

Tamara Heck, Düsseldorf

Analyse von sozialen Informationen für Autorenempfehlungen

Kollaborationen mit Kollegen sind für fast alle Wissenschaftler bedeutend. Forscher bauen sich im Laufe ihrer Karriere ein soziales Netzwerk mit Kontakten zu anderen Wissenschaftlern mit ähnlichen Interessen und Forschungsschwerpunkten auf. Ein Empfehlungssystem könnte einer Person dabei helfen, geeignete Kollegen und neue Kooperationspartner zu finden. Für einen Wissenschaftler ist seine Reputation sehr wichtig, die mit Zitations- und Referenzdaten analysiert werden kann. Solche Daten können dazu dienen, Cluster mit ähnlichen Forschern zu ermitteln, die wiederum für ein Empfehlungssystem verwendet werden können. Darüber hinaus enthalten neue Dienste des sozialen Webs, wie Social-Bookmarking-Systeme, weitere Informationen über Wissenschaftler, auf Basis derer Personenvorschläge gemacht werden können. Im Folgenden wird ein Versuch beschrieben, verschiedene soziale Informationen über Wissenschaftler zu nutzen, um diesen relevante Kooperationspartner vorzuschlagen.

Schlagwörter: Empfehlung, Wissenschaftler, Zitat, Kollaboratives Filtern, Web of Science, Scopus, CiteULike, empirische Untersuchung

Analysis of social information for author recommendation

Researchers in almost all scientific disciplines rely heavily on the collaboration of their colleagues. Throughout his or her career, any researcher will build up a social academic network consisting of people with similar scientific interests. A recommendation system could facilitate the process of identifying and finding the right colleagues, as well as pointing out possible new collaborators. As a researcher's reputation is of great importance, the social information gleaned from citations and reference data can be used to cluster similar researchers. Web services, such as social bookmarking systems, provide new functionalities and a greater variety of social information – if exploited correctly; these could lead to better recommendations. The following describes, by way of example, one approach to author recommendation for social networking in academia.

Keywords: researcher, recommendation system, citation index, bibliographic coupling, author-co-citation, empirical study, Web of Science, Scopus, CiteULike

Analyse de l'information sociale pour obtenir des recommandations d'auteurs

Les collaborations avec les collègues sont importantes pour la majorité des scientifiques. Au cours de leur carrière, les chercheurs se construisent un réseau social de contacts avec d'autres chercheurs ayant des intérêts et domaines de recherche similaires. Un système de recommandation pourrait aider une personne à trouver des collègues appropriés et de nouveaux partenaires potentiels. Pour un scientifique, sa réputation, qui peut être analysée en utilisant des données sur les citations et références, est très importante. Ces données peuvent être utilisées pour identifier des clusters avec des chercheurs similaires qui, à leur tour, peuvent être utilisés pour un système de recommandation. En outre, de nouveaux services du web social, comme les systèmes de partage de signets, contiennent quantités d'informations sur les scientifiques, sur base desquelles des personnes peuvent être proposées. L'auteur décrit une tentative d'utiliser une variété d'informations sociales sur les scientifiques, afin de leur proposer des partenaires de coopération pertinents.

Mots clés: scientifique, citation, recommandation, enquête, Web of Science, Scopus, CiteULike

1 Einführung

Die Zahl der Kooperationen unter Wissenschaftlern ist in den vergangenen Jahren stetig gewachsen. Forscher in akademischen Einrichtungen und forschenden Unternehmen suchen nach Partnern, unter anderem um Forschungsfragen voranzutreiben, Forschungsanträge einzureichen, Konferenzen und Workshops zu organisieren und Forschergruppen zu gründen (z. B. Cronin, Shaw, &

La Barre, 2003; Heck & Peters, 2010; Luukkonen, Persson, & Sivertsen, 1992). Ein geeigneter und relevanter Kooperationspartner sollte entweder die gleichen Interessen haben oder die eigenen Interessen und Kenntnisse eines Forschers ergänzen. Dabei legen Forscher in der Regel Wert auf Kooperationspartner mit guter Reputation (Cruz, Motta, Santoro, & Elia, 2009), d. h. Wissenschaftler mit vielen Publikationen, die entsprechend zitiert wurden (Cronin, 1994). Will man die Reputation eines Wissenschaftlers messen, so muss man seine Publikationen, seine Referenzen und seine Zitationen analysieren. Diese Daten geben Aufschluss über die Beziehungen eines Forschers zu anderen Forschern und können zur Darstellung sozialer Netzwerke genutzt werden, zum Beispiel Autorenetzwerke einer bestimmten Disziplin. Mit solchen sozialen Informationen lassen sich aber auch potentielle Kooperationspartner finden, zum Beispiel in Informationsdiensten wie Scopus und Web of Science (Li, Burnham, Lemley, & Britton, 2010; Meho, & Rogers, 2008; Meho, & Sugimoto, 2009). Dort finden sich Informationen über die Aktivitäten eines Wissenschaftlers, d. h. seine Publikationen, die Autoren, die er zitiert und diejenigen Autoren, von denen er zitiert wird. Zwei Forscher, die oft gemeinsam zitiert wurden, haben wahrscheinlich ähnliche wissenschaftliche Forschungsschwerpunkte. Diese Autoren-Co-Zitationen sind ein Hinweis auf die Ähnlichkeit der beiden Forscher bezüglich ihres wissenschaftlichen Interessengebietes (Leydesdorff, 2005; White & Griffith, 1981; White & Griffith, 1982). Folglich wären sie wahrscheinlich geeignete Kooperationspartner. Eine andere Methode, ähnliche Wissenschaftler zu finden, ist das Bibliographic Coupling (Kessler, 1963): Wenn zwei Forscher viele gemeinsame Referenzen in ihren Publikationen zitieren, ist dies ebenfalls ein Hinweis auf wissenschaftliche Ähnlichkeit und mögliche Kooperationsbereitschaft. Beide Methoden, Bibliographic Coupling und Co-Zitationsanalyse decken implizite Beziehungen zwischen zwei wissenschaftlichen Autoren auf, im Gegensatz zu den expliziten Co-Autoren-Beziehungen oder Analysen direkter Zitationen. Auf Basis von Ähnlichkeiten, die auf gemeinsame Referenzen oder Co-Zitationen beruhen, könnten einem Wissenschaftler in einem Empfehlungssystem somit neue, relevante Partner vorgeschlagen werden, die er vielleicht noch nicht kennt oder nicht sofort darauf stoßen würde. Somit werden diese beiden Ansätze für das folgende Autorenempfehlungssystem genutzt.

Der folgende Forschungsansatz soll Wissenschaftlern helfen, Kooperationspartner auf Basis von sozialen Informationen zu finden (Heck, Hanraths, & Stock, 2011; Heck, Peters, & Stock, 2011). Das Empfehlungssystem soll dabei auf unterschiedlichen sozialen Informationen und deren

Auswertung basieren. Zum einen werden Daten aus Scopus und dem Web of Science genommen, basierend auf den Methoden der Co-Zitationsanalyse und des Bibliographic Coupling. Zum anderen sollen diese Auswertungen mit sozialen Informationen aus nutzergenerierten Daten des Social-Bookmarking-Dienstes CiteULike angereichert werden. Hier kann die Methode des Collaborative Filtering (Höhfeld & Kwiatkowski, 2007) angewendet werden, bei dem die Ähnlichkeiten zweier Autoren auf gleichen Nutzern oder gleichen Tags basieren. Dadurch, dass soziale Informationen aus unterschiedlichen Diensten zusammenfließen und diese auf unterschiedlichen Ähnlichkeitsberechnungen beruhen, sollen die Expertenempfehlungen verbessert werden. Die beiden Methoden aus der Szientometrie, Co-Zitationsanalyse und Bibliographic Coupling, sollen durch die Methode des Collaborative Filtering, basierend auf Daten in CiteULike, ergänzt werden. Die beiden Kernfragen sind:

1. Sind soziale Informationen in einem Bookmarking-Dienst geeignet, um einen Zielautor potentielle Kooperationspartner vorzuschlagen?
2. Wie unterscheiden sich die Ergebnisse von denen, die durch Autoren-Co-Zitationen und Bibliographic Coupling gewonnen werden?

