

**Netzwerke in der Wissenschaft – Auswirkungen von Open Access
auf die Verbreitung von Forschungsergebnissen**

Wolfgang König, Steffen Bernius, Matthias Hanauske

Inhaltsverzeichnis

- 1 Einleitung: Open Access als neues Publikationsparadigma
- 2 Charakteristika von Zitationsnetzwerken
- 3 Preferential Attachment und Open Access
 - 3.1 Simulationsmodell
 - 3.2 Ergebnisse
- 4 Zusammenfassung

1. Einleitung: Open Access als neues Publikationsparadigma

Wesentliches Ziel der akademischen Forschung ist die Generierung neuer wissenschaftlicher Erkenntnisse und die auf Interaktion zwischen den Forschern basierende Teilung bzw. Verbreitung dieses Wissens. Hierbei lassen sich zwei zentrale Interaktionsmodi identifizieren: Kooperation bei der Erzeugung von Forschungsergebnissen (Ko-Autorenschaften) und Zitation von zuvor geleisteten Arbeiten im Rahmen der Manuskripterstellung [vgl. Knackstedt/Winkelmann 2006]. Aus diesen Interaktionsformen lassen sich jeweils komplexe soziale Netzwerkstrukturen ableiten, die zur strukturellen Beschreibung der Wissensverbreitung in einem Fachgebiet herangezogen werden können. Dies sind zum einen Ko-Autorenschaften-Netzwerke [vgl. z. B. Newmann 2001; Barabasi et al. 2002] und zum zweiten Zitationsnetzwerke [vgl. z. B. Redner 1998; Lehmann et al. 2003], welche im vorliegenden Beitrag fokussiert werden. Zitationen sind hierbei ein wichtiger Indikator für die Reputation eines Wissenschaftlers: Je häufiger seine Werke zitiert werden (bzw. je häufiger eine Zeitschrift zitiert wird, in welcher er publiziert), desto höher ist sein Ansehen innerhalb der Community. Um in diesem Reputationswettbewerb zu bestehen, sollte ein Autor also Strategien verfolgen, die eine Erhöhung der Zitate auf seine Arbeit wahrscheinlich machen.

In diesem Zusammenhang spielen aktuelle Entwicklungen auf dem Markt für wissenschaftliche Fachpublikationen eine wichtige Rolle. So wird das traditionelle Publikationsmodell, wonach wissenschaftliche Forschungsergebnisse ausschließlich in abonnementspflichtigen Fachzeitschriften veröffentlicht werden, zunehmend bedrängt von alternativen Veröffentlichungsquellen, die im Zeichen des Open-Access-Paradigmas stehen [vgl. z. B. Bernius/Hanauske 2007]. Open Access bedeutet, dass wissenschaftliche Artikel kostenfrei und für jeden Interessierten zugänglich im Internet zur Verfügung gestellt werden. Dies kann zum einen über Einreichung eines Artikels in einem Open Access Journal („Golden Road to Open Access“), zum anderen über die Selbstarchivierung des Artikels in institutionellen oder fachspezifischen Repositories („Green Road to Open Access“) erreicht werden [Harnad 2005]. Verschie-

dene Studien haben gezeigt, dass Open-Access-Artikel schneller und vor allem häufiger zitiert werden [vgl. z. B. Lawrence 2001; Hajjem et al. 2005; Eysenbach 2006; Moed 2007]. Befragungen unter Wissenschaftlern zeigen, dass Open Access zwar befürwortet wird – die Umsetzung des neuen Paradigmas geht in den meisten Fachgebieten aufgrund mangelnder Anreize für Autoren jedoch nur sehr schleppend voran.

Im vorliegenden Beitrag sollen daher folgende Forschungsfragen näher betrachtet werden: Wie wirkt sich Open Access auf die Struktur von Zitationsnetzwerken und damit auf die Verbreitung von neuen Forschungsergebnissen aus und inwiefern lassen sich daraus Anreize für Autoren ableiten, ihre Artikel nach dem Open-Access-Paradigma zu veröffentlichen? Hierzu werden zunächst die Eigenschaften von Zitationsnetzwerken – insbesondere im Hinblick auf deren typische „Long-Tail-Verteilung“ – beschrieben. In einem zweiten Schritt soll basierend auf Simulationsergebnissen gezeigt werden, wie sich die Position eines Autors in einem Zitationsnetzwerk in Abhängigkeit von seiner Publikationsstrategie verändern kann.

