

# Ein gradualisiertes Community-Modell zur Bildung wissenschaftlicher Gemeinschaften

Michael Hahne, Corinna Jung, Jérôme Kunegis,  
Andreas Lommatzsch, Andre Paus

*Für eine optimale Nutzung von Social Software müssen bei ihrer Entwicklung Besonderheiten von potenziellen Akteuren und Communities berücksichtigt werden. Das gradualisierte Communitymodell bietet hierfür einen Leitfaden. Der Artikel präsentiert Ergebnisse der empirischen Überprüfung dieses Modells sowie im Projekt „PIA+COMM“ entwickelte Funktionen zur Prüfung seiner praktische Plausibilität.*

## 1 Einleitung

In jüngster Zeit wird Social Software sehr viel diskutiert. Die Grundidee dabei ist, räumlich verteilte Akteure zusammen zu bringen und sie bei gemeinsamen Aufgaben zu unterstützen. Die jeweils verwendeten Anwendungen fördern und gestalten dabei die Entstehung von Communities, also sozialen Netzwerken zwischen Nutzern.

Im Projekt „PIA+COMM“ wird für Wissenschaftler aus dem Bereich Informatik und Elektrotechnik der TU Berlin eine Social Softwareanwendung entwickelt<sup>1</sup>, die die Entstehung und den Erhalt von Communities fördern soll. In einem iterativen Entwicklungsprozess wird kontinuierlich der Bedarf der Nutzer, die Usability und die Akzeptanz der Funktionalitäten evaluiert. Die Ergebnisse dieser Evaluationen werden bei der technischen Entwicklung berücksichtigt.

Dass räumlich verteilt agierende Akteure Social Software für Community-Aktivitäten nutzen, ist nicht selbstverständlich, sondern muss bei der Planung einer Anwendung als kritische Komponente

---

<sup>1</sup> Der folgende Artikel stellt Zwischenergebnisse aus dem BMBF geförderten Forschungsprojekt zur Entwicklung der Suchmaschine PIA+COMM vor.

einkalkuliert werden. Deshalb müssen die Bedürfnisse und Erwartungen der Nutzer berücksichtigt sowie angemessene Funktionen entwickelt werden, die das Entstehen von Community-Typen (s.u.) ermöglichen.

Welchen Kriterien die einzelnen Funktionen dabei entsprechen müssen, beschreibt das „gradualisierten Communitymodell“. Das Modell schlägt eine analytische Kategorisierung von Community-Typen mit unterschiedlichen Entstehungskriterien und Unterstützungspotenzialen vor (vgl. Abschnitt 2). Ziel ist es, die empirischen Untersuchungen und programmiertechnischen Entwicklungen anzuleiten und über das Projekt hinaus, einen Leitfaden für die Entwicklung von Social Software bereitzustellen.

Die theoretischen Annahmen des Modells haben wir empirisch untersucht, indem wir eine breit angelegte Bedarfsanalyse durchgeführt haben (vgl. Abschnitt 3). Dabei ging es erstens um die Frage, wann und warum Wissenschaftler an Communities teilnehmen. Beabsichtigt war, einen Überblick über vorhandene Communities zu erhalten sowie das Potenzial für den Einsatz von Social Software in diesem Bereich zu ermitteln. Zweitens wurde erfragt, wie Wissenschaftler ihre sozialen Beziehungen realisieren und ob sie hierzu bereits auf technische Unterstützung, insbesondere Social Software, zurückgreifen. Um konkrete Gestaltungskriterien für die Entwicklung der Community-Funktionen abzuleiten, haben wir drittens nach dem Bedarf und der Gestaltung von Social Software gefragt.

Der in der Empirie identifizierte Bedarf nach Community-Unterstützung wurde teilweise in Funktionen übersetzt, deren Nutzung wiederum untersucht wurde.

Der vierte Abschnitt stellt exemplarisch zwei bereits realisierte Funktionen vor sowie eine dritte experimentelle, die in der nächsten Implementierungsrunde umgesetzt wird.

## **2 Das gradualisierte Community-Modell**

Im Folgenden werden die Begriffe „Community“ und „Gemeinschaft“ gleichgesetzt<sup>2</sup>. Das „gradualisierte Community-Modell“ trägt

---

<sup>2</sup> Die größere Reichweite des englischen Begriffs ist uns bekannt, kann jedoch an dieser Stelle nicht erörtert werden (vgl. dazu ausführlich Wellman, B., and B. Leighton. 1979. "Networks, Neighborhoods, and Communities: Approaches to the Study of the Community Question." *Urban Affairs Quarterly* 14:363-390, Wellman, Barry.

der kontinuierlichen Transformation des Gemeinschaftsbegriffs seit der ursprünglichen Definition durch Tönnies Rechnung (Tönnies 1991/1887). Insofern gibt das Modell keine feste Definition von Community vor, sondern ordnet unterschiedlichen Community-Graden notwendige Kriterien zu. Die Grade repräsentieren dabei die Komplexität der Communities, d.h. wie stark die unterschiedlichen Communities dem Idealtypus einer Gemeinschaft (alle Kriterien erfüllt) entsprechen.

Die Gradualisierung erfolgt entlang zweier Dimensionen: Dauer, als kurz- und langfristige Beziehungen, und Anzahl, als dyadische (Zweierbeziehungen) und kollektive Beziehungen. Daraus ergeben sich vier Community-Typkombinationen: Dyadisch/kurzfristig (wie z.B. die meisten Internetforen), dyadisch/langfristig (z.B. Netzwerkbeziehungen zu Fachkollegen), kollektiv/kurzfristig (z.B. Arbeitskreise) und kollektiv/langfristig (z.B. Freundschaftscliquen, Familie, Arbeitskollegen). Potenzielle Gemeinschaften bilden den fünften und basalen Typ, da er notwendige Bedingung für die folgenden vier Typen ist. Diese zeichnen sich dadurch aus, dass ihre Mitglieder eine Gemeinsamkeit haben, noch bevor sie sich überhaupt wechselseitig wahrnehmen

Aus den Erkenntnissen der Sozialpsychologie und der Soziologie erscheinen uns folgende Kriterien geeignet, um diese fünf Community-Typen zu beschreiben (vgl. Tabelle 1)

Potenzielle Communities zeichnet aus, dass die Akteure eine der realen Gemeinschaft vorgängige Gemeinsamkeit (Kriterium 1) haben. Unterstützungspotenzial dafür liegt in Funktionen, bei denen Nutzer von den Erkenntnissen und Interessen anderer Nutzer profitieren, ohne mit diesen in Kontakt zu treten.

Bei dyadisch/kurzfristigen Communities kommt die wechselseitige Wahrnehmung (Awareness) als zweites Kriterium hinzu (Dourish and Bellotti 1992). Gemeint sind Informationen über Existenz, Kontaktmöglichkeit und geteilte Gemeinsamkeit mit anderen Akteuren. Technisch lässt sich dies durch die Identifizierung, Aufarbeitung und Bereitstellung dieser Informationen unterstützen. Weiteres Kriterium ist der Charakter der Interaktion (Durkheim 1977/1893; Mead 1967/1934; Streeck and Schmitter 1985) (Kriterium 3): Als Interaktionscharakter sind z.B. das Bedürfnis nach Zusammenarbeit oder

---

1979. "The Community Question: The Intimate Networks of East Yorkers." *American Journal of Sociology* 84:1201-1231.).

die Bereitschaft zu Solidarität denkbar. Technisches Unterstützungspotential besteht vor allem darin Austausch zu ermöglichen, indem man Kommunikationsmedien, wie z.B. Foren, Mailinglisten, Email, Instant Messenger, etc. zur Verfügung stellt.

