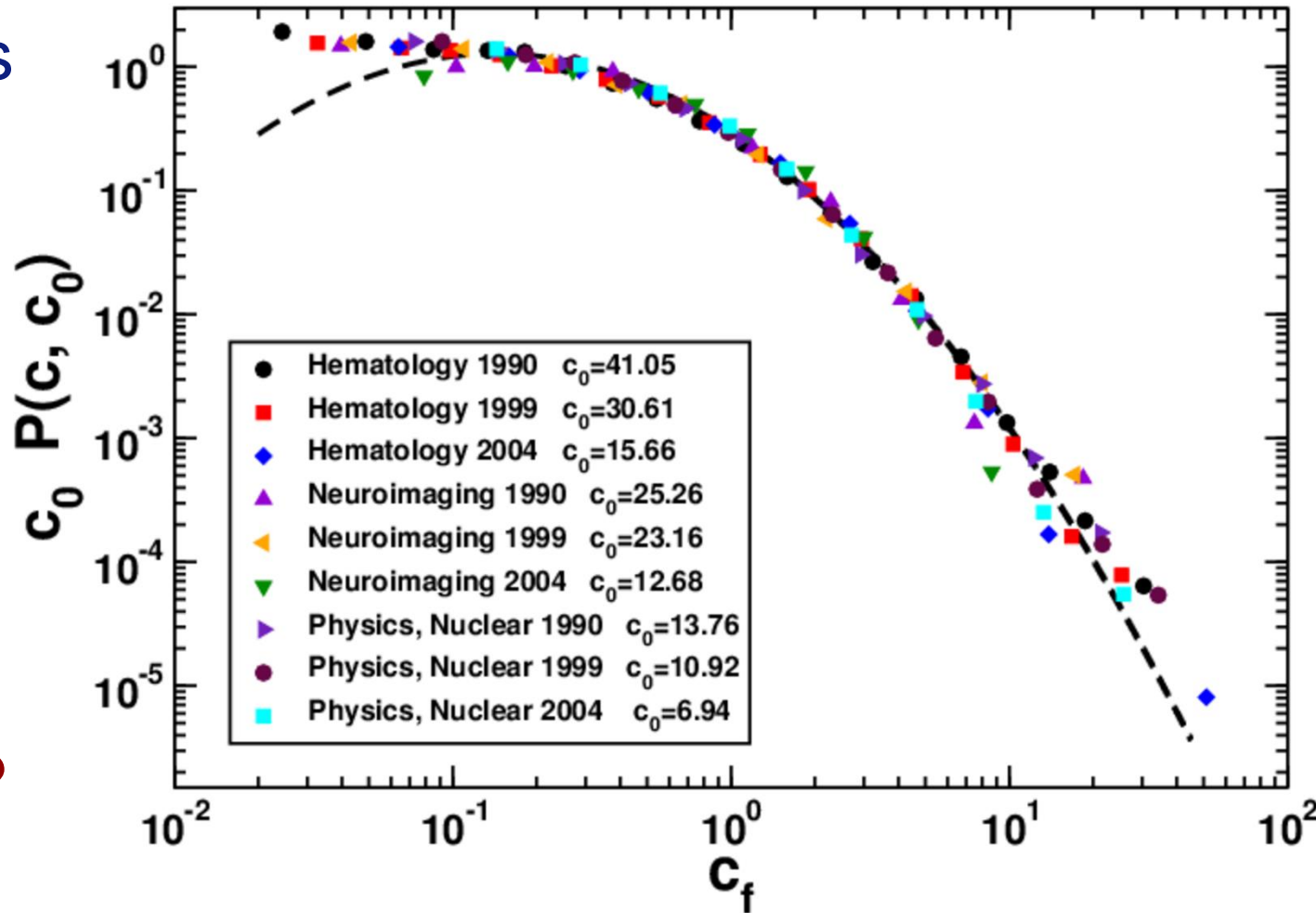
⇒ Tails look very similar at different times



Understanding Citations Through a Model

Similar Tails
Over TIME

Question:-
Can we
find model
which
reproduces
this feature?