2. Charakteristika von Zitationsnetzwerken

In einem Zitationsnetzwerk repräsentieren die Knoten die Artikel, während die Verbindungslinien zwischen den Knoten die Zitate eines Artikels auf einen anderen darstellen. Zitationsnetzwerke lassen sich durch drei wesentliche Eigenschaften charakterisieren und von anderen Netzwerken abgrenzen [vgl. z. B. Leicht et al. 2007]:

- die Kanten des Netzwerkes sind *gerichtet*;
- das Netzwerk ist *evolutorisch*, d. h. mit jedem neu veröffentlichten Artikel kommt ein neuer Knoten mitsamt den zugehörigen Kanten hinzu, was zu einem stetigen Wachstum des Netzes führt;
- das Netzwerk ist *azyklisch* (d. h. es weist keine Schleifen auf), da ein neues Paper immer nur die bereits existierenden Artikel zitieren kann (eine Ausnahme bilden Zitate auf „forthcoming papers“, die jedoch aufgrund ihres geringen Anteils vernachlässigt werden können).

Abbildung 1 schematisiert das Zitationsnetzwerk mit Einbeziehung der Knotenarten „Journals“ und „Autoren“, welche sich jeweils wiederum einzelnen Artikeln zuordnen lassen.

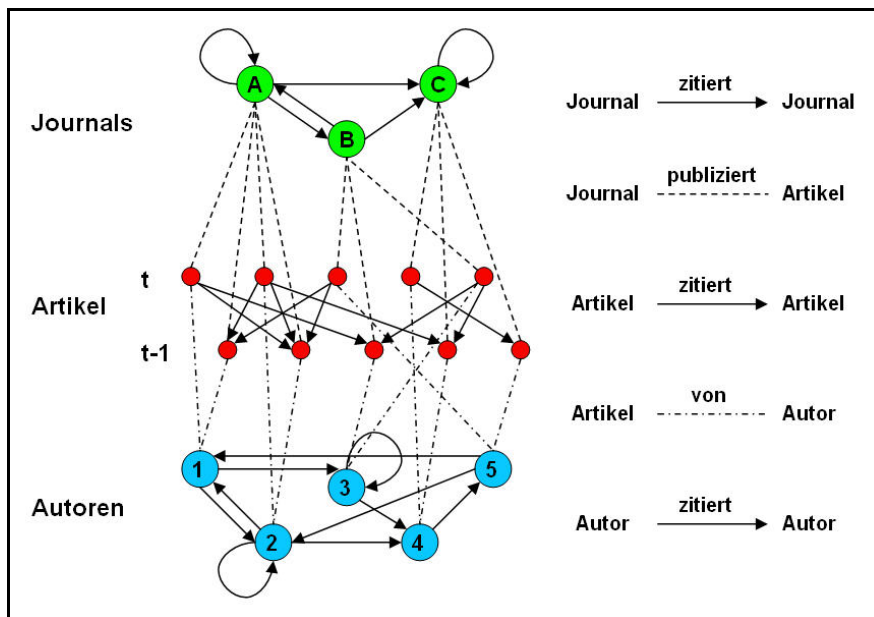


Abbildung 1: Schematische Darstellung eines Zitationsnetzwerks

Im Rahmen der Analyse realer Zitationsnetzwerke haben verschiedene Autoren unabhängig voneinander eine typische Form der Zitationsverteilung festgestellt [vgl. z. B. Redner 1998; Laherre/Sonette 1998; Lehmann et al. 2003]. Abbildung 2a veranschaulicht die von Redner für zwei verschiedene Zitationsnetzwerke gefundenen Ergebnisse. Auffallend ist hierbei der asymptotisch verlaufende „Long Tail“ der Verteilung: Die Mehrzahl der Artikel wird nur einmal oder gar nicht zitiert, während nur verhältnismäßig wenige Artikel viele Zitate aufweisen.

