

**Quantitative Analyse der Dynamik des wissenschaftlichen, wirtschaftlichen
und gesellschaftlichen Impacts von Forschungsaktivitäten
und -netzwerken (Q-Aktiv)**

Schlussbericht

nach NABF 2017

Zuwendungsempfänger:

ZBW - Leibniz-Informationszentrum Wirtschaft - Projektleiter: Prof. Dr. Klaus Tochtermann

*Christian-Albrechts-Universität zu Kiel; Lehrstuhl für Technologiemanagement - Projektleiter:
Prof. Dr. Carsten Schultz*

Deutsche Zentralbibliothek für Medizin (ZB MED) - Projektleiter: Prof. Dr. Konrad Förstner

*Förderprogramm: Quantitative Wissenschaftsforschung, BMBF-Förderschwerpunkt
"Wissenschafts- und Hochschulforschung"*

**Quantitative Analyse der Dynamik des wissenschaftlichen, wirtschaftlichen und
gesellschaftlichen Impacts von Forschungsaktivitäten und -netzwerken (Q-Aktiv)**

Förderkennzeichen: 01PU17013A; 01PU17013B; 01PU17013C

Laufzeit: 01.07.2018 - 30.06.2021

Sachbericht zum Verwendungsnachweis

Teil I Kurzbericht

I.1. Aufgabenstellung

Ziel des Projekts Q-Aktiv war eine Verbesserung der Methoden zur KI-basierten Prognose von Dynamiken und Wechselwirkungen zwischen Forschungsfeldern. Im Verbundvorhaben wurden insbesondere Konvergenzprozesse zwischen wissenschaftlichen Feldern im Bereich der Lebenswissenschaften fokussiert. Dabei versteht man in diesem Kontext unter Konvergenz die Annäherung unterschiedlicher wissenschaftlicher und technologischer Felder. Der Hauptzweck war die Entwicklung eines Maßes, welches ermöglichen sollte, wissenschaftliche Dynamiken für beliebige Themen im Bereich der Lebenswissenschaften zu messen. Um dies zu erreichen, wurde auf verschiedene maschinelle Lernverfahren zurückgegriffen und mit diesen durch Training spezifische Modelle generiert und angewandt. Mittels dieser Modelle sollte es möglich werden, sich annähernde oder sich entfernende Wissensfelder zu detektieren. Dabei wurden auf wissenschaftlichen Thesauri (z.B. MeSH Thesaurus) zur Darstellung der zu untersuchenden thematischen Felder zurückgegriffen.

I.2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Das Projekt wurde von drei Partner-Institutionen bearbeitet, wobei jede Partner-Institution eine spezifische Aufgabe ausführte. ZB MED stellte die Daten bereit, reicherte sie an und bereitete sie auf. Der Projektpartner ZBW entwickelte den Algorithmus des maschinellen Lernens und erstellte ein Modell, um in einem Graph die Ähnlichkeiten zwischen den Konzepten (z.B. MeSH Termen) der Publikationsmetadaten im Bereich der Lebenswissenschaften zu vermessen. Die Universität Kiel führte Analysen durch und wertete die Ergebnisse im Hinblick auf Identifikation von Konvergenzprozessen aus und ordnete diese Ergebnisse in den Kontext des F&E Managements ein.

I.3. Planung und Ablauf des Vorhabens

Das Verbundvorhaben war in sechs Arbeitspakete gegliedert, in denen fünf Meilensteine vorgesehen waren. Sämtliche Arbeiten wurden von Projektmitarbeiter:innen der ZBW, CAU und ZB MED durchgeführt.

I.3.1. Inhaltliche Planung

Für die Bearbeitung von zwei Hauptfallstudien, kurz "Cholesterol" und "SARS-CoV-2", und weiteren kleineren Validierungsstudien wurde ein iteratives Verfahren und ein interdisziplinärer Ansatz gewählt. Demnach wurden die vier großen Arbeitsprozesse, die Daten-Aggregation (AP2), die Entwicklung eines Algorithmus und Durchführung des Trainings eines Modells (AP3) und die Anwendung des entwickelten Algorithmus zur Analyse, Auswertung und Validierung der

Ergebnisse (AP4) und die Evaluation des Messinstrumentes sowie Prüfung des Einsatzes des Messinstrumentes in den Methoden der Frühaufklärung (AP5) in den beiden großen Phasen des Projektes wiederholt und größtenteils gemeinsam durchgeführt. Alle Arbeitspakete wurden von allen drei Projektpartnern kollaborativ erarbeitet – mit jeweiligem Schwerpunkt bei dem Verantwortlichen für das jeweilige Arbeitspaket.

Meilensteine: Während der Projektlaufzeit wurden fünf Meilensteine planmäßig erreicht: 1) Kick-Off; 2) Auswertung Bedarfsanalyse; 3) Zwischenevaluierung des Prognose-instrumentes und der Netzwerkanalysemethoden; 4) Finalisierung des Prognoseinstrumentes; 5) Finale Evaluierung des Prognoseinstrumentes und Projektabschluss.

I.3.2. Organisation und Ablauf

Die Projektarbeit wurde erst verspätet zum Oktober 2018 aufgenommen. Die Universität Kiel als Koordinator des Verbunds übernahm die Leitung des Verbundprojektes und koordinierte die Projekttreffen. Die beteiligten Projektmitarbeiter:innen konnten wöchentliche Treffen in der Gruppe zur besseren Koordination der zu erledigenden Arbeiten durchführen. Diese Treffen fanden auch schon vor der Pandemie aufgrund der Distanz digital statt.

I.4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Zu Beginn des Forschungsvorhabens wurden KI-gestützte Verfahren zur Modellierung von Technologydynamiken kaum methodisch unterstützt (Leker & Song, 2014). Die vorherigen Arbeiten nutzten die Methoden des gemeinsamen Auftretens der Begrifflichkeiten in den Themen oder Wissensfeldern, in welchen sich Konvergenzprozesse abzeichneten (Curran und Leker, 2009). Angesichts des großen Publikationsaufkommens im Bereich der Lebenswissenschaften in den letzten Jahren war es kaum möglich, die wissenschaftlichen Dynamiken und insbesondere Konvergenzprozesse ohne Unterstützung durch die Methoden des maschinellen Lernens zu untersuchen. Auch die einzelnen Akteure in Publikations-Netzwerken und zum Beispiel ihre Initiativen, mit anderen Akteuren zu kooperieren, wurden analysiert, aber nicht im Hinblick auf Dynamiken in wissenschaftliche Prozesse betrachtet (Chai und Shih, 2016).

I.5. Zusammenarbeit mit anderen Stellen

In der Umsetzung des Vorhabens konnten wir die de.NBI Cloud, ein VM-basierte Rechenumgebung mit flexibler Skalierung, der Universität Bielefeld nutzen. Durch de.NBI wurde uns eine virtuelle Maschine in der Cloud-Infrastruktur bereitgestellt, auf der wir die sehr großen Datenmengen ablegen und auch die Berechnungen mit einer hohen Prozess-Kapazität durchführen konnten. Gerade im Hinblick auf die räumliche Distanz der Projektpartner konnten wir auf dieser Grundlage produktiv zusammenarbeiten.

Zudem konnten wir uns zum Thema unserer ersten Fallstudie “Cholesterol” produktiven Input des Lehrstuhls von Prof. Dr. Stefanie Bröring für Technologie- und Innovationsmanagement im Agribusiness der Universität Bonn einbeziehen.

**Quantitative Analyse der Dynamik des wissenschaftlichen, wirtschaftlichen
und gesellschaftlichen Impacts von Forschungsaktivitäten
und -netzwerken (Q-Aktiv)**

Schlussbericht

nach NABF 2017

Zuwendungsempfänger:

ZBW - Leibniz-Informationszentrum Wirtschaft - Projektleiter: Prof. Dr. Klaus Tochtermann

*Christian-Albrechts-Universität zu Kiel; Lehrstuhl für Technologiemanagement - Projektleiter:
Prof. Dr. Carsten Schultz*

Deutsche Zentralbibliothek für Medizin (ZB MED) - Projektleiter: Prof. Dr. Konrad Förstner

*Förderprogramm: Quantitative Wissenschaftsforschung, BMBF-Förderschwerpunkt
"Wissenschafts- und Hochschulforschung"*

**Quantitative Analyse der Dynamik des wissenschaftlichen, wirtschaftlichen und
gesellschaftlichen Impacts von Forschungsaktivitäten und -netzwerken (Q-Aktiv)**

Förderkennzeichen: 01PU17013A; 01PU17013B; 01PU17013C

Laufzeit: 01.07.2018 - 30.06.2021

Sachbericht zum Verwendungsnachweis

Teil II Eingehende Darstellung

Inhaltsverzeichnis

Sachbericht zum Verwendungsnachweis	2
Teil I Kurzbericht.....	2
Teil II Eingehende Darstellung	5
II.1. Verwendung der Zuwendung	6
II.1.1. Verwertungsstrategische Ergebnisse	6
II.1.2. Inhaltlich-technische Ergebnisse	6
Einleitung und Zielsetzung des Verbundvorhabens	6
Anforderungsanalyse und Problemstellung (alle Verbundpartner)	6
Anwendungsfälle	7
Datengrundlagen (ZB MED).....	9
Entwicklung von Deep Learning Netzwerkanalysenmethoden und deren Umsetzung in Werkzeugen (ZBW)	12
Überführung der Analysemethoden in ein Prognoseinstrument für die strategische Frühaufklärung (CAU).....	14
Evaluation und Verwertung des Prognoseinstruments (CAU)	16
II.2 Wichtigste Positionen des zahlenmäßigen Nachweises	18
II.3 Notwendigkeit und Angemessenheit der geleisteten Arbeit.....	20
II.4 Voraussichtlicher Nutzen, insbesondere der Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans	20
II.5 Während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen.....	21
II.6 Erfolgte oder geplante Veröffentlichungen der Ergebnisse.....	22
Publikationsliste der im Projekt entstandenen Arbeiten.....	23
Vorträge der im Projekt erarbeiteten Ergebnisse	23
Projektrelevante Webseiten	24
Referenzen	25

II.1. Verwendung der Zuwendung

Die Zuwendung wurde eingesetzt, um die notwendigen Forschungsarbeiten durchzuführen, technische Voraussetzungen für die Entwicklung eines Indikators der wissenschaftlichen Dynamiken zu schaffen, das Instrument für die Messung der Trends in der Forschung und Wissenschaft zu entwickeln, zu erproben und zu validieren sowie die Entwicklungsergebnisse der wissenschaftlichen Community und der Praxis zur Verfügung zu stellen.

In mehreren Projekttreffen konnten wir die gemeinsame Zusammenarbeit sehr gut koordinieren.