Dyadisch/langfristige Communities haben neben den ersten drei noch weitere auf Langfristigkeit abzielende Kriterien: Die kontinuierliche Interaktion (Kriterium 4) sowie eine gemeinsame Aufgabe (Kriterium 5) (Mead 1967/1934) an der die Mitglieder sich orientieren bzw. aus der sie Anregungen für die Fortsetzung der Gemeinschaft ziehen. Die kontinuierliche Interaktion kann technisch durch Archivfunktionen unterstützt werden, die kontinuierliche Anschlussfähigkeit gewährleisten und die gemeinsame Aufgabe durch gemeinsame Bearbeitungswerkzeuge sowie über die Generierung immer neuer gemeinsamer Aufgaben.

Die kollektiv/kurzfristigen Communities werden durch Kriterien, die auf Kollektivität abzielen, ergänzt: Erstens durch auf die Gemeinschaft bezogene Interaktionen (Kriterium 6) (Krappmann 2000/1969; Mead 1967/1934; Schäfers 1994; Tajfel 1978; Tajfel 1982a; Tajfel 1982b; Weber 1976/1922). Zweitens durch ein positives Gemeinschaftsgefühl (Kriterium 7) (Goffman 1975/1963; Krappmann 2000/1969; Mead 1967/1934; Schäfers 1994; Tajfel 1978; Tajfel 1982a; Tajfel 1982b). Interaktionen der Gemeinschaft zielen auf Koordinationsfragen, auf wechselseitige Informierung und auf Austausch der Mitglieder untereinander ab. Technisch lassen sich diese Interaktionen fördern, indem man Mechanismen zur Koordination der Mitglieder sowie zum informellen Austausch zur Verfügung stellt. Das Gemeinschaftsgefühl könnte technisch durch Funktionen zum Konfliktmanagement (Krappmann 2000/1969) unterstützt werden.

Bei den kollektiv/langfristigen Communities addieren sich alle bisher genannten Kriterien und werden zusätzlich durch zwei neue ergänzt: Erstens ein gemeinschaftsbezogenes Stereotyp (verallgemeinerte Vorstellung) (Kriterium 8) [Tajfel 1978, 1982a], das es ermöglicht, jederzeit im Sinne der Gemeinschaft zu handeln, sich selbst in Bezug auf die Gemeinschaft zu verorten und sich von anderen Gemeinschaften abzugrenzen. Zweitens das so genannte Rollendifferenzial (Kriterium 9) (Krappmann 2000/1969; Schäfers 1994), dass jedes Mitglied in der Gemeinschaft eine Rolle findet, die den eigenen Fähigkeiten entspricht und zum Gelingen der gemeinsamen Aufgabe beiträgt. Unterstützung liegt für beide Kriterien in Reprä-

sentationsfunktionen. Beim Stereotyp geht es mehr um externe Repräsentation, um Abgrenzung gegenüber anderen Gemeinschaften (hierunter fällt z.B. die Darstellung der Community über eine Website oder ein repräsentatives Gebäude). Beim Rollendifferenzial um Repräsentation nach innen, das jeder in der Gemeinschaft weiss, wer wofür zuständig ist (hier geht es um Organigramme, Professionsbezeichnungen, einen größeren Schreibtisch oder Dienstwagen, usw.). Schließlich lässt sich das Rollendifferenzial durch Funktionen unterstützen, die es jedem Mitglied erlauben, möglichst uneingeschränkt seine Fähigkeiten einzubringen (vgl. Tabelle 1).

Typen	Minimalkriterien	Unterstützung
Potentielle Community	Pool von Akteuren mit gemeinsamen Interessen etc., die der Kenntnis der Akteure vorgängig sind (K 1)	Identifikation von Gemeinsamkeiten
Dyadisch kurzfristige Community	K1 und K2: wechselseitige Wahrnehmung K3: Charakter der Interaktion	Awareness Kommunikation
Dyadisch langfristige Community	K1 bis K3 und K4: kontinuierliche Interaktionen K5: gemeinsame Aufgabe/Objekte	Archiv Identifikation gemeinsamer Aufgaben kollaborative Werkzeuge
Kollektiv kurzfristige Community	K1 bis K3 und K6: auf die Gemeinschaft bezogene Interaktionen K7: positives Gemeinschaftsgefühl	Koordinationsmechanismen Konfliktmanagement, Vergleich
Kollektiv langfristige Community	K1 bis K7 und K8: gemeinschaftsbezogenes Stereotyp K9: Rollendifferenzial	Interne/ Externe Repräsentation Entfaltungsmöglichkeiten

**Tabelle 1:** Das gradualisierte Communitymodell

### 3 Empirie

Gemäß dem Ideal eines iterativen Systementwicklungsprozesses wird eine formative - das heißt begleitende - Evaluation der PIA+COMM-Plattform durchgeführt. Diese beinhaltet Bedarfsanalysen, Usability-Tests und Akzeptanzforschung. Konkret bedeutet das

eine an die jeweilige Entwicklungsstufe der Plattform angepasste sozialwissenschaftliche Evaluation, die den Veränderungen des Systems Rechnung trägt und deren Ergebnisse wieder in die programmiertechnische Entwicklung zurückfließen.

### **3.1 Vorgehensweise**

Wir haben für verschiedene Nutzergruppen – Studierende und wissenschaftliche Mitarbeiter – spezielle Leitfadeninterviews (vgl. Flick 1999; Flick 2002; Flick, Kardorff and Steinke 2000) entwickelt, mit denen wir detailliert auf den Bedarf an Communities und Social Software sowie das Nutzungsverhalten dieser beiden Gruppen eingehen und diese weiter ausdifferenzieren konnten.

Insgesamt wurden 23 Mitglieder der Fakultät Elektrotechnik und Informatik der TU-Berlin befragt. Davon war die überwiegende Mehrheit wissenschaftliche Mitarbeiter, da diese als Hauptbenutzergruppe des Systems avisiert waren. Darüber hinaus wurden einige Studierende befragt, um die Übertragbarkeit auf andere Nutzergruppen zu prüfen, bzw. unterschiedliche Bedarfslagen, Nutzungsverhaltensweisen zu identifizieren. Die Personen wurden nach dem Zufallsprinzip ausgewählt. Die Interviews waren Teil der zweiten Iterationsphase des Evaluationsprozesses zu PIA+COMM von insgesamt drei Iterationen und wurden nach der Einführung einer ersten Testversion des Systems, vier Monate nach Projektbeginn, durchgeführt.

Wir haben uns in dieser Phase des Projekts für eine qualitative Untersuchung entschieden, weil diese die beste Möglichkeit bietet, Bedarf und Nutzungsverhalten der Probanden zu untersuchen sowie auf deren Probleme, Wünsche und Motivationen im Zusammenhang mit Communities und Social Software einzugehen. Die Chancen qualitativer Methoden liegen in ihrer Flexibilität und Kontextbezogenheit, auch Fragen, die erst in der Interviewsituation aufkommen, können noch gestellt werden. Außerdem können auf Basis qualitativer Ergebnisse genau auf die Zielgruppe zugeschnittene Fragebögen für später geplante quantitative Befragungen konzipiert werden.

