



Autorennetzwerke: Netzwerkanalyse als Mehrwertdienst für Informationssysteme

Peter Mutschke

Informationszentrum Sozialwissenschaften
Lennéstr. 30, D-53113 Bonn
mutschke@bonn.iz-soz.de

Zusammenfassung:

Das Paper stellt Modelle vor, die Wissen über das Interaktionsgeschehen in wissenschaftlichen Communities und den sozialen Status ihrer Akteure für das Retrieval ausnutzen. Grundlage hierfür sind soziale Netzwerke, die sich durch Kooperation der wissenschaftlichen Akteure konstituieren (*Autorennetzwerke*). Die beschriebenen Studien zur Topologie von Autorennetzwerken zeigen, dass diese Netzwerke ein erhebliches Potential für Informationssysteme haben. Das Paper diskutiert Szenarios, die beschreiben, wie Autorennetzwerke und hier insbesondere das Konzept der Akteurszentralität für die Informationssuche in Datenbanken sinnvoll genutzt werden können.

Abstract:

The paper discusses models which intend to enhance information retrieval by taking into account knowledge about interaction within scientific communities and the social status of their actors. Basis for this are social networks that are constructed by coauthorships (*author networks*). Studies of the topology of author networks, described in the paper, show that these networks have a substantial potential for information systems. The paper discusses scenarios which describe how author networks and in particular the concept of actor centrality can be meaningfully used to improve searching in databases.



1 Einleitung

Virtuelle Bibliotheken enthalten eine Fülle an Informationen, die in ihrer Vielfalt und Tiefe von Standardsuchmaschinen nicht erschöpfend erfasst wird. Traditionelle Retrievalsysteme sind in der Regel strikt dokumentorientiert. Dem Benutzer ist es nicht möglich, die volle Komplexität der gespeicherten Information zu explorieren. Bibliographische Daten zum Beispiel bieten reichhaltige Informationen über die Entwicklung und Struktur einer wissenschaftlichen Community. Die Basisinformationen, wie z.B. Koautorenschaften, sind in den bibliographischen Datensätzen zwar enthalten, Dokumentgrenzen überschreitende Zusammenhänge werden von traditionellen Informationssystemen jedoch nicht erkannt und dem Benutzer somit auch nicht zugänglich gemacht. Dies betrifft v.a. Linkstrukturen zwischen Wissenschaftlern, die z.B. in Koautoren- oder Zitationsrelationen repräsentiert sind, insbesondere aber globalere Eigenschaften der Akteure, wie deren strategische Position in wissenschaftlichen Kommunikations- und Kooperationsstrukturen. Ein weiteres Problem ist das kontinuierliche Wachstum online verfügbarer Informationen, das in verstärktem Maße dazu führt, dass der Benutzer immer mehr mit irrelevanten Informationen überhäuft wird, sich andererseits aber relevante Informationen über heterogene Datenquellen verteilen.

Virtuelle Bibliotheken sind daher nur dann sinnvoll nutzbar, wenn sowohl hochwertige Suchservices, die vorhandene Informationsstrukturen voll ausschöpfen, als auch Dienste bereitgestellt werden, welche die Fülle und Komplexität der in Datenbanken abgelegten Informationen auf hochrelevante Objekte reduzieren. Ein Benutzer, der sich über ein bestimmtes wissenschaftliches Thema informieren möchte, wird nicht nur nach Literatur(nachweisen) suchen, sondern sich auch für menschliche Experten und die sozio-kognitive Struktur des Gebietes interessieren. Die zunehmende Informationsflut einerseits und die lauter werdende Forderung der Benutzer nach qualitativ hochwertigen Informationen andererseits (Krause 2003) legen daher die Entwicklung von Retrieval- und Analyseverfahren nahe, die über die herkömmlichen Retrievalmodelle hinausgehen.

Das Paper diskutiert Modelle, die Wissen über das Interaktionsgeschehen in wissenschaftlichen Communities und den sozialen Status ihrer Akteure für das Retrieval ausnutzen. Grundlage hierfür sind *soziale Netzwerke*, die sich durch Kooperation der wissenschaftlichen Akteure konstituieren und in den Dokumenten der Datenbasis z.B. als Koautorbeziehungen repräsentiert sind. Diese Netzwerke werden im folgenden *Autorennetzwerke* genannt.

Kernanliegen dieser Autorennetzwerkmodelle ist die Suche nach **Experten** und das **Ranking** von Dokumenten auf der Basis von Akteurszentralität, d.h. Autoren, deren strategische Positionierung in ihren sozialen Netzwerken auf eine besondere Relevanz des Autors schließen lässt. Hierfür wurden datenbankbasierte Komponenten in Java entwickelt, die Autorennetzwerke und Akteurszentralität im online-Zugriff auf (relationale) Datenbanken auf der Basis einer Recherche oder einer Gesamtkollektion berechnen. Die hier vorgestellten Modelle gehen auf das Projekt AKCESS am IZ zurück, in dem netzwerkanalytische Verfahren für die Suche nach Experten, unseres Wissens erstmalig, für Retrievalzwecke erprobt wurden (Mutschke 1994). Die im AKCESS-Projekt begonnenen Arbeiten wurden in dem Projekt DAFFODIL¹ weiterentwickelt und dort erstmals in einer größeren Retrievalumgebung eingesetzt (Mutschke 2001, Fuhr et al. 2002). Gegenwärtig werden sie in den Informationsverbund infoconnex² (Ballay et al. 2004) integriert, wo sie als zusätzliche Mehrwertdienste z.B. für das Dokumentenranking genutzt werden sollen.

Die Relevanz von Autorennetzwerken und Akteurszentralität für Retrievalzwecke konnte in einer Reihe von empirischen Studien belegt werden. Zwei am Informationszentrum Sozialwissenschaften (IZ) durchgeführte Studien unter Verwendung von AKCESS und Cognitive-Mapping-Verfahren wiesen einen starken statistischen Zusammenhang zwischen der Zentralität von Themen in Cognitive Maps und der von Akteuren in Koautorennetzwerken nach. Die Studien zeigen, dass Mainstream-Themen eines Forschungsfeldes in starkem Maße von zentral positionierten Akteuren besetzt werden (Mutschke & Renner 1995), wohingegen Innovationen eher von Akteuren mittlerer Zentralität auszugehen scheinen (Mutschke & Quan Haase 2001). Die Aussagekraft wissenschaftlicher Kooperationsnetzwerke wurde darüber hinaus in einer Reihe szientometrischer Studien untersucht. Stellvertretend seien hier die neueren Studien von Güdler (2003), White (2003), Newman (2001a-c, 2004) und Barabasi et al. (2002) genannt.

Graphentheoretische Konzepte wurden bisher nur in sehr eingeschränktem Umfang in Informationssystemen eingesetzt. Zu nennen sind hier insbesondere die Referralsysteme, wie z.B. REFERRALWEB (Kautz et al. 1997), wo Empfehlungspfade von einem Benutzer (oder Ausgangsakteur) zu einem menschlichen Experten in einer gegebenen Netzwerkstruktur evaluiert werden. Diese Systeme beschränken sich jedoch auf die Analyse von lokalen

¹ <http://www.daffodil.de>

² <http://www.infoconnex.de>

Akteursbeziehungen und ziehen die Eingebettetheit der Akteure in die Gesamtstruktur nicht in Betracht. Ähnliches gilt für die bei Kleinberg (1999) und in GOOGLE³ verwendeten Zentralitätskonzepte. Beide Systeme bestimmen die Relevanz von Web-Seiten aufgrund ihrer Vernetztheit. Allerdings schöpfen sowohl Kleinbergs Hubs und Authorities als auch Google's PageRank nicht die volle Netzwerkstruktur aus, sondern betrachten lediglich die Zahl der direkten Nachbarn eines Knotens im Netzwerk und reduzieren Zentralität damit auf ein rein lokales Attribut (vgl. Brandes & Cornelsen 2003).