II.1.1. Verwertungsstrategische Ergebnisse

Die Verbundpartner haben keine verwertungsstrategischen Ergebnisse im Sinne von Produkten oder Patenten erzielt, da sie als Bildungs- und Forschungseinrichtungen die primäre Aufgabe in den Lehrveranstaltungen und freien Dissemination der Forschungsergebnisse haben. Anmeldefähige Schutzrechte sind nicht entstanden und im Projektkontext kaum relevant.

II.1.2. Inhaltlich-technische Ergebnisse

Einleitung und Zielsetzung des Verbundvorhabens

Das Ziel des Verbundvorhabens war es, die Methoden zur Prognose von Dynamiken und Wechselwirkungen zwischen Forschung, Technologieentwicklung und Innovation zu verbessern. Der Fokus des Vorhabens lag auf der Entstehung neuer wissenschaftlichen Domänen, auf der Untersuchung der Netzwerke von Forschenden, Forschungsorganisationen und relevanten Unternehmen sowie auf der Analyse der Konvergenzprozesse im Bereich der Lebenswissenschaften. Das Ergebnis des Vorhabens ist ein entwickeltes, evaluiertes und validiertes Instrument zur Messung von wissenschaftlichen Dynamiken und Trends. Das Instrument kann für die Untersuchung der wissenschaftlichen Konvergenz- und Divergenzprozesse in der Wissenschaft und Praxis eingesetzt werden.

Nachfolgend sind die Ergebnisse detaillierter aufgeführt und die Federführung der drei Vorhabenpartner nachrichtlich vermerkt.

Anforderungsanalyse und Problemstellung (alle Verbundpartner)

Durch die im AP 1 (Bedarfsbestimmung für die Analyse der Dynamik in den F&E-Aktivitäten und F&E-Netzwerken) durchgeführten Experteninterviews konnten die Fragestellungen in den ausgewählten Fallstudien abgeleitet werden. Durch die Literaturrecherche wurde der Stand der Technik ermittelt und der methodische Entwicklungsbedarf abgeleitet. Die umfangreiche Literaturrecherche hat ergeben, dass es im Rahmen der wissenschaftlichen Dynamiken die Konvergenzprozesse eine besonders große Rolle hinsichtlich der zukünftigen Forschungs- und Entwicklungstrends sowie der Entwicklungen von Innovationen spielen. Wissenschaftliche Konvergenz ist durch die gemeinsame Nutzung wissenschaftlicher Methoden, Techniken,

Algorithmen und sogar Terminologie durch die konvergierenden Gebiete gekennzeichnet. Um ein tieferes Verständnis der Forschungsmethoden und -werkzeuge des jeweils anderen Forschungsfeldes zu erlangen, beginnen Wissenschaftler unterschiedlicher Bereiche zusammenzuarbeiten. Solche Kooperationen sind meist interdisziplinärer Natur (Sick *et al.*, 2019; Curran *et al.*, 2010). Einerseits fördern solche Kooperationen den Aufbau und die Entwicklung des aufstrebenden konvergierenden Feldes. Andererseits sind Missverständnisse zwischen solchen Kooperationspartnern sehr verbreitet. Aufgrund der großen funktionalen Distanz zwischen den in zwei wissenschaftlichen Domänen eingebetteten Informationen sind die Transaktionskosten der Zusammenarbeit der beteiligten Forschungsorganisationen besonders hoch (Nemet and Johnson, 2012; Yegros-Yegros *et al.*, 2015). Trotz der Schwierigkeiten, die durch die interdisziplinäre Zusammenarbeit auftreten, verbessert und beschleunigt wissenschaftliche Konvergenz die Entwicklung der betroffenen Forschungsfelder und bereichert die jeweilige Wissensgrundlage dieser Wissenschaftsbereiche mit wertvollen Erkenntnissen (Curran and Leker, 2009).

Durch die umfassende Aufbereitung des Standes der Forschung wurde festgestellt, dass eine der Herausforderungen der wissenschaftlichen Konvergenzforschung in der Analyse einer großen Menge bibliografischer Daten liegt. Daher sind maschinelle Lerntechniken, einschließlich Text-Mining-Methoden, die sich für das Erforschen der Entwicklung von Technologiefeldern als erfolgreich erwiesen haben und meist auf der Analyse von Wortvorkommen, Zitaten, Patentklassifikationen, Zusammenfassungen, Patentansprüchen oder Volltexten von Patentdaten basieren (Niemann *et al.*, 2017; Venugopalan und Rai, 2015) erforderlich. Jeong *et al.* (2018) schlagen vor, die wissenschaftliche Konvergenz nach Themenkategorien wissenschaftlicher Veröffentlichungen zu messen. Die Untersuchung der wissenschaftlichen Konvergenz als Konzept des Verschiebens und Zusammenführens von zuvor getrennten Forschungsbereichen erfordert zunächst die Identifizierung und Unterscheidung einzelner Forschungsfelder. Eine solche Analyse erfordert den Einsatz von Clustering-Techniken wie Rolling Clustering, das häufig für das Clustering von Patenten und wissenschaftlichen Veröffentlichungen genutzt wird (Niemann *et al.*, 2017; Upham *et al.*, 2010). Einige weit verbreitete Themenmodellierungsansätze, die auf Methoden des unüberwachten Lernens basieren, weisen jedoch eine Einschränkung in Bezug auf eine zu häufige und unspezifische Terminologie auf, die verschiedene Themen repräsentieren soll und daher wenig technologischen Inhalt bietet. Darüber hinaus weist die Sprache wissenschaftlicher Veröffentlichungen keine technische Ausrichtung auf, d. h. technologiespezifische Wortkombinationen, die man in Patentdaten finden kann (Venugopalan und Rai, 2015). Aufgrund der Entwicklung der wissenschaftlichen Konvergenz über die Zeit müssen zudem die Untersuchungsmethoden diesen dynamischen Aspekt berücksichtigen. An diesem Aspekt knüpften unsere weiteren Analysen und Entwicklungen des Messinstruments zur dynamischen Untersuchung der Konvergenzprozesse an.

Anwendungsfälle

Im Rahmen des Projektes konnten wir zwei große Themenbereiche, Cholesterol und SARS-CoV-2, bearbeiten. Außerdem wurden weitere kleinere Fallstudien zur Validierung des Ansatzes durchgeführt. Der eine Themenbereich wurde anhand der Fragestellungen, die aus den

durchgeführten Interviews abgeleitet wurden, sowie im Rahmen eines Workshops mit der Universität Bonn ausgewählt werden konnte. Im Rahmen dieses Workshops sowie im weiteren Austausch mit der Professur für Technologie- und Innovationsmanagement im Agribusiness konnten das ausgewählte Anwendungsfeld zu Funktionellen Lebensmitteln (Functional Food) auf kleinere spezifische Fallstudien eingegrenzt werden. Es wurde zunächst das Forschungsfeld um die Funktionellen Lebensmittel zur Behandlung der Cholesterol-bedingten Krankheiten bearbeitet. Im zweiten Schritt wurden die entwickelten Methoden an dem Forschungsfeld getestet, das sich durch interdisziplinäre Zusammenarbeit auszeichnet und somit weitere Validierungsmöglichkeiten angeboten hat. Der weitere Grund für die Auswahl der zweiten Fallstudie war der aktuelle Anlass, die Forschungsarbeiten im Forschungsbereich des neuen Coronavirus zu erforschen. Zudem zeichnet sich auch dieses Feld durch eine hohe Interdisziplinarität aus.

Cholesterol: Das Themenfeld „Funktionelle Lebensmittel“ und insbesondere die Forschungsarbeiten zu Cholesterol eignete sich für uns als Anfangsthema besonders gut, weil sich hier ein Forschungsfeld zeigt, dessen Konvergenz durch vorherige Forschung bereits qualitativ herausgearbeitet wurde. Hierfür wurde eine Fallstudie zur historischen und aktuellen Entwicklung dieses Anwendungsbereichs Cholesterol und durch Cholesterol verursachte Krankheitsbilder durchgeführt. Diese Fallstudie ermöglichte uns die Identifizierung der pharmazeutischen und pflanzlichen Stoffe, die zur Behandlung der Cholesterol-bedingten Krankheiten eingesetzt werden bzw. zur Prophylaxe in die Lebensmittel hinzugefügt werden können. Zur Erforschung von Wissensdynamiken und Konvergenzprozessen wurden in der Fallstudie „Funktionelle Lebensmittel“ sogenannte thematische Cluster anhand des Thesaurus MeSH gebildet, welche sowohl die pharmazeutischen und pflanzlichen Stoffe als auch eine Vielzahl an weiteren Begrifflichkeiten enthielten, die mit dem Thema Cholesterol im Zusammenhang standen. Mit unseren Methoden und mittels des MeSH-Thesaurus galt es daher das mit Cholesterol verbundene Publikations-Aufkommen zwar unvoreingenommen zu modellieren. Wir konnten Konvergenzprozesse aus der durchgeführten Fallstudie zur historischen und aktuellen Entwicklung dieses Anwendungsbereichs Cholesterol und der bisherigen Literaturergebnisse erwarten, die sich dann auch tatsächlich in den Daten in einer sich annähernden Kosinus-Distanz einzelner MeSH-Terme zeigten.

SARS-CoV-2: Mit der Fallstudie zu SARS-CoV-2 wurde auf ein aktuelles und gesellschaftlich besonders relevantes Thema im Verbundprojekt reagiert. Unter dem Eindruck der sich vollziehenden Pandemie, entschieden wir uns in der Gruppe für eine Anwendung unseres Modells auf ein unbekanntes Thema, dessen Entwicklung im Sinne einer Konvergenz oder Divergenz von Komponenten nicht erschlossen ist. Für die Bearbeitung beschäftigten wir uns mit wissenschaftlichen Artikeln sowie Preprints des Jahres 2020 (01.01.-31.12.) und untersuchten die interdisziplinären Anteile dabei.

Im Laufe des Projektes wurde das entwickelte Instrument auch auf andere Anwendungsfälle im Bereich der Lebenswissenschaften übertragen. So wurde das entwickelte Distanz-Maß in der Fallstudie „Interdisziplinarität im Bereich der biomedizinischen Algenforschung“ zur Messung der inhaltlichen Distanz zwischen den Schlüsselwörtern einer Publikation im Bereich Algenforschung

genutzt. Eine weitere Fallstudie fokussierte die Kooperationen zwischen Unternehmen der pharmazeutischen Branche und Universitäten vor dem Hintergrund der Notwendigkeit der Etablierung wissenschaftlicher Netzwerke zum Vorantreiben der Innovationen und wissenschaftlichen Dynamiken.

