Wirtschaftsinformatik
Information Systems

GOETHE
UNIVERSITÄT
FRANKFURT AM MAIN

Open Access to Scientific Literature

Increasing Citations as an Incentive for Authors to Make their Publications Freely Accessible

Steffen Bernius

Waikoloa, January 6th, 2009

Agenda

1. Introduction: The Open Access Paradigm
2. Motivation & Research Question
3. Research Design & Theoretical Grounding
4. Results and Discussion

Introduction: Open Access

Why are alternative publishing models discussed?

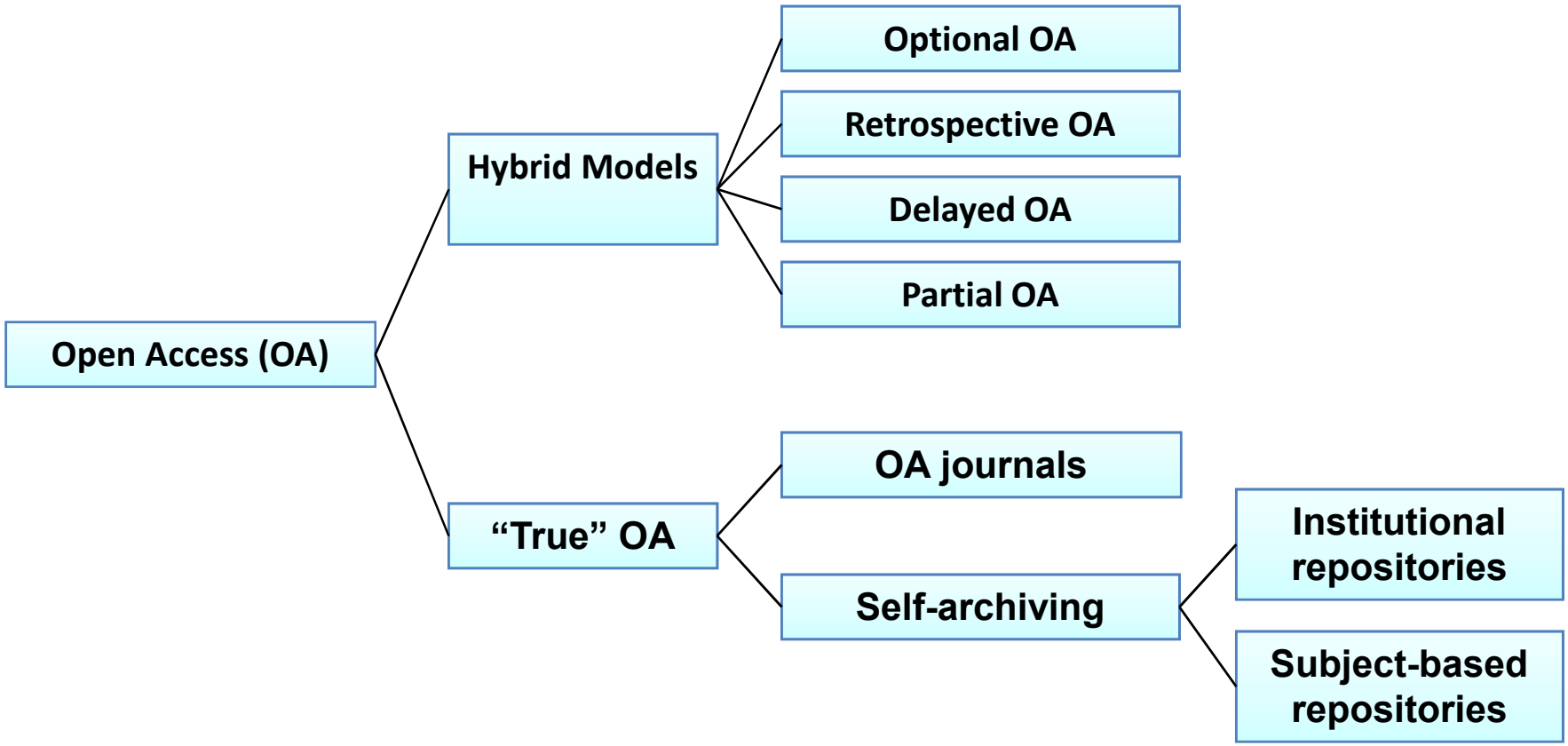
- New possibilities of content distribution enabled by the **internet**
- **"Serials crisis"** (rapidly rising journal prices + declining library budgets > cancellations of journal subscriptions)

What is the idea behind Open Access?

- Regain access to research findings in order to enhance scientific communication
- Initial definition by the Budapest Open Access Initiative (BOAI):
  - *By Open Access to scientific publications "we mean its **free availability on the public internet**,*
  - ***permitting any users to read**, download, copy, distribute, print, search, or link to the full texts of **these articles**, (…)*
  - ***without** financial, legal, or technical **barriers** other than those inseparable from gaining access to the internet itself."* [BOAI 2001]

Introduction: Open Access

- Ways of Open Access Publishing:



Agenda

1. Introduction: The Open Access Paradigm
2. **Motivation & Research Question**
3. Research Design & Theoretical Grounding
4. Results and Discussion

Motivation & Research Question

- Focus: Scientists as authors and readers of scientific literature
- Although authors **attitude toward OA is very positive**…
  - Wide and rapid dissemination of research findings
  - Reaching a broad readership

  [Schroter et al. 2005, Mann et al. 2008, Moed 2006]
- …they face a **lack of incentives** to make OA
  - "wait-and-see behavior" > critical mass problem
    [Mann et al. 2008]
  - "prisoner's dilemma" of authors in non-OA communities
    [Hanauske et al. 2007]
- On the other hand: Strong empirical evidence that **OA papers get cited more often** [e.g., Lawrence 2001, Harnad et al. 2004/2005, Kurtz et al. 2005, Eysenbach 2006]

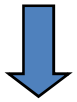
Research question: Which effect has a change in publication strategy for an individual author in a non-OA community ?