2 Soziale Informationen und soziale Netzwerke

In der Szientometrie werden Bibliographic Coupling und Autoren-Co-Zitationsanalyse oft dazu genutzt, Relationen zwischen Wissenschaftlern darzustellen und wissenschaftliche Netzwerke einzelner Disziplinen abzubilden. Will man ein Empfehlungssystem aufbauen, sollen die Partnervorschläge speziell für einen Zielautor ermittelt werden. Folglich wird nicht versucht, ein vollständiges Beziehungsnetzwerk aller Autoren eines wissenschaftlichen Faches abzubilden. Wichtig bei der Empfehlung von relevanten Kooperationspartnern ist die Datenbasis, auf der die Empfehlungen beruhen. Beruhen die Autoren-Ähnlichkeitsberechnungen auf Referenzen (Bibliographic Coupling) des Zielautors, wird dessen Perspektive berücksichtigt: Was hat der Zielautor zitiert und welche anderen Autoren zitieren dieselben Referenzen? Bei der Autoren-Co-Zitationsanalyse bedenkt man die Sicht eines dritten Autors: Welche beiden Autoren hat dieser zitiert? Der Zielautor hat hier keinen Einfluss auf die Co-Zitationsdaten, da er nicht bestimmen kann, wer ihn zitiert. Schwierig wird die Ermittlung der Referenz- und Co-Zitationsdaten bei jungen Wissenschaftlern, die Blazek

(2007) die „novice researchers“ nennt. Wer gerade mit seiner Forschungsarbeit begonnen und wenig publiziert hat, kann wenig Referenz- und Co-Zitationsrelationen aufweisen. Ähnlichkeitsberechnungen auf einer solch kleinen Datenbasis sind schwierig. Man kann hier von einem „Kaltstartproblem“ sprechen, das bei Empfehlungssystemen unter anderem das Phänomen bei neuen Ressourcen beschreibt: Zum Beispiel rückt ein neues Video in einem Online-Dienst selten bis gar nicht auf eine Empfehlungsliste für einen Nutzer (obwohl dieser es wahrscheinlich interessant finden würde), weil das Video noch kaum bis gar keine Bewertungen von anderen Nutzern hat. Somit werden Videos, die schon lange im Portal sind, gegenüber neuen Videos immer bevorzugt (Ahn, 2008). Hat ein Wissenschaftler also wenige Publikationen, kann mitunter eine Co-Zitationsanalyse gar nicht berechnet werden. Des Weiteren kommt es zudem bei den Zitationen zu einer Zeitverschiebung, d. h. eine Publikation wird erst zu einem späteren Zeitpunkt nach der Veröffentlichung zitiert. Bei einem „novice researcher“ könnten somit weitere soziale Informationen hilfreich sein, um ihm mehr und bessere Kooperationspartner vorzuschlagen.

Das Social Web ermöglicht es dem Nutzer, eigene Inhalte ins Netz zu stellen, diese zu verwalten und somit zur Anreicherung von Informationen über ihn und andere Personen beizutragen. Je mehr Nutzer in einem Dienst sind, desto mehr Inhalte werden in der Regel auch generiert. Die Frage ist, ob diese Inhalte auch für Expertenempfehlungssysteme genutzt werden können. Inhalte über wissenschaftliche Publikationen liefern Social-Bookmarking-Dienste, die sich auf die Verwaltung von Literatur spezialisiert haben, wie Mendeley, CiteULike, Bibsonomy und Connotea. Nutzer können in diesen Diensten bibliographische Angaben zu wissenschaftlichen Publikationen speichern, diese Ressourcen taggen und somit ihre Literatur online verwalten. Aber sie haben nicht nur Zugriff auf ihre eigene Literaturliste, sondern sehen auch die Ressourcen und Tags aller anderen Nutzer, können nach diesen suchen und relevante Ressourcen ihrer eigenen Liste hinzufügen. Somit entstehen Relationen zwischen den Nutzern, Ressourcen und Tags, die ausgewertet werden können. Es gibt zwei große Vorteile, solche nutzergenerierten Informationen zusätzlich zu den Daten aus Scopus und Web of Science zu nutzen: Zum einen hat man eine größere Varietät an Daten und zum anderen bezieht man die Perspektive der Nutzer mit ein. In einem Autorempfehlungssystem ergänzt die Nutzerperspektive die Perspektive des Zielautors (Bibliographic Coupling) und die der anderen Autoren (Autoren-Co-Zitationsanalyse). Man könnte also annehmen, wenn viele Nutzer die Publikationen zweier Autoren in

einem Bookmarking-System mit denselben Tags versehen, sind sich die Publikationen thematisch nah und folglich haben die Forscher womöglich ähnliche Interessensgebiete. Voraussetzung ist, dass die Tags die Ressourcen thematisch beschreiben. Folglich wären sich die Autoren beider Artikel in ihrem Forschungsgebiet ähnlich. Ein weiterer Ansatz ist die Anzahl gleicher Nutzer: Wenn viele Nutzer die gleichen Autoren, bzw. die Publikationen der Autoren, in ihren online erstellten Literaturlisten haben, sind sich diese Autoren wissenschaftlich ähnlich. Dies setzt natürlich voraus, dass der Nutzer ein spezielles Interessensgebiet hat, wozu er Literatur in einem Bookmarking-Dienst sammelt. Letzterer Ansatz ist sozusagen analog zur Autoren-Co-Zitationsanalyse. Der Unterschied ist, dass die Nutzer nicht zwangsläufig andere Wissenschaftler sein müssen. Des Weiteren geht das Bookmarken einer Publikation zeitlich schneller als eine Zitation. Zum Beispiel muss ein Artikel eines „novice researchers“ noch keine Zitationen haben, kann aber dennoch von Nutzern gebookmarkt sein. Die Nutzerperspektive ist aber nicht nur für einen „novice researcher“ wichtig. Auch ein schon bestehendes Kooperationsnetzwerk eines etablierten Wissenschaftlers kann durch diese erweitert werden und neue relevante Relationen aufdecken, die durch Autoren-Co-Zitationsanalyse und Bibliographic Coupling nicht ermittelt wurden.



Abb. 1: Explizite und implizite Relationen zwischen zwei Forschern basierend auf sozialen Informationen in Informationsdiensten und Social-Bookmarking-Systemen.

Potentielle Relationen zwischen zwei Wissenschaftlern basieren auf diversen sozialen Informationen (Abb. 1). Zum einen gibt es Daten aus den Informationsdiensten wie Scopus und Web of Science, zum anderen enthalten Social-Bookmarking-Systeme nutzergenerierte Daten über die Wissenschaftler. Es gibt sowohl explizite als auch implizite Relationen zwischen Autoren. Explizite Relationen sind unter anderem die Co-Autorenschaft zweier Wissenschaftler und die explizit angezeigte Freundschaft beider, zum Beispiel in einem sozialen Netzwerk (Ben Jabeur, Tamine, & Boughanem, 2010). Beides kann als Basis für Ähnlichkeitsberechnungen dienen. Im Folgenden sind jedoch implizite Relationen die Basis für die Berechnungen, weil der Hauptaspekt darin liegt, einem Zielautor möglichst neue, noch unbekannte potentielle Partner vorzuschlagen und keine schon bekannten Co-Autoren oder Freunde. Zwei Autoren sind demnach durch ihre gemeinsamen Referenzen verbunden (Bibliographic Coupling), die gemeinsamen Autoren, die sie beide zitieren (Autoren-Co-Zitationsanalyse) sowie gemeinsame Nutzer und Tags in einem Bookmarking-Dienst.