Datengrundlagen (ZB MED)

Für die zwei Hauptfälle und weitere kleinere Fallstudien wurden entsprechende Datensätze erarbeitet.

Nachdem mit dem Themenkomplex "Cholesterol" ein erster Untersuchungsgegenstand identifiziert wurde, stellte ZB MED den Datensatz für die Analysen zusammen. Dabei konnte auf die intern betriebene Datenbank ZB MED Knowledge Environment zurückgegriffen und die Metadaten zu wissenschaftlichen Artikeln sowie deren Annotierungen mit MeSH-Termen abgerufen werden. Auf Grundlage von Identifikatoren wurde außerdem eine Deduplizierung der Titel durchgeführt.

Im Zuge der Arbeit am ersten Fallbeispiel stellten wir das große Potenzial für die Analyse durch die Anreicherung der Metadaten durch autorenzentrierte Informationen fest. In den Diskussionen trat das Feld Wissenschaft als "soziales Feld" hervor, in dem einzelne Individuen Entscheidungen über Fragestellungen oder wissenschaftliche Kooperationen fällen, die dann auch zu Konvergenzen oder Divergenzen von Themen führen können. Die ausschließliche Fokussierung auf die Ko-Autorenschaft erschien aufgrund von fehlenden Informationen über die Autorenhintergründe unvollständig. Der soziale Aspekt der Autorenschaft über Kooperationen und thematische Entscheidungen werden durch die persönlichen Werdegänge von Forschenden beeinflusst. Es wurde daher die Hypothese aufgestellt, dass ein Netzwerkgraph über Themen und Autorenschaft durch weitere Informationen zu den Akteuren qualitativ gewinnen könnte.

Um die Hypothese zu prüfen, wurden weitere Informationen zu den Autor:innen der modellierten Artikel aus Wikidata, eine semantische Fakten-Datenbank der Wikimedia-Community, und der GND (Gemeinsame Normdatei) der deutschen Nationalbibliothek abgerufen und den Metadaten-Korpus hinzugefügt.

Mit der Entwicklung einer eigenen Python-Library ("Take-it-Personally" (TIP-library)) zur Anreicherung der Daten um akteurzentrierte Informationen hat ZB MED auch die Nachnutzung durch andere gesichert (siehe Abb. 1). Bis hiermit jedoch quantitativ zufriedenstellende Ergebnisse geliefert werden können, muss die Datengrundlage in Wikidata verdichtet werden. Allerdings waren für zeitgenössische Artikel bisher zu wenige Informationen über Autor:innen in Wikidata enthalten.

Die Anreicherung von Wikidata selbst durch weitere Informationen stellte sich damit als eine wichtige Voraussetzung für die flächendeckende Analyse von Autor:innen-Daten heraus. Mit dem OrcBot wurde ein Programm entwickelt (Abb. 2), das eben diese Lücke schließen sollte. OrcBot trägt automatisiert Informationen der Datenbank ORCID auf Basis des Publikationen-Identifiers DOI in Wikidata ein und verknüpft vorliegende Items von Autor:innen mit Artikel-Items. ORCID ist eine Datenbank, in die Wissenschaftler:innen selbst Angaben zu ihrer akademischen Biographie

und zu ihren Publikationen machen. Die häufig problembehaftete Disambiguierung von Autor:innen fällt damit aus, da die Autor:innen selbst Titel als die ihren kenntlich machen. Auf dieser Grundlage können Informationen aus ORCID zu den Items in Wikidata eingetragen werden, wenn diese mit ORCID-ID oder mit einer DOI eindeutig identifiziert sind.

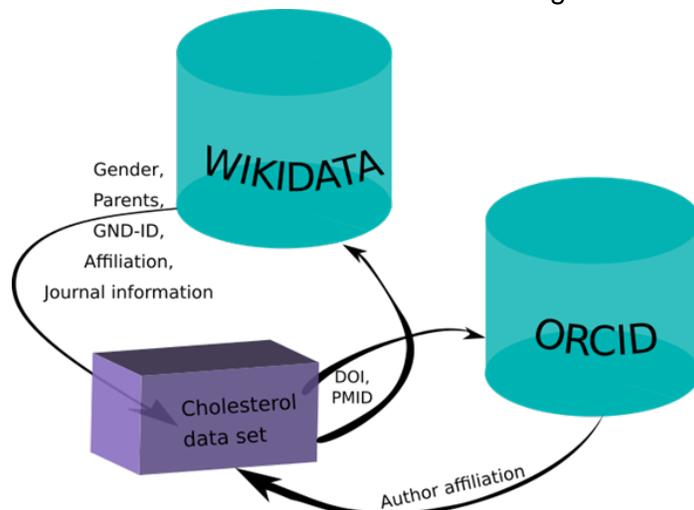


Abbildung 1: Anreicherung von Metadaten durch die Integration von Autorenzugehörigkeiten

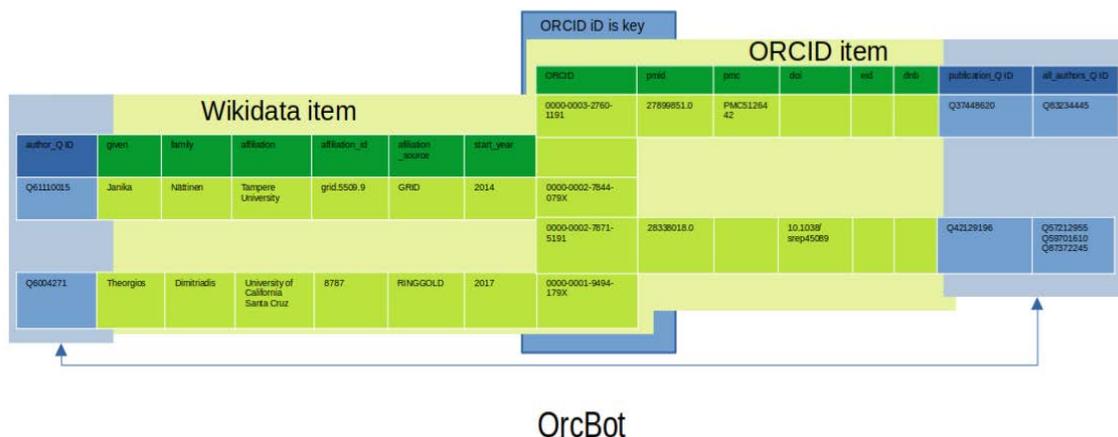


Abbildung 2: Schematische Darstellung von OrcBot

Auch für den zweiten Themenkomplex wurden von ZB MED die benötigten Daten zusammengestellt und aufbereitet. Auch hier waren der Ausgangspunkt Daten von ZB MED. Dazu wurden die für die Fallstudie SARS-CoV-2 benötigten Daten aus dem ZB MED Knowledge Environment abgerufen und mit Informationen aus den Datenbanken ORCID, Wikidata und Crossref (Referenzierte Literatur und ISSN) angereichert und aufbereitet. Besonders profitieren konnte das Verbundprojekt hierbei von der ZB MED Projektdatenbank COVID-19-Hub, die neben den Publikationen, die durch ein Peer-Review Verfahren geprüft sind, auch Preprints zum Thema enthält.

Den Preprints kommt in einer sich schnell entwickelnden Situation eines medizinischen Notfalls wie der COVID-19 Pandemien eine besondere Bedeutung zu, da ein regulärer Publikationsprozess inklusive wissenschaftlichem Review in der Regel mehrere Monate bis hin

zu einem oder zwei Jahren dauern und so auf neue Entwicklungen nicht reagieren kann. Gerade diese Eigenschaft der schnellen Aufnahme von veränderten Lagen und der Reaktion darauf in der Forschung macht die Preprints für die im Projekt bearbeitete Fragestellung einer Konvergenz von Forschungsfeldern interessant.

Dabei stellte sich die Einbeziehung der Preprints vor eine besondere Herausforderung dar, da das in der ersten Fallstudie „Forschung an der cholesterolsenkenden pharmazeutischen Präparaten und Naturstoffen“ entwickelte Ähnlichkeitsmaß auf einem Knowledge Graphen aufbaut, in dem die Publikationen in Relation zu annotierten Keywords (MeSH-Terme) modelliert werden. Weder war das entsprechende Schlagwort (COVID-19) in den MeSH-Thesaurus etabliert, weswegen zu Beginn der Pandemie die Publikationen allgemein unter „Beta-Coronavirus“ oder „Respiratory Syndrom“ kategorisiert wurden, noch sind Preprints in der Regel überhaupt mit MeSH-Termen annotiert. Als Abhilfe konnte auf das automatische Annotationsverfahren der ZB MED Kollegin Lisa Langnickel zurückgegriffen werden, die daher auch auf der entsprechenden Publikation als Ko-Autorin erscheinen wird.

Für die Analyse wurde zunächst ein Analysezeitraum vom 01.01.2020 bis 31.08.2020 festgelegt. Zum Jahresende wurde jedoch entschieden, die Datengrundlage auf das gesamte Jahr 2020 auszudehnen, um auch die langfristigen Momente der Entwicklung methodisch einzuholen. Die Ergebnisse für die erste Phase ließen darauf schließen, dass die betreffenden Forschungsgruppen generell die Untersuchungen im eigenen Feld durchgeführt haben und wenige Übernahmen von Forschungsergebnissen aus anderen Forschungsbereichen stattgefunden haben oder interdisziplinäre Forschungsansätze verfolgt wurden. Dies zeigte sich etwa daran, dass selten MeSH-Terme verwendet wurden, die aus gänzlich anderen MeSH-Hierarchie-Bereichen stammen.

Dennoch deutet sich für die spätere Entwicklung eine Trendumkehr an, die die Ausweitung des Untersuchungszeitraumes nahelegt. Eine Publikation, die auch die späteren Entwicklungen miteinbezieht, ist in Vorbereitung (siehe Abschnitt II.6).

Für die Fallstudie „Interdisziplinarität im Bereich der biomedizinischen Algenforschung“ wurde ein Datensatz erstellt, der auf Metadaten von etwa zwanzigtausend Forschungspublikationen basiert, die aus der Web of Science-Datenbank abgerufen wurden, und bibliografischen Metadaten der beteiligten Forschungseinrichtungen. Die Daten wurden direkt vom Konsortium Kompetenzzentrum Bibliometrie in FIZ Karlsruhe erhoben, dessen bibliometrische Dienstleistungen im Rahmen der BMBF-Förderlinie dem Verbundvorhaben zur Verfügung gestellt wurden. Die Studie „Interdisziplinarität im Bereich der biomedizinischen Algenforschung“ konzentrierte sich auf Algen und algenbezogene Technologien und Anwendungen zwischen 1980 und 2017. Daher wurde der Datensatz so ausgewählt, dass alle Publikationsmetadaten mit dem Wort „Algen“ im Titel oder Abstract, seinen Synonymen und angrenzenden Begriffen innerhalb dieser Zeitspanne extrahiert wurden.