Agenda

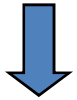
1. Introduction: The Open Access Paradigm
2. Motivation & Research Question
3. **Research Design & Theoretical Grounding**
4. Results and Discussion

Research Design

Developing a computational model of the traditional scientific publishing market based on the interactions between authors, journals, and libraries



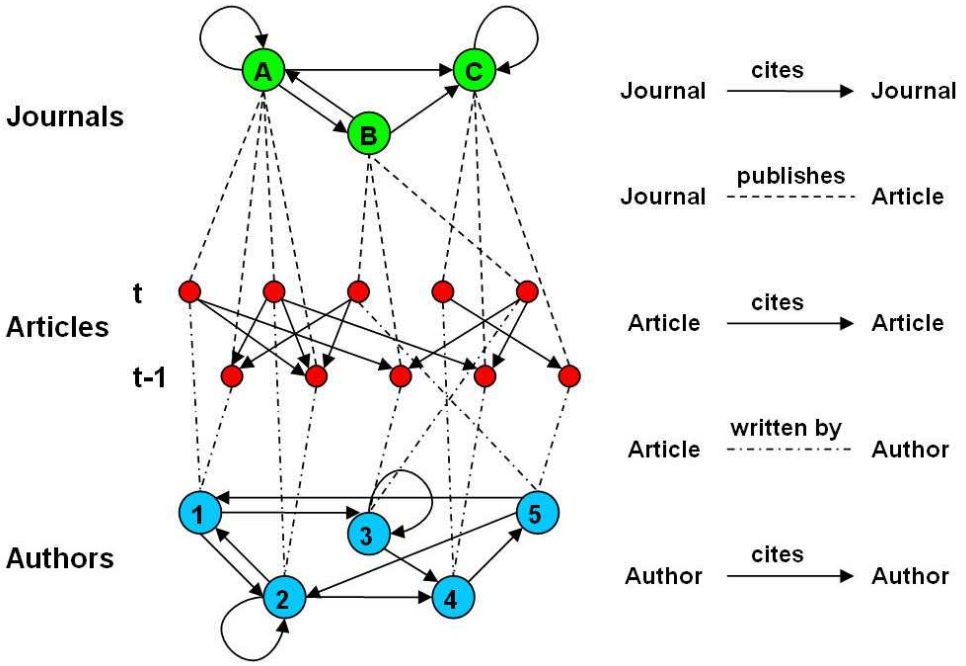
Simulating the emergence of complex citation networks



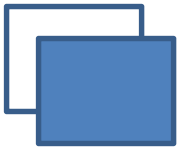
Experimentation: Measure effects of a switch from traditional publishing to Open Access on the author level

Characteristics of Citation Networks

- Structure of the network:



- Empirical findings [e.g., Redner 1998, Laherre/Sonette 1998, Lehmann et al. 2003]
  - Scale-free network
  - Long-tail distribution: majority of papers is seldom cited, only few papers are cited frequently



Characteristics of Citation Networks

- Preferential Attachment
  - plays a critical role during network evolution
  - Matthew Effect [Merton 1968]
  - Cumulative Advantage [Cole/Cole 1973]

- Modeling the emergence of scale-free (citation) networks
  - Barabasi/Albert Model (1999): Formal description of the evolution of scale free networks based on the preferential attachment mechanism
  - Restrictions of B/A Model:
    [Pujol et al. 2005]
    - mainly based on graph theoretic concepts
    - the individual agent is a mindless actor (“node”) and always has complete information about the structure of the whole network
    - Preferential attachment is modeled “directly” – not as a result of the behavior of individual agents

OA citation advantage

- Causal explanations of higher citation rates of OA papers
  [Overview: Craig et al. 2007]

  - **“Open Access postulate”** > OA articles are cited more often simply because they are more likely to be read