3 Expertenempfehlungen

Empfehlungs- bzw. Recommender Systeme wollen dem Nutzer im Web helfen, schneller an für ihn relevante Informationen und Ressourcen zu gelangen. Auf Basis der vorherigen Aktivitäten eines Nutzers, d. h. zum Beispiel seiner Bewertungen bestimmter Ressourcen, sollen seine Prioritäten erfasst und vorausgesagt werden. Diese dienen dazu, unbekannte Ressourcen und Informationen vorzuschlagen, die der Nutzer für interessant und relevant halten könnte. Recommender Systeme verwenden verschiedene Daten, Methoden und Algorithmen, um Ressourcen wie Produkte, Musik, Videos, Nachrichtenartikel und Experten vorzuschlagen (Hotho, Jäschke, Schmitz, & Stumme, 2006). Ein Ziel sind personalisierte Vorschläge, die genau auf die Zielperson zugeschnitten sind (Berkovsky, Kuflik, & Ricci, 2007; Ramezani, Bergman, Thompson, Burke, & Mobasher, 2008; Shepitsen, Gemell, Mobasher, & Burke, 2008). Die Hauptaspekte sind hierbei das Auffinden der relevanten Ressourcen sowie der richtige Ranking-Algorithmus zur Sortierung der Ergebnisse (Desrosiers & Karypis, 2011). Unterschieden werden im Allgemeinen zwei Methoden von Empfehlungssystemen, wobei diese kombiniert werden können: Zum einen gibt es den inhaltsbasierten Ansatz, der Ähnlichkeiten zwischen den Ressourcen ermittelt, zum Beispiel zwischen zwei Zeitungsartikeln, die das gleiche

Thema behandeln. Dem Nutzer werden dann unbekannte Ressourcen empfohlen, die denen von ihm als gut bewerteten Ressourcen ähnlich sind. Zum anderen gibt es den Ansatz des Collaborative Filtering (CF), der nicht nur die eigenen Bewertungen eines Nutzers mit einbezieht, sondern auch die Ressourcen-Bewertungen der anderen Nutzer in einem Dienst (Goldberg, Nichols, Oki, & Terry, 1992; Herlocker, Konstan, Borchers, & Riedl, 1999; Parra & Brusilovsky, 2009; Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994). Ein Vorteil von CF ist, dass es sich nicht auf Ähnlichkeiten von Ressourcen und deren Inhalte stützt, da die inhaltliche Ähnlichkeit alleine nicht immer geeignet ist, um gute Empfehlungen zu geben. CF setzt sozusagen auf die kollektive Meinung: Haben viele Nutzer die Ressource als gut bewertet, wird sie einem Zielnutzer vorgeschlagen. Auch beim folgenden Ansatz wird dieses Prinzip genutzt: Haben viele Nutzer zwei bestimmte Autoren in ihrer Online-Literaturliste, ist das ein Zeichen dafür, dass sich diese Autoren womöglich ähnlich sind.

Empfehlungssysteme basieren auf Nutzerbewertungen, mit Hilfe derer Annahmen über die Prioritäten eines Nutzers getroffen werden. Es gibt verschiedenen Möglichkeiten, Annahmen über das Nutzerverhalten zu treffen (Desrosiers & Karypis, 2011). Im folgenden Beispiel kommt der unäre Ansatz zum Tragen, d. h. der Nutzer gibt keine explizite Bewertung ab, sondern es wird angenommen, wenn er die Ressource nutzt, d. h. zum Beispiel wenn er einen Artikel seiner Online-Literaturliste hinzufügt, ist diese Ressource für ihn relevant. Somit kann dieser Ansatz für Empfehlungen in Social-Tagging-Systemen, in denen es meist keine expliziten Nutzerbewertungen gibt, genutzt werden (Marinho et al., 2011).

Expertenempfehlungen gibt es unter anderem für Nutzer sozialer Webdienste: Auf Basis gemeinsamer wissenschaftlicher Literatur oder gemeinsamen Tags in einem Bookmarking-Dienst können ähnliche Nutzer einander vorgeschlagen werden, zum Beispiel um Communities of Practice zu bilden (Heck & Peters, 2010). Au Yeung et al. (2009) schlagen Expertenutzer im Bookmarking-Dienst Del.icio.us vor. Ein Experte wird dabei über die Qualität seiner eingestellten Ressourcen sowie über die Tendenz, die guten Ressourcen als Erster einzustellen, definiert. Voraussetzung für solche Berechnungen ist jedoch, dass die Nutzer bzw. Wissenschaftler selbst in den Diensten angemeldet und aktiv sind. Der FolkRank (Hotho et al., 2006) betrachtet alle Relationen eines Social-Tagging-Dienstes und gewichtet diese nach dem Prinzip des PageRanks. Somit entsteht ein Netzwerk aus Nutzern, Ressourcen und Tags, die untereinander gewichtete Verbindungen haben und diese Gewichtungen weitergeben. Damit lassen sich dann Ressourcen, Tags oder Nutzer vorschlagen.

Experten-Recommend-Systeme gibt es auch für Mitarbeiter in Unternehmen. Der Dienst ICARE (Petry et al., 2008) soll Mitarbeitern helfen, für bestimmte Problemlösungen geeignete Kollegen vorzuschlagen. Basis für die Empfehlungen sind nicht Publikationen, sondern verschiedene Aspekte wie die Stellung im Unternehmen, die Erreichbarkeit und die Reputation auf bestimmten Fachgebieten. Reichling und Wulf (2009) nutzen für die Vorschläge in ihrem System Dokumente von Mitarbeitern, die sie für ihre Arbeit verfasst haben, sowie Nutzerprofile, die Jobbeschreibung, Aufgaben und Lebenslauf eines Mitarbeiters beinhalten.

Ein weiterer Aspekt von Empfehlungssystemen ist die Evaluation des Dienstes, um den Nutzen für den Zielnutzer zu analysieren. Denn Recommender Systeme sollten nicht nur relevante Vorschläge machen und zuverlässig laufen, sondern dem Nutzer auch einen effektiven Mehrwert bieten (Herlocker, Konstan, Terveen, & Riedl, 2004). Wird der Mehrwert von einem Nutzer nicht erkannt, wird er den Recommender Dienst als Unterstützung für sein Wissensmanagement nicht einsetzen (McNee et al., 2006). Einige Studien setzen daher neben den statistischen Tests auf gezielte Nutzerevaluationen, um ihre Systeme zu testen (Krohn-Grimberghe, Nanopoulos, & Schmidt-Thieme, 2010). Im folgenden Ansatz haben die Zielautoren die Ergebnisse bewertet und eingestuft.

4 Ein Beispiel für ein Autorenempfehlungssystem

Der folgende Ansatz soll einem Zielautor, einem Forscher an einem wissenschaftlichen Institut, Kooperationspartner vorschlagen (Heck, 2012). Dabei wird nicht nur eine Methode verwendet, sondern drei verschiedene Ähnlichkeitsberechnungen auf Basis drei unterschiedlicher Dienste. Es wird angenommen, dass die Methoden sich gegenseitig ergänzen und zu besseren Autorenvorschlägen führen

4.1 Collaborative Filtering (CF) in CiteULike

Der Dienst CiteULike ist in den vergangenen Jahren stark gewachsen (Linde, & Stock, 2011) und hat mittlerweile über 6 Millionen wissenschaftliche Artikel in seiner Datenbank, die zum größten Teil von Nutzern eingestellt und mit Tags versehen wurden. Die Folksonomy Struktur weist Nutzer-Ressource-Tag-Relationen auf, die als Tupel definiert werden kann (Marinho et al., 2011; Peters, 2009): $F := (U, T, R, Y)$, wobei die Mengen U die Elemente