Für die Fallstudie Kooperationen zwischen Universitäten und Pharmaunternehmen, die sich mit der Rolle deren Zusammenarbeit für den Erfolg klinischer Studien befasste, wurden ebenfalls Metadaten der gemeinsamen Publikationen zwischen Universitäten und der 56 größten pharmazeutischen Unternehmen erhoben. Darüber hinaus wurden die Patentanmeldung-Daten aus der Patentdatenbank PATSTAT extrahiert. Eine weitere Quelle der zu analysierenden Daten war die Datenbank klinischer Studien ClinicalTrials.gov.

Entwicklung von Deep Learning Netzwerkanalysenmethoden und deren Umsetzung in Werkzeugen (ZBW)

Die Untersuchung von bibliographischen Daten gewährt tiefe Einblicke in Wissenschaftsdynamiken. Dazu gehören die Entstehung von wissenschaftlichen Erkenntnissen, ihre spätere Diffusion und Konvergenz von distinkten wissenschaftlichen Feldern, sowie die Substitution von wissenschaftlichen Feldern mit konvergierten Bereichen.

Neues Wissen entsteht innerhalb von wissenschaftlichen Gemeinschaften – oft durch die Kollaboration von unterschiedlichen Akteuren (Phelps 2012; Wuchty 2007). Die Zusammenarbeit von unterschiedlichen Wissenschaftler:innen fördert die Diffusion von Wissen aus einem Wissenschaftsfeld in ein anderes. Eine intensive interdisziplinäre Zusammenarbeit führt daher dazu, die Grenzen der distinkten wissenschaftlichen Bereiche aufzuweichen und zum Teil sogar zum Entstehen neuer Wissenschaftsfelder (Curran *et al.*, 2010). In der Anforderungsanalyse (AP1) haben wir die folgenden Aufgaben für Netzwerkanalyse mit Deep Learning Methoden identifiziert: 1. Das unüberwachte Lernen von Konzeptrepräsentationen und 2. Die Vorhersage von zukünftigen Verbindungen in einem Netzwerk. Im Folgenden sind die Ergebnisse dieser beiden Aufgaben beschrieben, die im Kontext von *AP 3: Entwicklung von Deep Learning Netzwerkanalysemethoden* erzielt worden sind.

Unüberwachtes Lernen von Konzeptrepräsentationen in dynamischen Netzwerken

Große Korpora wissenschaftlicher Publikationen bergen ein großes Potential für automatisierte Analyse wissenschaftlicher Dynamiken. Techniken des maschinellen Lernens profitieren von großen Datensätzen, die damit auch für die Analyse von wissenschaftlichen Dynamiken essentiell sind. Eine Herausforderung in der Analyse war ein aussagekräftiges Ähnlichkeitsmaß zu finden. Die meisten vorliegenden Studien nutzen dazu eine textbasierte Ähnlichkeit auf Basis der Co-Zitationen (Niemann 2017; Upham 2010) oder anderer szientometrischer Methoden (Jeong 2016, Jeong 2018). Im Gegensatz dazu nutzen wir die Konzepte, mit denen die einzelnen Titel der digitalen Literatur-Korpora annotiert sind, um darauf aufbauend Ähnlichkeiten zwischen den Konzepten zu errechnen. Mit Techniken des maschinellen Lernens lassen sich niedrigdimensionierte kontinuierliche Vektoren erstellen, die die einzelnen Konzepte repräsentieren, die einen Artikel beschreiben. Wir nennen diese Vektoren "Konzeptrepräsentationen". Zwischen diesen Konzeptrepräsentationen kann schließlich ein Wert für das Ähnlichkeits-Maß berechnet werden.

Für das unüberwachte Lernen von Konzeptrepräsentationen haben wir einen neuen Ansatz auf Basis von Graph Convolution entwickelt, der als Konferenzbeitrag zur Jahrestagung der Gesellschaft für Informatik (INFORMATIK 2019) veröffentlicht wurde. Gegenüber vorherigen Ansätzen bietet dieser den konzeptuellen Vorteil, dass die Titel der Publikationen in der Konzeptrepräsentation abgebildet werden können. Hierzu haben wir eine maßgeschneiderte Kostenfunktion entwickelt, die für Gruppen von durch Ko-Autorenschaft verknüpften

Publikationen die passenden Konzepte rekonstruiert werden müssen. Die Repräsentation dieser Gruppe von Publikationen bildet dann die Repräsentation des zugehörigen Konzeptes. Die Ähnlichkeit zwischen zwei Konzepten wird dann anhand der Kosinusähnlichkeit der Repräsentationen (Vektoren von reellen Zahlen) gemessen.

In Experimenten auf dem EconBiz Datensatz (2.1 Millionen Titel aus den Wirtschaftswissenschaften) haben wir diesen neuen Ansatz, der sowohl Text als auch Netzwerkstruktur berücksichtigt mit etablierten Ansätzen wie DeepWalk/node2vec (Perozzi *et al.* 2013, Grover & Leskovec 2016) als reine netzwerkbasierter Methode sowie Singulärwertzerlegung als rein textbasierte Methode verglichen. Die Ergebnisse zeigen, dass der neue Ansatz bei Konzeptähnlichkeit eine vergleichbare Qualität erreicht wie der DeepWalk Algorithmus, und diesen sogar übertrifft, wenn es um Klassifikation anhand der gelernten Repräsentationen geht.

Im Projektverlauf haben wir diesen Ansatz für den wirtschaftswissenschaftlichen Datensatz auch für die Fallstudie "Interdisziplinarität im Bereich der biomedizinischen Algenforschung" nachgenutzt. Wegen der besseren Skalierbarkeit, haben wir für Analysen auf größeren Datensätzen (Cholesteroll, COVID-19) auf den DeepWalk Algorithmus zurückgegriffen.

Um eine Analyse von dynamischen Netzwerken zu ermöglichen, war es für beide Ansätze notwendig, einen inkrementellen Lernmechanismus zu entwickeln. Mit diesem inkrementellen Lernmechanismus wird die zeitliche Komponente abgebildet und die Daten werden nach und nach (auf Basis des Publikationsdatums) zum Training (bei DeepWalk) und zur Auswertung (bei unserem Graph Convolution Ansatz) des Modells herangezogen.

Mithilfe dieser entwickelten Methoden konnten wir die Forschungsdynamiken in drei Fallbeispielen untersuchen: Cholesteroll, Algenforschung, COVID-19. Jedes dieser Fallbeispiele erforderte eine Zusammenstellung geeigneter Datensätze, welche im Abschnitt *AP2: Akquise und Aufbereitung der Datensätze für die Analyse* im Detail beschrieben sind. Die Ergebnisse der Fallstudien sind im folgenden Abschnitt *AP4: Überführung der Analysemethoden in ein Prognoseinstrument für die strategische Frühaufklärung* beschrieben.

Im Allgemeinen können die entwickelten Methoden auf beliebige weitere Fallbeispiele angewandt werden. Mit einem einzelnen Trainingsdurchgang auf einem Datensatz mit großer Abdeckung können auch mehrere Fallbeispiele gleichzeitig untersucht werden.

Vorhersage und Empfehlung von zukünftigen Themenkombinationen

Neben dem Lernen von Konzeptrepräsentationen ist auch die Vorhersage von neuen Verknüpfungen zwischen Konzepten (Themenkombinationen) von Interesse. Dies wird im Allgemeinen als "Link Prediction" bezeichnet. Dazu haben wir die heterogenen Informationsnetzwerke so vorverarbeitet, dass ein Netzwerk aus Konzepten entsteht, in dem zwei

Konzepte genau dann verbunden sind, wenn mindestens eine Veröffentlichung existiert, die in der Datengrundlage mit beiden Konzepten annotiert ist.

Die Aufgabe ist es, anhand dieser Daten, vorherzusagen, ob in der Zukunft eine neue, vorher nicht vorhandene, Verknüpfung zwischen zwei Konzepten entsteht. Auch hier haben wir die aktuell vorherrschenden Methoden DeepWalk (Perozzi *et al.* 2013) und Graph Convolution (Kipf und Welling 2016) für die Link Prediction angepasst und angewendet. Zudem haben wir die klassischen Netzwerkanalysemethoden Common Neighbors (Newman 2001), Preferential Attachment (Barabasi 2002), und Adamic/Adar (2003) zum Vergleich herangezogen.

In unseren Experimenten zu Link Prediction in Konzeptnetzwerken haben wir festgestellt, dass die Vorhersage von zukünftigen Verknüpfungen zwischen Konzepten sehr schwierig ist. Ansätze basierend auf DeepWalk/node2vec waren nicht zielführend. Graph Convolution-basierte Ansätze wie Graph Autoencoder (Kipf und Welling 2016) und SEAL boten eine höhere Vorhersagekraft. Die höchste Vorhersagekraft wurde jedoch, überraschenderweise, von den klassischen Methoden erzielt.

Unter rigoroser Evaluierung zeigte sich jedoch, dass auch die besten Ansätze, absolut betrachtet, nur eine niedrige Vorhersagegüte lieferten. Unsere Ergebnisse eignen sich jedoch als Grundlage für weiterführende Forschung, oder als Kernbaustein eines Empfehlensystems, bei dem die Ausgabe als Empfehlung von Forscher:innen herangezogen und geprüft werden kann.

Werkzeugkasten zur Netzwerkanalyse

Abschließend wurden die entwickelten Deep Learning Methoden für die dynamische Netzwerkanalyse in einem nachnutzbaren Softwarepaket: "QGraph" zusammengefasst, auffindbar unter <https://gitlab.com/Q-Aktiv/qgraph>. Dieser Werkzeugkasten ermöglicht die Anwendung der Methoden auf weitere Datensätze um die Dynamiken in weiteren Wissensgebieten zu analysieren.

Überführung der Analysemethoden in ein Prognoseinstrument für die strategische Frühaufklärung (CAU)

Die im AP 3 (Entwicklung von Deep Learning Netzwerkanalysemethoden und deren Umsetzung in Werkzeugen) entwickelten Methoden wurden durch alle drei Verbundpartner in einen spezifischen Indikator zur Messung der inhaltlichen Distanz zwischen verschiedenen MeSH-Konzepten als thematischen Clustern (siehe Abbildung 3) über die Zeit überführt. In AP 4 (Überführung der Analysemethoden in ein Prognoseinstrument für die strategische Frühaufklärung) erfolgte die Erprobung des Instruments bei der Bearbeitung der Fallstudie „Forschung an der cholesterolsenkenden pharmazeutischen Präparaten und Naturstoffen“. Unsere Analysen mit dem entwickelten Indikator, die durchgeführten qualitativen Validitätstests sowie quantitativen Tests ergeben, dass der Indikator für die Messung von wissenschaftlichen Dynamiken, inkl. Konvergenz- und Divergenzprozessen gut geeignet ist. Wir haben die

Kosinusdistanz als Indikator für die Ähnlichkeit zweier Konzepte verwendet, die wissenschaftliche Felder, Gebiete oder Themen sein können.