Die Interviews dauerten zwischen 60 und 90 Minuten. Sie wurden aufgezeichnet, transkribiert und anonymisiert. Die Transkription erfolgte nach den Richtlinien von Flick (Flick 1999; Flick 2002; Flick, Kardorff and Steinke 2000). Ausgewertet wurden sie – wie auch die Transkriptionen der Usability-Tests – nach der Methode des offenen und axialen Kodierens von Strauss und Corbin (Böhm, Le-

gewie and Muhr 1992; Strauss and Corbin 1990) in Kombination mit Vorschlägen von Philipp Mayring zur Qualitativen Inhaltsanalyse (Mayring 1983; Mayring 2000).

### **3.2 Ergebnisse**

Hinsichtlich der ersten eingangs gestellten Frage, welche Formen von Communities in der scientific community existieren, wurde im Rahmen unserer Erhebung auf die Wichtigkeit lokaler Communities im Fachbereich und innerhalb der Universität hingewiesen. Zu diesen gehören institutionalisierte Treffen innerhalb der Fachbereiche wie z.B. Kolloquien, vor allem aber auch die informellen Gespräche, etwa auf dem Flur oder in der Cafeteria. In informellen Gesprächen können Probleme detaillierter diskutiert und Beziehungen gefestigt werden. Projekte (inter- und intradisziplinäre) sind eine weitere Form von Communities. Es geht dabei um die konkrete Zusammenarbeit an gemeinsamen Aufgaben. Es gibt aber auch eine Reihe von translokalen Communities. Zu diesen gehören u.a. (regelmäßige) Zusammentreffen von Wissenschaftlern auf Konferenzen, Online-Communities und der sporadische elektronische Austausch mit Wissenschaftlern, zu denen sonst keine nähere Beziehung besteht. Konferenzen bieten Wissenschaftlern oft neuen Input von außen, vereinzelt bieten sich auch Möglichkeiten zu Kooperationen. Schließlich sind Konferenzen ein Ort zur Kontaktpflege mit ehemaligen Kollegen.

Beim elektronischen Austausch mit Wissenschaftlern geht es darum, zügig spezifisches Wissen oder schlechter zugängliche Wissensbestände, zu erhalten. Online-Communities werden von den befragten Informatikern genutzt, jedoch ist ihr Stellenwert relativ klein. Verwendung finden Newsgroups, Mailinglisten, Webforen und Webgroups. Andere Formen von Onlinecommunities, wie z.B. Chaträume, werden im Zusammenhang mit der Arbeit von den Befragten nicht appliziert. Wir können festhalten, dass sich in wissenschaftlichen Online-Communities des betrachteten Bereiches die Interaktion auf sehr wenige Interaktionssequenzen beschränkt. Motivation ist in der Regel die schnelle Beschaffung von Informationen.

Zur zweiten Frage, wie Wissenschaftler Beziehungen realisieren und inwieweit sie dafür technische Hilfsmittel in Anspruch nehmen, können wir festhalten, dass es bereits ein großes Repertoire an Praktiken zur Pflege wissenschaftlicher Beziehungen und Kooperation

gibt. Dieses reicht von ortsgebundener Face-to-Face-Kommunikation über E-Mailverkehr bis hin zum Besuch von Konferenzen.

Am häufigsten genanntes technisches Hilfsmittel war die E-Mail. Sie wird z.B. zur Terminabsprache verwendet oder wenn Asynchronität der Interaktion gewünscht ist. Der Kollege kann dann seine E-Mails beantworten, wann er Zeit hat

Weitere technische Hilfsmittel sind die bereits erwähnten Anwendungen: Newsgroups, Mailinglisten, Webforen und Webgroups<sup>3</sup>. In unseren Untersuchungen zeigte sich, dass die Communities vor Ort wegen des informellen Austauschs zur Gewinnung neuer Ideen für die Befragten die größte Bedeutung haben. Grund dafür ist die Tatsache, dass Wissen eine soziale Ressource ist, die am besten informell transportiert wird. Es vermehrt sich durch wechselseitige Zirkulation. Die Wahrscheinlichkeit einer „guten Idee“ ist daher größer, wenn Personen ihr Wissen untereinander austauschen. Dieses Phänomen ist in den Sozialwissenschaften als Emergenz bekannt (Aristoteles 1994; Luhmann 1987).

Die dritte Frage zur Gestaltung von Social Software wurde ebenfalls beantwortet. Aus Platzgründen werden hier nur die relevanten Punkte genannt. Der Bedarf nach Online-Communities korrespondiert mit dem Bedarf an problembezogenen Informationen. Wissenschaftler nutzen wissenschaftliche Online-Communities dann, wenn sie ein Problem haben, aber weder interne Kollegen helfen können noch eine klassische Internetrecherche ein Ergebnis bringt.

In diesem Zusammenhang besteht der Bedarf nach sozialer Unterstützung der Informationsrecherche indem z.B. Pools von Begriffen zur Verfügung gestellt werden, deren Ähnlichkeit durch andere Nutzer bestimmt wird. Dieser Pool könnte mit Hilfe semantischer Abstraktion zusammengestellt werden. Suchende könnten nach Eingabe eines Begriffs weitere Begriffe in bestimmten Rubriken nachschlagen. Darüber hinaus besteht aber auch Bedarf, Kontakt zu Menschen herzustellen, die als Experten im gesuchten Bereich gelten.

Alle Befragten äußerten Bedarf an übersichtlicheren Verwaltungsfunktionen für eigene und vom System gefundene Dokumente. Besonders wurde eine Verwaltung mit kollaborativer Verschlagwortung

---

<sup>3</sup> Der wichtigste Unterschied zwischen diesen vier verschiedenen Anwendungen besteht darin, dass Mailinglisten und Newsgroups vor allem von erfahreneren Wissenschaftlern genutzt werden, während die jüngeren lieber Webforen und Webgroups verwenden.

als wünschenswert genannt, um Dokumente und andere Inhalte sinnvoller ordnen und mit anderen gemeinsam verwenden zu können.

Für die gemeinsame Arbeit in Projekten wurden Anwendungen gewünscht wie z.B. Whiteboard- oder Content-Management-Systeme. Erstere weisen auf den Bedarf von Informatikern hin, sich kollaborativ über visuelle Modelle und Schaubilder zu verständigen.

Schließlich wurden Termin- und Organisationskalender als potenziell wünschenswerte Funktionen genannt. Hier ist jedoch darauf aufmerksam gemacht worden, dass kollaborative Terminkalender oft recht schnell nicht mehr verwendeten wurden, da sie hinsichtlich Privacy, Usability und Mobility meist mangelhaft sind.

### 3.3 Diskussion der Ergebnisse

Die beschriebenen Ergebnisse werden auf die beiden Dimensionen des graduierten Community-Modells abgebildet. Es zeigt sich, dass viele der identifizierten Communities nicht eindeutig einer bestimmten Position zugeordnet werden können.

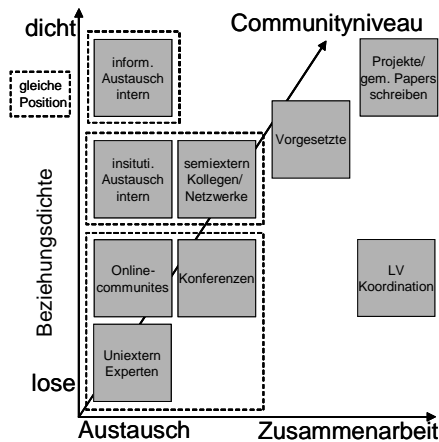


Abbildung 1: Das überarbeitete graduierte Communitymodell

Es gibt z.B. Fälle von Communities mit externen Experten oder Online-Communities, die im Bereich der dyadisch/kurzfristigen und im Bereich der dyadisch/langfristigen Communitytypen eingeordnet werden müssten, weil sich Beispiele finden, die sowohl langfristige

als auch kurzfristige Interaktionsbeziehungen beschreiben. Daher haben wir die Dimensionen angepasst und uns dafür entschieden die praktischen Interaktionsarten als Differenzierungskriterium heranzuziehen. Diese Entscheidung lässt eine klare Zweiteilung in Gemeinschaften zum Wissensaustausch und Gemeinschaften zur kollaborativen Zusammenarbeit zu (vgl. Abbildung 1).