Nutzernamen, T die Elemente Tags und R die Elemente Ressourcen beinhalten und Y die Relationen zwischen diesen drei Elementen enthält, $Y \subseteq U \times T \times R$. Im folgenden Ansatz werden als vierte Instanz die Autoren der gebookmarkten Artikel hinzugenommen: $F_E := (U, T, R, A, Y)$ mit $Y \subseteq U \times T \times R \times A$, wobei A der Menge der Autoren beinhaltet. Für diese Autoren werden ähnliche Autoren gesucht, die ihnen als potentieller Partner vorgeschlagen werden sollen. Nutzernetzwerke, d. h. die Beziehungen zwischen den Nutzern in CiteULike interessieren hierbei nicht. Wichtig sind nur diejenigen Nutzer, die zumindest einen Artikel des Zielautors in CiteULike gebookmarkt haben. U_a (bzw. U_b) sind demnach alle Nutzer, die mindestens einen Artikel des Zielautors a bzw. des Autors b in ihrer Online-Literaturliste haben, R_a sind die Ressourcen, in diesem Fall wissenschaftliche Artikel, die die gleichen Tags wie die eines Artikel des Zielautors haben, und T_a bzw. T_b sind alle Tags, die mindestens einem Artikel des Zielautors a bzw. des Autors b hinzugefügt wurden. Zwei Möglichkeiten sind denkbar, um das Datenset für die Ähnlichkeitsberechnungen auszuwählen: Entweder werden alle Nutzer gesucht, die mindestens einen Artikel des Zielautors gebookmarkt haben. Oder es werden alle Ressourcen gesucht, die mindestens einen gemeinsamen Tag mit einem Artikel des Zielautors haben. Der Nachteil im ersten Fall könnte sein, dass man zu wenige Nutzer hat und die Berechnungen dadurch schwierig sein könnten (Lee & Brusilovsky, 2010a). Die Anzahl der Tags ist in der Regel höher. Zudem weisen Tags eine thematische Ähnlichkeit auf, weshalb sie geeigneter erscheinen. Haben zwei Artikel gleiche Tags, die die Ressource thematisch beschreiben, sind sich die Artikel thematisch ähnlich, folglich könnten sich auch die Autoren beider Artikel in ihrem Forschungsinteresse ähneln. Daher wird der zweite Ansatz genommen, mit der Bedingung, dass alle Ressourcen genommen werden, die mindestens zwei gemeinsame Tags mit den Artikeln des Zielautors haben. In diesem Beispiel wurden den Artikel der Zielautoren sehr allgemeine Tags wie „nanotube“ oder „spectroscopy“ hinzugefügt. Zwei gemeinsame Tags erhöhen die Wahrscheinlichkeit auf das gemeinsame Forschungsgebiet der Wissenschaftler. Ein Nachteil der Methode ist, dass Nutzer dem Datenset möglicherweise verloren gehen, die zwar Artikel des Zielautors gebookmarkt, diese aber nicht mit Tags versehen haben. Die CiteULike-Datenbasis ist somit wie folgt festgelegt:

$$R_a := \{r \in R \mid t \in T_a, (r, t) \in Y \text{ with } |T_a| \geq 2\}. \quad (1)$$

Die Ähnlichkeit zwischen den Autoren a und b wird mit dem Kosinus-Maß berechnet. Basierend auf der Daten-

basis werden die tagbasierte Variante (A) und die nutzerbasierte Variante (B) für die Ähnlichkeitsberechnungen zwischen dem Zielautor und den anderen Autoren angewendet:

$$\begin{aligned} \text{A) } \text{sim}(a, b) &:= \frac{|T_a \cap T_b|}{\sqrt{|T_a| \cdot |T_b|}}, \\ \text{B) } \text{sim}(a, b) &:= \frac{|U_a \cap U_b|}{\sqrt{|U_a| \cdot |U_b|}}. \end{aligned} \quad (2)$$

4.2 Autoren-Co-Zitationsanalyse und Bibliographic Coupling

Relationen zwischen Forschern können in Bezug auf ihre Publikationen auf verschiedenen Aspekten basieren: Zwei Autoren veröffentlichen eine gemeinsame Publikation (Co-Autorenschaft), ein Autor zitiert den anderen (Zitationen), zwei Autoren werden von einem dritten Autor zitiert (Co-Zitationen) und zwei Autoren nutzen dieselben Referenzen (Bibliographic Coupling). Da ein Empfehlungssystem einem Zielautor in erster Linie Autoren vorschlagen soll, die er noch nicht kennt, werden Co-Zitationen und gemeinsame Referenzen herangezogen, um potentielle Kooperationspartner zu erhalten. Denn bei Co-Autorenschaft und Zitationen ist es ersichtlich, dass der Zielautor seine Kollegen schon kennt. Bibliographic Coupling (Kessler, 1963) und Co-Zitationsanalyse (Leydesdorff, 2005; Schneider & Borlund, 2007a; Schneider & Borlund 2007b; Small, 1973; White & Griffith, 1981; White & Griffith, 1982) zeigen ungerichtete Relationen zwischen zwei wissenschaftlichen Publikationen auf. Im folgenden Ansatz werden diese beiden Methoden auf der Autorenebene eingesetzt.

Beim Bibliographic Coupling auf Autorenebene zählt man die Anzahl der gemeinsamen Referenzen, die zwei Forscher in ihren Publikationen verwendet haben. Haben beide viele gemeinsame Referenzen, deutet dies darauf hin, dass sie ähnliche Forschungsinteressen haben. Die Datenbasis für einen Zielautor wird aus dem Web of Science (WoS) ermittelt, da der Dienst es über den Link „related records“ erlaubt, Publikationen – in diesem Fall werden die Autoren genommen – mit gemeinsamen Referenzen zu finden und diese nach der Anzahl gemeinsamer Referenzen ordnen zu lassen (Cawkell, 2000; Stock, 1999). Das Datenset besteht demnach aus drei Mengen: D mit den Elementen Dokumente, A mit den Elementen Autoren und Ref mit den Elementen Referenzen, wobei D und Ref die gleichen Elemente beinhalten, nämlich die Publikationen in WoS. Es wird angenommen,

dass zwei Autoren sich ähnlich sind, wenn sie viele gemeinsame Referenzen pro Dokument verwenden. Daher ist nicht die Gesamtanzahl gemeinsamer Referenzen wichtig, sondern die Anzahl der gemeinsamen Referenzen pro Dokument. Ausgehend von dieser Annahme werden als Datenbasis alle Dokumente in WoS genommen, die mindestens eine bestimmte Anzahl gemeinsamer Referenzen (Anzahl kann nach Zielautor variieren) mit den Artikeln des Zielautors a haben:

$$D_{BC} := \{d \in D \mid |Ref_{dj} \cap Ref_{da}| \geq n, n \in \mathbb{N}\} \quad (3)$$

wobei Ref_{dj} die Anzahl der Referenzen in einem Dokument $d \in D$ von Autor $j \in A$ ist, und Ref_{da} die Anzahl der Referenzen in einem Dokument $d \in D$ von Zielautor $a \in A$. Da die Gesamtzahl der Referenzen pro Autor, die für die Berechnungen mit dem Kosinus-Maß benötigt wird, manuell aus WoS ermittelt werden muss, wird das Datenset D_{BC} begrenzt: Pro Zielautor werden mindestens 30 Autoren für die Ähnlichkeitsberechnungen genommen. Das Kosinus-Maß errechnet sich aus der Anzahl gemeinsamer Referenzen zwischen zwei Autoren geteilt durch die Wurzel des Produktes der Anzahl der Referenzen pro Autor (siehe Formel 2).

Autoren-Co-Zitationsanalyse (ACC) (Leydesdorff, 2005; Schneider & Borlund, 2007a; Schneider & Borlund, 2007b) ermittelt die Anzahl der Publikationen, die zwei Autoren co-zitieren und somit beide in ihren Referenzen haben. Als Datenbasis dient der Informationsdienst Scopus. WoS zählt in den Referenzlisten nur den ersten Autor und keine Co-Autoren der zitierten Literatur auf, d. h. die Co-Autoren würden für die Berechnungen wegfallen. Da jedoch alle Autoren, die neben dem Zielautor zitiert wurden, ermittelt werden sollen, wird auf Scopus zurückgegriffen. Somit werden alle Autoren für die Co-Zitationsanalyse berücksichtigt (Persson, 2001; Zhao & Strotmann, 2007). Die Datenbasis enthält die Mengen D mit den Elementen Dokumente, A mit den Elementen Autoren und C mit den Elementen Zitationen, wobei D und C dieselben Elemente beinhalten. Es werden alle Dokumente genommen, die mindestens eine Publikation von Zielautor a zitiert haben:

$$D_{ACC} := \{d \in D \mid \exists c \in C_{da}\} \quad (4)$$

wobei C_{da} die Anzahl der zitierten Dokumente von Zielautor a ist. Das Datenset pro Zielautor wird wiederum auf mindestens 30 Autoren begrenzt, da die Anzahl der zitierten Dokumente pro Autor in Scopus manuell ermittelt werden muss. Das Kosinus-Maß für die Ähnlichkeitsberechnungen berechnet sich aus der Anzahl der ge-

meinsamen Dokumente, die zwei Autoren zitieren geteilt durch die Wurzel des Produkts der Anzahl der zitierten Dokumente pro Autor (siehe Formel 2).

Bezogen auf die Analyse wissenschaftlicher Netzwerke hat sich herausgestellt, dass Bibliographic Coupling und Co-Zitationsanalyse in Kombination am besten funktionieren (Boyack & Klavans, 2010; Gmur, 2003). Für die Autoreneempfehlungen wird angenommen, dass bessere Ergebnisse erzielt werden, wenn verschiedene Methoden und Datensätze herangezogen werden.