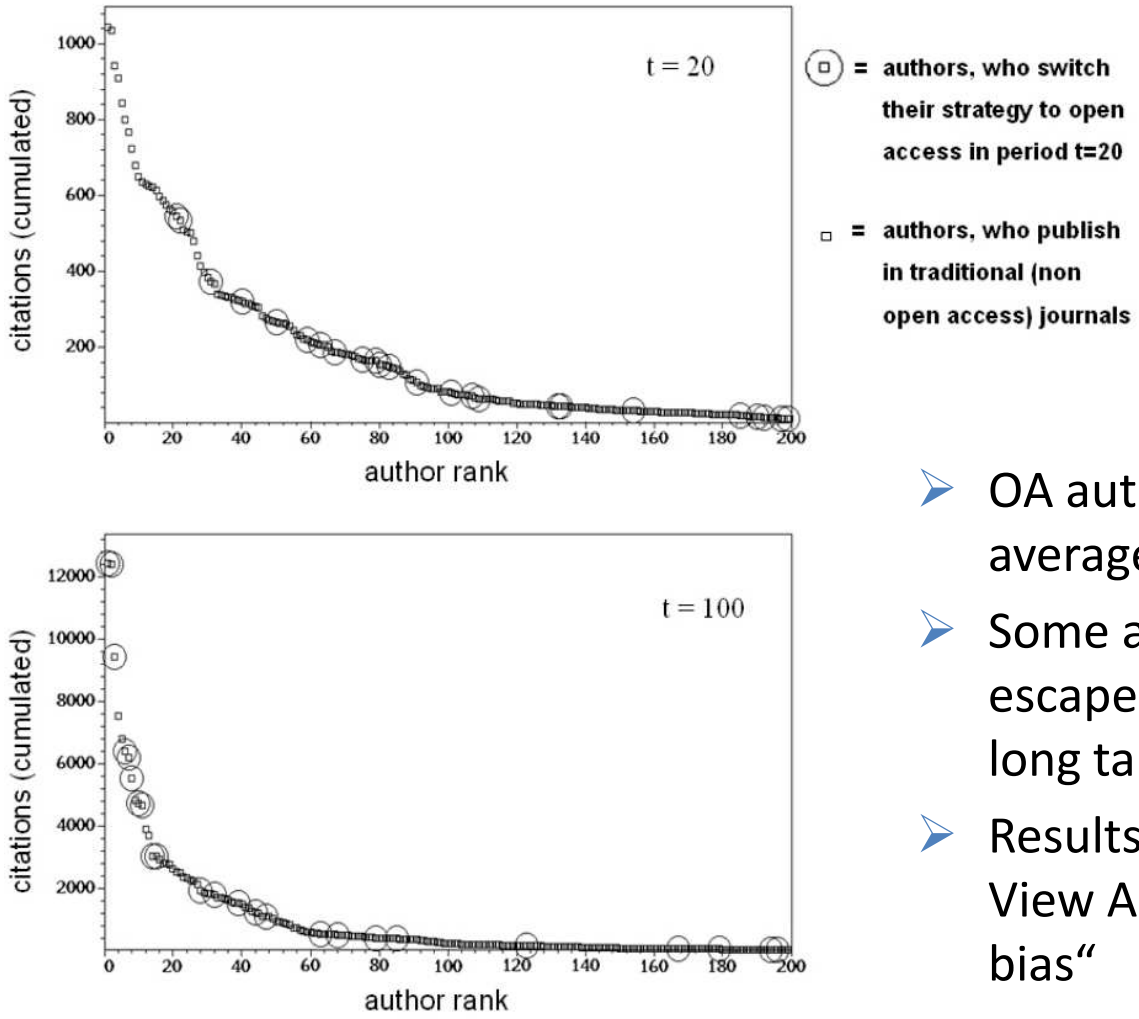
  - **“Early View postulate”** > refers to self-archiving; early posting of an article in a pre-print repository allows time for earlier citation

  - **“Selection Bias postulate”** > prominent authors are more likely to make their articles available under an OA model, and they are more likely to do so with their most important (and thus most citable) articles.

Agenda

1. Introduction: The Open Access Paradigm
2. Motivation & Research Question
3. Theoretical Grounding & Methodology
4. **Results and Discussion**

Simulation Results



- OA authors improve their average rank from 100 to 68
- Some authors even manage to escape from the bottom of the long tail
- Results not caused by „Early View Advantage“ or „selection bias“

Summing-Up

Practical Conclusions:

- By widening the scope from the article level to the author level potential benefits of OA become visible
- Especially in non-OA communities, unrestricted access to his/her work can broaden an author's impact – at least in terms of increasing citations
- "First movers" may benefit (preferential attachment)

Theoretical contribution:

- Controlling for Early View Effect and Selection Bias
- Simulating the emergence of complex networks as a consequence of plausible sociological micro-processes

Research in progress:

- Evaluation of alternative publication models from the viewpoint of other market actors (publishers, libraries)

Wirtschaftsinformatik
Information Systems

GOETHE
UNIVERSITÄT
FRANKFURT AM MAIN

Thank you for your attention.

Steffen Bernius
Goethe University Frankfurt
Institute of Information Systems
www.is-frankfurt.de

bernius@wiwi.uni-frankfurt.de

Back Up >

Literature (1/2)

- A.-L. Barabasi and R. Albert, “Emergence of scaling in random networks”, Science 286, 1999, pp. 509-512.
- S. Bernius and M. Hanauske, “Open Access”, Wirtschaftsinformatik, 47(6), 2007, pp. 456-459.
- T. Brody, S. Harnad, and L. Carr, “Earlier Web Usage Statistics as Predictors of Latter Citation Impact”, Journal of the Association for Information Science and Technology (JASIST), 57(8), pp. 1060-1072.
- Budapest Open Access Initiative (BOAI), http://www.soros.org/openaccess/read.shtml
- J.R. Cole and S. Cole, “Social Stratification in Science”, The University of Chicago Press, London, Chicago, 1973.
- I.D. Craig, A.M. Plume, M.E. McVeigh, J. Pringle, and M. Amin, “Do Open Access Articles have a Greater Research Impact? A critical review of the literature”, Journal of Informetrics, 1 (3), 2007, pp. 239-248.
- G. Eysenbach, “Citation Advantage of Open Access Articles”, PLoS Biology, 4 (2006) 5, pp. 692-698.
- J. Guedon, “The “Green” and “Gold” Roads to Open Access: The Case for Mixing and Matching”, Serials Review, 30(4), 2004, pp. 315-328.
- C. Hajjem, S. Harnad, and Y. Gingras, “Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact”, IEEE Data Engineering Bulletin, 18(4), 2005, pp. 39-47.
- M. Hanauske, S. Bernius, and B. Dugall, “Quantum Game Theory and Open Access Publishing”, Physica A, 382(2), 2007, pp. 650-664.
- S. Harnad, “Fast-Forward on the Green Road to Open Access: The Case Against Mixing Up Green and Gold”, Ariadne, 42 (2005), http://www.ariadne.ac.uk/issue42/har-nad/intro.html.
- S. Harnad, T. Brody, F. Vallières, L. Carr, S. Hitchcock, Y. Gingras, C. Oppenheim, C. Hajjem, and E.R. Hilf, “The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update”, Serials Review, 34(1), 2008, pp. 36-40.
- S. Harnad and T. Brody, “Comparing the Impact of Open Access(OA) vs. Non-OA Articles in the Same Journals”, D-Lib Magazine, 10(6), 2004, http://www.dlib.org/dlib/june04/harnad/06harnad.html.
- O. Hein, M. Schwind, and W. König, „Scale-Free Networks – The Impact of Fat Tailed Degree Distribution on Diffusion and Communication Processes”, Wirtschaftsinformatik, 48(4), 2006, pp. 267-275.
- D.W. King and C. Tenopir, “Evolving Journal Costs: Implications for Publishers, Libraries, and Readers”, Learned Publishing, 12(4), 1999, pp. 251-258.