Bei der Evolution des Netzwerkes spielt hierbei der Mechanismus des „Preferential Attachment“ eine zentrale Rolle. Dieses dynamische Phänomen wurde in der Soziologie erstmals 1968 von Merton unter der Bezeichnung „Matthew Effect“ bekannt – in Anlehnung an das Zitat aus dem Matthäusevangelium „Denn wer da hat, dem wird gegeben werden, daß er Fülle habe; wer aber nicht hat, von dem wird auch genommen, was er hat“. Auf ein Netzwerk angewendet, bedeutet dies, dass neue Knoten sich mit einer höheren Wahrscheinlichkeit mit jenen bestehenden Knoten verbinden,

die bereits viele Kanten aufweisen [vgl. z. B. Hein et al. 2006]. Ein Artikel, der mehr Zitate aufweist als ein Konkurrenzartikel, wird also auch mit einer höheren Wahrscheinlichkeit erneut zitiert. Cole und Cole (1973) bezeichnen diese Eigenschaft von Literaturnetzwerken auch als „Cumulative Advantage“ [Cole/Cole 1973].

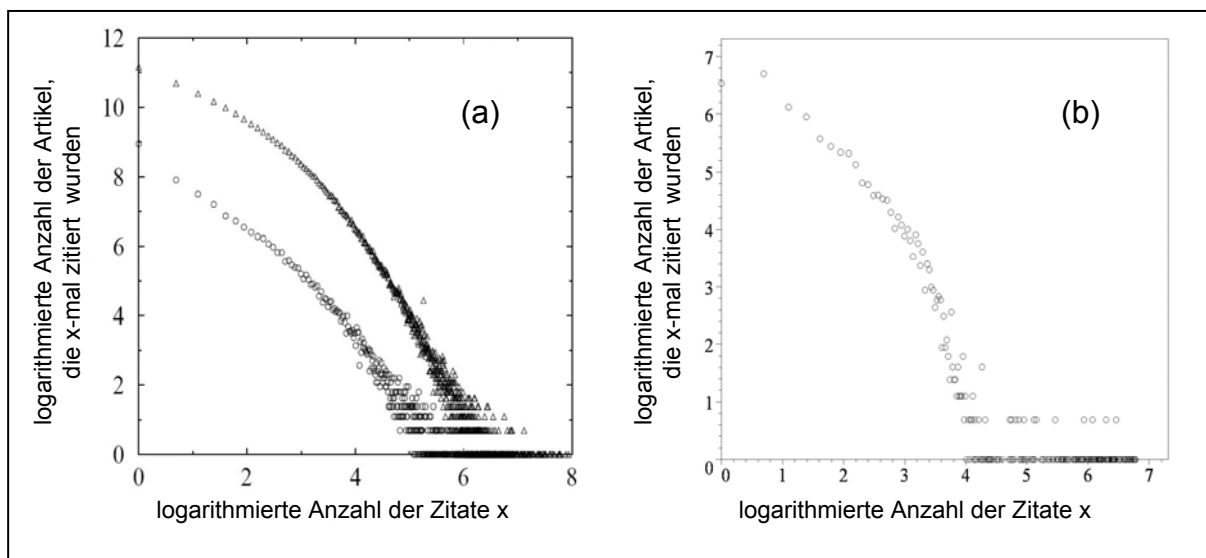


Abbildung 2: (a) Zitationsverteilungen zweier realer Zitationsnetzwerke in doppelt-logarithmischer Darstellung (Δ = ISI-Datenbank; O = Zeitschrift Physical Review D) [Redner 1998]; (b) Zitationsverteilung, welche sich aus dem Simulationsprogramm ergibt (siehe Kap. 3).