Simple Models Fail



Price Model (1965)

Cumulative Advantage (Preferential Attachment)
new documents cite old documents in
proportion to the number of current citations

Find number of citations **c** proportional to square
root of age of document

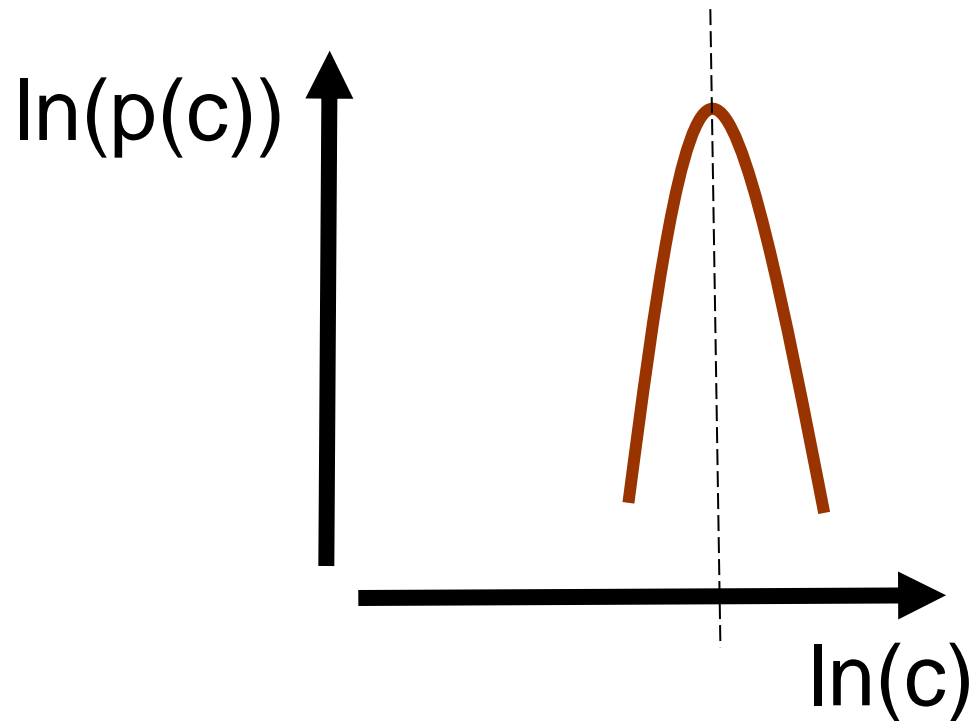
$$c \sim \sqrt{t - t_{\text{pub}}}$$

Price Model

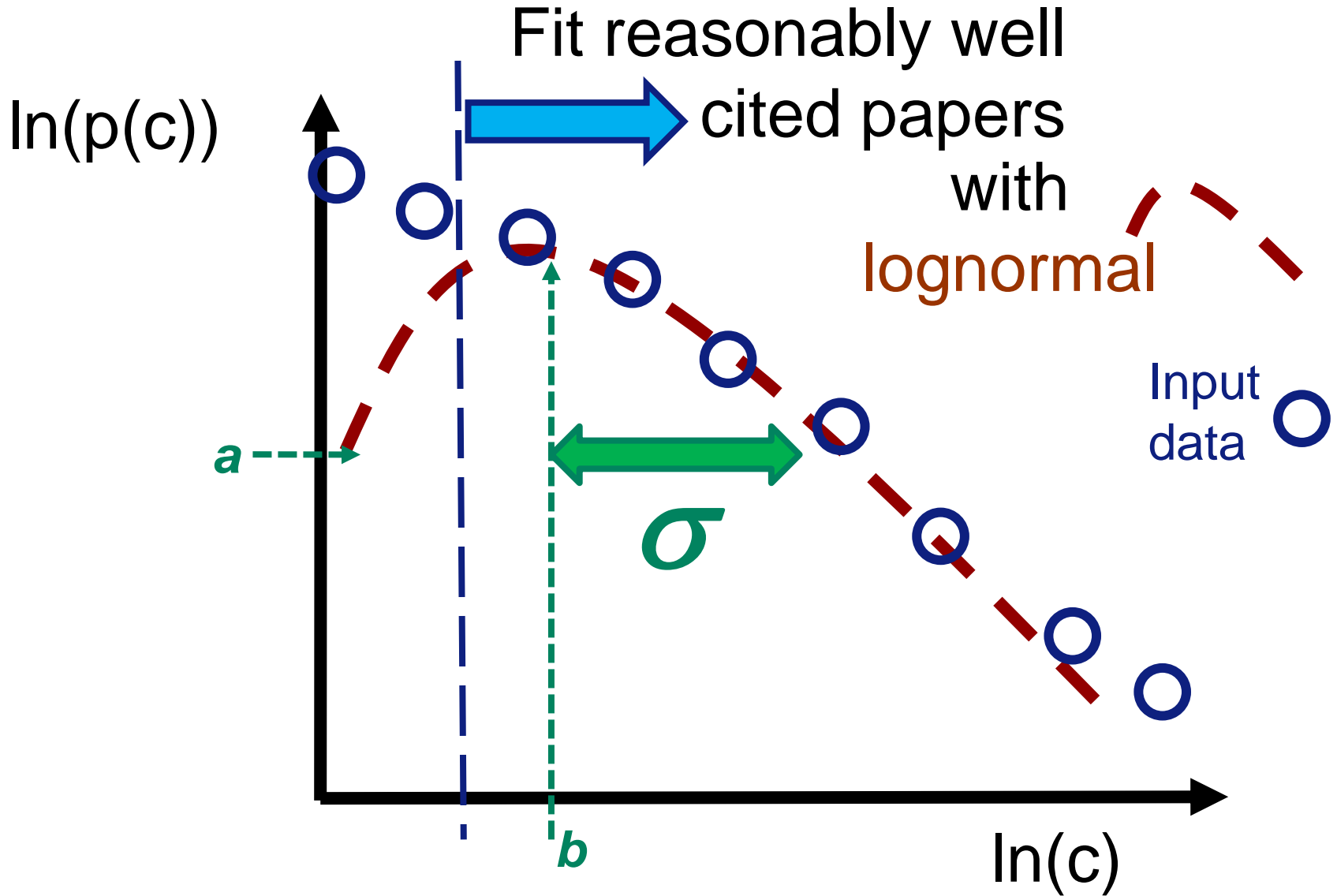


No fat tails for publications in same year

All papers of similar age gain similar number of citations (square root of age)