Literature (2/2)

- M.J. Kurtz, G. Eichhorn, A. Accomazzi, C.S. Grant, M. Demleitner, and S.S. Murray, “Worldwide use and impact of the NASA Astrophysics Data System digital library”, Journal of the American Society for Information Science and Technology, 56(1), 2005, pp. 36-45
- M.J. Kurtz, G. Eichhorn, A. Accomazzi, C.S. Grant, M. Demleitner, E. Henneken, and S.S. Murray, “The effect of use and access on citations”, Information Processing and Management: an International Journal, 41(6), 2005, pp. 1395-1402.
- J. Laherre and D. Sonette, „Stretched exponential distributions in nature and economy: ‚fat tails’ with characteristic scales”, The European Physics Journal B, 2 (1998), pp. 525-539.
- C. Lamb, “Open Access Publishing Models: Opportunity or Threat to Scholarly and Academic Publishers?”, Learned Publishing, 17(2), 2004, S. 143-150.
- S. Lawrence, “Free online availability substantially increases a paper's impact”, Nature 411 (31 May 2001), S. 521-522.
- S. Lehmann, B. Lautrup, and A.D. Jackson, „Citation networks in high energy physics”, Physical Review E, 68 (2003), p. 026113 (8 pages).
- E.A. Leicht, G. Clarkson, K. Shedden, and M.E.J. Newman, “Large-Scale Structure of Time Evolving Citation Networks”, The European Physics Journal B, 59(1), 2007, pp. 75-83.
- F. Mann, B. von Walter, T. Hess, and R.T. Wigand, “Open Access Publishing in Science: Why it is highly appreciated but rarely used”, Communications of the ACM, Vol. 51 (accepted for publication), 2008.
- R.K. Merton, “The Matthew Effect in Science”, Science, 159(3810), 1968, pp. 56-63.
- A.M. Odlyzko, “The Rapid Evolution of Scholarly Communication”, Learned Publishing, 15 (2002), pp. 7-19.
- J.M. Pujol, A. Flache, J. Delgado, and R. Sangüesa, “How can social networks ever become complex? Modelling the emergence of complex networks from local social exchanges”, Journal of Artificial Societies and Social Simulation (JASSS), 8(4), 2005, http://jasss.soc.surrrey.ac.uk/8/4/12.html.
- S. Redner, “How popular is your Paper? An empirical study of the citation distribution”, The European physical Journal B, 4 (1998), pp. 131-134.
- S. Schroter, L. Tite, and R. Smith, “Perceptions of open access publishing: interviews with journal authors, BMJ; 330 (2005), pp. 756-759.
- C. Tenopir and D.W. King, “Towards Electronic Journals: Realities for Scientists, Librarians and Publishers”, Special Libraries Association, 2000.
- S. White and C. Claire, “Trends in Scholarly Journal Prices 2000-2006”, LISU, Loughborough, 2007.

Dilemma of authors in non-OA communities

General open access payoff matrix

A\B	o	ø
o	$(r+\delta, r+\delta)$	$(r-\alpha, r+\beta)$
ø	$(r+\beta, r-\alpha)$	(r, r)

Assumptions:
- main objective of scientists is the maximization of their reputation
- reputation is primarily gained by publishing in top journals
- in non-OA communities the top journals do not support OA

Parameters:
o = player makes OA
ø = player does not make OA
r = actual reputation of the two scientists
α = reputation decrease, if choosing OA strategy alone
β = reputation increase, if the other player chooses OA strategy alone
δ = potential benefit, if both players choose OA

Different game settings always lead to dominant strategies of ø with (ø ø) as Nash equilibrium:

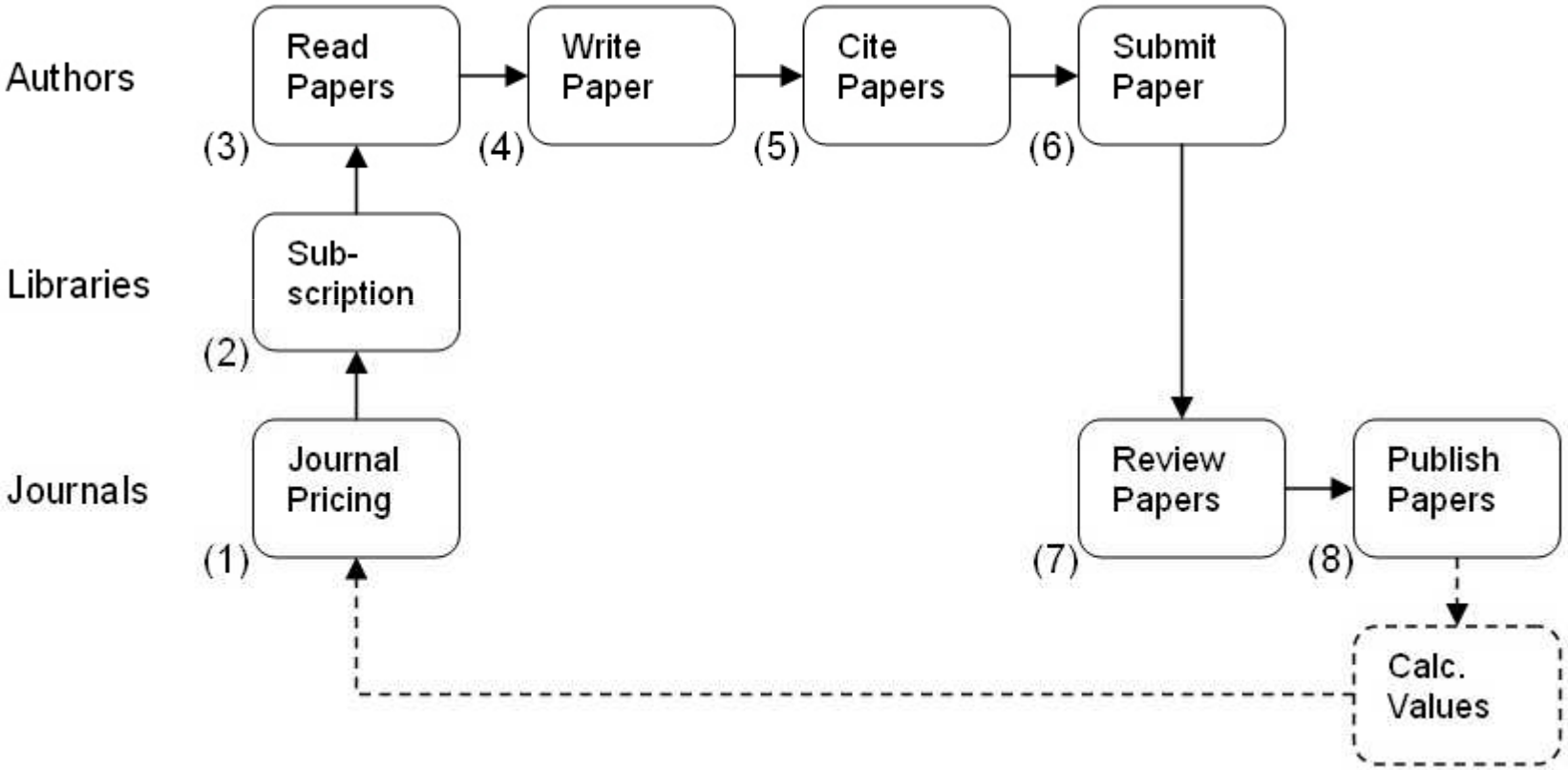
A\B	o	ø
o	(0,0)	(−1,1)
ø	(1,−1)	**(0,0)**

Zero sum game (reputation as a relative quantity; r = 0, δ = 0, α = β = 1)

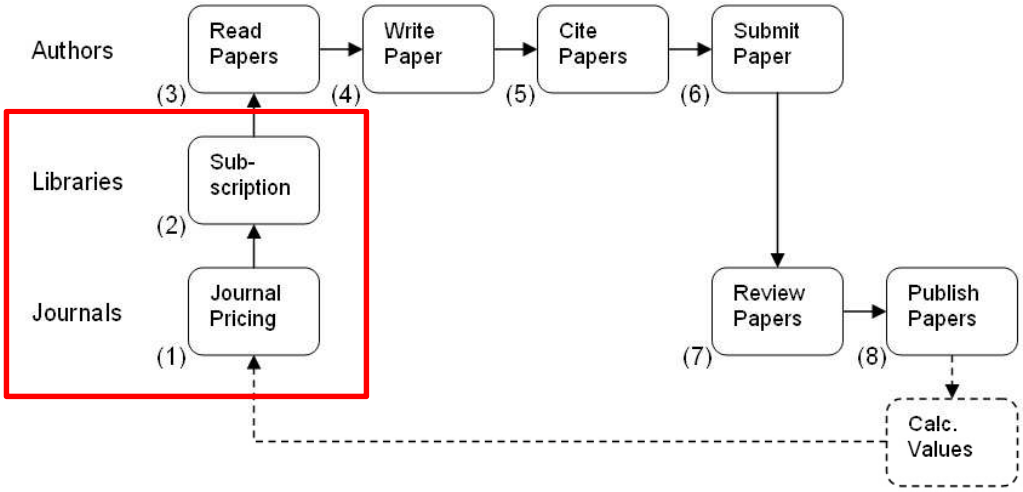
A\B	o	ø
o	(4,4)	(1,5)
ø	(5,1)	**(3,3)**

Prisoners' dilemma (reputation increases, if both choose OA; e.g. with r = 3, δ = 1, α = β = 2)

Simulation Phases (1/5)



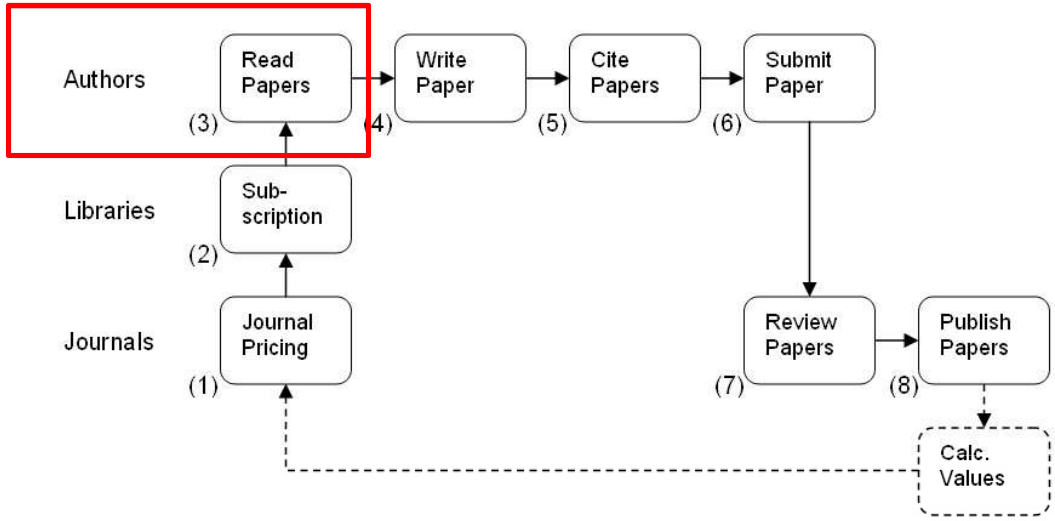
Simulation Phases (2/5)