Auch bei der zweiten Dimension kurzfristig/langfristig haben wir eine treffendere Unterscheidung, nach der Dichte der Beziehung, gefunden: Dabei gibt es lose Beziehungen, die z.B. nur einige Male im Jahr oder zu einer Konferenz stattfinden und dichte Beziehungen, die meist ortsgebunden und häufig informell sind.

<b>Typen</b>	<b>Minimalkriterien</b>
Potentielle Community	K1: Pool von Akteuren mit gemeinsamen Interessen etc., die vor der Kenntnis der Akteure voneinander entstehen
Lose Austausch Beziehung	K1 und K2: wechselseitige Wahrnehmung K3: Charakter der Interaktion: Zusammenarbeit oder Austausch
Lose Kooperations Beziehung	K1 bis K3 und K4: auf die Gemeinschaft bezogene Interaktionen K5: gemeinsame Aufgabe/Objekt
Dichte Austauschbeziehung	K1 bis K3 und K6: positives Gemeinschaftsgefühl K7: kontinuierliche Interaktion
Dichte Kooperationsbeziehung	K1 bis K7 und K8: gemeinschaftsbezogenes Stereotyp K9: Rollendifferenzial

**Tabelle 2:** Das überarbeitete gradualisierte Communitymodell

Die Beziehungen zu den Vorgesetzten stehen zwischen diesen Typen, was ihre besondere Bedeutung hervorhebt. Aus der Perspektive der Mitarbeiter ist der Vorgesetzte sowohl eine Ressource für fachliches, organisatorisches und strategisches Wissen als auch ein Partner bei der Arbeit in Projekten und Lehrveranstaltungen.

Hinsichtlich der Kriterien wurden der zweite und dritte Beziehungstyp vertauscht, da das Community-Niveau der dichten Beziehungen wegen des vorhandenen Gemeinschaftsgefühls höher eingeschätzt wurde. Das Gemeinschaftsgefühl wird nun den Austauschbeziehungen zugerechnet, da es vor allem für dichte Beziehung beschrieben wurde. Das Kriterium „gemeinsame Aufgabe/Objekt (K5)“ wurde zu den Kooperationsbeziehungen hinzugefügt, da Kooperation immer eine gemeinsame Aufgabe impliziert. Somit ergibt sich die überarbeitete Kriterientabelle (Tabelle 2) für das „gradualisierte Community-Modell“.

Zuletzt werden die Bedarfsangaben auf die identifizierten Unterstützungspotenziale übertragen und separat in Tabelle 3 dargestellt

Es gibt ein großes Bedürfnis nach Funktionen im Bereich der Communities, bei denen vom Wissen anderer Nutzer profitiert werden kann. Die Befragten haben Interesse an Dokumenten und Suchbegriffen, die von anderen Nutzern positiv bewertet wurden. Derartige Funktionen passen gut zur Unterstützung von potenziellen Communities. Wie in Abschnitt 4 skizziert, bietet PIA+COMM eine Unterstützung bei der Dokumentensuche an, bei der den Nutzern Suchbegriffe vorgeschlagen werden.

Ferner besteht Bedarf daran, Experten kennen zu lernen. Dies setzt die wechselseitige Wahrnehmung voraus, die technisch durch Funktionen wie User-Pages, automatisierte Vorschläge von Experten oder durch die grafische Darstellung verwandter Experten mit Hilfe eines Userradars unterstützt werden kann (vgl. Abschnitt 4).

Für den Austausch mit interessanten Interaktionspartnern wünschen sich die Befragten unterschiedliche Kommunikationskanäle, die dem Charakter der Interaktion angemessen sind. Social Software sollte stets eine Vielzahl an Kommunikationsmedien zur Verfügung stellen; sowohl öffentliche als auch private.

Kommunikationsmedien, die mehrere Personen gleichzeitig ansprechen wie z.B. Newsgroups, Mailinglisten oder Foren genügen zudem der vorgeschlagenen Archivfunktionalität und ermöglichen somit kontinuierliche Interaktion.

Der größte Bedarf besteht an kollaborativen Bearbeitungswerkzeugen. Im Gegensatz zu Austauschbeziehungen sind den Befragten nur wenige Anwendungen für kollaborative Arbeit bekannt und diese wurden vorwiegend als defizitär beschrieben. Für PIA+COMM wur-

de eine experimentelle Funktion zur Identifikation gemeinsamer Dokumente entwickelt, die bestehenden Communities neue Anregungen für gemeinsame Arbeit bieten und dadurch die Kontinuität der Gemeinschaft gewährleisten soll (vgl. Abschnitt 4).

<b>Unterstützungspotentiale</b>	<b>Konkreter Unterstützungsbedarf</b>
Identifikation von Gemeinsamkeiten	Pool von Dokumenten und Begriffen anderer Nutzer
Awareness	Community-Radar, User-Page, Experten vorschlagen
Kommunikation	E-Mail, Newsgroups, Foren, Mailinglisten
Archiv	Newsgroups, Foren Mailinglisten
Identifikation gemeinsamer Aufgaben/Objekte, kollaborative Bearbeitungswerkzeugen	Dokumente für die Gemeinschaft als ganzes finden, Multiuser Dokumentensystem, Terminkalender, Contentmanagementsystem, etc.
Koordinationsmechanismen	Moderatoren, Informationskontrolle
Konfliktmanagement, Vergleich	Informationskontrolle, Unterstützung der informellen Kommunikation
Externe Repräsentation	Informationskontrolle, Contentmanagement
Interne Repräsentation, Entfaltungsmöglichkeiten	Kolloquien, gutes und offenes Arbeitsklima

**Tabelle 3: Unterstützungspotentiale**

Für die verbleibenden beiden Community-Typen konnte wesentlich weniger Unterstützungsbedarf identifiziert werden. Dies liegt unseres Erachtens vor allem daran, dass die dichten Beziehungsformen zu großen Teilen von informeller Kommunikation getragen werden und technische Unterstützung für die meisten Befragten eher nicht vorstellbar ist<sup>4</sup>. Daher ist es erforderlich sich mit Mechanismen zu behelfen, die die entsprechenden Kriterien in den vorhandenen Communities tragen. Hinsichtlich der Koordinationsmechanismen und des Konfliktmanagements besteht z.B. die Notwendigkeit eines

---

<sup>4</sup> In anderen Bereichen, wie zum Beispiel in Online-Spielegemeinschaften kann hingegen sehr gut nachgezeichnet werden, wie Technik die Entstehung auch der höheren Community-Niveaus sehr erfolgreich unterstützen kann [vgl. Hahne 2006], obwohl es sich dort nicht um eine einzelne Social Software handelt, die diese Unterstützung leistet, sondern um ein soziotechnisches Ensemble [Schubert 2006] aus den Spielern und verschiedensten technischen Unterstützungsfunktionen. Weitere Vorschläge für solche komplexeren Formen der Gemeinschaftsunterstützung finden sich außerdem z.B. im Zusammenhang mit Aktivitätstheorie bei Kuutti [Kuutti 1996]

Moderators in Online-Communities, sowie der Bedarf nach Funktionen zur Informationskontrolle, d.h. die Möglichkeit den unberechtigten Zugang zu persönlichen Informationen zu regulieren. Ansonsten wird Konfliktmanagement in realen Communities in erster Linie durch informelle Kommunikation realisiert. Hinsichtlich der Repräsentation der eigenen Community ist oft die Rede von der Website z.B. eines Fachbereichs. Hierzu besteht von Seiten der Betreuer der Bedarf nach Content-Management-Systemen, die es erlauben, die Informationen der unterschiedlichen Community-Mitglieder schnell zu integrieren. Hinsichtlich der persönlichen Entfaltungsmöglichkeiten und der internen Repräsentation nannten die Befragten vor allem ein gutes Arbeitsklima und die fachbereichsspezifischen Kolloquien als Hilfen, um die höheren Community-Kriterien zu ermöglichen. Tabelle 3 fasst die Unterstützungsmöglichkeiten noch einmal zusammen.