3. Preferential Attachment und Open Access

Um zu zeigen, wie ein Zusammenspiel von Open Access und „Preferential Attachment“ die Struktur des Zitationsnetzwerks – genauer die Position einzelner Autoren innerhalb dieses Netzwerks – beeinflussen kann, wird auf ein Simulationsmodell des Marktes für wissenschaftliche Fachinformationen zurückgegriffen, welches im Rahmen des DFG-Projektes „Wissenschaftliche Informationsversorgung und alternative Preisbildungsmechanismen (WIAP)“ entwickelt wurde. Wesentliche Zielsetzung des Projektes ist die Bewertung alternativer Publikationsmechanismen hinsichtlich ihrer gesamtwirtschaftlichen Effizienz und Durchführbarkeit.

3.1 Simulationsmodell

Im Simulationsmodell werden die zentralen Marktakteure (Wissenschaftler, Bibliotheken, Verlage bzw. Journals) sowie die Interaktionen innerhalb und zwischen diesen Spielergruppen abgebildet. Als marktkoordinierende Mechanismen wirken die Reputation von Autoren und Journals, der Preis wissenschaftlicher Zeitschriften sowie die Nutzung der Zeitschriften durch die Autoren. Abbildung 3 stellt schematisch die pro Periode zu durchlaufenden Phasen dar, welche nachfolgend kurz erläutert werden.¹

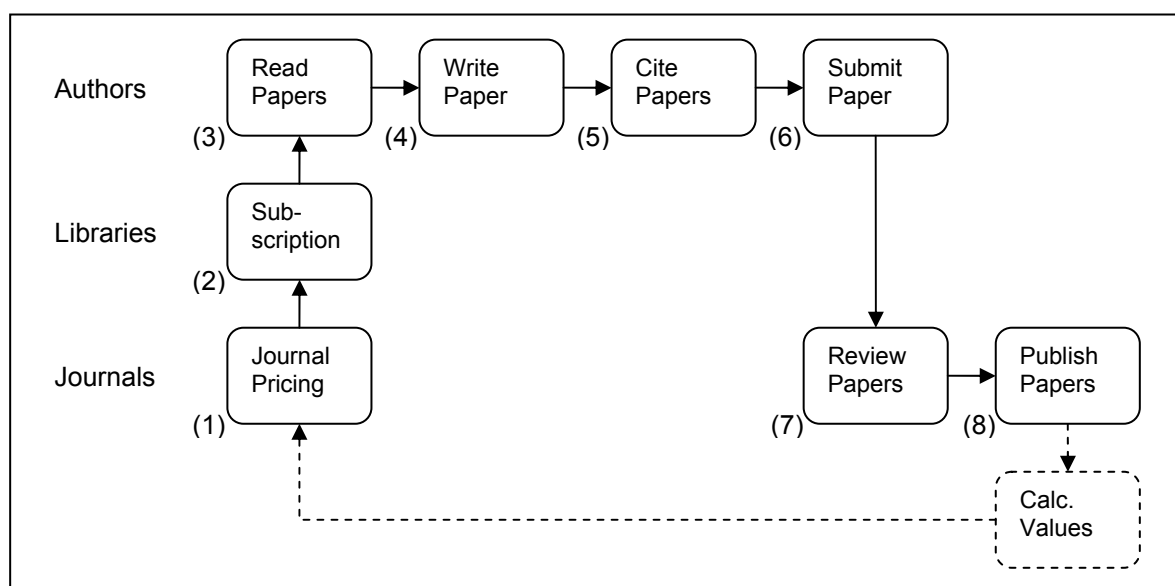


Abbildung 3: Simulationsphasen innerhalb einer Periode

Zu Beginn einer Periode setzen die Verlage den Preis ihrer Zeitschriften fest (1), wobei unterschiedliche Strategien vorgegeben werden können (konstante Preiserhöhung, Preissenkung bei Rückgang der Subskriptionen, Preissteigerung bei Rückgang der Subskriptionen, etc.). Den Bibliotheken ist wiederum jeweils ein Budget zugeteilt, welches ihnen pro Periode zur Subskription von Zeitschriften zur Verfügung steht (2). Hierbei kann ebenfalls festgelegt werden, ob das Budget im Simulationsverlauf konstant bleibt, sinkt oder steigt. Auf Autorenebene beginnt der Produktionsprozess eines neuen Artikels mit dem Lesen einer bestimmten Anzahl von Papers (3), welche jedoch in Zeitschriften publiziert sein müssen, die von der dem Autor zugeordneten

¹ Für eine detaillierte Beschreibung des Simulationsablaufs sowie der einzelnen Simulationsparameter und -phasen verweisen wir auf die Homepage des Projektes (<http://wiap.wiwi.uni-frankfurt.de>).