(1) At the start of a simulation period the publishers set the price of their journals. In this regard they can choose between different pricing strategies:
- constant price increase over time,
- price decrease if subscriptions decrease
- price increase when journal subscriptions decrease.

(2) The Libraries have a fixed budget per period, which they can use to subscribe to journals. In this context, it can also be specified, whether this budget is constant, decreases, or increases over time.

Simulation Phases (3/5)

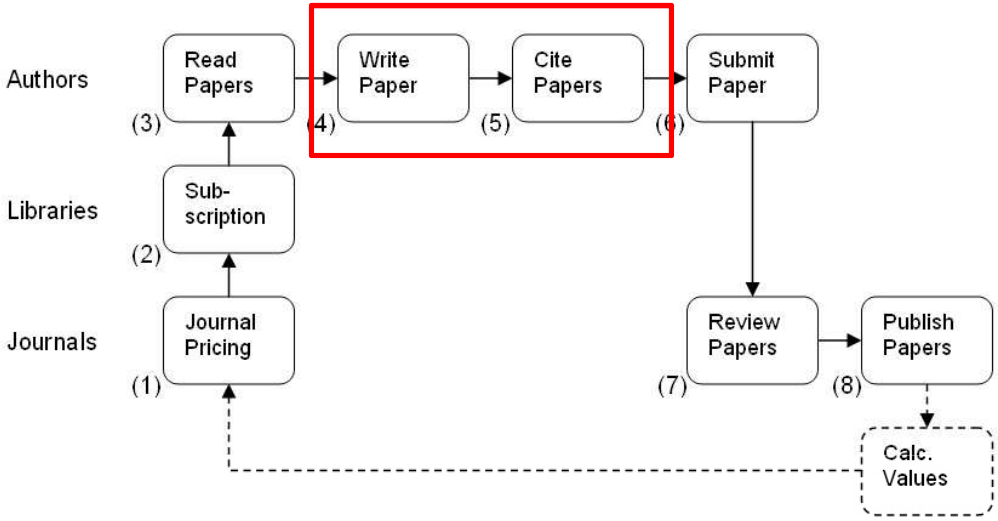


(3) At the authors' level the production process of a new article starts with reading a specific number of papers.

Assumptions:

- Authors can only read those articles, which are published in a journal licensed by the author's library, or which are archived in an open access repository.
- When selecting articles for reading, authors mainly resort to journals with a relatively high reputation (which are publishing high quality papers).

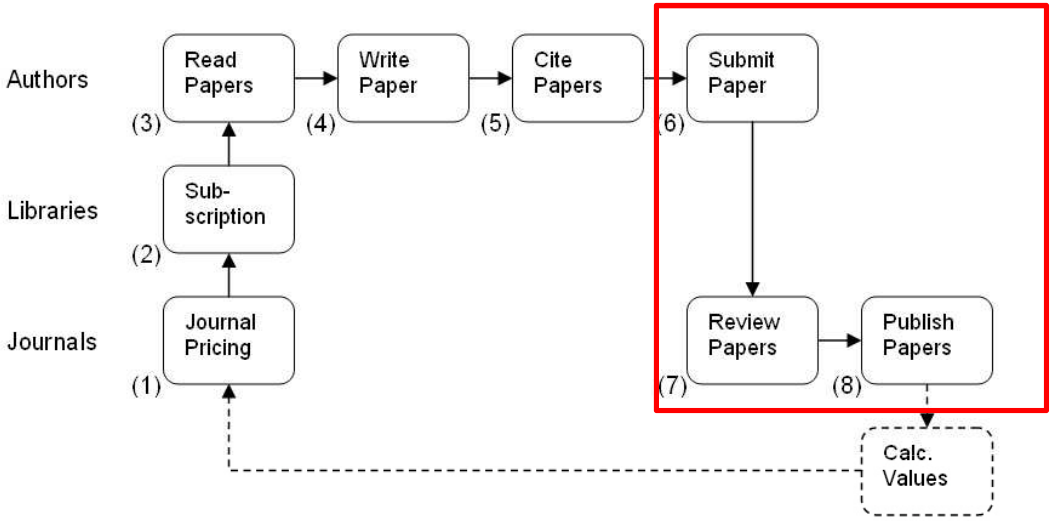
Simulation Phases (4/5)



(4) Every author writes one article per period, so that the article network (and in consequence the citation network) grows constantly over time.
The quality of a new written article varies from author to author (implemented by using a stochastic process based on a geometric Brownian motion).

(5) In the citation phase the author cites a specific number of papers from the current period as well as papers she has read in past periods.
The probability of an article to be cited decreases with the age and increases with the number of citations the article has received in the past ("preferential attachment").