Der Ansatz der vom Autor entwickelten Autorennetzwerk-Komponenten ist es dagegen, Zentralität über die gesamte Netzwerkstruktur zu evaluieren, um somit zu Aussagen über die globale Bedeutung eines Akteurs für die betrachtete Community zu gelangen. Die theoretischen und methodischen Grundlagen der hier diskutierten Autorennetzwerkmodelle werden in Kapitel 2 beschrieben. Kapitel 3 skizziert generelle Eigenschaften von Autorennetzwerken anhand empirischer Untersuchungen zur Topologie von Autorennetzwerken in verschiedenen Forschungsfeldern. In Kapitel 4 werden Szenarios diskutiert, die beschreiben, wie Autorennetzwerke und hier insbesondere das Konzept der Akteurszentralität für die Informationssuche in Datenbanken sinnvoll genutzt werden können. Das Paper schließt mit einer Zusammenfassung und einem Ausblick.

2 Theoretische und methodische Grundlagen

Der theoretische Begründungszusammenhang für die hier vorgestellten Autorennetzwerkmodelle ist die *Theorie sozialer Netzwerke* (vgl. Jansen 2003, Wasserman & Faust 1994), die für die Handlungsmöglichkeiten individueller oder korporativer Akteure vor allem deren *Eingebettetsein* in soziale Kontexte verantwortlich macht. Um das Handeln von Individuen verstehen und erklären zu können, muss man aus der Perspektive der Netzwerkanalyse also das Ganze untersuchen, d.h. das Netzwerk, in das die Individuen eingebettet sind. Aus dieser Perspektive stehen nicht die individuellen, sondern die *relationalen* Merkmale der Akteure und ihr sozialer Status in der Gesamtstruktur im Mittelpunkt der Analyse. Der Grundansatz der Netzwerkanalyse ist dabei, nicht nur die direkten Beziehungen der Akteure, sondern gerade auch die *indirekten* Beziehungsmuster zu berücksichtigen, um die Eingebettetheit der Akteure zu evaluieren. Aus der Netzwerkperspektive ist vor allem der Status der Akteure, und hier

³ <http://www.google.de>

insbesondere deren Zentralität in der Netzwerkstruktur interessant. Zentrale Akteure sind Akteure, die an vielen Beziehungen im Netzwerk beteiligt sind, daher einen leichteren Zugang zu Netzwerkressourcen (v.a. Informationen) haben und Interaktionsprozesse in der Community kontrollieren. Untersuchungen von wissenschaftlichen Communities ergaben, dass Netzwerkpositionen mit hoher Zentralität dabei oftmals um ihrer selbst willen zum Ziel von positiven Kooperationsangeboten anderer Akteure werden (Newman 2001a-c, 2004, Barabasi et al. 2002, Jansen 2003). Dieses, *preferential attachment* genannte Phänomen bewirkt sich selbst verstärkende Anlagerungsprozesse, die zu einer „Stratifizierung der Akteure“ führen (Jansen 2003, 31).

Es bietet sich daher an, soziale Netzwerke und das Konzept der Akteurszentralität auch für die Informationssuche in wissenschaftlichen Literaturdatenbanken zu nutzen. Dies ist das Ziel der hier beschriebenen Autorennetzwerk-Retrievalmodelle. Ein soziales Netzwerk (hier: Autorennetzwerk) wird in unserem Modell in Anlehnung an die klassische Graphentheorie (s. Palmer 1985) als ein Graph $G = (V, E)$ beschrieben, der aus einer Menge V von Knoten (*vertices*) und einer Menge E von Kanten (*edges*) besteht. In unserem Autorennetzwerkmodell werden die Knoten durch Autoren und die Kanten durch Koautorenschaften repräsentiert.⁴

Die technische Basis für die Generierung von Autorennetzwerken ist die am IZ entwickelte infoconnex-Software⁵, mit der gleichzeitig mehrere heterogene Datenquellen abgefragt werden können, sowie eine mit der infoconnex-Suchmaschine kommunizierende, vom Autor entwickelte Autorennetzwerk-Komponente. Die Autorennetzwerk-Komponente unterstützt zur Zeit die IZ-Datenbanken SOLIS⁶ und FORIS⁷ sowie die Datenbank FIS-Bildung⁸ vom Deutschen Institut für Pädagogische Forschung (DIPF)⁹.

⁴ Die Struktur eines Netzwerkes erschließt sich über sog. Wege (*walks*) oder Pfade (*paths*), mit denen v.a. die Beziehungen zwischen entfernteren Akteuren beschrieben werden können. Ein *walk* in dem Graph von $s \in V$ nach $t \in V$ ist eine Sequenz von verbundenen Knoten, beginnend mit s und endend in t , so dass s und t verbunden sind. Ein Graph ist *verbunden*, wenn jeder Knoten durch jeden anderen über einen *walk* erreicht werden kann. Ein maximal verbundener Teilgraph wird *Komponente* genannt. Ein *path* ist ein *walk*, in dem alle Knoten und Kanten distinkt sind. Die *Länge* eines Pfades wird durch die Anzahl seiner Kanten bestimmt. Die Nähe zwischen zwei Punkten in dem Graphen wird durch die *kürzeste Pfaddistanz* ausgedrückt. Das ist die Länge des kürzesten Pfades (*geodesic*) zwischen zwei Punkten.

⁵ <http://www.infoconnex.de>

⁶ <http://www.social-science-geis.de/Information/SOLIS/index.htm>

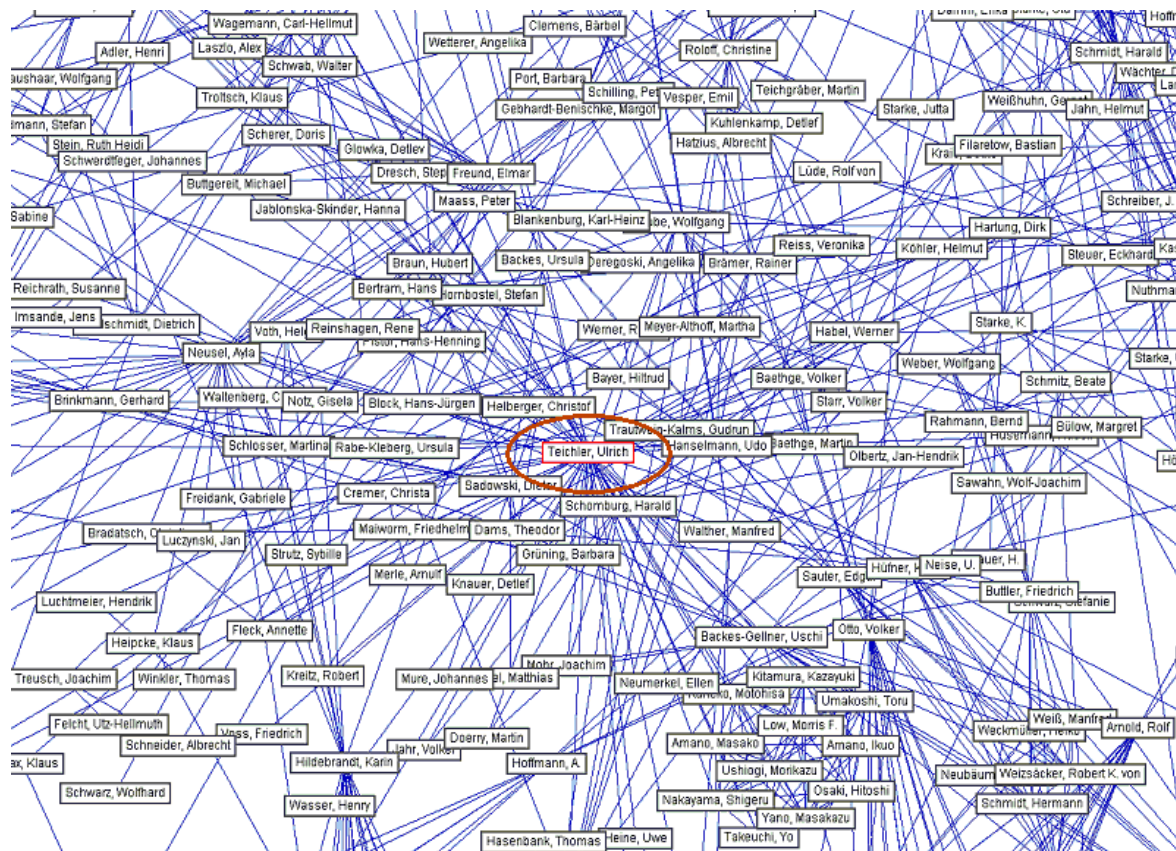


Abbildung 1: Ein Hochschulforschernetzwerk in SOLIS (Ausschnitt). Ulrich Teichler (s. Kreis) ist zentralster Akteur.