4.3 Experimentelle Ergebnisse

Mit den oben genannten Methoden und Berechnungen erhält jeder Zielautor nun vier verschiedene Sets (Cluster) mit ähnlichen Autoren: Zwei Cluster basieren auf der CF-Methode in CiteULike und entweder gemeinsamen Tags (CULT-Cluster) oder gemeinsamen Nutzern (CULU-Cluster). Ein Cluster wurde mit Bibliographic Coupling in WoS berechnet (BC-Cluster) und ein viertes Cluster ergibt sich aus der Autoren-Co-Zitationsanalyse in Scopus (ACC-Cluster). Ein Zielautor erhält somit eine Liste ähnlicher Autoren sortiert nach dem Kosinus-Maß. In den Clustern lassen sich zudem Ähnlichkeitsberechnungen zwischen allen Autoren ermitteln, basierend auf das jeweilige zuvor festgelegte Datenset. Diese Berechnungen mit dem Kosinus-Maß werden in Graphen visuell dargestellt und zeigen Autorenrelationen auf. Diese Graphen werden ebenfalls von den Zielautoren bewertet. Für die Visualisierung wird die Software Gephi¹ benutzt. Bisher wurden die Daten von sechs Physikern aus dem Forschungszentrum Jülich analysiert und bewertet. Von den Zielautoren wurden die Publikationen aus den Jahren 2006–2011² genommen, um die Empfehlungen auf den aktuellen Forschungsschwerpunkt zu basieren. Für jeden Zielautor gibt es vier Cluster (Tab. 1) sowie vier Graphen, die die Ähnlichkeiten zwischen allen Autoren eines Clusters darstellen (siehe Abb. 2–5). Die Knoten der Graphen sind die Autorennamen, die Größe der Knoten entspricht der Anzahl der Relationen eines Autors zu anderen Autoren. Die Kanten stellen die Relationen zwischen den Autoren dar und basieren auf dem Kosinus-Maß. Da die CiteULike-Graphen relativ groß sind, wurden teilweise Schwellenwerte gesetzt, um die Graphen übersichtlich zu halten und die Autorennamen lesbar zu machen. Des Weiteren waren die Daten für einige Autoren und deren Relationsberechnungen nicht aussagekräftig genug, so-

Datenbasis	Methode	Ähnlichkeitsberechnung	Name
CiteULike	Collaborative Filtering (CF) tag-basiert	Kosinus-Maß	CULT-Cluster
CiteULike	Collaborative Filtering (CF) nutzer-basiert	Kosinus-Maß	CULU-Cluster
Web of Science (WoS)	Bibliographic Coupling	Kosinus-Maß	BC-Cluster
Scopus	Autoren-Co-Zitationsanalyse	Kosinus-Maß	ACC-Cluster

Tabelle 1: Datenbasis und Methoden zur Berechnung von Autoren-ähnlichkeiten für ein Empfehlungssystem.

dass diese aus den Graphen genommen wurden. Dies war der Fall, wenn es Autorenpaare mit der höchsten Ähnlichkeit von $sim=1$ gab und diese nur auf einem Nutzer oder einem Tag beruhte. Einer (Autor 6) der sechs Zielautoren hatte keine Artikel in CiteULike, die von Nutzern gebookmarkt wurden. Es wurden zwar Artikel gefunden, jedoch wurden diese von dem Systembetreiber, dem Springer-Verlag, selbst in den Dienst gestellt. Zudem hatten Artikel eines weiteren Autors (Autor 5) keine Tags. In diesem Fall wurde nach allen Nutzern gesucht, die seine Artikel in CiteULike gebookmarkt haben.

Bei den Kosinus-Werten fällt auf, dass die Werte beim Bibliographic Coupling in der Regel weitaus geringer sind als die Werte bei der Autoren-Co-Zitationsanalyse und CF. Dies liegt daran, dass die Autoren viel mehr Referenzen als Zitationen oder Nutzer und Tags haben. Wenn die Autoren wenige Nutzer und Tags in CiteULike hatten, waren die Kosinus-Werte für CF dementsprechend hoch. Es sollte bei den Kosinus-Werten in den ver-

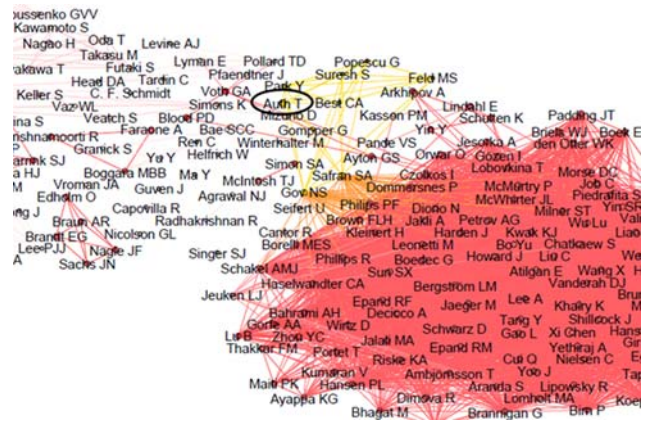


Abb. 2: Ausschnitt aus einem CULT-Graphen: Kosinus-Intervall 0,5–1. Kreis zeigt Zielautor.

1 <http://gephi.org/>
 2 Studie durchgeführt im Mai 2011

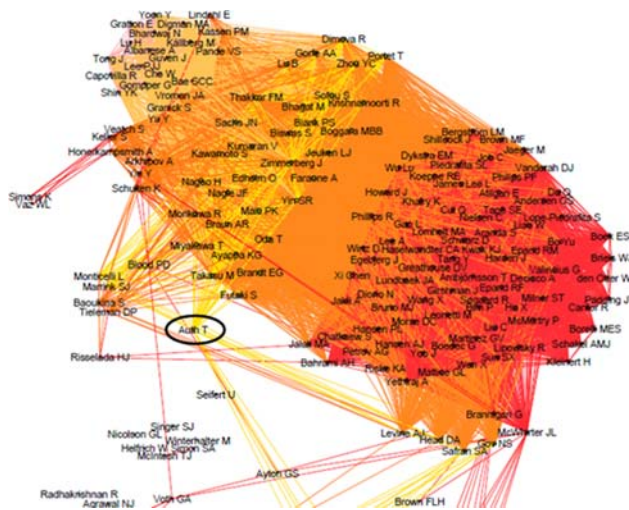


Abb. 3: Ausschnitt aus einem CULU-Graphen: Kosinus-Intervall 0,5–1. Kreis zeigt Zielautor.

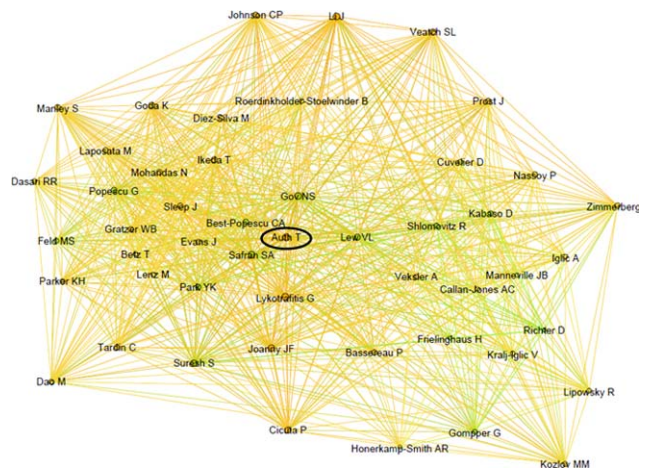


Abb. 5: ACC-Graph: Kreis zeigt Zielautor. Grüne Knoten sind relevante Autoren.