## **4 Funktionen**

Nachdem im vorhergehenden Teil die theoretischen Grundlagen der Bildung von Communities sowie die ermittelten Anforderungen an ein System zur Bildung und Unterstützung von Communities im wissenschaftlichen Bereich vorgestellt wurden, stellt der folgende Abschnitt die Funktionen vor, die im Anschluss an das gradualisierte Community-Modell für PIA+COMM entwickelt wurden.

PIA+COMM-System ist eine so genannte intelligente Suchmaschine, deren Entwickler von der Annahme ausgehen, dass Wissenschaftler sich für verschiedene Themengebiete interessieren und zu diesen regelmäßig Suchanfragen stellen. Das System bietet dazu die Möglichkeit, Accounts anzulegen, in denen Informationsprofile ("Suchanfragen") definiert und gespeichert werden können. Aus den expliziten Dokumentenbewertungen und Verhaltensmustern der Benutzer innerhalb dieser Informationsprofile lernt PIA+COMM deren implizite Präferenzen.

Neben der individuellen Suche nach Dokumenten stellt auch das Wissen von anderen Benutzern eine wichtige Informationsquelle für Wissenschaftler dar. PIA+COMM wird daher um Funktionen zur Etablierung und Unterstützung von Communities erweitert. In der ersten Phase des Projektes stehen dabei die Unterstützung potenzieller Communities ("Identifikation von Gemeinsamkeiten"), die Sichtbarmachung von Communities ("Awareness, Kommunikationsmög-

lichkeit") und das Unterstützen eines gemeinsamen Arbeitens ("kollaboratives Lösen von Aufgaben") im Vordergrund. Für jeden der drei Bereiche werden in den folgenden Abschnitten ausgewählte Dienste vorgestellt.

#### 4.1 Identifizieren von Gemeinsamkeiten der Benutzer

Die Basisfunktionalität für das Ermitteln von potenziellen Communities besteht im Identifizieren von Gemeinsamkeiten, d.h. ähnlichen Interessen von Benutzern. Ähnlichkeiten werden auf der Grundlage bewerteter Dokumente berechnet. Dazu haben die Nutzer die Möglichkeit, Dokumente mit einem Zahlenwert zwischen 1 („finde ich gut“) und 5 („finde ich schlecht“) zu bewerten.

Im folgenden Abschnitt wird der in PIA+COMM verwendete Algorithmus zur Berechnung von Ähnlichkeiten erläutert. Im Anschluss daran, werden Funktionalitäten vorgestellt, die auf dem entwickelten Algorithmus aufbauen.

#### 4.2 Berechnen der Ähnlichkeit

Die Grundidee der Ähnlichkeitsberechnung ist, Dokumentenbewertungen zu vergleichen, die von zwei verschiedenen Nutzern abgegeben wurden. Die übliche Vergleichsmethode besteht darin, den Korrelationskoeffizienten zwischen den Bewertungen zu berechnen (Shardanand and Maes 1995 Gleichung 4).

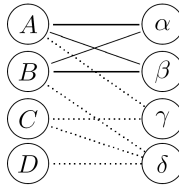
$$Sim(x, y) = \frac{\sum_{x,y} (U_x - \bar{U}_x)(U_y - \bar{U}_y)}{\sum_x (U_x - \bar{U}_x) \sum_y (U_y - \bar{U}_y)}$$

Der Korrelationskoeffizient nimmt bei kompletter Übereinstimmung der Bewertungen den Wert 1 an. Ansonsten ist der Korrelationswert geringen als 1. Die Korrelation ist negativ wenn die Benutzer gegensätzliche Bewertungen abgegeben haben.

Auf diese Weise kann eine Ähnlichkeit immer dann berechnet werden, wenn Benutzer mindestens ein Dokument gemeinsam bewertet haben. Jedoch sollten besser mehrere Dokumente gemeinsam bewertet worden sein, da sonst der Korrelationskoeffizient nicht aussagekräftig ist. Da dies bei der großen Anzahl der Dokumente in PIA+COMM (über 150.000) aber nur selten der Fall ist, wird eine andere Methode verwendet. In dieser Methode können auch Ähn-

lichkeiten zwischen Benutzern gefunden werden, die nur indirekt über andere Nutzer und Dokumente miteinander verbunden sind.

Dazu wird ein bipartiter Bewertungsgraph (Sedlacek 1968) aufgebaut: Benutzer und Dokumente sind Knoten, und jede Bewertung zwischen Benutzer und Dokument wird als Kante dargestellt.



**Abbildung 2:** Bipartiter Bewertungsgraph: Die vom Korrelationskoeffizienten ignorierten Kanten sind als unterbrochene Linien dargestellt.

In Beispiel (Abbildung 2) sind A, B, C und D Benutzer,  $\alpha$ ,  $\beta$ ,  $\gamma$  und  $\delta$  sind Dokumente. Linien (auch gestrichelte) repräsentieren Bewertungen. Normalerweise fließen nur die durchgängig gezeichneten Bewertungen in die Berechnung des Korrelationskoeffizienten von A und B ein. Gäbe es diese Bewertungen nicht, wäre eine Berechnung nicht möglich. Jedoch kann auch aus den gestrichelt gezeichneten Bewertungen etwas über die Ähnlichkeit zwischen A und B ausgesagt werden.

Dazu wird der Graph zwischen den zwei Benutzern als Netzwerk von elektrischen Widerständen aufgefasst, wobei die Bewertung einer Kante als der Leitwert (Kehrwert) des Widerstandes interpretiert wird. Der Gesamtleitwert wird als Indikator für die Ähnlichkeit betrachtet. Negative Bewertungen werden nicht als Widerstände mit negativen Leitwerten modelliert, sondern als ein Widerstand mit positivem Leitwert, gefolgt von einer Komponente, welche das Potenzial umkehrt.

Um den Gesamtleitwert zwischen zwei Benutzern  $a_1$  und  $a_2$  zu berechnen, wird erst ein lineares Gleichungssystem gelöst. Es hat folgende Variablen:

- Für jeden Knoten  $i$  eine Variable  $x_i$ , die das elektrische Potenzial an diesem Knoten darstellt.

- Für jede Kante (a, b) eine Variable  $y_{ab}$ , die den Strom durch diese Kante repräsentiert. Dabei ist a immer ein Benutzer und b ein Dokument.

Folgende Gleichungen werden aufgestellt:

- Für jeden Benutzer a, ausgenommen  $a_1$  und  $a_2$ , eine Gleichung  $\sum_b y_{ab} = 0$ . Dabei wird über alle zu a adjazenten Knoten b summiert.
- Für die zwei Benutzer  $a_1$  und  $a_2$  die zwei Gleichungen  $a_1=0$  und  $a_2=0$ .