Für die Generierung von Autorennetzwerken werden zwei Propagierungsvarianten unterstützt: (1) die Propagierung auf der Basis einer gegebenen Dokumentmenge und (2) die Propagierung von „persönlichen“ Netzwerken eines gegebenen Autors. Bei Variante (1) führt der Benutzer eine Recherche mit der infoconnex-Suchmaschine aus, z.B. zum Thema ‚Hochschule‘ in SOLIS¹⁰. Aus den in den Ergebnisdokumenten enthaltenen Angaben zu den Autoren werden Koautoren-Paare gebildet und in eine

⁷ <http://www.social-science-geis.de/Information/FORIS/index.htm>. Da FORIS als Forschungsprojektdatenbank keine Koautoren-Relationen enthält, werden hier stattdessen Ko-Projektmitarbeiter-Relationen verwendet.

⁸ www.fis-bildung.de

⁹ Eine Propagierung von Autorennetzwerken über die infoconnex-Suchmaschine ist zur Zeit allerdings nur für Datenbanken möglich, die über JDBC erreichbar sind und abfragbare Koautoren-Relationen in ihrer Datenbank vorhalten. Eine Erweiterung des Protokolls, die auch eine Unterstützung von Nicht-JDBC-Datenquellen erlaubt, ist in Vorbereitung.

¹⁰ Schlagwort = Hochschule OR Studium

interne Repräsentation des Autorennetzwerkes eingestellt¹¹. Abbildung 1 zeigt einen Ausschnitt des Autorennetzwerkes zum Thema ‚Hochschule‘ in SOLIS.

Bei Variante (2) wird von einem bestimmten Autor ausgegangen und dessen „persönliches“ Autorennetzwerk per Breitensuche generiert. Diese Variante entspricht den in der Netzwerkanalyse bekannten Ego-zentrierten Netzwerken. Ego-zentrierte Netzwerke sind Netzwerke, die die Beziehungen eines betrachteten Akteurs Ego mit allen seinen Ko-Akteuren, den Alteri, sowie deren Vernetzung untereinander darstellen. Ego-zentrierte Netzwerke beschränken sich in der traditionellen Netzwerkanalyse allerdings auf direkte Beziehungen von Ego zu den Alteri. Um auch das weitere strukturelle Umfeld eines Autors evaluieren zu können, erweitern wir dieses Modell um die indirekten Beziehungen von Ego, d.h. es werden auch Akteure in das Netzwerk von Ego aufgenommen, die zwei oder mehr Schritte von Ego entfernt sind. Bei einer Schrittweite von 2 werden also nicht nur die direkten Koautoren von Ego erfasst, sondern auch deren Ko-Akteure (sofern sie nicht selbst direkt mit Ego verbunden sind). Der Breitensuchprozess terminiert an einem vom Benutzer festgelegten Tiefenschwellwert. Abbildung 2 visualisiert das Koautorennetzwerk (Tiefe=1) des Hochschulforschers Harald Schomburg in SOLIS.

¹¹ Alternativ wird die Generierung von Koautoren-Anfragen unterstützt, d.h. es werden für die Ergebnisdokumente Koautoren-Anfragen generiert und über die infoconnex-Suchmaschine an die Datenquellen gestellt. Die Ergebnistupel können dann direkt in die interne Netzwerkrepräsentation eingestellt werden. Diese Variante kommt zum Tragen, wenn die Suchmaschine nur Kurzinformationen zu den Trefferdokumenten zurückliefert, die keine Angaben zu den Autoren enthalten.

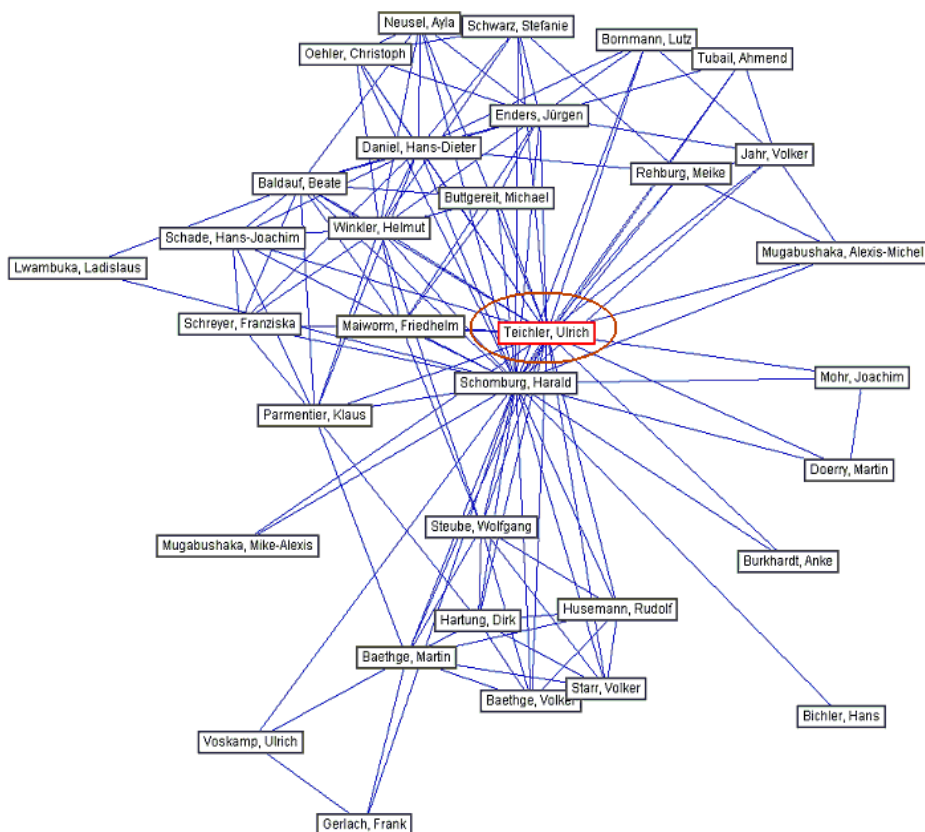


Abbildung 2: Das Koautorennetzwerk von Harald Schomburg in SOLIS. Zentralster Akteur im unmittelbaren Umfeld von Harald Schomburg ist Ulrich Teichler.

Für die Berechnung von Akteurszentralität werden drei grundlegende Zentralitätsmaße aus der Netzwerkanalyse verwendet (vgl. Freeman 1979, Wasserman & Faust 1994): (1) Das einfachste Maß zur Charakterisierung der Zentralität eines Akteurs ist die Zahl seiner direkten Nachbarn (*degree centrality*). Zentral ist nach diesem Zentralitätskonzept der Akteur, der viele direkte Beziehungen hat. (2) *Closeness* misst die Nähe eines Akteurs zu entfernteren Akteuren im Netzwerk¹². Zentral sind demnach Akteure, die über viele *kurze* Verbindungen zu *allen* anderen Akteuren im Netzwerk verfügen. Aus der Theorieperspektive der Netzwerkanalyse sind closeness-zentrale Akteure relativ selten auf die Vermittlung durch andere Akteure angewiesen. Aufgrund ihrer kurzen Verbindungen im Netzwerk kommen Informationen ohne große Verzerrungen bei ihnen an. (3) *Betweenness* misst die Zahl der