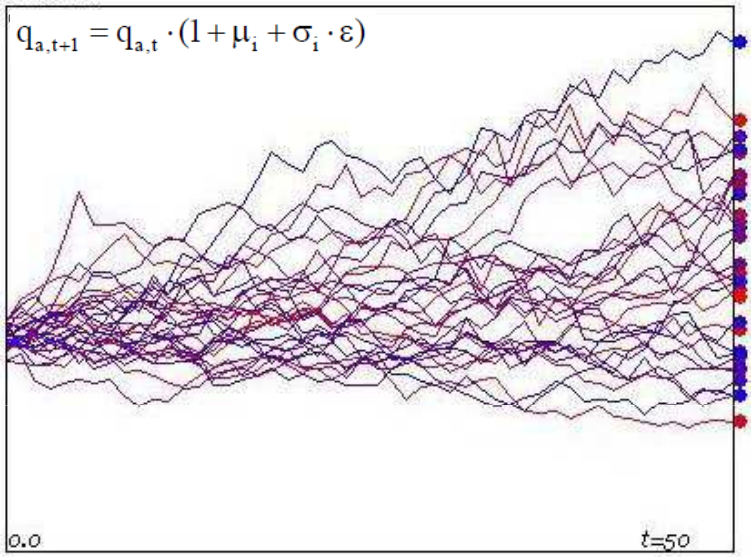
Simulation Phases (5/5)



(6) The last phase at the authors' level is the submission of the new paper to an appropriate journal. The decision where to submit depends firstly on the relation between the quality of the article and the average quality of the articles published in a specific journal, and secondly on the expected gain in reputation if the article is accepted.

(7+8) Subsequently the review phase starts and the journals accept a given number of articles, which then are immediately published.

Calculation of author/paper qualities



Example of quality evolution over simulation time:

- 30 Authors/Agents, 50 periods,
- σ_i = 0,05 (Volatility of the article quality of author i)
- μ_i = 0,005 (Average increase of article quality of author i)
- with ε as a distributed random number

> the quality of a new written article depends on the quality of the articles previously written by the author

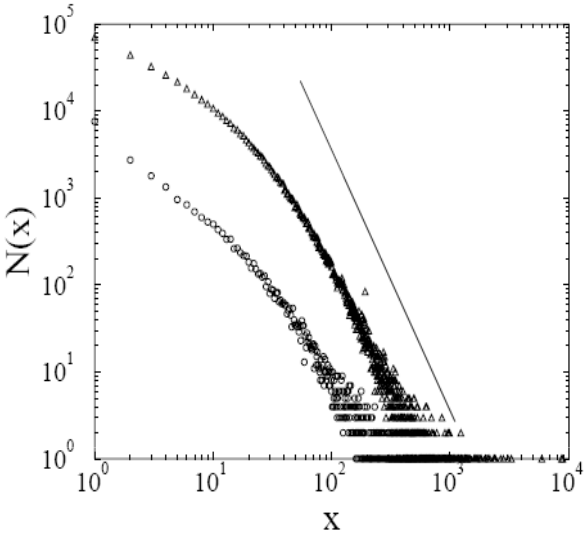
Model Parameters and Variables

Parameters and variables		
Parameters (fixed values)	T	Simulation duration
	I	Number of authors
	J	Number of Journals
	L	Number of Libraries
	σ_i	Volatility of the article quality of author i
	μ_i	Average increase of article quality of author i
	ε	distributed random number
	na_i	capacity per period of author i
	na_j	Issue capacity of journal j
	nc^{new}	Citation rate of read papers of period t
	nc^{old}	Citation rate of read papers of past periods
Input variables (independent variables)	$PriceStrat_{j,t}$	Pricing Strategy of journal j
	$SubStrat_{l,t}$	Subscription strategy of library l at time t
	$PubStrat_{i,t}$	Publication strategy of author i at time t
	$B_{a,t=0}$	Starting budget of Library l
Intermediate variables (according to Barton 2004)	q_a	Quality of article a
	$b_{l,t}$	Budget of library l at time t
	$\tau_{a,t}$	Age of article a at time t
	$k_{a,t}$	Cumulated cites of article a at time t
	$\Pi_{a,t}$	Citation probability of article a at time t
	$r_{j,t}$	Reputation of journal j at time t
Output variables (dependent variables)	$C_{i,t}$	Cumulated cites of author i at time t
	$c_{j,t}$	Cites of journal j at time t
	$P_{j,t}$	Price of journal j at time t
	$U_{j,t}$	Usage of journal j at time t

Model Assumptions

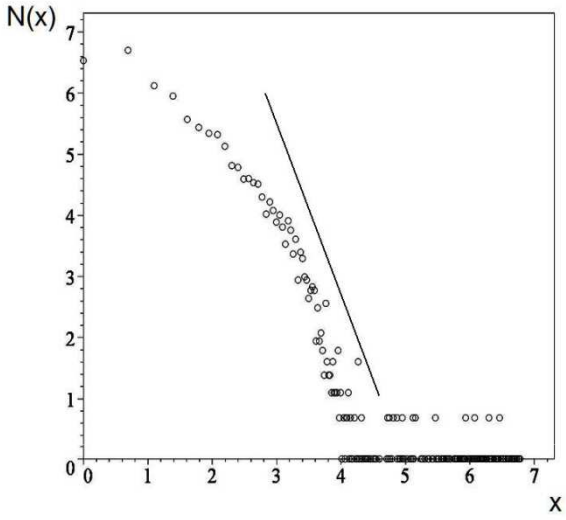
Assumptions of the Simulation Model	
general	- when reading articles authors prefer journals with relatively high reputation - the quality of an article depends on the quality of the articles previously written by the author - journals decision to publish an article depends only on the article's quality - reputation of journals depends only on the quality of the articles published - no distinction between print and online journals
"technical"	- every author reads a certain number of articles per period - authors write exactly one article per period - only single author papers - every new article cites a certain number of other articles - authors know the quality of their articles (better/worse than another article) - authors know the reputation/quality hierarchy of the journals - every journal releases one issue per period - every issue has a maximum capacity of articles - the article's quality can be exactly determined by the referees - the review process is accomplished within one period - authors can only read those papers, which are OA or which are published in journals licensed by their library - rejected articles are not submitted to another journal - accepted articles get published in the period of acceptance

Empirical validation



Citation distribution of two **real citation networks** (Δ= ISI database, o= Physical Review D) in double logarithmical representation; with x = number of citations, and N(x) = number of articles with x citations.