Bibliothek abonniert sind. Die Qualitäten der neu geschriebenen Artikel variieren von Autor zu Autor (4). In der Zitationsphase referenziert der Autor eine bestimmte Anzahl bereits gelesener Papers aus der aktuellen sowie aus vergangenen Perioden (5). Die Zitationswahrscheinlichkeit eines Artikels sinkt hierbei mit dessen Alter und steigt hingegen mit der Anzahl der Zitate, die das Paper bereits aufweist (Preferential Attachment).

Die letzte Phase auf Autorenebene ist die Einreichung des neu erstellten Artikels in ein geeignetes Journal (6). Die Entscheidung, wo ein Autor einreicht, hängt dabei zum einen vom Verhältnis zwischen der Qualität seines eigenen Artikels und der durchschnittlichen Qualität der in einem Journal veröffentlichten Artikel und zum anderen von dem erwarteten Reputationszugewinn bei Annahme des Artikels ab. Anschließend wird die Begutachtungsphase (7) durchlaufen, in der die Journals eine vorgegebene Anzahl von Beiträgen annehmen, welche sie dann publizieren (8). Übersteigt die Anzahl der eingereichten Beiträge die Annahmekapazität, wählt das Journal Artikel anhand ihrer Qualität aus, bis die jeweilige Ausgabe gefüllt ist. Am Ende einer Periode werden dann die neuen Werte (bspw. für Zitate pro Autor oder Journal, Reputation der Autoren und Journals und Nutzung der Zeitschriften) berechnet und in die nächste Periode übergeben.

3.2 Ergebnisse

Die folgenden Ergebnisse entstammen Simulationen mit 200 Autoren, 9 Journals (welche jeweils bis zu 10 Papers pro Periode bzw. Ausgabe publizieren) und 3 Bibliotheken. Die Simulationsdauer beträgt 100 Perioden (es werden also bspw. bei 6 Ausgaben pro Jahr gut 16 Jahre simuliert). Per Annahme zitieren alle Autoren in ihren neu geschriebenen Artikeln genau 10 andere Papers aus dem Netzwerk. Ab Periode 20 wechseln insgesamt 25 Autoren zur Open-Access-Strategie, d. h. sie veröffentlichen ihre Beiträge zusätzlich in einem Repository, auf dessen Artikel alle Autoren im Netzwerk uneingeschränkt zugreifen können. Als weitere Rahmenbedingungen wurde zum einen ein konstant sinkendes Budget der Bibliotheken (1% pro Periode) festgelegt und zum anderen die Preisstrategie der Verlage so gewählt, dass diese den Preis eines Journals in Abhängigkeit von dessen Abonnements variieren

(Preissenkung bei rückläufigen Subskriptionen, ansonsten konstante Preiserhöhung). Hierbei ist anzumerken, dass sinkende Zeitschriftenpreise in der Realität (zurzeit noch) selten zu beobachten sind: Die meisten Verlage erhöhen den Preis eines Journals bei einem Rückgang der Subskriptionen, um den Deckungsbeitrag zu halten. Im Simulationsablauf wird jedoch deutlich, dass diese Strategie bei sinkenden Bibliotheksbudgets langfristig zum Scheitern verurteilt ist, da immer mehr Journals abbestellt werden bis im Extremfall nur noch eines am Markt existiert.

Bei dieser Konfiguration ergibt sich am Ende der Simulation die in Abbildung 2b dargestellte Zitationsverteilung. Ein Vergleich mit den Verteilungen realer Zitationsnetzwerke (Abbildung 2a) zeigt, dass die Simulationsergebnisse die empirischen Daten sehr gut widerspiegeln (die Struktur der sich im Simulationsablauf entwickelnden Zitationsverteilung ist hierbei allenfalls marginal von der Einbeziehung von Open-Access-Autoren abhängig, sodass der Vergleich mit den empirischen Daten zulässig ist).