¹² Summe der Länge der kürzesten Pfade zwischen einem betrachteten Knoten $v \in V$ und allen $t \in V$ in G . Um Closeness auch in unverbundenen Graphen evaluieren zu können, wurden nach Tallberg (2000) die Closeness-Werte mit der Größe der jeweiligen Komponente gewichtet, so dass Knoten in größeren Komponenten einen vergleichsweise höheren Zentralitätswert erhalten.

durch einen Akteur verbundenen Akteure¹³. Zentral nach *Betweenness* sind die Akteure, die *zwischen* vielen Akteurspaaren im Netzwerk auf deren kürzesten Verbindungen positioniert sind. Ein *betweenness*-zentraler Akteur verbindet also viele Akteure im Netzwerk, wird deshalb häufig von anderen Akteuren als „Makler“ benutzt und kann deshalb viele Aktivitäten im Netzwerk kontrollieren.¹⁴

Degree gilt in der Netzwerkanalyse als Maß für die soziale Aktivität eines Akteurs, Closeness als Maß für seine soziale Effizienz (im Sinne von Unabhängigkeit) und *Betweenness* als Maß für die Kontrolle von sozialen Beziehungen. Doch ist Netzwerkzentralität nicht das mathematische Kunstprodukt eines mehr oder weniger zufällig entstehenden Kooperationsgeflechtes? Untersuchungen der Topologie von Autorennetzwerken in naturwissenschaftlichen Domänen zeigen allerdings, dass man es bei der Sozialstruktur von wissenschaftlichen Communities keinesfalls mit Zufallsstrukturen zu tun hat (vgl. Newman 2001b+c, 2004, Barabasi et al. 2002). Die Ergebnisse dieser Strukturanalysen lassen sich auch für den sozialwissenschaftlichen Bereich bestätigen, wie wir im folgenden Kapitel zeigen.

3 Topologie von Autorennetzwerken

Wenn man sich die Struktur von Autorennetzwerken in verschiedenen (sozialwissenschaftlichen) Forschungsfeldern anschaut, dann findet man – unabhängig von der Größe der Netzwerke - sehr ähnliche topologische Eigenschaften vor (s. Tab. 1): eine relativ kurze charakteristische Pfadlänge¹⁵ von durchschnittlich sieben Links, d.h. alle Akteure werden durchschnittlich über maximal sieben Zwischenschritte erreicht, und ein relativ hoher Clustering-Koeffizient¹⁶ von durchschnittlich 63%. Diese Strukturmerkmale

¹³ Summe der kürzesten Pfade zwischen allen $s \in V$ und allen $t \in V$, die den betrachteten Knoten $v \in V$ als Vorgänger- oder Nachfolgerknoten auf dem kürzesten, s und t in G verbindenden Pfad haben. In unserem Modell werden durchweg normalisierte Werte verwendet. Auf die Darstellung des Formelwerks zu diesen Algorithmen sei hier aber verzichtet. Es findet sich in der einschlägigen Literatur, z.B. in dem Standardwerk von Wasserman & Faust (1994). Für eine effiziente Berechnung von Zentralität auch in großen Netzwerken wurde eine Adaption des *Betweenness*-Algorithmus von Brandes (2001) verwendet, die auch Closeness berücksichtigt.

¹⁴ vgl. Jansen 2003, 131

¹⁵ Durchschnittliche Länge aller kürzesten Pfade

¹⁶ Grad, in dem die Nachbarn eines betrachteten Akteurs ebenfalls untereinander vernetzt sind

weisen das Organisationsmuster dieser Netzwerke als *Small-World-Architektur* (Watts 1999) aus. Small Worlds zeichnen sich gerade dadurch aus, dass die Länge der Verbindungen in einem Netzwerk mit dessen Anwachsen nur unwesentlich, nämlich logarithmisch, zunimmt und gleichzeitig die lokale Vernetzungsrate relativ hoch ist.

SOLIS/FORIS-Kollektion	Größe	Charakt. Pfadlänge	Clustering-Koeffizient
(1) Internet	75	5.19	0.45
(2) Gewalt	469	6.76	0.75
(3) Hochschule	2140	7.63	0.73
(4) >= 1998	29.749	8.75	0.63
(5) ALLE	104.364	6.68	0.60
Durchschnitt		7.00	0.63

Tabelle 1: Topologie von Autorenetzwerken unterschiedlicher Größe¹⁷

Ein Erklärungsmodell für dieses Phänomen ist der bereits genannte und empirisch bestätigte Preferential-Attachment-Mechanismus, also die Neigung wissenschaftlicher Akteure primär mit denen zu kooperieren, die bereits viele Koautoren haben (Newman 2001c, Barabasi et al. 2002). Eine besondere Ausprägung des Preferential-Attachment-Effekts in Scientific Communities ist die Existenz von *Superknoten*, d.h. von Akteuren, von denen verhältnismäßig viele Verbindungen abgehen. Ein Blick auf die Verteilung der Zahl der Akteure auf die Zahl der Verbindungen am Beispiel ‚Hochschule 1997-2001‘ (s. Abb. 3) zeigt, dass sehr viele Akteure nur sehr wenige Kooperationsbeziehungen haben, während einige wenige Akteure sehr viele Verbindungen unterhalten und somit herausragend zentral sind. Ähnliche Verteilungen finden sich auch in anderen Forschungsfeldern. Dieses Ergebnis zeigt, dass Autorenetzwerke in wissenschaftlichen Communities offensichtlich eine ausgeprägte *Zentrum-Peripherie-Topologie* haben (vgl. auch Jansen 2003, 33f.).

¹⁷ (1) Schlagwort=‚Internet‘, (2) Schlagwort=‚Gewalt‘, (3) Schlagwort=‚Hochschule‘, (4) Erscheinungsjahr >= 1998, (5) Gesamtkollektion SOLIS und FORIS. Die Maßzahlen in der Tabelle beziehen sich jeweils auf die größten Komponenten in den untersuchten Netzwerken.

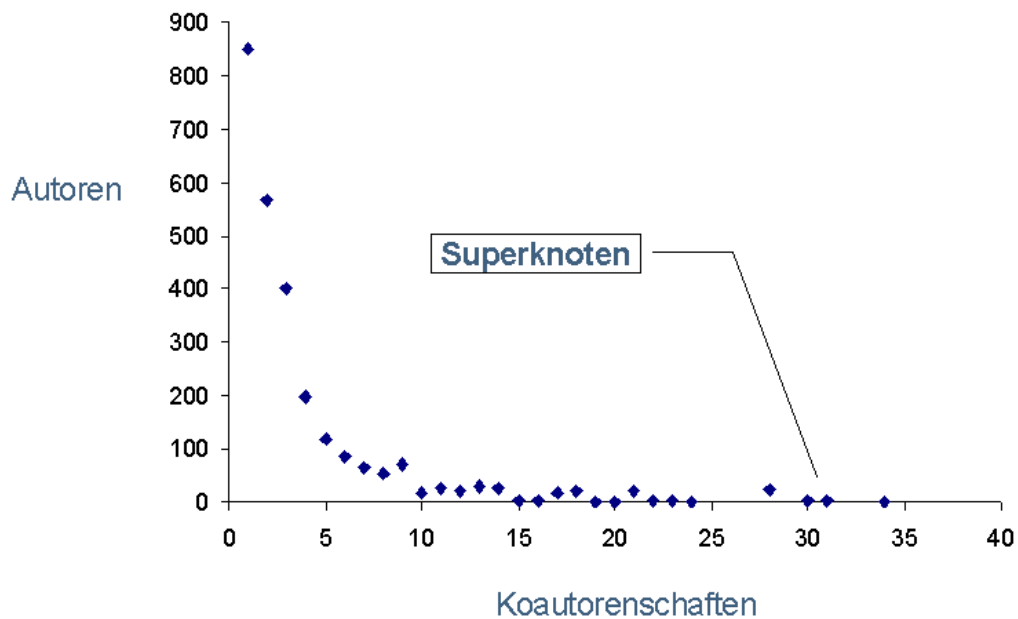


Abbildung 3: Verteilung Akteure – Beziehungen im Forschungsfeld ‚Hochschule 1997-2001‘

Diese Ergebnisse zeigen, dass sich in wissenschaftlichen Communities nicht nur signifikant hohe Vernetzungsraten nachweisen lassen, sondern auch ein bestimmtes Organisationsmuster von Vernetzung, nämlich Small-World-Architekturen „aristokratischen“ Typs, die sich durch eine relativ hohe lokale Clusterdichte, global relativ kurze Verbindungen zwischen den Knoten im Netzwerk und durch eine starke Zentralisierung auszeichnen. Wir haben es bei wissenschaftlichen Kooperationsnetzwerken also keinesfalls mit „Zufallsgraphen“ zu tun, sondern mit immer wiederkehrenden Organisationsmustern, die als grundlegendes Architekturprinzip die Struktur der Netzwerke prägen: der Small-World-Struktur. Dies legt die Annahme nahe, dass auch die strategische Position der Akteure in diesen Netzwerken, mithin deren Zentralität, keine Zufallserscheinung ist, sondern das Ergebnis von Small-World- und Preferential-Attachment-Prozessen.