$y_{ab} = r_{ab} (x_b - x_a)$ , dabei ist  $r_{ab}$  die Bewertung zwischen a und

b.

Nach Lösen des Gleichungssystems kann der Gesamtleitwert zwischen den Benutzern  $a_1$  und  $a_2$  berechnet werden mit:

$$G_{ab} = \sum_b y_{a_1,b}$$

Anders als beim Korrelationskoeffizienten ist der berechnete Wert nicht auf  $[-1, +1]$  beschränkt.

Diese Methode hat folgende Eigenschaften: Erstens ergeben parallele Kanten, wie bei der Berechnung des Korrelationskoeffizienten, eine größere Ähnlichkeit, da parallel geschaltete Widerstände einen geringeren Widerstand und somit einen größeren Leitwert ergeben.

Zweitens fließen lange Kantenverbindungen, die bei der Berechnung des Korrelationskoeffizienten ignoriert werden, in das Endergebnis ein, haben jedoch ein geringeres Gewicht als kurze Verbindungen. Das Gewicht von Verbindungen ist umgekehrt proportional zur Länge der Verbindung.

Drittens werden hintereinander liegende negative Kanten in ihrem Vorzeichen multipliziert. Somit kann z.B. ermittelt werden, dass zwei Benutzer, die dasselbe Dokument schlecht fanden, sich ähneln.

### *Dokumenten-Features*

Die im vorherigen Abschnitt beschriebene Methode setzt voraus, dass es mindestens eine indirekte Verbindung über bewertete Dokumente gibt. Besteht eine derartige Verbindung nicht, können Ähnlichkeiten auch über so genannte Features berechnet werden. Features sind Eigenschaften von Dokumenten wie z.B. enthaltene Wörter, Wortgruppen (Phrasen), aber auch Eigenschaften wie der Autor, zitierte Beiträge.

Im Folgenden wird skizziert, wie Wörter und Wortgruppen aus Dokumenten extrahiert werden, die zur Ähnlichkeitsberechnung herangezogen werden. Dazu wird für jedes Wort und jede Phrase (Wortgruppe) das TF-IDF-Gewicht (Wu and Salton 1981) berechnet.

Zur Beschreibung von Dokumenten wird für jedes Dokument die Häufigkeit ("TF": Termfrequenz) dieser Wörter im Text ermittelt. Dazu wird ein Wortschatz aus allen Worten der vorhandenen Dokumente gebildet. Ausgenommen sind Wörter, die kaum einen Informationsgehalt besitzen, wie beispielsweise "die", "der" oder "und". Worte, die in weniger als drei Dokumenten des gesamten Repositoriums vorkommen werden ebenfalls ausgeschlossen. Außerdem werden alle Wörter auf ihren Wortstamm zurückgeführt, beispielsweise "programming" und "programmed" auf den gemeinsamen Wortstamm "program", um die Sinnverwandtschaft der Terme berücksichtigen zu können. Der Umfang des so generierten Wortschatzes und damit die Dimension der Worthäufigkeitsvektoren liegt in der Größenordnung von 10.000 bis 100.000 Worten.

Worte die in sehr vielen Dokumenten vorkommen, wie beispielsweise "Einleitung", "Wissenschaft" und "Zusammenfassung" eignen sich weniger zur Unterscheidung von relevanten und irrelevanten Dokumenten als Worte, die sich nur in wenigen Dokumenten finden. Daher erhalten alle Wörter einen Gewichtungsfaktor, der umgekehrt proportional zur Anzahl der Dokumente ist, in denen das Wort auftaucht ("IDF": inverse Dokumentfrequenz)

Zur Verbesserung der Feature-Extraktion werden zusätzlich Synonyme erkannt und wie ein einziges Wort behandelt. So kann ermittelt werden, dass ein älteres Dokument über "distributed artificial intelligence" und ein neueres Dokument über „agent technology“ dasselbe Thema behandeln. Außerdem werden nicht nur enthaltene Wörter und Phrasen, sondern auch Metadaten wie z.B. die Sprache des Dokuments, das Veröffentlichungsdatum, der Autor oder die Referenzen als Features erfasst.

### **4.3 Anfrageoptimierung mit Profilen von ähnlichen Benutzern**

Nachdem im vorhergehenden Abschnitt die algorithmischen Grundlagen für die Berechnung von Ähnlichkeiten zwischen Benutzern erläutert wurde, stellen die folgenden Abschnitte die Funktionalität

vor, mit denen Communities auf der Grundlage dieser Ähnlichkeiten identifiziert und gefördert werden können.

Dieser Abschnitt stellt eine Funktionalität vor, bei der Benutzer mit ähnlichen Interessen (potenzielle Community), gegenseitig von ihren Kenntnissen profitieren können, indem sie Suchbegriffe anderer Nutzer vorgeschlagen bekommen, die bereits zu dem entsprechenden Thema eine Suche durchgeführt haben. Die Grundidee ist, dass Begriffe aus Anfragen von Benutzern mit ähnlichen Interessen gut geeignet sind, eigene Suchanfragen zu verfeinern. Es werden dazu Begriffe aus ähnlichen Anfragen ermittelt, die noch nicht Bestandteil der zu optimierenden Anfrage sind. In PIA+COMM wird dies folgendermaßen realisiert:

Zu der zu optimierenden Anfrage werden ähnliche Anfragen ermittelt. Man verwendet dafür das in Abschnitt 4.1 vorgestellte Verfahren in leicht modifizierter Form: Bewertungen und Übereinstimmungen von gemeinsamen Begriffen in Suchanfragen werden nicht für Benutzer, sondern für einzelne Anfragen betrachtet. Das heißt, bei der Berechnung werden die verschiedenen Anfragen eines Benutzers als untereinander unabhängige Entitäten betrachtet.

Zur Berechnung der Ähnlichkeiten zwischen Anfragen erstellt man den bipartiten Bewertungsgraph aus Anfragen und Suchbegriffen. Die Kanten werden durch Gewichtungen der Suchbegriffe definiert. Wie in Abschnitt 4.1 beschrieben, werden die Suchbegriffe ermittelt, die in enger Beziehung zu der optimierenden Anfrage stehen. Das heißt, es werden die Knoten im bipartiten Bewertungsgraph ermittelt, die Suchbegriffe repräsentieren und mit einem hohen äquivalenten Leitwert von der Anfrage aus zu erreichen sind.

Von den als besonders gut zur Anfrage passenden Suchbegriffen werden diejenigen ermittelt, die noch nicht zur aktuell betrachteten Anfrage gehören.

Der Benutzer kann die Vorschläge ablehnen, akzeptieren oder in modifizierter Form übernehmen. Durch das Akzeptieren von Suchbegriffen wird die "Ähnlichkeit" zwischen Benutzern erhöht, so dass beim erneuten Nutzen der Funktion die Benutzer bevorzugt ausgewählt werden, von denen ein Benutzer bereits Begriffe übernommen hat ("Verstärkung der Communitybeziehung").

Es ist geplant, dem Benutzer eine Erklärung für die generierten Vorschläge bereitzustellen. Diese ermöglicht es, Kontakt mit Personen aufzunehmen, von denen Vorschläge zur Profiloptimierung

stammen. Es wird damit die Möglichkeit geboten, eine potenzielle Community in eine explizite Community zu überführen.