Um die Eingangs adressierte Forschungsfrage, wie sich Open Access auf die Position einzelner Autoren im Zitationsnetzwerk auswirkt, zu beantworten, muss von der Betrachtung der „Zitate pro Paper“ auf die Ebene der „Zitate pro Autor“ gewechselt werden. Durch Aufsummierung der Zitate auf die Artikel eines Autors über die Simulationsdauer lässt sich eine Rangfolge der Wissenschaftler bilden. Je mehr Zitate ein Autor im Vergleich zu anderen Mitgliedern der Community aufweist, desto besser platziert er sich in der Rangliste. Trägt man die Autoren geordnet nach Rängen gegen die kumulierte Zitationsanzahl ab (siehe Abbildung 4), ergibt sich ähnlich wie auf Artikelebene eine typische „Long-Tail-Verteilung“: Der Großteil der Zitate verteilt sich auf nur wenige Autoren, während verhältnismäßig viele Autoren kaum zitiert werden.

Die obere Grafik in Abbildung 4 zeigt die Zitationsverteilung für Autoren zur Simulationsperiode $t=20$, also zu jenem Zeitpunkt, an dem ein Teil der Autoren (hier: 25 von 200; in der Abbildung durch Kreise hervorgehoben) ihre Strategie hin zu Open Access wechseln. Die betreffenden Autoren werden hierbei zufällig aus dem Netzwerk ausgewählt, sodass der durchschnittliche Rang dieser Gruppe zu diesem Zeitpunkt über mehrere Simulationsläufe gemittelt 100 beträgt. Im weiteren Ablauf der

Simulation wächst das Zitationsnetzwerk an und die Long-Tail-Verteilung kristallisiert sich stärker heraus. Entscheidend hierfür ist wiederum der Mechanismus des „Preferential Attachment“, der dazu führt, dass wenn ein Autor mehr Zitate aufweist als ein Kollege, er auch eine höhere Wahrscheinlichkeit besitzt, erneut zitiert zu werden.