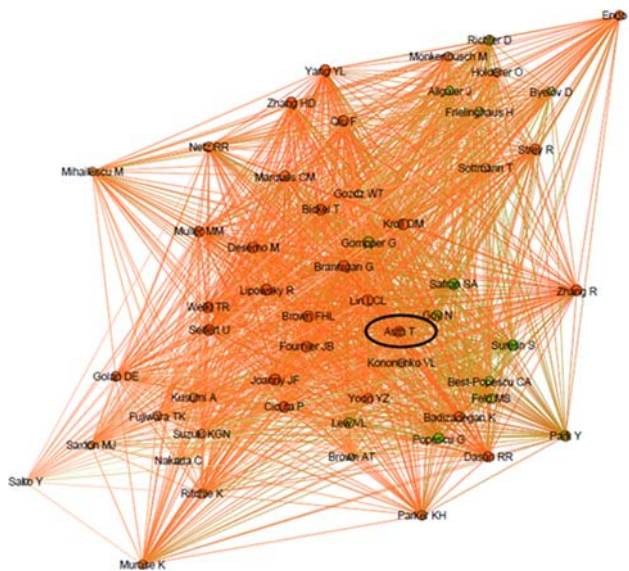


Abb. 4: BC-Graph: Kreis zeigt Zielautor. Grüne Knoten sind relevante Autoren.

schiedenen Clustern daher beachtet werden, dass ein direkter Vergleich der absoluten Werte nicht unbedingt geeignet ist, um ähnliche von weniger ähnlichen Autoren zu unterscheiden.

5 Evaluation durch die Zielautoren

Die jeweils vier Cluster und Graphen für die Autorempfehlungen wurden von den Physikern in einem Interview bewertet. Die Befragung gliederte sich in drei Teile: Der

erste Teil beinhaltete allgemeine Fragen zum Rechercheverhalten und den Kooperationen der Forscher. Diese geben Aufschluss darüber, wie wichtig Kooperationspartner für den einzelnen sind, welche Kooperationen es gibt und wie geeignete Kollegen gesucht und gefunden werden. Im zweiten Teil ging es um die Clusterbewertung. Dazu wurden von den vier Clustern die nach dem Kosinus-Maß jeweils zehn ähnlichsten Autoren genommen und dem Zielautor als Liste in einer alphabetischen Reihenfolge vorgelegt. Die Bewertung lief in sechs Schritten ab (Tab. 2). Im dritten Teil der Befragung ging es um die Evaluation der Graphen (Tab. 2).

Von Teil 1 werden die wichtigsten Aspekte kurz zusammengefasst: Wie schon in einer vorherigen Studie festgestellt (Heck & Peters, 2010), arbeiten viele der befragten Forscher in kleinen Forschungsgruppen, die in der Regel nicht mehr als fünf Mitglieder haben. Neue Kooperationspartner werden oft auf Konferenzen oder über Kollegen gefunden. Je nachdem, wie lange der Wissenschaftler schon auf seinem Gebiet arbeitet, hat er mehr oder weniger starke Kontakte zu seinem wissenschaftlichen Netzwerk, d. h. es ist für einen „novice researcher“ schwieriger, weil er sich in seinem Gebiet unter den anderen Kollegen noch nicht so stark etabliert hat. Für die Wahl eines Kooperationspartners gibt es verschiedene Faktoren. Die räumliche Nähe kann wichtig sein, wenn es zum Beispiel für ein Experiment nötig ist, gewisse Gerätschaften o. ä. auszutauschen. Ansonsten ist das Forschungsgebiet ausschlaggebend: Entweder gibt es ein sehr hohes gemeinsames Interesse, oder aber die Arbeiten der Kollegen ergänzen sich. Als Gründe, mit einem Kollegen keine Kooperation einzugehen, wurden zum einen eine sehr hohe thematische Überschneidung im Forschungsgebiet (Konkurrenzgedanke) und zum anderen

Fragen zur Autorenliste	Fragen zu den Graphen
1. Kennen Sie diesen Autor (persönlich)?	7. Entspricht dieser Graph Ihrer Einschätzung nach der Realität bezüglich den Verteilungen der Autoren und ihrer Nähe zueinander?
2. Wenn ja, haben Sie mit ihm schon mal kooperiert?	8. Entsprechen die Autorengruppen der Realität und sehen Sie sich richtig zugeordnet?
3. Sind Sie dem Autor bezüglich Ihrer Forschung ähnlich?	9. Würde dieser Graph Ihnen helfen, geeignete Partner zu finden, z. B. für Kooperationen oder wenn Sie eine Konferenz organisieren möchten ?
4. Bitte schätzen Sie die Relevanz des Autors für Ihre aktuelle persönliche Forschung ein (1 = nicht relevant, 10 = sehr relevant)	10. Bitte bewerten Sie den Graph bezüglich der Frage Nr. 3 (1 = nicht hilfreich, 10 = sehr hilfreich)
5. Würden Sie mit dem Autor kooperieren (wenn nicht, was sind die Gründe)?	
6. Fehlen Ihnen Autoren, die für Ihre aktuelle Forschung wichtig und relevant sind?	

Tabelle 2: Fragen an die Probanden zu den Clustern (Autorenlisten) und den Graphen.

eine aktuell zu geringe thematisch Überschneidung angegeben (siehe Frage 5, Tab. 2).

Der zweite Evaluationsteil betrifft die Bewertung der Autoren auf der Autorenliste. Bei Frage 4 (Tab. 2) sollten die Autoren nach ihrer Relevanz für die aktuelle Forschung des Zielautors bewertet werden. Für die Analyse der Verteilung wichtiger Autoren in den einzelnen Clustern wurden alle Autoren genommen, die der Zielautor mit fünf und besser bewertet hat, sowie alle Autoren, die dem Zielautor für die aktuelle Forschung wichtig sind und nicht auf der Liste standen (Frage 6, Tab. 2). Diese Autoren werden als relevant definiert, d. h. sie sind potentieller Kooperationspartner, und werden für die folgenden Berechnungen mit einbezogen (Co-Autoren ausgeschlossen). Jeder Zielautor hat eine unterschiedliche Anzahl relevanter Autoren, bei den sechs Zielautoren variierte die Zahl zwischen 18 und 55. Abbildung 6 zeigt die Verteilung relevanter Autoren in den einzelnen Diensten. Die Anzahl der wichtigen Personen ist für Scopus, WoS und CiteULike fast gleich, jedoch gibt es wenige Überschneidungen, d. h. nur wenige Autoren sind in mehr als einem Dienst zu finden. Nur zwölf Autoren waren in allen drei Diensten. Dies deutet darauf hin, dass unterschiedliche wichtige Personen in den Diensten gefunden werden und die höchste Anzahl erreicht wird, wenn alle drei Dienste und Methoden für die Ähnlichkeitsberechnungen genutzt werden. Die Zahlen zum Beispiel für Autor 1 unterstützten die These (Tab. 3). Sechs der insgesamt 29 relevanten Autoren von Autor 1 wurden nur in CiteULike gefunden, das ACC- und das BC-Cluster haben nur einen einzigen gemeinsamen Autor.

Abbildung 7 zeigt die Abdeckung relevanter Autoren für die einzelnen Zielautoren unter den ersten 20 Plätzen (nach dem Kosinus-Maß) in einem Cluster. Zum Beispiel sind bei Autor 1 30 Prozent der Top 20 Autoren im ACC-Cluster relevant. Im BC-Cluster sind es nur 20 Prozent, im CULU-Cluster 30 Prozent sowie im CULT-Cluster 25 Prozent. Zwischen den Zielautoren gibt es große Unterschiede: Bei den Autoren 2, 3 und 4 (6 ausgeschlossen, da keine CiteULike Daten vorhanden) scheinen das ACC-Cluster und das BC-Cluster am besten abzuschneiden, wohingegen bei den Autoren 1 und 5 die CULU- bzw. CULT-Cluster gleich gut oder besser sind. Bei Autor 5 (keine Tags in CiteULike, nur Nutzer) sind alle Top 20 Autoren als relevant eingestuft. Die teils großen Unterschiede können zum einen daher rühren, dass einige Probanden ihr Forschungsgebiet leicht verändert haben und aktuell andere Schwerpunkte haben. Dieser Umstand kann mit Daten, die auf vergangene Publikationen beruhen, leider

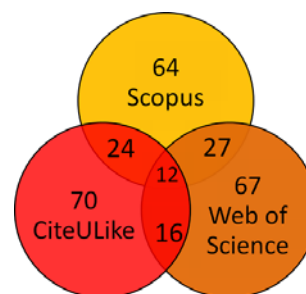


Abb. 6: Anzahl relevanter Autoren aller Zielautoren, die in den drei Diensten gefunden wurden.