## 4.4 Anzeigen ähnlicher Benutzer

Um die gegenseitige Wahrnehmung von Benutzern mit gemeinsamen Interessen (Community-Kriterium K2) zu unterstützen und die Bildung von Communities zu fördern, bietet das System PIA+COMM die Möglichkeit, nach ähnlichen Benutzern zu suchen und die Beziehungen zu Benutzern mit ähnlichen Interessen zu visualisieren. Dies geschieht durch das Benutzerradar, das seinem Namen seiner visuellen Ähnlichkeit mit Radarbildschirmen verdankt (Abbildung 3).

In der Mitte befindet sich der Benutzer, der andere Benutzer sucht. Die anderen Benutzer werden außen herum angezeigt. Die Platzierung der anderen Benutzer ergibt sich aus der berechneten Ähnlichkeit mit dem Ausgangsbenutzer und der Ähnlichkeit der Benutzer unter sich. Die Distanz der anderen Benutzer zum zentralen Benutzer ist umgekehrt proportional zur berechneten Ähnlichkeit zwischen den Benutzern. Außerdem werden andere Benutzer untereinander so positioniert, dass ähnliche Benutzer möglichst nah beieinander liegen. Zusätzlich werden Dokumente und Dokumenten-Features angezeigt, die die visualisierten Benutzer gemeinsam haben. Features und Dokumente sind mit Benutzern verknüpft, die die Dokumente oder Features (indirekt) bewertet haben.

Über das Radar können Informationen über die Dokumente und Benutzer abgerufen sowie Kontakt zu diesen hergestellt werden.

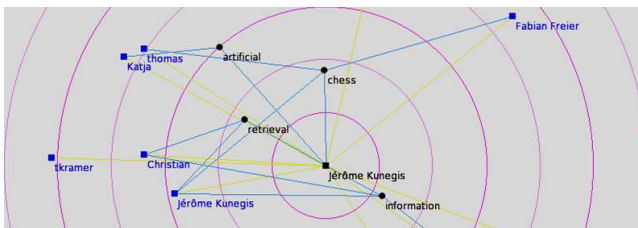


Abbildung 3: Der Benutzerradar

## 4.5 Community-basierte Dokumentenempfehlungen

In einer der nächsten Systemversionen wird es die Möglichkeit geben, virtuelle Communities zum Wissensaustausch und zur Zusammenarbeit zu gründen. Jedoch neigen virtuelle Communities dazu, schneller zu zerfallen als Gemeinschaften im realen Leben (Thiedeke 2000). Ein Grund dafür ist, dass in virtuellen Gemeinschaften (z.B. Sport, Familie, Gemeinde) größere Interaktivität notwendig ist, d.h. dass die Teilnehmer sich stärker um die Inhalte der Gemeinschaft kümmern müssen, was in realen Gemeinschaften durch lokal koprsente Gegenstände und Ereignisse entlastet wird. In unserem Community-Modell ist daher zur Bildung langfristiger Communities eine Orientierung an gemeinsamen Aufgaben oder externen Objekten notwendig (Kriterium 5).

Wir unterstützen diese Langfristigkeit, indem wir kontinuierlich aktuelle Dokumente filtern und anbieten, die für die Community als ganzes von besonderer Relevanz sind. Dadurch erzeugen wir auch für virtuelle Gemeinschaften externe Objekte, die zu neuerlichen Interaktionen führen sollen.

Im Gegensatz zum gewöhnlichen Information-Retrieval steht zu dieser Filterung jedoch keine Suchanfrage zur Verfügung. Die breitgefächerten Interessen einer Community ließen sich außerdem nur beschränkt durch eine einzelne Suchanfrage beschreiben.

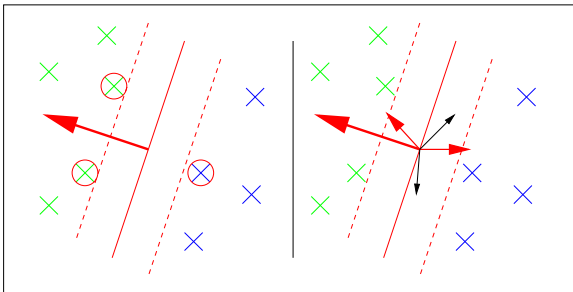
Um automatisiert für die Community relevante Dokumente zu filtern, sollen vor allem die Dokumentbewertungen der Mitglieder analysiert werden. Diese dienen als Trainingsmenge für einen maschinellen Lernalgorithmus, der die Bewertungen auf neue Dokumente verallgemeinert.

Zur Bildung eines Klassifikators, soll eine Potential-Supportvektormaschine (P-SVM) auf den bewerteten Dokumenten trainiert werden. Um die Dokumente von einem maschinellen Lernalgorithmus verarbeiten lassen zu können, werden die Dokumente als TF-IDF-Vektoren repräsentiert. Die Trainingsmenge aller bewerteten Dokumente lässt sich in einer Matrix zusammenfassen, wobei jeweils eine Spalte ein Dokument repräsentiert, während in jeder Zeile die gewichteten Worthäufigkeiten eines Wortes festgehalten werden.

Supportvektormaschinen (SVM) (Burgess 1998; Schölkopf and Smola 2002; Vapnik 1995) beruhen darauf, die Klassifikationsgrenze so zu legen, dass der Abstand (Margin) zwischen der Trennebene

und den ihr beiderseits nächstgelegenen Dokument-Vektoren (den Supportvektoren) maximiert wird, wie in Abbildung 4 veranschaulicht wird. In unserem Anwendungsfall sind die Supportvektoren die Dokumente, die am dichtesten an der Klassifikationsgrenze zwischen relevant und irrelevant liegen, also am schwierigsten einzuordnen sind. Die Trennebene wird durch ihren Normalenvektor repräsentiert, der durch eine lineare Kombination der Supportvektoren gebildet wird. Alle anderen Dokumente brauchen nicht berücksichtigt zu werden.

Um die SVM bei gekrümmten Klassifikationsgrenzen einsetzen zu können, werden die Daten durch eine geeignet Funktion nichtlinear so in höher-dimensionale Räume projiziert, dass sie sich dort linear separieren lassen. Durch den Einsatz von Kernelfunktionen kann die SVM dabei effizient realisiert werden, obwohl die Dimension der Projektion der Daten beliebig groß werden darf. Bei nicht separierbaren, verrauschten Daten, lässt sich durch eine parametrisierte Toleranz gegenüber falsch klassifizierten Trainingsdaten der Anteil an Fehlklassifikationen bei neuen Daten minimieren.



**Abbildung 4:** Schematische Darstellung der Arbeitsweisen einer SVM (links) und einer P-SVM (rechts). Die Datenpunkte zweier Klassen (veranschaulicht durch grüne und blaue Kreuze) sollen durch eine lineare Klassifikationsgrenze (durchgezogene, rote Linie) getrennt werden. Bei der SVM wird der orthogonal zur Klassifikationsgrenze stehende Gewichtsvektor (breiter, roter Pfeil) als Linearkombination der Supportvektoren (rote Kreise) so gebildet, dass der Margin (gestrichelte Linien) maximiert wird. Bei der P-SVM werden unter allen Features (kurze Pfeile), die besonders relevanten Supportfeatures (rote, kurze Pfeile) bestimmt und aus ihnen der Gewichtsvektor gebildet.

Die SVM beruht auf fundierten theoretischen Konzepten und minimiert eine obere Abschätzung des Generalisierungsfehlers. Darüber hinaus hat sie sich als state-of-the-art Algorithmus insbesondere bei hochdimensionalen und spärlich besetzten Daten etabliert (Sebastiani 2002).