Nach Jansen kann aus einer zentralen Position in einem Zentrum-Peripherie-Netzwerk sogar der wissenschaftliche Erfolg einzelner Akteure oder Forschungseinrichtungen abgeleitet werden: „Die Akteure im Zentrum der Sozialstruktur ... sitzen im Zentrum des Informationsaustausches und der Kooperation. Dies erlaubt es ihnen, verschiedene wissenschaftliche Ergebnisse, Methoden und Ansätze zu kombinieren und zu bewerten und trägt zu ihrem künftigen wissenschaftlichen Erfolg bei“ (Jansen 2003, 34). Dabei integrierten zentrale Akteure das Fachgebiet nicht nur, weil ihre wissenschaftlichen Leistungen anerkannt seien, sondern auch weil sie

Kontakte zu den peripheren Gruppen unterhielten. Ziel der hier vorgestellten Autorennetzwerkmodelle ist es, diese Akteure zu finden und für die Verbesserung der Suche in Datenbanken zu nutzen. Im Folgenden werden einige Szenarios hierzu vorgestellt.

4 Nutzung von Autorennetzwerken in Informationssystemen

Wie wir gesehen haben, weisen Autorennetzwerke bestimmte Organisationsmuster auf, die auch die strategische Position der Akteure in diesen Netzwerken, mithin deren Zentralität, entscheidend prägen: Autorennetzwerke haben eine ausgeprägte Small-World-Architektur mit einer starken Tendenz zur Zentralisierung (Zentrum-Peripherie-Muster). Zentrale Akteure in wissenschaftlichen Communities sind mithin genau die Akteure, die in ihrer Community hoch integriert sind und viele positive Kooperationsangebote anderer Akteure auf sich ziehen. Diese herausragende strategische Position legt die Annahme nahe, dass zentrale Autoren sowohl thematisch zentrale als auch qualitativ hochwertige Publikationen für die Community liefern. Es bietet sich deshalb an, Wissen über die Zentralität der Akteure in diesen Strukturen auch für die Informationensuche in Datenbanken zu nutzen. Ziel der Nutzung von Autorennetzwerken in Informationssystemen ist es, Akteurszentralität als *Suchstrategie* zu benutzen, um relevante Informationen in Datenbanken unter Vagheitsbedingungen besser auffinden zu können.

4.1 Ranking von Dokumenten nach Akteurszentralität

Allein die schiere Masse an Informationen, die einem Benutzer über bibliographische Datenbanken zugänglich sind, legt intelligente Verfahren nahe, die Ergebnismengen sinnvoll strukturieren und aus der Fülle der Daten die hochrelevanten Informationen herausfiltern. Für das Dokumentretrieval sind hierfür statistische Rankingverfahren (wie z.B. das Vektorraummodell) entwickelt worden, die jedoch allein auf inhaltliche Entscheidungskategorien abgestimmt sind, d.h. auf das Vorkommen der Anfrageterme in den Dokumenten (vgl. Ferber 2003). Das Vektorraummodell z.B. liefert Dokumente, die den Suchbegriffen des Benutzers *inhaltlich* am ähnlichsten sind. Relevant sind hier Dokumente, die eine hohe Selektivität bzgl. der *inhaltlichen* Suchbegriffe der Anfrage haben. Die Relevanz der Autoren, also der eigentlichen Akteure wissenschaftlichen Outputs, bleibt in den

herkömmlichen Standardretrievalverfahren jedoch vollkommen unberücksichtigt.

Eine Berücksichtigung der Relevanz der Akteure wissenschaftlicher Arbeit beim Dokumentenranking ist allerdings in dreifacher Hinsicht interessant:

- **Hohe Selektivität:** Die Zahl der Akteure der in Datenbanken vorkommenden Autoren ist dramatisch hoch und sie nimmt mit dem Wachstum der Datenbanken kontinuierlich zu: Allein in SOLIS sind fast 150.000 Autorennamen verzeichnet. Das Verhältnis von Zahl der Dokumente in SOLIS (ca. 300.000) zu der Zahl der Autoren ist in SOLIS also in etwa 2:1. Dieses Verhältnis macht Autoren zu einem scharfen Selektionskriterium. Mit Autorennamen kann der Suchraum dramatisch eingeschränkt werden. Akteurszentralität wird nicht durch hochfrequente Autoren gestört - wie es etwa bei einem Begriffsnetzwerk der Fall wäre, wo es viele Begriffe gibt, die in sehr vielen Dokumenten vorkommen.
- **Status in der Sozialstruktur einer Community:** Die Akteure wissenschaftlicher Arbeit sind keineswegs isolierte Individuen, die quasi „gleichrangig“ mit anderen agieren. Sondern sie sind eingebettet in Forschungskontexte, die aus einer Vielzahl von (wie auch immer) miteinander interagierenden Individuen bestehen, d.h. sie forschen und schreiben auf dem Hintergrund bestehender Kommunikations- und Kooperationsstrukturen, oftmals sogar in direkter Interaktion mit anderen Wissenschaftlern, oder zumindest mit Rekurs auf die wissenschaftlichen Arbeiten anderer Autoren. Dies bedeutet, dass sie bestimmte Rollen und Positionen im Interaktionsgeschehen einnehmen. Zentrale Autoren sind genau die Akteure, die den besten Zugang zu Ressourcen (Informationen) haben, Kommunikations- und Kooperationsprozesse steuern und somit die inhaltliche Konzeptualisierung der Community entscheidend mitprägen. Es liegt daher nahe, bei der Suche nach relevanten Dokumenten auch den sozialen Status der Autoren in der Community zu berücksichtigen.
- **Resultset-Strukturierung:** Benutzer werden oftmals mit Rechercheergebnissen beträchtlichen Umfangs konfrontiert, die eine gezielte Suche nach relevanten und qualitativ hochwertigen Informationen erschweren oder sogar unmöglich machen. Eine sinnvolle inhaltliche Einschränkung der Suche stellt den Benutzer oftmals vor unüberwindliche Probleme, ist oftmals auch gar nicht erwünscht oder mit unerfreulichen Informationsverlusten verbunden. Eine sinnvolle Strukturierung eines größeren Resultsets, die Informationen über die Struktur eines Forschungsfeldes berücksichtigt, wie z.B. seine

Kooperationsstruktur und den Status der Akteure in dieser Struktur, könnte die Qualität einer Recherche angesichts der schier Masse an Informationen möglicherweise beträchtlich erhöhen.

Die Berücksichtigung von Akteurszentralität bei der Suche oder beim Ranking hätte also nicht nur den Vorteil, den Suchraum drastisch einschränken zu können. Sie würde vielmehr auch die Chance erhöhen, Dokumente zu finden, die von *relevanten* Autoren verfasst wurden, also Veröffentlichungen von Autoren, die in ihrer Community eine bedeutende, vielleicht sogar herausragende Rolle spielen, und deshalb in der Community stärker rezipiert werden als Publikationen von Autoren, die eher an der Peripherie der Community angesiedelt sind. Das auf Akteurszentralität basierende Rankingmodell zielt konzeptuell also darauf ab, Veröffentlichungen peripherer Autoren von denen zentraler Autoren zu unterscheiden und Dokumente zu favorisieren, bei denen von einer erhöhten Relevanz für die Community ausgegangen werden kann, weil sie von Autoren verfasst wurden, die in dem Forschungsfeld eine zentrale Rolle spielen (oder gespielt haben).