	Scopus	WoS	CiteULike
Scopus	11(3)	1	9
WoS	1	5(3)	2
CiteULike	9	2	15(6)

Table 3: Anzahl relevanter Autoren in den einzelnen Diensten von Autor 1. Zahl in Klammer = Anzahl Autoren, die nur in einem Dienst gefunden wurden.

nicht ausgeglichen werden. Ein Vorteil, Daten aus CiteULike statt Co-Zitationsdaten zu verwenden, wäre hier, dass die Nutzer die Publikationen schon mit dem Erscheinen bookmarken können. Bis eine Publikation zitiert wird, kann allerdings einige Zeit vergehen.

Im dritten Evaluationsteil bewerteten die Probanden die Graphen (siehe Fragen Tab. 2.). Durchschnittlich bekamen diese von den Zielautoren folgenden Wertungen (Relevanz von 1 bis 10; Autoren 5 und 6 bewerteten nur die AC- und BC-Graphen, da zu wenig Daten in CiteULike vorhanden waren):

- Autoren-Co-Zitationsanalyse (ACC-Graph): 5,08
- Bibliographic Coupling (BC-Graph): 8,70
- Nutzerbasiertes CF (CULU-Graph): 2,13
- Tagbasiertes CF (CULT-Graph): 5,25

Die Verteilung der Autoren und Autorengruppen in den Graphen war nach Angaben der Probanden bis auf wenige Ausnahmen richtig. Auch die eigene Platzierung in den Graphen fanden die Zielautoren richtig und ihrem Forschungsschwerpunkt entsprechend eingeordnet. Zwei Autoren gaben an, dass die BC- und CULT-Graphen am Relevantesten wären und eine Kombination dieser beiden Graphen die wichtigsten Autoren für ihre Forschung abdecken würden. Beide seien geeignet, um potentielle Kooperationspartner zu finden oder zum Beispiel eine

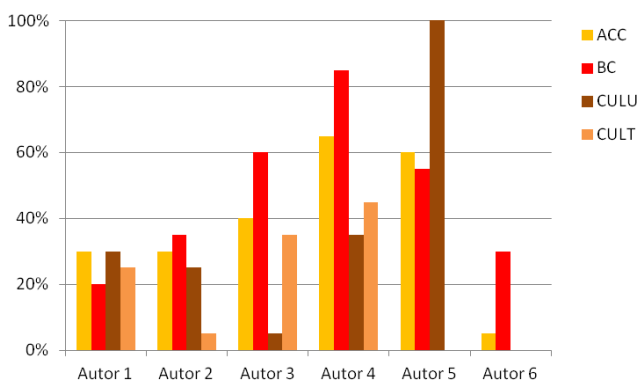


Abb. 7: Abdeckung relevanter Autoren unter den Top 20 Autoren eines Clusters. Ranking nach dem Kosinus-Maß.

Konferenz zu organisieren. Bei den ACC- und BC-Graphen merkten einige Probanden an, dass diese zwar relevante Autoren zeigten, diese aber auch schon bekannt und die Relationen zu offensichtlich seien. Um mehr neue potentielle Partner zu finden, seien größere Graphen interessant, die auch mehr Forschungsgebiete abdeckten und deren Verknüpfungen untereinander zeigten. Gerade für den „novice researcher“ wäre dies von Vorteil, um seine Forschung genauer einordnen zu können. Beim Anschauen der Graphen fielen allen Befragten Autoren auf, die für sie relevant sind, die ihnen aber bei der Befragung im zweiten Evaluationsteil nicht eingefallen sind. Daher wurden die Graphen und die Visualisierung der Autoren und Autorengruppen als sehr hilfreich empfunden, um Kooperationspartner zu suchen.

Ein wichtiger Aspekt bei den Graphen ist die Übersichtlichkeit und die sichtbare Abgrenzung von Autorengruppen entsprechend ihres Forschungsgebietes. Bei den Zielautoren, die nur wenige Tags und Nutzer in CiteULike hatten, war dieser Aspekt problematisch. Denn die Graphen zeigten keine eindeutigen Autorengruppen, sondern eher große Autorenfelder, und es war für den Probanden daher schwieriger, sich in den Graphen zurechtzufinden und relevante Autoren bzw. Forschergruppen zu erkennen. Hier wurde von den Befragten angemerkt, dass in den Graphen zusätzliche Informationen angezeigt werden könnten, wie zum Beispiel die Tags, die Auskunft über die Forschungsschwerpunkte der Autoren geben könnten.

6 Fazit

Mit dem Wachsen des Social Web wird es neue und mehr Informationen über Nutzer und auch Forscher geben. Diese sozialen Informationen können unter anderem für Expertenempfehlungssysteme genutzt werden. Die Befragung hat gezeigt, dass gerade die „novice researcher“ ständig neue Kooperationspartner suchen. Viele Forscher arbeiten in kleinen Teams zusammen und daher ist es wichtig, sich untereinander zu vernetzen. Ein Empfehlungssystem, das einem Wissenschaftler potentielle Partner vorschlägt, könnte hier unterstützen. Es gibt zurzeit verschiedenen Methoden und Vorgehensweisen, um einem Zielnutzer personalisierte Empfehlungen zu Ressourcen, Tags und auch Nutzern zu geben. Der vorgestellte Ansatz nutzt hier das Collaborative Filtering und die Daten aus einem Social-Bookmarking-System um einem Zielautor potentielle Kooperationspartner vorzuschlagen. Darüber hinaus kommen die Autoren-Co-Zitationsanalyse und das Bibliographic Coupling hinzu,

zwei Ansätze aus der Szientometrie. Alle drei Methoden nutzen unterschiedliche soziale Informationen aus drei verschiedenen Diensten. Erste Ergebnisse mit Probanden haben gezeigt, dass mit der Kombination der Ansätze und der Datensätze die meisten relevanten Autoren gefunden werden. Die Visualisierung der Autorenrelationen in Graphen wurde als sehr hilfreich empfunden, da Autorengruppen und deren Forschungsgebiete aufgedeckt und neue Kooperationspartner gefunden wurden. Eine Aufgabe ist es nun, diese Methoden miteinander zu verknüpfen. Da die Kosinus-Werte auf Grund der unterschiedlichen Daten sehr voneinander abweichen, ist eine einfache Summierung der Werte für einen Autor nicht geeignet. Technische Aspekte für ein laufendes Recommender System wären unter anderem die Effektivität und die Genauigkeit der Vorschläge. Daneben sollte auch der Nutzer in die Evaluation des Systems mit einbezogen werden – denn nur wenn er dem System vertraut und darin einen Mehrwert sieht, wird er dieses auch nutzen.

7 Danksagung

Danke an Wolfgang G. Stock, Isabella Peters, Oliver Hanraths und die Physiker, die sich an der Evaluation beteiligt haben. Das Projekt wurde gefördert vom Strategischen Forschungsfonds der Heinrich-Heine-Universität Düsseldorf.