Während die SVM darauf beruht, Supportvektoren, also Spalten in der Matrixdarstellung zur Klassifikation zu verwenden, kann die PSVM für die Klassifikation relevante Supportfeatures selektieren, also Zeilen in der Matrixdarstellung (siehe. Abbildung 4). In unserer Anwendung werden daher informative Terme zur Unterscheidung von relevanten und irrelevanten Dokumenten bestimmt.

Vorteile dieser Featureextraktion bestehen zum einen in einer Beschleunigung des Verfahrens, da nur die Worthäufigkeit weniger Worte berücksichtigt zu werden braucht. Zum anderen können diese extrahierten Terme auch zur automatisierten Charakterisierung von Communities dienen.

Die P-SVM [Hochreiter 2006] beruht auf ähnlichen Konzepten wie die SVM und hat sich in vielen Benchmark-Tests (z.B. mit Tipster-Datensätzen (Harman 1993)) im Vergleich mit SVMs bewährt. Ihr Konzept der Supportfeatures empfiehlt sie insbesondere bei im Vergleich zur Dimensionalität der Vektoren kleinen Trainingsdatensätzen, wie sie bei PIA+COMM auftreten. Da sie in unserer Anwendung nicht auf einer individuellen Anfrage beruht, sondern die Dokumente aller Mitglieder analysiert, ermöglicht sie einen Blick auf die Interessen der Community im Ganzen und kann so einen Beitrag zur Stärkung der Community leisten.

## **5 Schluss und Zusammenfassung**

Herkömmliche Social Softwareanwendungen unterstützen wissenschaftliche Communities bisher nur auf den unteren Community-Stufen. Das gradualisierte Community-Modell ist aber ein Schema, das auch zur der Adressierung höherer Community-Niveaus herangezogen werden kann und Hilfestellungen für die Entwicklung neuer Funktionen bietet. In der empirischen Untersuchung hat sich insbesondere gezeigt, dass man erstens zwischen Beziehungen unterscheiden muss, die auf den Austausch von Wissen abzielen und solchen, bei denen es um kollaborative Zusammenarbeit geht. Zweitens muss

man zwischen losen und dichten Beziehungen differenzieren. Für jeden Community-Typ konnte ein Set von Kriterien identifiziert werden, welches adressiert werden muss, wenn die entsprechenden Communities gefördert werden sollen. Schließlich wurde für die Kriterien eine Reihe von Unterstützungspotenzialen identifiziert.

PIA+COMM stellt sich der Herausforderung Funktionen zu entwickeln, um vielfältige wissenschaftliche Communities zu ermöglichen. Dazu haben wir einige Funktionen vorgestellt, die wir bereits implementiert haben.

Im weiteren Verlauf unseres Projekts werden wir dieses Funktionsspektrum kontinuierlich erweitern und auf seine Effektivität beim Community-Building untersuchen.

## Literatur

- Aristoteles. 1994.  
*Metaphysik.*  
Reinbek, Rowohlt.
- Böhm, A., H. Legewie, and T. Muhr. 1992.  
*Kursus Textinterpretation: Grounded Theory.*  
Berlin: TU-Berlin, Bericht aus dem IfP Atlas 92-93.
- Burges, C. J. C. 1998. "A tutorial on support vector machines for pattern recognition." *Data Mining and Knowledge Discovery* 2:1-47.
- Dourish, P., Bellotti, V.. 1992.  
"Awareness and Coordination in Shared Workspaces." Pp. 107-114 in *Proc. of CSCW '92 - Sharing Perspectives*, edited by J. Turner and R. Kraut. Toronto, Canada: ACM Press.
- Durkheim, Emile. 1977/1893.  
*Über die Teilung der sozialen Arbeit.*  
Frankfurt/M., Suhrkamp.
- Flick, U. 1999.  
*Qualitative Forschung Theorie, Methoden, Anwendung in Psychologie und Sozialwissenschaften.*  
Reinbek, Rowohlt.
- Flick, U., E.v. Kardorff, and I. Steinke. 2000. *Qualitative Forschung - Ein Handbuch.* Reinbek: Rowohlt.
- Goffman, Erving. 1975/1963.  
*Stigma. Über Techniken der Bewältigung beschädigter*

- Identität.*  
Frankfurt a.M.: Suhrkamp.
- Harman, D. K. 1993.  
*The First Text Retrieval Conference (TREC-1).*  
NIST Special Publication.
- Krappmann, Lothar. 2000/1969.  
*Soziologische Dimensionen der Identität. Strukturelle Bedingungen für die Teilnahme an Interaktionsprozessen.*  
Stuttgart: Klett-Cotta.
- Luhmann, Niklas. 1987.  
*Soziale Systeme.*  
Frankfurt a. M.: Suhrkamp.
- Mayring, P. 1983.  
*Qualitative Inhaltsanalyse. Grundlagen und Techniken*  
(2. Aufl.), Weinheim: Deutscher Studien Verlag.
- Mead, George Herbert. 1967/1934.  
*Mind, Self, and Society.*  
Chicago, London: The University of Chicago Press.
- Schäfers, Bernhard (Ed.). 1994.  
*Einführung in die Gruppensoziologie.*  
Heidelberg, Wiesbaden: Quelle & Meyer.
- Schölkopf, B., and A.J. Smola. 2002.  
*Learning with Kernels - Support Vector Machines, Regularization, Optimization, and Beyond.*  
Cambridge: MIT Press.
- Sebastiani, F. 2002.  
*Machine learning in automated categorization.*  
*ACM Computing Surveys.* 34. New York, ACM Press.
- Sedlacek, J. 1968.  
*Einführung in die Graphentheorie.*  
Leipzig: B. G. Teubner.
- Shardanand, U., and P. Maes. 1995  
*Social Information Filtering: Algorithms for Automating "Word of Mouth".*  
Proceedings of the SIGCHI conference on Human factors in computing systems:210 - 217.
- Strauss, A., and J. Corbin. 1990.  
*Basics of Qualitative Research.*  
London: Sage.
- Streeck, Wolfgang, and Philippe C. Schmitter. 1985. "Community, Market, State - and Associations? The Prospective

- Contribution of Interest Governance to Social Order." Pp. 1-29 in *Private Interest Government*, edited by Wolfgang Streeck and Philippe C. Schmitter. London: Sage.
- Tajfel, Henri. 1978. *Differentiation between Social Groups*. London: Academic Press.
- Thiedeke, Udo. 2000.  
*Virtuelle Gruppen: Begriff und Charakteristik*.  
 Pp. 23 - 74 in *Virtuelle Gruppen*,  
 Wiesbaden: Westdeutscher Verlag.
- Tönnies, Ferdinand. 1991/1887.  
*Gemeinschaft und Gesellschaft*  
 Leipzig: Fues.
- Vapnik, V. 1995.  
*The nature of statistical learning theory*.  
 New York, Springer.
- Weber, Max. 1976/1922.  
*Wirtschaft und Gesellschaft. Grundriss der verstehenden Soziologie*  
 5. Auflage. Tübingen: Mohr.
- Wellman, B. , and B. Leighton. 1979.  
*Networks, Neighborhoods, and Communities: Approaches to the Study of the Community Question*. Urban Affairs Quarterly 14:363-390.
- Wellman, Barry. 1979.  
*The Community Question: The Intimate Networks of East Yorkers*.  
 American Journal of Sociology 84:1201-1231.
- Wu, Harry, and Gerard Salton. 1981.  
*A comparison of search term weighting: term relevance vs. inverse document frequency*.  
 Proceedings of the 4th annual international ACM SIGIR conference on Information storage and retrieval: theoretical issues in information retrieval:30 - 39.