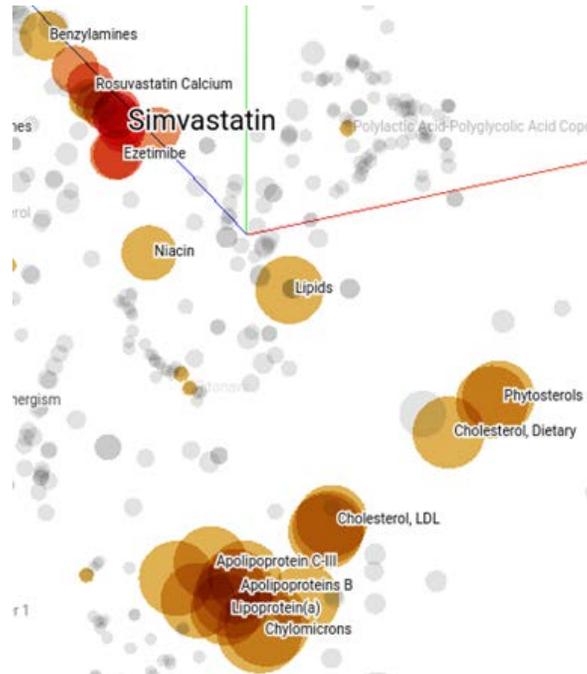


Abbildung 3: t-SNE-Visualisierung der Distanzen zwischen den Themen im mehrdimensionalen Raum am Beispiel des Cholesterols-Themas

Dieses Maß liefert kontextuell wertvolle Einblicke in die wissenschaftliche Dynamik in intra- und interdisziplinären Forschungsbereichen und kann für die Identifizierung der wissenschaftlichen Konvergenz in relevanten Forschungsfeldern eingesetzt werden. In unserer Fallstudie „Forschung an der cholesterolsenkenden pharmazeutischen Präparaten und Naturstoffen“ haben wir festgestellt, dass im Bereich einzelner Wissenschaftsbereiche (Pharma oder Lebensmittel- und Ernährungsbereiche) derzeit ein Konvergenzmuster der Wissensdynamik vorherrscht. Allerdings resultiert ein hoher Spezialisierungsgrad innerhalb eines Fachgebiets in einer schnellen Wissensverteilung innerhalb des Fachgebiets, da Forschende desselben Fachgebiets geringere Transaktionskosten haben, die in der interdisziplinären Forschung durch fehlende gemeinsame Sprache, gemeinsame Normen und Forschungstechniken, sowie durch das Fehlen von Hintergrundwissen aus anderen Domänen verursacht werden können (Yegros-Yegros *et al.*, 2015). Daher deuten unsere Ergebnisse darauf hin, dass in einigen Feldern zwischen zwei verschiedenen und weit entfernten wissenschaftlichen Domänen divergierende Prozesse überwiegen. Divergenz ist charakteristisch für die Evolution der Wissensgenerierung zwischen funktional entfernten und sich spezifizierenden wissenschaftlichen Domänen (Curran und Leker, 2009).

Nach der durchgeführten Evaluation in der Fallstudie „Forschung an der cholesterolsenkenden pharmazeutischen Präparaten und Naturstoffen“ kann das Instrument unter anderem für die strategische Frühaufklärung verwendet werden. Das Instrument ermöglicht eine Verfolgung wissenschaftlicher Trends und insbesondere von Konvergenz- und Divergenzprozessen. Durch

die Anwendung des entwickelten Instrumentes auf wissenschaftliche Publikationsdaten lassen sich zukunftsrelevante Themen der Wissensbereiche frühzeitig erkennen.

Außerdem konnte die Universität Kiel aus dem entwickelten Ähnlichkeitsmaß andere Indikatoren ableiten, wie z.B. ein Distanzmaß, welches den inhaltlichen und disziplinären Unterschied zwischen zwei oder mehreren Themen (Konzepten) innerhalb eines Forschungsprojektes messen kann. Dieses Maß hat die Universität Kiel durch die Panel- und Längsschnittanalysen in der Fallstudie „Interdisziplinarität im Bereich der biomedizinischen Algenforschung“ evaluiert. Die Panelanalysen wurden am Beispiel der Notwendigkeit der Interdisziplinarität für die Forschung und Entwicklung der Erfindungen und Innovationen im Bereich der Algenforschung (als ein Wissensfeld im Bereich der Lebenswissenschaften) durchgeführt. Die Ergebnisse der Panelanalysen zeigten, dass das im Projektvorhaben entwickelte Instrument zur inhaltlichen Distanz/Ähnlichkeitsmessung auch für die komplexen Fragestellungen geeignet ist, die kausalen Zusammenhänge zwischen dem Grad der Interdisziplinarität, welche die Vorstufe der Konvergenzprozesse widerspiegelt, und dem wissenschaftlichen Impact der interdisziplinären Forschungsarbeiten, zu erforschen.

Unsere Ergebnisse zeigen, dass sich verschiedene Komponenten von Wissen aus unterschiedlichen Domänen ergänzen können und insgesamt eine integrierte Sicht auf ein komplexes Problem ermöglichen, das nicht von einzelnen Disziplinen getrennt gelöst werden kann. Auf der anderen Seite hindert der zu hohe Grad der Interdisziplinarität den wissenschaftlichen Forschungs- und Entwicklungoutput. Unsere Ergebnisse deuten darauf hin, dass ein umgekehrter U-förmiger Zusammenhang zwischen der Diversität von Forschungsergebnissen und der F&E-Leistung besteht. Eine hohe Diversität der Forschungsergebnisse ist vorteilhaft, entferntes Wissen aus verschiedenen Wissenschaftsbereichen zu kombinieren. Eine Integration von extrem weit entfernten Wissen erhöht jedoch die Komplexität und Unsicherheit, was sich negativ auf die F&E-Leistung auswirkt. Der negative Effekt der Diversität der Wissensbasis knüpft an die bisherigen Forschungsarbeiten in diesem Bereich an. Unsere Befunde deuten darauf hin, dass Forscher:innen, die sich auf wenige wissenschaftliche Gebiete konzentrieren (geringe Diversität der Wissensbasis), fokussiertere Forschungsergebnisse erzielen können, die von der wissenschaftlichen Gemeinschaft mehr geschätzt werden, was zu einer höheren F&E-Leistung führt. In Übereinstimmung mit der bisherigen Literatur implizieren unsere Ergebnisse, dass durch zu hohe Diversität der Wissensbasis das Ergebnis von integriertem Fernwissen in einem Forschungsprojekt sehr riskant und ungewiss sein kann. Ein solches Ergebnis kann als äußerst unkonventionell bewertet werden und Wissenschaftler mit einem spezialisierten Studienfach kämpfen mit der Aufnahme von kognitiv herausfordernden neuen Wissens-elementen. Da die Beschränkung spezialisierter Wissenschaftler auf ihr eigenes Forschungsparadigma für ihren wissenschaftlichen Ruf und ihre Finanzierungszwecke von Vorteil sein kann, bestehen wenig Anreize, die Grenzen ihres Studienfachs zu überschreiten, was sich nachteilig auf ihre Forschungsergebnisse auswirken und zu fehlenden bahnbrechenden Ergebnissen führen kann.

Evaluation und Verwertung des Prognoseinstrumentes (CAU)

Mittels des im Verbundvorhaben in AP 3 (Entwicklung von Deep Learning Netzwerkanalysemethoden und deren Umsetzung in Werkzeugen) entwickelten Instrumentes

können Distanzen zwischen den Inhalten wissenschaftlicher Themen und deren Veränderung über die Zeit gemessen werden. Das entwickelte Instrument zur Messung der wissenschaftlichen Dynamiken kann für die Identifizierung von Themen (Felder, Domänen) eingesetzt werden, die inhaltlich Überschneidungen und Verknüpfungen haben oder entwickeln werden. Außerdem können beim Einsatz des Instruments über einen zeitlichen Rahmen die Themen identifiziert werden, welche sich über die Zeit annähern und deren Verbindung sich intensiviert, insbesondere wenn diese Themen zu unterschiedlichen Fachbereichen gehören. Abbildung 4 illustriert schematisch die Nutzung des Instruments im Rahmen des Gesamtnutzungskonzepts.