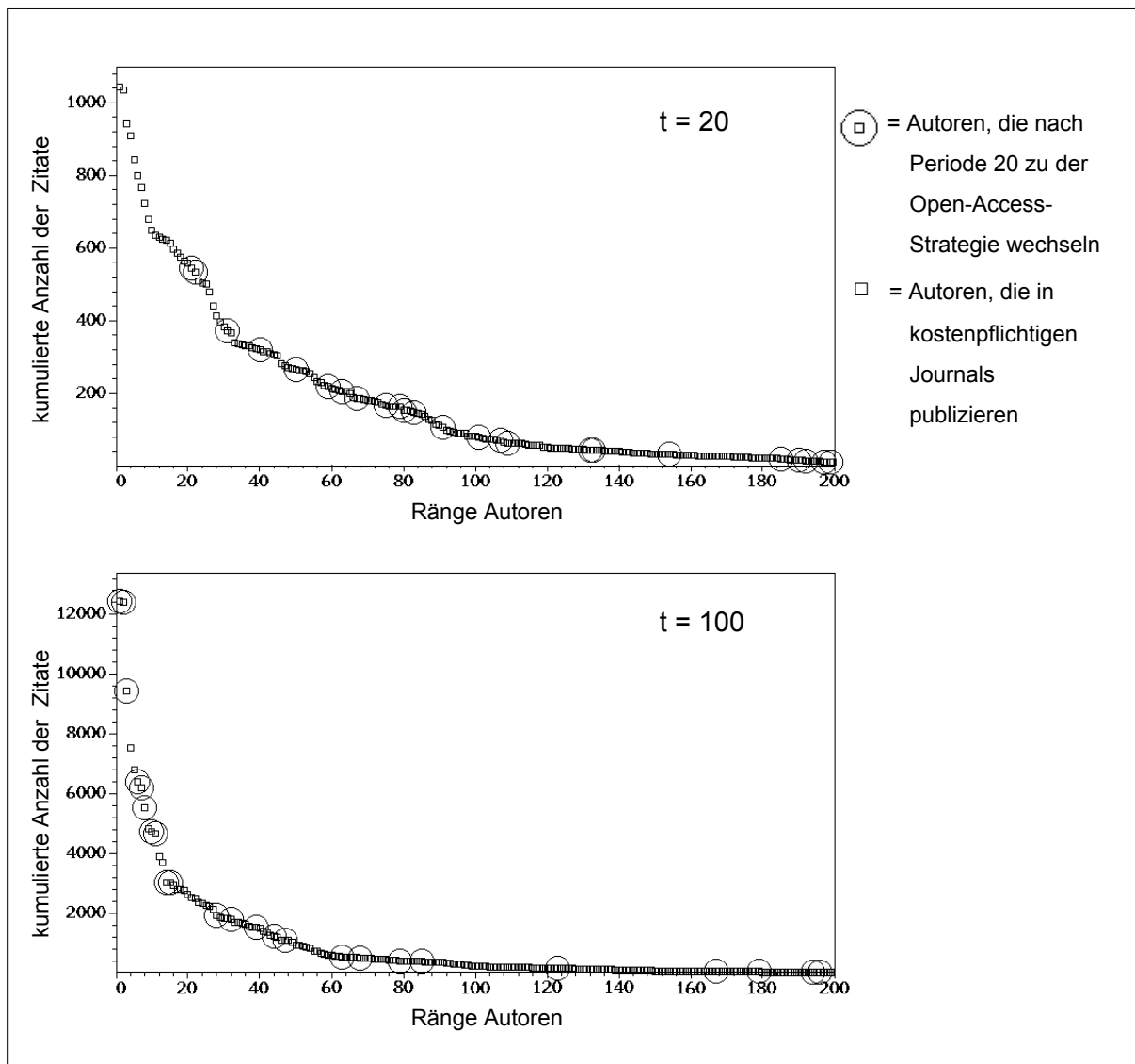


Abbildung 4: Entwicklung der Autoren-Rangordnung über die Zeit bei einem typischen Simulationslauf. Autoren, die nach Periode $t=20$ zu einer Open-Access-Strategie wechseln (obere Grafik) werden im Vergleich zu den anderen Autoren im Netzwerk häufiger zitiert und verbessern mit fortschreitender Simulationsdauer ihre Position innerhalb der Community (untere Grafik: Periode $t=100$) – der durchschnittliche Rang der Open-Access-Autoren sinkt von 100 (bei $t=20$) auf 68 (bei $t=100$).

Anhand der unteren Grafik in Abbildung 4 lässt sich nun erkennen, dass sich die Open-Access-Autoren im Zeitablauf in der Rangliste nach vorne bewegen – die Anzahl der Kreise in den niedrigeren Rängen steigt zulasten der höheren Ränge. Im Rahmen der beschriebenen Simulationsspezifikation steigerten sie ihren durchschnittlichen Rang auf 68 zum Endzeitpunkt der Simulation ($t=100$). Einigen Open-Access-Autoren gelingt es sogar, von den hinteren Plätzen des „Long Tail“ (und somit sozusagen der „wissenschaftlichen Unsichtbarkeit“) zu entkommen.

4. Zusammenfassung

Die im Rahmen der Simulation gefundenen Ergebnisse bestätigen die empirisch gemessenen Zitationssteigerungen bei Open-Access-Artikeln bzw. Open-Access-Autoren. Der unbeschränkte Zugriff auf diese Papers führt dazu, dass sie schneller und häufiger gelesen und infolgedessen auch häufiger zitiert werden. Open Access kann also zu einer beschleunigten Verbreitung von Wissen innerhalb einer Community führen. Als Anreiz für Autoren konnte ein weiterer Faktor identifiziert werden: „Preferential Attachment“ kann im Zusammenspiel mit der höheren Visibilität von Open-Access-Artikeln zu einem Anstieg der Reputation eines Autors (gemessen in Zitaten) innerhalb einer Community führen. Dieser Vorteil verschwindet freilich, wenn alle Autoren zu Open Access wechseln. Da dies (vor allem in „Non-Open-Access-Communitys“) jedoch mittelfristig noch nicht zu erwarten ist, könnten „First Mover“ in der Übergangsphase profitieren.