Das hier vorgeschlagene Rankingmodell grenzt sich von traditionellen Verfahren nicht nur hinsichtlich des Relevanzkriteriums ab (hier die Bedeutung der Autoren, dort die inhaltliche Ähnlichkeit zwischen Dokument und Anfrage), sondern vor allem hinsichtlich der gewählten Analyseebene: Während traditionelle Rankingverfahren beim Dokument ansetzen (um etwa die Häufigkeit eines Anfrageterms in einem Volltext festzustellen)¹⁸, wird bei dem hier vorgeschlagenen Zentralitätsmodell die *strukturelle* Bedeutung des Autors in der *globalen* Kommunikations- und Kooperationsstruktur seiner Community betrachtet. Dies macht Akteurszentralität nicht einfach nur zu einem weiteren Selektionskriterium neben anderen, sondern aus Benutzersicht potentiell auch zu einem Qualitätssicherungskonzept: Der Benutzer erhält nicht einfach nur Dokumente, die zu seiner Anfrage am besten passen, sondern – so die Kernthese des Modells – Dokumente, von denen zu erwarten ist, dass sie auch in der wissenschaftlichen Diskussion des Faches eine wichtige Rolle spielen (oder gespielt haben), weil sie von Autoren verfasst wurden, die eine herausragende Stellung in der Community einnehmen.

¹⁸ Bei der inversen Dokumenthäufigkeit wird zwar auch die Relevanz eines Terms in der Gesamtkollektion gemessen. Dies ist jedoch ein reines Häufigkeitskonzept, das die Bedeutung des Terms für die Community unter Berücksichtigung ihrer strukturellen Beschaffenheit nicht erfasst.

Grundsätzlich gibt es für ein auf Akteurszentralität basierendes Rankingmodell zwei Operationalisierungsvarianten: Akteurszentralität wird auf der Basis der Ergebnismenge zu einer Recherche (Ex-Post-Ranking) oder für eine Gesamtkollektion (ggf. für einzelne Subkollektionen) berechnet und für die Indexierung der Dokumente verwendet (Ex-Ante-Ranking¹⁹). Bei der Ex-Post-Ranking-Variante, dem gegenwärtig implementierten Modell, wird Akteurszentralität auf der Basis der Ergebnismenge zu einer Recherche berechnet. Der Benutzer führt also eine infoconnex-Recherche in den jeweils angeschlossenen Datenbanken durch. Auf der Basis der Ergebnisdokumente (aus u.U. mehreren heterogenen Datenquellen) wird ein Autorennetzwerk propagiert und die Zentralität der Autoren berechnet. Die Ergebnisdokumente werden nach der Zentralität ihrer Autoren (und Erscheinungsjahr als sekundäres Sortierkriterium) absteigend sortiert an den Benutzer zurückgeliefert. Für unsere Beispielanfrage aus Kapitel 2 („Hochschule“ in SOLIS) würde demnach ein Ranking nach Autorenzentralität unter mehr als 8000 Trefferdokumenten die Arbeiten von Ulrich Teichler „empfehlen“ (vgl. Abb. 1).²⁰

Ein besonderer Vorteil des Ex-Post-Verfahrens ist, dass mit dieser Methode ein Modell eines datenquellenübergreifenden Rankings vorliegt, das von datenquellenspezifischen Rankings unabhängig ist und ohne ein Mischen proprietärer Rankings aus unterschiedlichen Datenquellen auskommt.²¹ Das Modell bietet dem Benutzer zugleich einen direkten Zugang zu zentralen Akteuren seines Faches im Sinne einer Expertensuche. Mit dem Konzept des Autorennetzwerkes besteht für den Benutzer darüber hinaus die Möglichkeit, zu Koautoren zentraler Autoren zu verzweigen, um nach weiteren relevanten Dokumenten im Umfeld zentraler Autoren zu suchen (s. Kap. 4.2).

¹⁹ Näheres zu dieser Variante findet sich in (Mutschke 2004).

²⁰ Alternativ könnte auch Autorenzentralität als sekundäres und z.B. Erscheinungsjahr als primäres Sortierkriterium benutzt werden oder eine alphabetisch sortierte Ausgabe der ersten n zentralsten Autoren erfolgen. Eine weitere interessante Nutzungsoption wäre, dem Benutzer die Möglichkeit zu geben, z.B. zu Dokumenten von Autoren mittlerer Zentralität zu verzweigen, von denen nach Mutschke & Quan Haase (2001) ja am ehesten „neue“ Ideen auszugehen scheinen. Wie sich diese Variante auf die Recherchequalität auswirken würde, müsste allerdings noch untersucht werden. Gleiches gilt für „Einzelautoren“ (Autoren mit einem Degree von 0), die durch das Zentralitätsmodell ebenfalls ausgewiesen werden.

²¹ In Mutschke & Stahl (2004) werden einige Anwendungen einer datenbankübergreifenden Autorennetzwerkanalyse auf der Basis von SOLIS, FORIS und FIS Bildung vorgestellt.

Für das Ex-Post-Ranking von Rechercheergebnissen nach Akteurszentralität wurde im Rahmen einer heuristischen Evaluation ein erster Retrieval-Test für zehn Boolesche Schlagwortanfragen auf der Basis von SOLIS durchgeführt (s. Tab. 2). Die Ergebnismengen wurden nach folgenden drei Verfahren gerankt: Sortierung nach Erscheinungsjahr absteigend, die Standardausgabe von SOLIS (PY); Ranking nach Closeness-Zentralität der Autoren im Autorennetzwerk der Ergebnismenge (ACL); Ranking nach inverser Dokumenthäufigkeit (IDF, vgl. Ferber 2003)²².

Query	Ergebnismenge sortiert nach			Information. Mehrwert
	PY	IDF	ACL	ACL (%)
Jugend – Gewalt	0.25	0.60	0.55	92
Rechtsextremismus – Ostdeutschland	0.35	0.45	0.60	122
Frau – Personalpolitik	0.35	0.60	0.65	100
Widerstand – Drittes Reich	0.40	0.65	0.95	138
Zwangsarbeit – II. Weltkrieg	0.55	0.65	0.70	92
Eliten – BRD	0.40	0.70	0.85	107
Armut – Stadt	0.30	0.35	0.55	157
Arbeiterbewegung – 19./20. Jahrh.	0.55	0.55	0.90	164
Wertewandel – Jugend	0.40	0.50	0.30	50
Terrorismus - Demokratie	0.20	0.35	0.60	129
Durchschnitt	0.38	0.54	0.67	115

Tabelle 2: Precision von zehn Resultsets aus SOLIS (PY = Erscheinungsjahr, IDF = Inverse Dokumenthäufigkeit, ACL = Autor-Closeness)

Der Test ergab, dass mit dem auf Akteurszentralität basierenden Rankingmodell (ACL) eine deutliche höhere Precision erzielt werden kann als mit der Standardausgabe von SOLIS (nach Erscheinungsjahr absteigend, PY) und traditionellen Rankingverfahren wie der inversen Dokumenthäufigkeit (IDF). Ein weiteres interessantes Ergebnis dieses Tests ist, dass Rankings mit Autorenzentralität offenbar ganz andere Dokumente favorisieren als auf Termvorkommen abstellende Rankingverfahren. Die Top-20-Dokumente der nach Akteurszentralität gerankten Menge überschneiden sich so gut wie gar nicht mit der nach IDF gerankten Menge. Der informationelle Mehrwert des ACL-Rankings, gemessen an dem relativen Anteil der relevanten Dokumente unter den ACL-Top-20, die in den IDF-Top-20 nicht auftauchten, betrug –

²² Für das IDF-Ranking wurde der contains-Operator von ORACLE TEXT in der *accumulate*-Scoring-Variante auf der Basis von Titeln und Abstracts verwendet. Die Precision der Ergebnismengen wurde anhand der Zahl hochrelevanter Dokumente unter den ersten 20 Dokumenten der gerankten Trefferliste evaluiert. Als Evaluationskriterium für die Relevanz eines Dokumentes wurde das Vorkommen der Anfragerterme, einschließlich ihrer Synonyme, im Titel des Dokumentes verwendet.

bezogen auf die Zahl der relevanten Dokumente unter den IDF-Top-20 – mehr als 100%. Das Autorenzentralität-Rankingmodell scheint dem Benutzer also eine ganz andere Sicht auf die Datenbank zu bieten als herkömmliche Retrievalmodelle. Es liegt daher nahe, ein auf Zentralität in Autorennetzwerken basierendes Retrievalmodell dem Benutzer als alternativen Zugang zu einer Datenquelle neben herkömmlichen Verfahren zur Verfügung zu stellen.