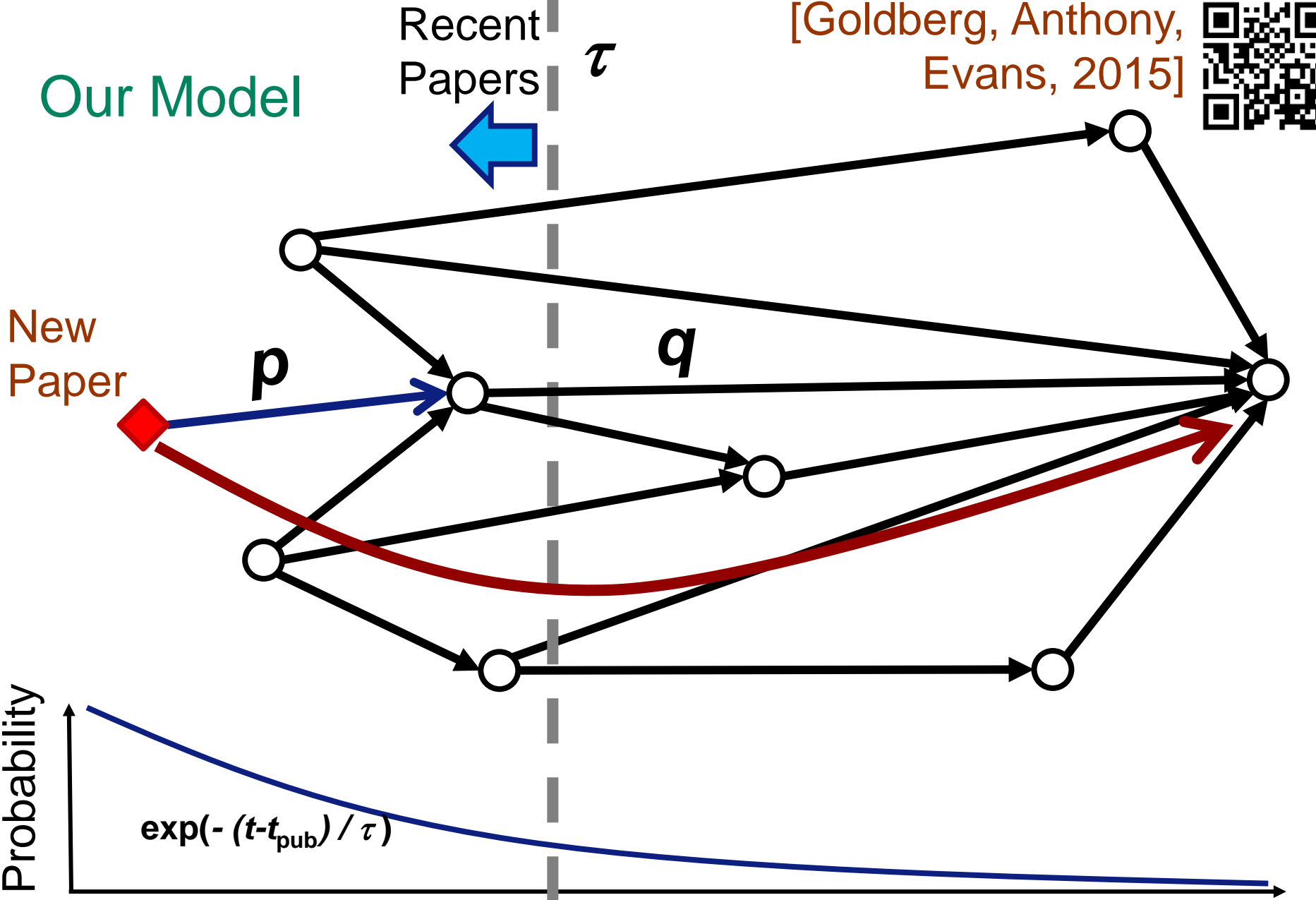


Tail Width Measure



Our Model

[Goldberg, Anthony, Evans, 2015]