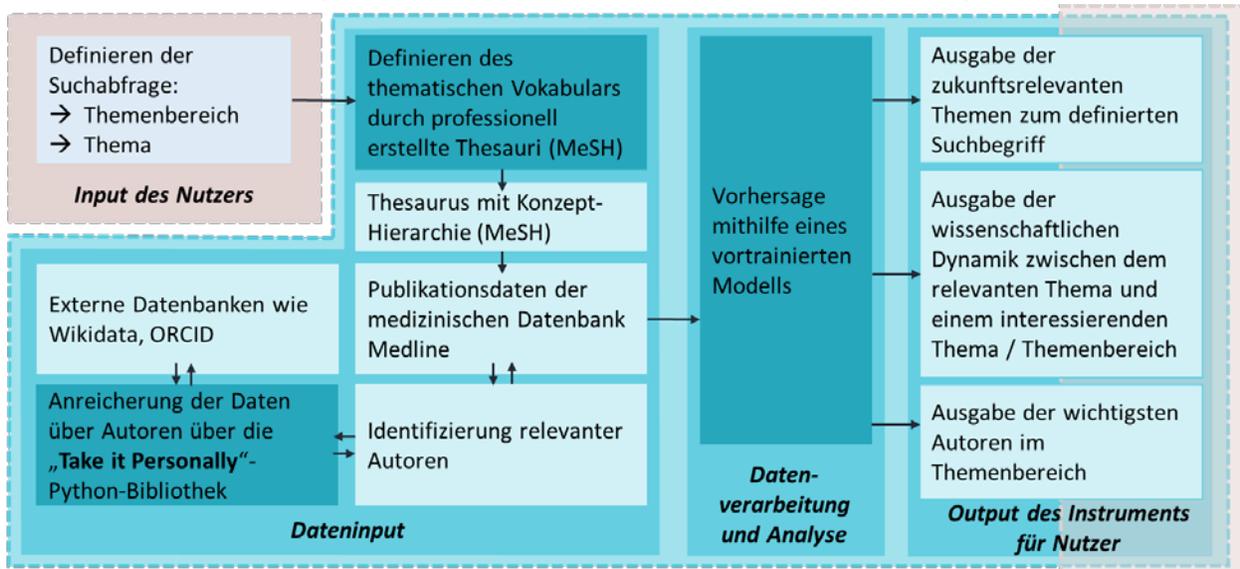


Abbildung 4: Schematische Darstellung des Instrumentes

Die inhaltliche Distanz zwischen den relevanten Themen und die abgeleiteten Trends stellen eine quantitative Grundlage zum Treffen qualitativ besserer Entscheidungen dar. Im Bereich der Wissenschaft und Forschung können beispielsweise Entscheidungen über die Organisation der Wissenschaftseinrichtungen getroffen werden, wenn es darum geht, eine Homogenität einzelner Abteilungen zu erzielen oder die interdisziplinäre Zusammenarbeit zwischen den fachfremden Fachbereichen zu unterstützen. Mittels des Instrumentes können zukunftsrelevante Themen der Wissensbereiche frühzeitig erkannt werden. Somit können interessierte Akteure wie Forschungseinrichtungen, aber auch die Politik und die Wirtschaft, die Fördermittel in die Bereiche investieren, bei welchen sich der Konvergenzprozess mit anderen Fachbereichen abzeichnet und welche von der fortschreitenden Konvergenz besonders profitieren können. Das Instrument kann die Fokussierung auf einen zukunftssträchtigen Themenbereich ermöglichen, in welchem organisationsinterne und organisationsübergreifende Cluster und Projekte initiiert werden können. Die Forschungseinrichtungen, aber auch Unternehmen, bei welchen das Instrument zum Einsatz kommen könnte, könnten sich auf die Wissensgebiete profilieren, bei welchen Konvergenzen zu erwarten sind, und starke Signale der steigenden Kompetenzen an die interessierten wissenschaftlichen und praxisnahen Netzwerke aussenden. Mittels des entwickelten Instrumentes könnten außerdem die Forschenden mit höchster Kompetenz im angehenden Konvergenzfeld identifiziert werden. Durch die Identifizierung der wichtigsten

Forschenden können Forschungseinrichtungen und Unternehmen potentielle Kooperationspartner identifizieren.

Des Weiteren hat die Universität Kiel das entwickelte Distanzmaß in der Fallstudie "Auswirkungen des Grades der Interdisziplinarität auf die technologische Neuheit von Medizintechnikunternehmen" für den Einsatz in der Managementforschung an Unternehmen auf Realwelt-Daten evaluiert und validiert. Diese Fallstudie eignete sich hervorragend zur Validierung der Prognosegüte von zukünftigen Veränderungen der Relevanz der Themen, welche die 70 weltweit größten Medizintechnikunternehmen in ihren Forschungs- und Entwicklungsaktivitäten bearbeitet haben. Das entwickelte Konstrukt des Grades der Interdisziplinarität wird in Zusammenhang mit der Unternehmensleistung gebracht, was eine Grundlage für die quantitative Analyse des Technologiemanagements bietet. Somit erfolgte die mehrdimensionale Bewertung des neu entwickelten Ähnlichkeits-/Distanzmaßes. Darüber hinaus wurde ein Vorgehensmodell als Studiendesign für die Prognoseinstrumente (Indikatoren für thematische Ähnlichkeiten oder Distanzen) ausgearbeitet und erfolgreich erprobt.

Das Hauptergebnis der Fallstudie zeigt einen positiven Zusammenhang zwischen den Projekten des Unternehmens mit einem hohen Interdisziplinaritätsgrad und der technologischen Neuheit des Unternehmens. Unser Ergebnis steht im Einklang mit der bisherigen Forschung, die darauf hindeutet, dass interdisziplinäre Forschung Wissenschaftlern und Forschungsorganisationen mehrere Vorteile bietet. Der Befund unserer Studie erweitert auch die bisherige Literatur, indem es Belege dafür liefert, dass auch Unternehmen in der Lage sind, von interdisziplinären Forschungsprojekten zu profitieren. Darüber hinaus haben wir die Bedingungen analysiert, unter denen der Grad der Interdisziplinarität die technologische Neuheit von Medizintechnikunternehmen verbessern kann. Unsere Ergebnisse zeigen, dass eine hohe wissenschaftliche Aufnahmefähigkeit in interdisziplinären Projekten die Rekombination, Absorption und Integration von Wissen fördert. Daher ist es für wissenschaftsintensive Unternehmen entscheidend, ihre wissenschaftliche Absorptionsfähigkeit aufzubauen und zu entwickeln. Da medizintechnische Produkte immer komplexer werden, anspruchsvolle Funktionalitäten aufweisen und aus mehreren technologischen Komponenten bestehen, benötigen Unternehmen dieser Branche die Wissensintegration aus vielen unterschiedlichen wissenschaftlichen und technologischen Bereichen. Auf diese Weise kann die wissenschaftliche Absorptionskapazität dazu beitragen, die erforderlichen Verbindungen zwischen entfernten Wissensdomänen herzustellen und als solche die kognitiven Barrieren abzubauen, die ein Team von Forschern mit unterschiedlichem wissenschaftlichen Hintergrund haben kann.

II.2 Wichtigste Positionen des zahlenmäßigen Nachweises

Der größte Teil der Zuwendung des Projekts umfasst die Personalkosten für die Mitarbeiterinnen und Mitarbeiter der drei Verbundpartner während der gesamten Projektlaufzeit. Das

Projektpersonal hat die Forschungsarbeit und inhaltlich-technischen Aufgaben übernommen und abgeschlossen. Diese Arbeiten beinhalten 1) die konzeptionell und wissenschaftliche Ausarbeitung der im Projekt gestellten Fragestellungen; Durchführung von Experteninterviews und Fallstudien; 2) Aggregation, Integration und Deduplizierung der notwendigen Daten aus verschiedenen Datenquellen; 3) Integration und Annotation der wissenschaftlichen Publikationsdaten mit externen Thesauri (z.B. MeSH-Thesaurus, Standardthesaurus Wirtschaft der ZBW); 4) Management der benötigten Datenbasis und bei Bedarf laufende Aktualisierung der Datenbasis; 5) Erstellen eines neuronalen Netzwerks auf der Basis der zur Verfügung stehenden Daten unter Berücksichtigung der Relationen zwischen den Entitäten der Publikationsmetadaten (Ko-Autorenschaften, Annotationen, Publikationsjahre, Organisationen der Autorenschaft, Fachzeitschriften, Referenzen); 6) Entwicklung eines Algorithmus auf Basis der Deep Learning Methoden, die Optimierung der Hyperparameter beim Training des Modells und Validierung des Modells; 7) Integration der Modelloutputs in das Instrument zur Messung der wissenschaftlichen Dynamiken; 8) Validierung der Prognosegüte des Instruments und Durchführung von Längsschnittanalysen; 9) Entwicklung weiterer Messinstrumente und deren Evaluation; 10) Integration des Instruments in die Methoden der strategischen Frühaufklärung; 11) Evaluation und Verwertung der Ergebnisse.

Das Projektpersonal wurde bei seinen Aufgaben und Arbeiten während der gesamten Projektlaufzeit von studentischen Hilfskräften unterstützt, welche insbesondere an Datenaufbereitung, und -integration, Entwicklung des Algorithmus, Datenannotation und logischer Organisation beteiligt waren.

Die größten Positionen unter den Sachmittelausgaben sind die Mittel für die Beschaffung einer Lizenz zur Nutzung von Patentrohdaten der Patentdatenbank PATSTAT, welche durch die Universität Kiel erworben wurde. Diese PATSTAT-Lizenz wurde benötigt, um die Unternehmensdaten für die Realwelt-Fallstudien zu erheben und das im Projekt entwickelte Instrument zu validieren (AP 5). Zudem entstanden Kosten für die professionelle Sprachprüfung wissenschaftlicher Artikel, welche ebenfalls durch die Universität Kiel bezahlt wurden.

Zu der weiteren großen Position in den Sachmitteln gehören die Reisemittel, welche für die Dienstreisen zur Teilnahme an den nationalen und internationalen Konferenzen (siehe Abschnitt II.6) ausgegeben wurden. Auf den nationalen und internationalen Fachtagungen wurden die Ergebnisse des Verbundprojektes vorgestellt und mit der wissenschaftlichen Gemeinschaft diskutiert. Durch die konstruktiven Diskussionen wurden die Ergebnisse weiter verbessert und optimiert. Ein Teil der Sachmittel wurde für die Organisation und Durchführung des Workshops mit der Universität Bonn ausgegeben. Außerdem nahmen zwei Mitarbeiterinnen des Verbundprojekts an der European Summer School for Scientometrics zum Zweck der Weiterbildung im Bereich der bibliometrischen Analysen teil. Durch diese Weiterbildung haben die Mitarbeiterinnen neue szientometrische Methoden erlernt, die unmittelbar bei der Projektbearbeitung für die Autorenanalysen und MeSH-Konzept-Untersuchungen der Fallstudie "Forschung an der cholesterolsenkenden pharmazeutischen Präparaten und Naturstoffen" angewendet wurden. Die Mitarbeiterin der Universität Kiel nahm an dem Weiterbildungskurs "Data Science as a Research Method" teil. Die Teilnahme ermöglichte ihr das Erlernen der neuen

datenwissenschaftlichen Methode, die bei der Bearbeitung der Fallstudie zu SARS-CoV-2 Anwendung fand.

II.3 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Das Verbundvorhaben integrierte Perspektiven und Fragestellungen aus mehreren Fachbereichen (Lebenswissenschaften, Informationswissenschaften, Informatik und Innovationsforschung). Aus diesem Grund war das Vorhaben so komplex, dass die Bearbeitung der aufgestellten Fragestellungen sowie die Erreichung der wissenschaftlichen Ziele ohne die Zuwendung nicht möglich gewesen wäre. Das Verbundprojekt verlief nach der Projektplanung laut Antrag. Alle im Projekt gesetzten Ziele wurden erreicht. Alle gesetzten Meilensteine wurden erfolgreich abgeschlossen. Aufgrund der Coronapandemie mussten die Transferinstrumente angepasst werden. Die Ergebnisse des Verbundprojektes wurden der Öffentlichkeit in verschiedenen Publikationen zugänglich gemacht. Der Kostenrahmen wurde eingehalten. Der Zeitrahmen wurde eingehalten.