4.2 Zentrale Akteure im Netzwerk eines Autors

Zentrale Akteure lassen sich aber nicht nur auf der Basis einer Dokumentmenge ermitteln, sondern auch im sozialen Netzwerk eines bestimmten Autors. Die Kernthese hier ist, dass Autoren über ihre Kooperationsbeziehungen in ihrem näheren oder weiteren strukturellen Umfeld mit Experten ihres Faches verbunden sind. Das Konzept dieses Szenarios ist es also, diese Experten zu finden. Solch ein Szenario erscheint sinnvoll, wenn der Benutzer einen Autor kennt und zentrale Akteure im Umfeld dieses Autors finden möchte. Der Benutzer startet also mit einem Autorennamen (Ego). Die Generierung des Autorennetzwerkes von Ego erfolgt gemäß dem in Kapitel 2 beschriebenen Propagierungsmodell per Breitensuche, bis der vom Benutzer vorgegebene Tiefenschwellwert erreicht ist²³. Auf der Basis des so propagierten Autorennetzwerkes wird die Zentralität der Akteure berechnet. Der Benutzer erhält eine Liste der Akteure im Umfeld von Ego, nach deren Zentralität absteigend sortiert. Im Falle des Ego-zentrierten Netzwerkes von Harald Schomburg z.B. (s. Abb. 2 in Kapitel 2) wäre Ulrich Teichler der zentralste Akteur (nach Ego).

²³ In Mutschke (2004) werden einige Scaling-Verfahren für ego-zentrierte Netzwerke vorgestellt, die ein tieferes Eindringen in die Sozialstruktur ohne Angabe eines Tiefenschwellwertes auf effiziente Weise unterstützen.

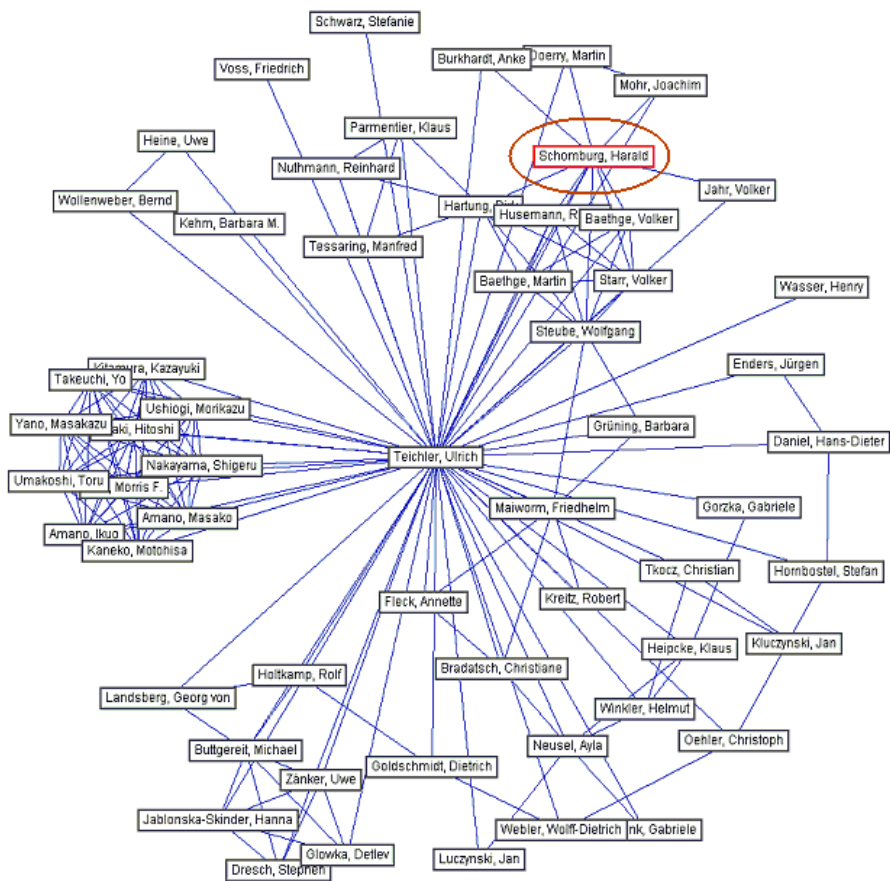


Abbildung 4: Koautorennetzwerk von Ulrich Teichler zum Thema ‚Hochschule‘ in SOLIS. Zentralster Autor im unmittelbaren Umfeld von Teichler ist Harald Schomburg.

Dieses Modell einer ego-zentrierten Expertensuche kann auch mit einer inhaltlichen Suche verknüpft werden. In diesem Fall hätten die inhaltlichen Suchbedingungen eine Filterfunktion für den Propagierungsprozess. Dieses Szenario ist relevant, wenn ein Benutzer einen bestimmten Autor als relevant identifiziert hat (z.B. Ulrich Teichler für das Forschungsfeld ‚Hochschule‘) und zu zentralen Koautoren dieses Autors verzweigen möchte. So würde eine Fokussierung auf zentrale Autoren im unmittelbaren Kooperationsumfeld von Ulrich Teichler zum Thema ‚Hochschule‘ zu Harald Schomburg führen (s. Abb. 4), der zwar nicht zu den zentralsten Autoren in der Hochschulforscher-Community insgesamt gehört, aber relevante Arbeiten zu diesem Thema veröffentlicht hat, die im „Dunstkreis“ von Teichler entstanden sind.

5 Zusammenfassung und Ausblick

Die in diesem Bericht beschriebenen empirischen Studien zu Autorennetzwerken in wissenschaftlichen Communities haben gezeigt, dass diese Netzwerke ein erhebliches Potential für Informationssysteme haben. Die

Untersuchungen zur Topologie von Autorennetzwerken (Small-World-Architektur, Zentrum-Peripherie-Muster, Preferential-Attachment) unterstreichen die Leistungsfähigkeit des Zentralitätskonzeptes auch für soziale Netzwerke in wissenschaftlichen Kooperationsstrukturen. Autorennetzwerke lassen sich daher nicht nur zum Browsen in den Kooperationsstrukturen eines Forschungsfeldes verwenden, sondern auch für die Evaluation der Relevanz der Autoren aufgrund ihrer strategischen Position in der globalen sozialen Struktur ihrer Community (Zentralität).

Das Konzept der datenbankbasierten Analyse von Autorennetzwerken als Mehrwertdienst für Informationssysteme, und hier primär das Konzept der Zentralität der Autoren in solchen Netzwerken, wurde am Beispiel einiger Anwendungsszenarios exemplifiziert. Autorenzentralität lässt sich sinnvoll nutzen für das Ranking von Rechercheergebnissen und für die Suche nach Experten in einem bestimmten Forschungsfeld oder im strukturellen Umfeld einer gegebenen Person. Hierfür wurden verschiedene Varianten vorgestellt. Die Ergebnisse der heuristischen Evaluation geben erste Hinweise, dass Rankings auf der Basis von Autorenzentralität die Retrievalqualität erheblich verbessern können. Weitere Untersuchungen hierzu sind allerdings noch erforderlich. Der Test zeigte auch, dass die Nutzung von Autorenzentralität beim Ranking den Benutzer offenbar zu ganz anderen relevanten Dokumenten hinführt als traditionelle termbezogene Rankingverfahren wie z.B. das Vektorraummodell. Es bietet sich daher an, dem Benutzer ein auf Autorenzentralität basierendes Ranking als Rechercheoption neben den herkömmlichen Retrievalmodellen an die Seite zu stellen.