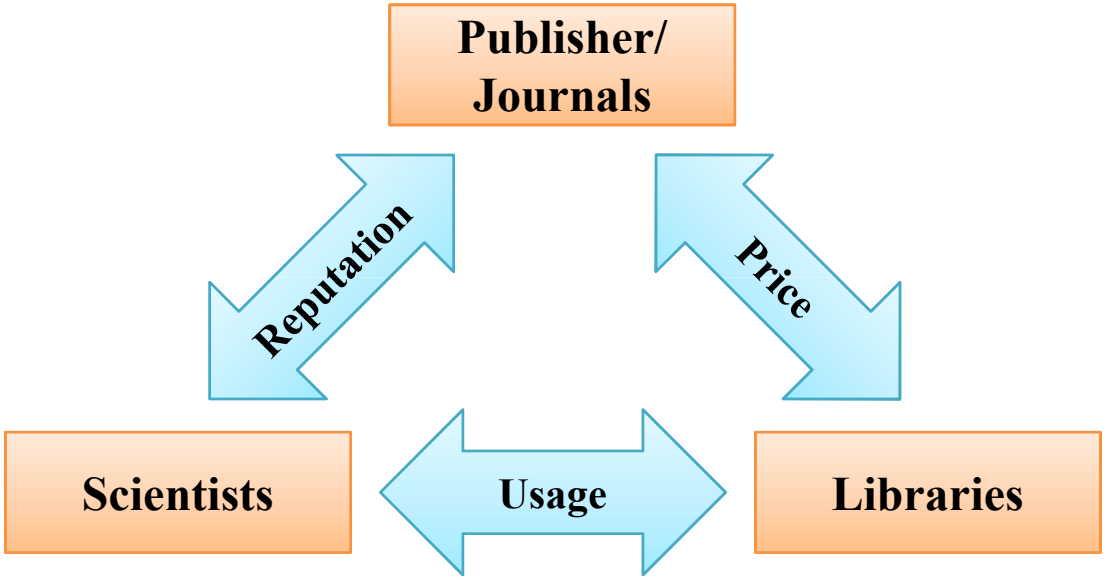
[Redner 1998]



Distribution of the citation network evolving during the **simulation**; double logarithmical representation with x = number of citations and N(x) = number of articles with x citations; for visual reference a straight line of slope -3 is also shown

The network emerging in the simulation matches an important characteristic of real citation networks: a large part of the citation distribution (particularly the asymptotic tail) can be described by a power-law of the form $N(x) \sim x^{-\alpha}$ with α close to 3.
The bigger the simulation network gets over time, the better it describes real distributions.

Market Players and Coordinating Mechanisms



[Bernius et al. 2009]

Journal Prices

2.10 Median journal prices 2000-06

	2000	2001	2002	2003	2004	2005	2006	*% change 2000-06*	No incl.
U. of Chicago	54	61	76	78	75	115	119	*119.7%*	16
Blackwell	127	145	167	193	219	242	263	*107.1%*	210
Sage	179	212	220	286	328	342	359	*100.8%*	162
Taylor & Francis	155	174	191	210	235	265	300	*93.5%*	373
Springer	147	163	180	185	210	221	243	*65.5%*	48
Oxford Journals	105	114	123	140	149	158	173	*65.1%*	48
Wiley	318	381	436	383	420	471	513	*61.2%*	39
Cambridge UP	79	86	95	99	100	111	122	*54.5%*	29
Lippincott	187	226	264	260	248	260	283	*51.5%*	17
Elsevier	314	337	360	394	415	436	464	*47.4%*	246

[White/Creaser 2007]

Self-Archiving: the example of arXiv.org

12/2006

New paper

Archiving
on OA-Server

Journal
Submission

Cornell University Library
arXiv.org

PHYSICA A
STATISTICAL MECHANICS AND ITS APPLICATIONS
Recognized by the European Physical Society

- Declaration: „submitted to …“
- „Public Review“
- Updating the original version is possible
- Declaration: „accepted by …“

- Peer Review
- Acceptance : April 13th 2007
- Revision
- Date of Publication: August 17th 2007

04/2007

08/2007

Business Models of OA Journal Publishers

Publisher	Type	# Journals	Main income sources	Publication fees [in US$]
Biomed Central	commercial	195	publication fees	0-2145 (mean: 1350)
PLoS	non-commercial	7	publication fees, grants	2000-2500 (mean: 2257)
Hindawi Publishing Corporation	commercial	150	publication fees, print subscriptions	0-500
Medknow Publications	commercial	65	print subscriptions, advertising	0

[Bernius et al. 2009]

IS Conferences and OA (2003-2007)

Konferenz	Jahr	Open Access		Elektronischer Zugriff über . . .
		ja	nein	
ICIS	2003–2006		×	AIS Electronic Library
AMCIS	2003–2006		×	AIS Electronic Library
ECIS	2003–2007	×		http://is2.lse.ac.uk/asp/aspecis
PACIS	2003–2005		×	AIS Electronic Library
	2006	×		Konferenzseite
	2007		×	–
HICSS	2003–2007	×		IEEE Digital Library
MKWI/WI	2003		×	–
	2004		×	–
	2005	×		Konferenzseite
	2006		×	–
	2007	×		Konferenzseite

[Bernius/Hanauske 2007]