II.4 Voraussichtlicher Nutzen, insbesondere der Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans

Die Verbundpartner fokussieren als öffentliche Bildungs- und Forschungseinrichtungen ihre Aufgaben der Bildung des wissenschaftlichen Nachwuchses sowie der Forschung an gesellschaftlich relevanten Fragestellungen. Im Mittelpunkt des Verbundprojektes stand daher nicht die Verwertung durch die Partner selbst. Aus Sicht der Wirtschaft können jedoch durch die umfassenden Analysen der Faktoren der Entstehung und zeitlichen Entwicklung der wissenschaftlichen Konvergenz die potenziellen anschließenden Technologie-, Markt- und Industriekonvergenzen präziser vorausgesagt werden. Als Ergebnis daraus können sich relevante Stakeholder aus der Politik und Wirtschaft auf bevorstehenden technologischen und marktorientierten Veränderungen in den betreffenden Wirtschaftsbranchen gezielter vorbereiten. Um dieses Ergebnis der breiten Öffentlichkeit zugänglich zu machen, wurde im Laufe des Projektes ein Artikel mit praxisorientierten Inhalten u.a. in der Fachzeitschrift Hochschulmanagement veröffentlicht. Die Verwertung durch die wissenschaftliche Community und potentielle Anwender wurde ferner durch zahlreiche Publikationen in angesehenen Konferenzen und Zeitschriften gefördert (vergleiche II.6 Erfolgte oder geplante Veröffentlichungen der Ergebnisse).

Unsere Projektergebnisse werden in den nächsten Jahren in Lehrveranstaltungen an der Universität Kiel und an der Technische Hochschule Köln verwertet. Außerdem werden auf Basis

der erzielten Ergebnisse Folgeanträge ausgearbeitet, um die Relevanz der neuen Methoden des maschinellen Lernens in der Lebenswissenschaften- und Innovationsforschung zu steigern. Beispielsweise sind die Ergebnisse in das erfolgreich beantragte Vorhaben BlueHealthTech der Universität Kiel in der BMBF Bekanntmachung WIR! – Wandel durch Innovation in der Region eingeflossen und werden dort für die Strategieentwicklung an der Schnittstelle der Meeresforschung und der Gesundheitsforschung angewendet. Weiterhin können die im Projekt erhobenen und aus mehreren Datenquellen integrierten Datensätze nachgenutzt werden, solange keine rechtlichen Gründe im Sinne des Urheberrechts dagegensprechen. Der im Projekt entwickelte Programm-Code wird unter Open-Source-Lizenzen der Öffentlichkeit zur Nachnutzung zugänglich gemacht. Zusätzlich können die entwickelten Algorithmen sowie das Werkzeug-Schema zum Gebrauch des Messinstruments zur Bestimmung der wissenschaftlichen Dynamiken nachgenutzt werden.

Beide im Rahmen der Datenanreicherung entwickelten Verfahren, Take-it-Personally und OrcBot wurden im wissenschaftlichen Kontext vorgestellt und diskutiert (vergleiche Abschnitt II.1.2 Inhaltlich-technische Ergebnisse. Datengrundlagen; II.6 Erfolgte oder geplante Veröffentlichungen der Ergebnisse). Mit Hilfe der TIP-Library können bibliographische Metadaten mit Informationen aus Wikidata angereichert werden, etwa die Affiliation, Eltern, Geschlecht. OrcBot ermöglicht, die Datenbank Wikidata im umfassenden Maßstab mit weiteren qualitativ hochwertigen Daten aus ORCID anzureichern. TIP-Library und OrcBot sind in GitHub Repositorien dokumentiert und können für andere Aufgaben angepasst werden. Mit OrcBot trägt Q-Aktiv in einem Nebenergebnis dazu bei, die Datenbank Wikidata inhaltlich massiv zu erweitern.

II.5 Während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Projektlaufzeit haben wir die relevanten Forschungsfelder genau verfolgt. Im Bereich der Forschung an wissenschaftlichen Dynamiken und Konvergenzprozessen wurden die Arbeiten von Coccia (2018 und 2020) zu Faktoren der Evolution von Forschungsfeldern in unserer Forschung berücksichtigt. Die neuen Befunde zu den Konvergenzprozessen auf der Markt- und Industrieebene werden in den Forschungsarbeiten von Aaldering *et al.* (2019) und Sick *et al.* (2019) aufgezeigt. Neue Methoden zur Integration von Zitationsdaten bei der Generierung und Analyse eines Publikationsnetzwerks sowie zur Identifizierung von Themenclustern, basierend auf Latent Dirichlet Algorithmus und auf Grundlage des Fast-Newman-Algorithmus wurden im Artikel von Zhou *et al.* (2019) diskutiert. Mit den vorgestellten Methoden von Zhou *et al.* (2019) lassen sich verschieden Evolutionsphasen in den Prozessen der wissenschaftlichen Konvergenz identifizieren. Hacklin *et al.* (2021) haben eine andere Methode zur Messung der Ähnlichkeit der

zu konvergierenden wissenschaftlichen Themen vorgeschlagen, welche auf dem Jaccard Index als Ähnlichkeitsmaß basiert. Diese neuen Methoden haben wir bei der Entwicklung und Evaluierung unseres Ähnlichkeitsmaßes berücksichtigt und unser Instrument in die bisherige Literatur eingeordnet. Auf methodischer Seite (AP 3) gab es während der Projektlaufzeit zahlreiche neue Entwicklungen im Bereich von neuronalen Netzen für strukturierte Daten. Dazu zählen insbesondere Graph Convolutional Networks (Kipf & Welling 2017b), Stochastic Graph Convolution with Variance Reduction (Chen *et al.* 2018), Graph Attention Networks (Velickovic 2018), sowie und SEAL (Zhang & Chen 2018), Deep Graph InfoMax (Velickovic *et al.* 2019) und GraphSAINT (Zeng 2020). Für eine detaillierte Zusammenfassung der Entwicklungen verweisen wir auf: M. Bronstein, "Geometric deep learning" und W. Hamilton: "Graph Representation Learning". Auf der Seite der Vorhersage von Verknüpfungen in Netzwerken (Link Prediction, AP3), haben wir neben Graph Autoencoder (Kipf und Welling 2016) auch SEAL in unserem Methodenpool aufgenommen. Außerdem haben wir auf Techniken des Subgraph Sampling (wie in GraphSAINT) zurückgegriffen. Diese neuen Methoden haben wir in der Entwicklung unserer eigenen Methoden (AP 3) für die dynamische Netzwerkanalyse mit einbezogen. So haben wir für das Lernen von Knotenrepräsentationen die Technik der Graph Convolution (Kipf & Welling 2017b) mit Varianzreduktion (Chen *et al.* 2017) als Grundlage für unsere eigens entwickelte Kostenfunktion verwendet.

Die Vielzahl der Studien deuten darauf hin, dass die Relevanz der Analyse der wissenschaftlichen Dynamiken und insbesondere der wissenschaftlichen Konvergenz in den letzten Jahren stark gestiegen ist. Dies zeigt, dass dieses Thema in Bezug auf die wissenschaftliche Konvergenz nicht nur von reinem wissenschaftlichem Interesse ist, sondern auch praktische Bedeutung hinsichtlich des Monitoring und der Prognose von Technologie-, Markt- und Industriekonvergenz-Prozessen hat.

Aus Sicht der Aufbereitung und Abrufung der Daten haben sich keine grundlegend neuen Veränderungen oder Erkenntnisse ergeben, die das Vorgehen verändert hätten.

II.6 Erfolgte oder geplante Veröffentlichungen der Ergebnisse

Die Ergebnisse des Vorhabens wurden auf zahlreichen nationalen und hochkompetitiven internationalen Konferenzen und in den Workshops vorgestellt und veröffentlicht sowie in den hochkarätigen Journals publiziert. Die hohe wissenschaftliche Qualität ist die Grundlage für die erfolgreiche Verwertung der Projektergebnisse. Das Publizieren und Vorstellen der Ergebnisse in den Fachzeitschriften und auf Konferenzen schafften somit die Basis für die nachhaltige Verwertung und Dissemination der Ergebnisse. Dies gewährt eine langfristige Sichtbarkeit der im Vorhaben behandelten Themen und Fragestellungen und stellt den substanziellen Teil der Verwertungsstrategie dar.

Publikationsliste der im Projekt entstandenen Arbeiten

- Lukas Galke, Eva Seidlmayer, Gavin Lüdemann, Lisa Langnickel, Tetyana Melnychuk, Konrad U. Förstner, Klaus Tochtermann, Carsten Schultz (2021), **COVID-19++: A Citation-Aware Covid-19 Dataset for the Analysis of Research Dynamics**, *Big Data Analytics for COVID-19 Workshop @ IEEE Big Data 2021*, IEEE.
- Tetyana Melnychuk, Carsten Schultz, and Alexander Wirsich (2021), **The effects of university–industry collaboration in preclinical research on pharmaceutical firms’ R&D performance: Absorptive capacity’s role**. *J. Prod. Innov. Manag.*, 38: 355-378. <https://doi.org/10.1111/jpim.12572>
- Tetyana Melnychuk, Lukas Galke, Eva Seidlmayer, Konrad U. Förstner, Klaus Tochtermann, Carsten Schultz (2021): **Früherkennung wissenschaftlicher Konvergenz im Hochschulmanagement**. *Hochschulmanagement* 16(1), 24–28.
- Lukas Galke, Tetyana Melnychuk, Eva Seidlmayer, Steffen Trog, Konrad U. Förster, Carsten Schultz, Klaus Tochtermann (2019): **Inductive Learning of Concept Representations from Library-Scale Bibliographic Corpora**. INFORMATIK.
- Eva Seidlmayer/Jakob Voß/Tetyana Melnychuk/Lukas Galke/Klaus Tochtermann/Carsten Schultz (2020), **ORCID for Wikidata – Data Enrichment for scientometric Applications**, in CEUR-WS, CEUR Workshop Proceedings (presented at the 19th International Semantic Web Conference (ISWC 2020), Lucie-Aimée Kaffee, Oana Tifrea-Marcuska, Elena Simperl, Denny Vrandečić), urn:nbn:de:0074-2773-1.
- Eva Seidlmayer, Lukas Galke, Tetyana Melnychuk, Carsten Schultz, Klaus Tochtermann, Konrad U. Förstner (2019), **Take it Personally — A Python library for enrichment in informetrical applications**. Posters&Demos @ SEMANTICS 2019.