Für die zukünftige Forschung gilt es, die hier vorgestellten Ergebnisse durch eine umfassende Validierung und Parametrisierung der Simulation zu überprüfen und das Simulationsmodell geeignet zu erweitern. So wurden zum einen noch keine reinen Open-Access-Zeitschriften, welche neben Repositorys den zweiten wichtigen Weg zu Open Access darstellen, im Modell berücksichtigt. Zum anderen wurde die Annahme getroffen, dass die Journals einer parallelen Veröffentlichung zustimmen – ein Umstand, der in vielen Fachbereichen derzeit noch schwer denkbar ist.

5. Literaturverzeichnis

- BARABASI, A. L.; JEONG, H.; NEDA, Z.; RAVASZ, E.; SCHUBERT, A.; VICSEK, T.:
Evolution of the social network of scientific collaborations. In: *Physica A* 311 (2002) 3-4, S. 590-614.
- BERNIUS, S.; HANAUSKE, M.:
WI – Schlagwort: Open Access. In: *WIRTSCHAFTSINFORMATIK* 49 (2007) 6.
- COLE, J. R.; COLE, S.:
Social Stratification in Science. The University of Chicago Press, London, Chicago, 1973.
- EYSENBACH, G.:
Citation Advantage of Open Access Articles. In: *PLoS Biology* 4 (2006) 5, S. 692-698.
- HAJJEM C.; HARNAD, S.; GINGRAS, Y.:
Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. In: *IEEE Data Engineering Bulletin* 28 (2005) 4, S. 39-47.
- HARNAD, S.:
Fast-Forward on the Green Road to Open Access: The Case Against Mixing Up Green and Gold. In: *Ariadne* 42 (2005),
<http://www.ariadne.ac.uk/issue42/harnad/intro.html>.
- HEIN, O.; SCHWIND, M.; KÖNIG, W.:
Scale-Free Networks - The Impact of Fat Tailed Degree Distribution on Diffusion and Communication Processes. In: *WIRTSCHAFTSINFORMATIK* 48 (2006) 4, S. 267-275.
- KNACKSTEDT, R.; WINKELMANN, A.:
Online-Literaturdatenbanken im Bereich der Wirtschaftsinformatik: Bereitstellung wissenschaftlicher Literatur und Analyse von Interaktionen der Wissens-
teilung. in: *WIRTSCHAFTSINFORMATIK* 48 (2006) 1, S. 47-59.
- LAHERRE, J.; SONETTE, D.:
Stretched exponential distributions in nature and economy: ‚fat tails‘ with characteristic scales. In: *European Physics Journal B*, 2 (1998), S. 525-539.

LAWRENCE, S.:

Free online availability substantially increases a paper's impact. In: *Nature* 411 (2001), S. 521-522.

LEHMANN, S.; LAUTRUP, B.; JACKSON, A. D.:

Citation networks in high energy physics. In: *Physical Review E* 68 (2003), 026113.

LEICHT, E. A.; CLARKSON, G.; SHEDDEN, K.; NEWMAN, M. E. J.:

Large-Scale Structure of Time Evolving Citation Networks. arXiv:0706.0015v2 [physics.soc-ph], 6 Jun 2007. <http://de.arxiv.org/abs/0706.0015>.

MERTON, R. K.:

The Matthew Effect in Science. In: *Science*. Bd. 159, Nr. 3810 (1968), S. 56-63.

MOED, H. F.:

The effect of "Open Access" upon citation impact: An analysis of ArXiv's Condensed Matter Section. In: *Journal of the American Society for Information Science and Technology* (2007; im Druck). <http://arxiv.org/abs/cs.DL/0611060>.

NEWMAN, M. E. J.:

Who is the best connected scientist? A study of scientific co-authorship networks. In: *Physical Review E*, 64 (2001), 016131.

REDNER, S.:

How popular is your Paper? An empirical study of the citation distribution. In: *European Physics Journal B*, 4 (1998), S. 131-134.