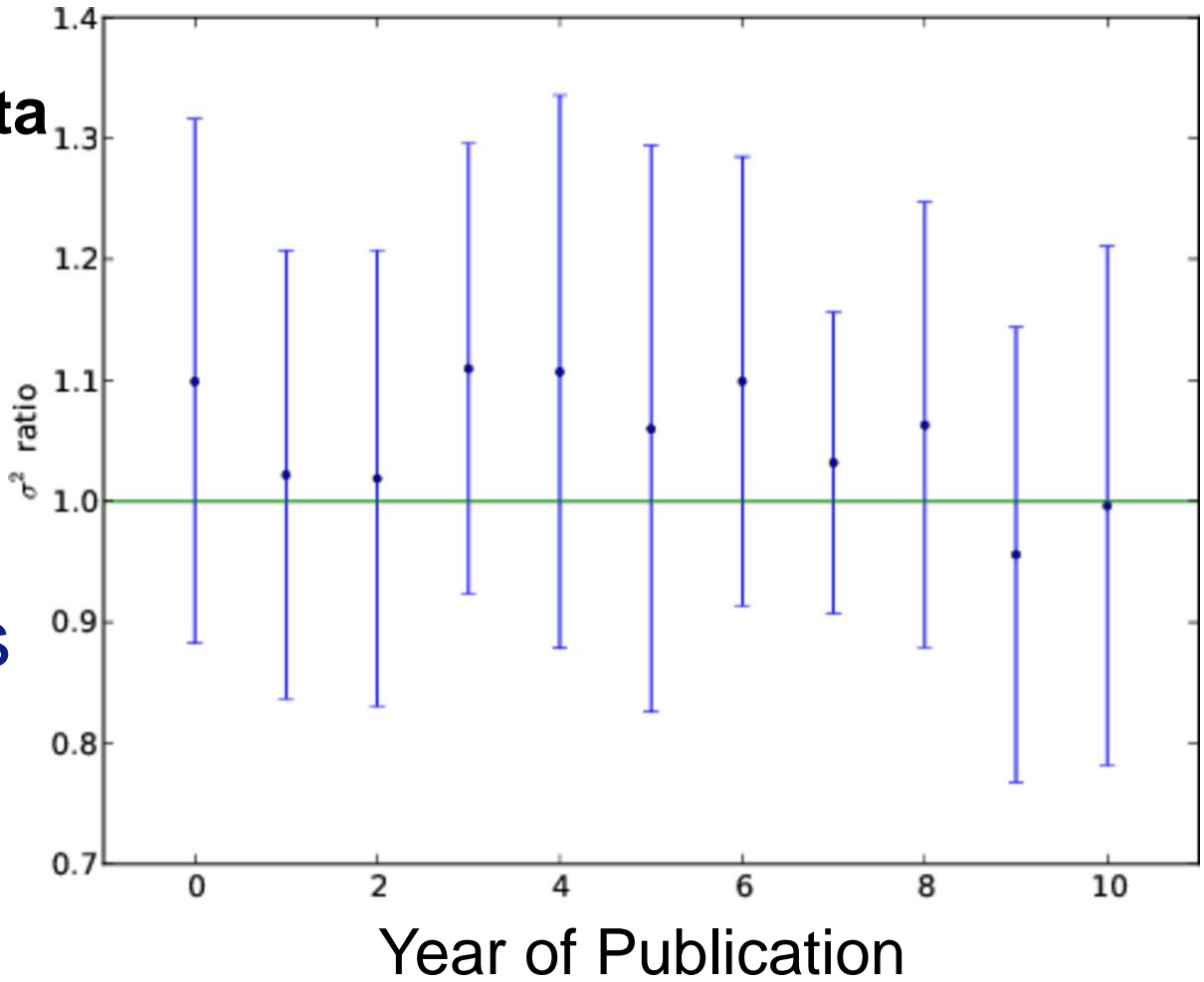
Results

[Goldberg, Anthony,
Evans, 2015]



$$\sigma / \sigma_{\text{data}}$$

Find
reasonable
fat tails for
publications
of all ages





Results

For **hep-th** with 2000 papers/year, $\langle c \rangle \approx 12.0$
we find

$$p=0.55, q=0.2, \tau_{1/2} = 200 \text{ papers}$$

\Rightarrow 70% of citations are 'copied' from recent papers

c.f. Simkin and Roychowdary (2005)

Clough, Gollings, Loach, Evans (2015)

TYPICAL PAPER

In Our Model

R. Smith

University of Fiction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam vehicula leo id nisi volutpat convallis. Aliquam quis nunc orci.

Accedite et ecce arbor in. Puella si potes invenire. Accedite et ecce tenebrae. Nam post te. Audio vocem invocabat nomen meum. Alta voce in tenebris. Arboribus primum occurrit vocem audio. Sed quia repente serum tincidunt. Silva perditus sum solus.

Puella non est, semper est idem. Non currit contra me, iterum atque iterum atque iterum.