Vorträge der im Projekt erarbeiteten Ergebnisse

- Lukas Galke, Eva Seidlmayer, Gavin Lüdemann, Lisa Langnickel, Tetyana Melnychuk, Konrad U. Förstner, Klaus Tochtermann, and Carsten Schultz (2021), **COVID-19++: A Citation-Aware Covid-19 Dataset for the Analysis of COVID-19++Research Dynamics**. IEEE BigData 2021 workshop BDA COVID-2021, online workshop, Dec 15-18, 2021.
- Tetyana Melnychuk (2021), **How Degree of Interdisciplinarity Determines Firms’ Technological Novelty: Investigating Internal and External Learning Capabilities**. *DRUID 2021 Conference*. Copenhagen, Denmark. October 18-20 2021.
- Daniel Laufs, Tetyana Melnychuk, and Carsten Schultz (2021), **The Effects of Interdisciplinarity on R&D Performance: The Efficacy of Knowledge Recombination**. *Academy of Management Proceedings* 2021(1) <https://doi.org/10.5465/AMBPP.2021.16131abstract>
- Tetyana Melnychuk, Carsten Schultz (2021), **Effects of degree of interdisciplinarity on firms’ technological novelty: The moderating role of collaborations with universities and market participants, scientific absorptive capacity and financial performance**. *28th Innovation and Product Development Management Conference*. Milan, Italy (Online Conference), 6-8 June 2021.

- Eva Seidlmayer (2020), **ORCID for Wikidata: A workflow for matching author and publication items in Wikidata**, Projektpräsentation, online, Semantic Web in Library (SWIB), 2020-11-26.
- Eva Seidlmayer (2020), **ORCID for Wikidata – Data enrichment for scientometric applications**, Paper Diskussion, online, 19th International Semantic Web Conference (ISWC 2020), 2020-11-02.
- Eva Seidlmayer (2020), **ORCID for Wikidata**, Projektpräsentation, WikiCite Virtual Conference, 2020-10-28.
- Eva Seidlmayer (2020), **ORCID for Wikidata: A workflow for matching author and publication items in Wikidata**, Projektpräsentation, online, Semantic Web in Library, 2020-11-26.
- Tetyana Melnychuk, Daniel Laufs, Carsten Schultz (2020), **How the degree of interdisciplinarity determines R&D performance: The moderating effects of team diversity and knowledge base diversity**. *27th Innovation and Product Development Management Conference*. Online Conference, 7 - 9 June 2020.
- Tetyana Melnychuk, Alexander Wirsich, Carsten Schultz (2019), **Effects of university-industry collaborations in basic research on different stages of pharmaceutical new product development**. *26th Innovation and Product Development Management Conference*. Leicester, Great Britain, 9 - 11 June 2019.
- Tetyana Melnychuk, Lukas Galke, Eva Seidlmayer, Michael Wustmans, Klaus Tochtermann, Konrad Förstner, Stefanie Bröring, Carsten Schultz (2019), **Analyzing scientific dynamics – Does machine learning help to predict scientific convergence based on bibliographic data?** *R&D Management Conference 2019: The Innovation Challenge: Bridging Research, Industry and Society*. Paris, France, 17 – 21 June 2019.

Geplante Veröffentlichungen

- Tetyana Melnychuk, Lukas Galke, Eva Seidlmayer, Konrad U. Förstner, Stefanie Bröring, Klaus Tochtermann, Carsten Schultz (2021), **Development of Similarity Measures from Graph-Structured Bibliographic Metadata: An Application to Identify Scientific Convergence**.
- Tetyana Melnychuk, Lukas Galke, Eva Seidlmayer, Konrad U. Förstner, Klaus Tochtermann, Carsten Schultz (2021), **COVID-19 Research Cannibalism and the Lack of Interdisciplinary**.

Projektrelevante Webseiten

- <https://q-aktiv.github.io/>
- https://www.wihoforschung.de/wihoforschung/de/bmbf-projektfoerderung/foerderlinien/quantitative-wissenschaftsforschung/q-aktiv/q-aktiv_node.html
- <https://www.zbmed.de/forschen/laufende-projekte/q-aktiv/>
- <https://www.zbw-mediataalk.eu/de/2019/03/praxisbericht-open-science-diese-tools-foerdern-die-zusammenarbeit/>

- <https://www.b-i-t-online.de/neues/5386>

Referenzen

Aaldering, L. J., Leker, J., & Song, C. H. (2019). **Uncovering the dynamics of market convergence through M&A**. *Technological Forecasting and Social Change*, 138, 95-114. doi:<https://doi.org/10.1016/j.techfore.2018.08.012>

Adamic, L. A., & Adar, E. (2003). **Friends and neighbors on the Web**. *Soc. Networks* 25(3), 211-230

Albert-Laszlo, B., Jeong, H., Néda, Z., Ravasz, E., Schubert, A. & Vicsek, T. (2002). **Evolution of the social network of scientific collaborations**. *Physica A: Statistical mechanics and its applications* 311(3-4), 590-614.

Bronstein, M. M., Bruna, J., Cohen, T., & Velickovic, P. (2021). **Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges**. *CoRR* abs/2104.13478

Chai, S. & Shih, W. (2016). **Bridging science and technology through academic–industry partnerships**. *Research Policy*, 45(1), 148–158.

Chen, J., Zhu, J., & Song, L. (2018) **Stochastic Training of Graph Convolutional Networks with Variance Reduction**. *ICML 2018*, 941-949

Coccia, M. (2018). **General properties of the evolution of research fields: a scientometric study of human microbiome, evolutionary robotics and astrobiology**. *Scientometrics*, 117(2), 1265-1283. doi:10.1007/s11192-018-2902-8

Coccia, M. (2020). **The evolution of scientific disciplines in applied sciences: dynamics and empirical properties of experimental physics**. *Scientometrics*, 124(1), 451-487. doi:10.1007/s11192-020-03464-y

Curran, C.-S., Bröring, S., & Leker, J. (2010). **Anticipating converging industries using publicly available data**. *Technological Forecasting and Social Change*, 77(3), 385-395. doi:<https://doi.org/10.1016/j.techfore.2009.10.002>

Curran, C.-S., & Leker, J. (2009). **Employing STN AnaVist to forecast converging industries**. *International Journal of Innovation Management*, 13(04), 637-664. doi:10.1142/s1363919609002455

Jeong, D., Lee, K., & Cho, K. (2018). **Relationships among international joint research, knowledge diffusion, and science convergence: the case of secondary batteries and fuel**

cells. Asian Journal of Technology Innovation, 26(2), 246-268.
doi:10.1080/19761597.2018.1522961

Leker, J., Song, C. H. (2014): **Die Prognose von Konvergenzentwicklungen zur Identifikation attraktiver Innovationsfelder.** In: Schultz, C., Hölzle, K. (Hrsg.): Motoren der Innovation – Zukunftsperspektiven der Innovationsforschung (S. 3-22). 1. Auflage, Wiesbaden: Springer Gabler.

Grover, A., Leskovec J. (2016). **node2vec: Scalable Feature Learning for Networks.** KDD 2016, 855-864

Hacklin, F., Wallin, M. W., Björkdahl, J., & von Krogh, G. (2021). **The Making of Convergence: Knowledge Reuse, Boundary Spanning, and the Formation of the ICT Industry.** IEEE Transactions on Engineering Management, 1-13. doi:10.1109/TEM.2021.3087365

Hamilton, W. L. (2020). **Graph Representation Learning.** Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool Publishers.

Kipf, T. N., Welling, M. (2016). **Variational Graph Auto-Encoders.** Bayesian Deep Learning Workshop (NIPS 2016).

Kipf, T. N., Welling, M. (2017). **Semi-Supervised Classification with Graph Convolutional Networks.** ICLR (Poster) 2017.

Jeong, D., Cho, K., Park, S., Hong, S. (2016). **Effects of knowledge diffusion on international joint research and science convergence: Multiple case studies in the fields of lithium-ion battery, fuel cell and wind power.** Technological Forecasting and Social Change, 108, 15-27.

Jeong, D., Lee, K., Cho, K. (2018). **Relationships among international joint research, knowledge diffusion, and science convergence: the case of secondary batteries and fuel cells.** Asian Journal of Technology Innovation, 26(2), 246-268.

Nemet, G. F., & Johnson, E. (2012). **Do important inventions benefit from knowledge originating in other technological domains?** Research Policy, 41(1), 190-200.
doi:<https://doi.org/10.1016/j.respol.2011.08.009>

Newman, M. E. J. (2001). **Clustering and preferential attachment in growing networks.** Physical review E 64(2), 025102.

Niemann, H., Moehrle, M. G., & Frischkorn, J. (2017). **Use of a new patent text-mining and visualization method for identifying patenting patterns over time: Concept, method and test application.** Technological Forecasting and Social Change, 115, 210-220.
doi:<https://doi.org/10.1016/j.techfore.2016.10.004>

Perozzi B., Al-Rfou, R., & Skiena, S. (2014). **DeepWalk: online learning of social representations**. KDD 2014: 701-710

Phelps, C., Heidl, R., & Wadhwa, A. (2012). **Knowledge, Networks, and Knowledge Networks: A Review and Research Agenda**. Journal of Management, 38(4), 1115-1166

Sick, N., Preschitschek, N., Leker, J., & Bröring, S. (2019). **A new framework to assess industry convergence in high technology environments**. Technovation, 84-85, 48-58. doi:<https://doi.org/10.1016/j.technovation.2018.08.001>

Upham, S., Rosenkopf, L., & Ungar, L. (2010). **Innovating knowledge communities**. Scientometrics, 83. doi:10.1007/s11192-009-0102-2

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). **Graph Attention Networks**. ICLR (Poster)

Venugopalan, S., & Rai, V. (2015). **Topic based classification and pattern identification in patents**. Technological Forecasting and Social Change, 94, 236-250. doi:<https://doi.org/10.1016/j.techfore.2014.10.006>

Velickovic, P., Fedus, W., Hamilton, W. L., Liò, P., Bengio, Y., & Hjelm, R. D., (2019). **Deep Graph Infomax**. ICLR (Poster)

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). **The Increasing Dominance of Teams in Production of Knowledge**. Science, 316(5827), 1036-1039.

Yegros-Yegros, A., Rafols, I., & D'Este, P. (2015). **Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity**. PLOS ONE, 10(8), e0135095-e0135095. doi:10.1371/journal.pone.0135095

Zhang, M., & Chen, Y. (2018). **Link Prediction Based on Graph Neural Networks**. NeurIPS 2018: 5171-5181

Zhou, Y., Dong, F., Kong, D., & Liu, Y. (2019). **Unfolding the convergence process of scientific knowledge for the early identification of emerging technologies**. Technological Forecasting and Social Change, 144, 205-220. doi:<https://doi.org/10.1016/j.techfore.2019.03.014>

Zeng, H., Zhou, H., Srivastava, A., Kannan, R., & Prasanna, V. K. (2020). **GraphSAINT: Graph Sampling Based Inductive Learning Method**. ICLR 2020