Literatur

- Ahn, H. J. (2008). A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. In: *Information Sciences* 178, 37–51.
- Au Yeung, C. M., Noll, M., Gibbins, N., Meinel, C., & Shadbolt, N. (2009). On measuring expertise in collaborative tagging systems. In: *Proceedings of Web Science Conference: Society On-Line*.
- Ben Jabeur, L., Tamine, L., & Boughanem, M. (2010). A social model for literature access: towards a weighted social network of authors. In: *Proceedings of RIAO '10 International Conference on Adaptivity, Personalization and Fusion of Heterogeneous Information*, 32–39.
- Berkovsky, S., Kuflik, T., & Ricci F. (2007). Mediation of user models for enhanced personalization in recommender systems. In: *User Modeling and User-Adapted Interaction* 18 (3), 245–286.
- Blazek, R. (2007). *Author-Statement Citation Analysis Applied as a Recommender System to Support Non-Domain-Expert Academic Research (Doctoral Dissertation)*. Fort Lauderdale, FL: Nova Southeastern University.
- Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation. Which citation approach represents the research front most accurately? In: *Journal of the American Society for Information Science and Technology* 61(12), 2389–2404.
- Cawkell, T. (2000). Methods of information retrieval using Web of Science. Pulmonary hypertension as a subject example. In: *Journal of Information Science* 26(1), 66–70.
- Cronin, B. (1984). *The Citation Process. The role and significance of citations in scientific communication*. London, UK: Taylor Graham.
- Cronin, B., Shaw, D., & La Barre, K. (2003). A cast of thousands: Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. In: *Journal of the American Society for Information Science and Technology* 54(9), 855–871.
- Cruz, C. C. P., Motta, C. L. R., Santoro, F. M., & Elia, M. (2009). Applying reputation mechanisms in communities of practice. A case study. In: *Journal of Universal Computer Science* 15(9), 1886–1906.
- Desrosiers, C., & Karypis, G. (2011). A comprehensive survey of neighborhood-based recommendation methods. In: F. Ricci, L. Rokach, B. Shapira, & P.B. Kantor (Eds.), *Recommender Systems Handbook* (pp. 197–144). New York, NY: Springer.
- Gmur, M. (2003). Co-citation analysis and the search for invisible colleges. A methodological evaluation. In: *Scientometrics* 57(1), 27–57.
- Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. In: *Communications of the ACM* 35(12), 61–70.
- Heck, T. (2012). Analyse von Folksonomy-basierten Netzwerken als komplementärer Ansatz für Autoreneempfehlungen in der Wissenschaft. In: *Social Media & Web Science. Das Web als Lebensraum. Proceedings der 2. DGI-Konferenz, 64. Jahrestagung der DGI* (pp. 179–193). Frankfurt a. M.: DGI.
- Heck, T., & Peters, I. (2010). Expert recommender systems: Establishing Communities of Practice based on social bookmarking systems. In: *Proceedings of the 10th International Conference on Knowledge Management and Knowledge Technologies* (pp. 458–464).
- Heck, T., Hanraths, O., & Stock, W.G. (2011). Expert recommendation for knowledge management in academia. In: *Proceedings of the 74th ASIS&T Annual Meeting, Vol. 48*.
- Heck, T., Peters, I., & Stock, W.G. (2011). Testing collaborative filtering against co-citation analysis and bibliographic coupling for academic author recommendation. In: *Proceedings of the 3rd ACM RecSys'11 Workshop on Recommender Systems and the Social Web*.
- Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J. (1999). An algorithmic framework for performing collaborative filtering. In: *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 230–237). Berkeley, CA, USA .
- Herlocker, J. L., Konstan, J. A., Terveen L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. In: *ACM Transactions on Information Systems* 22(1), 5–53.
- Höfeld, M., & Kwiatkowski, M. (2007). Empfehlungssysteme aus informationswissenschaftlicher Sicht – State of the Art. In: *Information. Wissenschaft und Praxis* 58(5), 265–276.
- Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006). Information retrieval in folksonomies: Search and ranking. In: *Lecture Notes in Computer Science* 4011, 411–426.

- Kessler, M. M. (1963). Bibliographic coupling between scientific papers. In: *American Documentation* 14, 10–25.
- Krohn-Grimberghe, A., Nanopoulos, A., & Schmidt-Thieme, L. (2010). A novel multidimensional framework for evaluating recommender systems. In: *Proceedings of the ACM RecSys 2010 Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces*.
- Lee, D. H., & Brusilovsky, P. (2010a). Social networks and interest similarity. The case of CiteULike. In: *Proceedings of the 21st ACM Conference on Hypertext & Hypermedia* (pp. 151–155). New York, NY, USA: ACM.
- Leydesdorff, L. (2005). Similarity measures, author cocitation analysis, and information theory. In: *Journal of the American Society for Information Science and Technology* 56(7), 769–772.
- Li, J., Burnham, J. F., Lemley, T., & Britton, R. M. (2010). Citation analysis. Comparison of Web of Science, Scopus, SciFinder, and Google Scholar. In: *Journal of Electronic Resources in Medical Libraries* 7(3), 196–217.
- Linde, F., & Stock, W.G. (2011). *Information Markets*. Berlin, Germany, New York, NY: De Gruyter Saur.
- Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding Patterns of International Scientific Collaboration. In: *Science, Technology, & Human Values* 17(1), 101–126.
- Marinho, L. B., Nanopoulos, A., Schmidt-Thieme, L., Jäschke, R., Hotho, A., Stumme, G., & Symeonidis, P. (2011). Social tagging recommenders systems. In: F. Ricci, L. Rokach, B. Shapira & P.B. Kantor (Eds.): *Recommender Systems Handbook* (pp. 615–644). New York, NY: Springer.
- McNee, S. M., Kapoor, N., & Konstan, J.A. (2006). Don't look stupid. Avoiding pitfalls when recommending research papers. In: *Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work* (pp. 171–180).
- Meho, L. I., & Rogers, Y. (2008). Citation counting, citation ranking, and h-index of human-computer interaction researchers. A comparison of Scopus and Web of Science. In: *Journal of the American Society for Information Science and Technology* 59(11), 1711–1726.
- Meho, L. I., & Sugimoto, C. R. (2009). Assessing the scholarly impact of information studies. A tale of two citation databases – Scopus and Web of Science. In: *Journal of the American Society for Information Science and Technology* 60(12), 2499–2508.
- Parra, D., & Brusilovsky, P. (2009). Collaborative filtering for social tagging systems. An experiment with CiteULike. In: *Proceedings of the 3rd ACM Conference on Recommender Systems* (pp. 237–240). New York, NY: ACM.
- Peters, I. (2009). *Folksonomies. Indexing and Retrieval in Web 2.0*. Berlin: De Gruyter Saur.
- Ramezani, M., Bergman, L., Thompson, R., Burke, R., & Mobasher, B. (2008). Selecting and applying recommendation technology. In: *Proceedings of the International Workshop on Recommendation and Collaboration, in Conjunction with 2008 International ACM Conference on Intelligent User Interfaces*.
- Persson, O. (2001). All author citations versus first author citations. In: *Scientometrics* 50 (2), 339–344.
- Petry, H., Tedesco, P., Vieira, V., & Salgado, A. C. (2008). ICARE. A context-sensitive expert recommendation system. In: *Proceedings of Workshop on Recommender Systems on the 18th European Conference on Artificial Intelligence* (pp. 53–58).
- Reichling, T., & Wulf, V. (2009). Expert recommender systems in practice. Evaluating semi-automatic profile generation. In: *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (pp. 59–68).
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (pp. 175–186). New York, NY: ACM.
- Schneider, J. W., & Borlund, P. (2007a). Matrix Comparison, Part 1: Motivation and important issues for measuring the resemblance between proximity measures or ordination results. In: *Journal of the American Society for Information Science and Technology* 58(11), 1586–1595.
- Schneider, J. W., & Borlund, P. (2007b). Matrix Comparison, Part 2: Measuring the resemblance between proximity measures or ordination results by use of the Mantel and Procrustes statistics. In: *Journal of the American Society for Information Science and Technology* 58(11), 1596–1609.
- Shepitsen, A., Gemmell, J., Mobasher, B., & Burke, R. (2008). Personalized recommendation in social tagging systems using hierarchical clustering. In: *Proceedings of the 2008 ACM Conference on Recommender Systems* (pp. 259–266). New York, NY: ACM.
- Small, H. (1973). Cocitation in scientific literature. New measure of relationship between 2 documents. In: *Journal of the American Society for Information Science* 24(4), 265–269.
- Stock, W. G. (1999). Web of Science. Ein Netz wissenschaftlicher Informationen – gesponnen aus Fußnoten [Web of Science. A web of scientific information – cocooned from footnotes]. In: *Password* 7+8, 21–25.
- White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. In: *Journal of the American Society for Information Science* 32(3), 163–171.
- White, H. D., & Griffith, B. C. (1982). Authors as markers of intellectual space. Co-citation in studies of science, technology and society. In: *Journal of Documentation* 38(4), 255–272.
- Zhao, D., & Strotmann, A. (2007). All-author vs. first author co-citation analysis of the Information Science field using Scopus. In: *Proceedings of the 70th Annual Meeting of the American Society for Information Science and Technology*, 44(1), 1–12.



Tamara Heck
tamara.heck@uni-duesseldorf.de

Tamara Heck ist wissenschaftliche Mitarbeiterin der Abteilung für Informationswissenschaft an der Heinrich-Heine-Universität Düsseldorf. Sie studierte Neuere und Ältere Anglistik, Amerikanistik und Informationswissenschaft an der Universität Düsseldorf. In ihrer Promotionsarbeit in der Informationswissenschaft untersucht sie Ansätze zur Erstellung von Experten Recommender Systemen für den kooperativen Wissensaustausch und dem Aufbau von Experten-netzwerken.