References

- Random Recent Core Paper 1
- Random Recent Core Paper 2
- Popular Recent Core Paper 3
- Popular Recent Core Paper 4
- Copied from Core Paper 1
- Copied from Core Paper 1
- Copied from Core Paper 2
- Copied from Core Paper 3
- Copied from Core Paper 3
- Copied from Core Paper 3
- Copied from Core Paper 4
- Copied from Core Paper 4
- Copied from Core Paper 4
- Copied from Core Paper 4



A Better Null Model for Citations

- Mixes local and global knowledge
- Includes ageing
- Variable bibliography length (out-degree)

- Produces correct temporal behaviour for width of citation distribution
- Matches a decade of data from **hep-th**
⇒ A better null-model

Open Questions

- Bibliography length has an effect but ours was not realistic
 - correct mean but not fat-tailed
- Other academic paper data sets?
- Other types of citation data
- Short memory?
 - 5 weeks for hep-th data set

THANKS

Work done with: Sophia Goldberg and Hannah Anthony

- [Slides: 10.6084/m9.figshare.2061870](https://doi.org/10.6084/m9.figshare.2061870)



- Goldberg, S., Anthony, H. & Evans, T.S.,
Modelling Citation Networks,
Scientometrics, **2015**, *105*:, 1577-1604
[DOI: [10.1007/s11192-015-1737-9](https://doi.org/10.1007/s11192-015-1737-9) [arXiv:1408.2970](https://arxiv.org/abs/1408.2970)]



- Goldberg, S., Anthony, H. & Evans, T.S.,
Do We Need Global and Local Knowledge of the Citation
Network?
In *Proceedings of ISSI 2015*, **2015**, 282-283 see
[http://10.6084/m9.figshare.1452953](https://doi.org/10.6084/m9.figshare.1452953)

Bibliography

- Clough, J. R.; Gollings, J.; Loach, T. V. & Evans, T. S., Transitive reduction of citation networks, *J.Complex Networks*, **2015**, 3, 189-203 [[10.1093/comnet/cnu039](https://doi.org/10.1093/comnet/cnu039) [arXiv:1310.8224](https://arxiv.org/abs/1310.8224)]
- Clough, J.R. & Evans, T.S. What is the dimension of citation space? *Physica A* **2016** (to appear) [[arXiv:1408.1274](https://arxiv.org/abs/1408.1274)]
- Evans, T.S, Hopkins, N. & Kaube, B., Universality of Performance Indicators based on Citation and Reference Counts, *Scientometrics*, **2012**, 93, 473-495 [DOI: [10.1007/s11192-012-0694-9](https://doi.org/10.1007/s11192-012-0694-9) [arXiv:1110.3271](https://arxiv.org/abs/1110.3271)]
- Evans, T.S., Complex Networks, *Contemporary Physics*, **2004**, 45, 455-474 [[10.1080/00107510412331283531](https://doi.org/10.1080/00107510412331283531) [arXiv:cond-mat/0405123](https://arxiv.org/abs/cond-mat/0405123)]
- Garfield, E.; Sher, I. H. & Torpie, R. J., The use of citation data in writing the history of science, DTIC Document, 1964.
- Goldberg, S.; Anthony, H. & Evans, T.S., Modelling Citation Networks, *Scientometrics*, **2015**, 105:, 1577-1604 [DOI: [10.1007/s11192-015-1737-9](https://doi.org/10.1007/s11192-015-1737-9) [arXiv:1408.2970](https://arxiv.org/abs/1408.2970)]
- Goldberg, S.; Anthony, H. & Evans, T.S., Do We Need Global and Local Knowledge of the Citation Network? In Salah, A. A.; Tonta, Y.; Salah, A. A. A.; Sugimoto, C. & Al, U. (Eds.) , *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference, Istanbul, Turkey, 29 June to 3 July, 2015*, **2015**, 282-283 see <http://10.6084/m9.figshare.1452953>
- Price, D. J. d. S. The scientific foundations of science policy. *Nature*, **1965**, 206, 233-238
- Radicchi, F.; Fortunato, S. & Castellano, C. Universality of citation distributions: Toward an objective measure of scientific impact, *PNAS*, **2008**, 105, 17268--17272
- Simkin, M. V. & Roychowdhury, V. P., Copied citations create renowned papers? *Annals of Improbable Research*, **2005**, 11, 24-27
- Simkin, M. V. & Roychowdhury, V. P., Stochastic modeling of citation slips, *Scientometrics, Springer*, **2005**, 62, 367-384