Weitere Anwendungsmöglichkeiten für die Nutzung von Autorennetzwerken in Informationssystemen wären z.B. die Erkennung von Brückenpersonen und Forschercliquen oder die Erweiterung von Ergebnismengen über Koautorenschaften. Autorennetzwerke ließen sich aber nicht nur für die Suche in Datenbanken nutzen, sondern auch für die Unterstützung von Zusammenarbeit zwischen Wissenschaftlern. Ein Alerting-Dienst könnte z.B. auch so konfiguriert werden, dass ein Benutzer nur über Neuerungen aus seinem *Netzwerk* benachrichtigt wird (social filtering). Auch hier sind weitere Szenarios denkbar.

Eine offene Frage ist nach wie vor, wie man Autorennetzwerke in einem Informationssystem so visualisiert, dass ein (unkundiger) Benutzer mit der Visualisierung auch *arbeiten* kann. Die herkömmlichen Graphenlayout-Verfahren (spring embedder usw.) reichen hier sicherlich bei weitem nicht aus (zumal diese Verfahren vermutlich auch kaum noch zu verbessern sind).

Naheliegender ist, die Graphenvisualisierung mit Browsing-Funktionalität auszustatten (Anzeigen von Veröffentlichungen der jeweiligen Koautoren bei Klick auf einen Link usw.) und Autoren-Graphen mit den thematischen Schwerpunkten der Autoren zu annotieren. Ähnliches gilt für Verfahren der Komplexitätsreduktion von Graphen. In Mutschke (2004) werden hierzu einige Ansätze vorgestellt. Aber auch hier ist noch weitere Forschung notwendig. Ein weiteres offenes Problem betrifft die Entwicklung von Kriterien für die *Adäquatheit* von Netzwerkmodellen im Information Retrieval. Relevante Konzepte hierfür wären die Größe der Netzwerke, ihre interne Kohärenz, ihre Robustheit und vor allem ihr Community-Charakter.

Autorennetzwerke sind allerdings nur *ein* Beispiel für die Nutzung von Netzwerkanalyseverfahren in Informationssystemen. Weitere interessante Anwendungsperspektiven wären die netzwerkanalytische Betrachtung von Kooperationsbeziehungen zwischen Institutionen, von Shared-Interest-Netzwerken (z.B. anhand gemeinsam verwendeter Begriffe bei Autoren und Institutionen), von Begriffsnetzwerken (um z.B. die Zentralität von Thesaurustermen für das Ranking von Dokumenten zu benutzen), oder von Linkstrukturen in Clearinghouse-Daten.

6 Literatur

- Ballay, A., Markham, R., Mutschke, P., Stempfhuber, M. (2004): Der Informationsverbund Bildung-Sozialwissenschaften-Psychologie (infoconnex). Education, Research and New Media. Chances and Challenges for Science. 10. IuK-Frühjahrstagung, Darmstadt, 15. – 18. März 2004.
- Barabasi, A.-L. et al. (2002): Evolution of the social network of scientific collaborations. *Physica A* 311, 590-614.
- Brandes, U. (2001): A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology* 25, 163-177.
- Brandes, U., Cornelsen, S. (2003): Visual Ranking of Link Structures. *Journal of Graph Algorithms and Applications* 7 (2), 181–201.
- Freeman, L. C. (1979): Centrality in social networks: Conceptual clarification. *Social Networks* 1, 215-239.
- Ferber, R. (2003): Information Retrieval. Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web. Dpunkt.verlag, Heidelberg.
- Fuhr, N., Schaefer, A., Klas, C.-P., Mutschke, P.: Daffodil (2002): An Integrated Desktop for Supporting High-Level Search Activities in Federated Digital Libraries. In: Agosti, M., Thanos, C. (eds.): *Research and Advanced Technology for Digital Libraries*. 6th European Conference, EDCL 2002, Proceedings. Lecture Notes in Computer Science, Vol. 2458. Springer-Verlag, Berlin - Heidelberg - New York, 597-612.
- Jansen, D. (2003): Einführung in die Netzwerkanalyse. Leske + Budrich. Opladen

- Krause, Jürgen (2003): Suchen und "Publizieren" fachwissenschaftlicher Informationen im WWW. In: Audiovisuelle Medien online: Tagung: "Audiovisuelle Wissensmedien online"; Informationsveranstaltung der IWF Wissen und Medien gGmbH, Göttingen, 03.12. - 04.12.2002. Wien: Lang. (IWF: Menschen, Wissen, Medien).
- Kautz, H., Selman, B., Shah, M. (1997): The Hidden Web. *AI Magazine* 18 (2), 27-36.
- Kleinberg, J. M. (1999): Authoritative sources in a hyperlinked environment. *Journal of the Association for Computing Machinery* 46 (5), 604-632.
- Mutschke, P. (1994): Processing Scientific Networks in Bibliographic Databases. In: Bock, H.H., et al. (eds.): *Information Systems and Data Analysis. Prospects-Foundations-Applications. Proceedings 17th Annual Conference of the GfKI 1993*. Springer-Verlag, Heidelberg Berlin, 127-133.
- Mutschke, P., Renner, I. (1995): Wissenschaftliche Akteure und Themen im Gewaltdiskurs. Eine Strukturanalyse der Forschungslandschaft. In: Mochmann, E., et al. (eds.): *Gewalt in Deutschland. Soziale Befunde und Deutungslinien*. Oldenbourg Verlag, München, 147-192.
- Mutschke, P., Quan Haase, A. (2001): Collaboration and Cognitive Structures in Social Science Research Fields: Towards Socio-Cognitive Analysis in Information Systems. *Scientometrics* 52 (3), 487-502.
- Mutschke, P. (2001): Enhancing Information Retrieval in Federated Bibliographic Data Sources Using Author Network Based Stratagems. In: Constantopoulos, P., Sölvberg, I.T. (eds): *Research and Advanced Technology for Digital Libraries: 5th European Conference, ECDL 2001, Proceedings. Lecture Notes in Computer Science; Vol. 2163*. Springer-Verlag, Berlin – Heidelberg – New York, 287-299.
- Mutschke, P. (2004): Autorennetzwerke: Verfahren der Netzwerkanalyse als Mehrwertdienste für Informationssysteme. IZ-Arbeitsbericht Nr. 32. Bonn.
- Mutschke, P., Stahl, M. (2004): Kooperationsnetzwerke und Akteurszentralität im Forschungsfeld Bildung. ASI-Jahrestagung "Bildung und Bildungssysteme im internationalen Vergleich", Eichstätt, 09.10.-10.10.2003 (erscheint).
- Newman, M.E.J. (2001a): Who is the best connected scientist? A study of scientific coauthorship networks. *Phys. Rev. E* 64, 016131.
- Newman, M.E.J. (2001b): The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* 98, 404-409.
- Newman, M.E.J. (2001c): Clustering and preferential attachment in growing networks, *Phys. Rev. E* 64, 025102.
- Newman, M.E.J. (2004): Coauthorship networks and patterns of scientific collaboration, *Proc. Natl. Acad. Sci. USA*, in press.
- Palmer, E. (1985): *Graphical Evolution*. Wiley, New York.
- Tallberg, C. (2000): Centrality and random graphs. Technical Report 7, Stockholm University, Department of Statistics.
- Wasserman, S., Faust, K. (1994): *Social Network Analysis: Methods and Applications*, Cambridge University Press, New York.
- Watts, D.J. (1999): *Small Worlds*. Princeton: Princeton University Press.

Peter Mutschke

White, Howard D. (2003): Pathfinder networks and Author Cocitation Analysis: A Remapping of Paradigmatic Information Scientists. *Journal of the American Society for Information Science and Technology* 54 (5), 423-434.