**MedienPädagogik**
Zeitschrift für Theorie und Praxis der Medienbildung
**www.medienpaed.com** **ISSN 1424-3636**

**Themenheft 72: Mixed Methods im Spannungsfeld. Weiterentwicklung digitaler Forschungsmethoden zwischen KI, NLP, Big Data und transdisziplinärer Forschung.**
Herausgegeben von Jana Heinz, Andrea Hense und Christian Janßen

# Kartierung der europäischen Bildungsforschungslandschaft durch Topic Modelling

## Labelling, Iteration und Analyse medienpädagogischer Begriffscluster

Alexander Christ[1], Christoph Schindler[1] und Jens Röschlein[1]

[1] DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

**Zusammenfassung**
*Latent-Dirichlet-Allocation (LDA) ist eine Machine-Learning-Methode zur Themenmodellierung (engl. Topic Modelling) grosser Textkorpora (d. h. zur Bestimmung latenter Begriffs- und Dokumentencluster, sog. Themen). LDAs stellen eine transparente, ressourcenschonende und intuitiv verständliche Alternative zu LLMs dar. Im Sinne der Mixed-Methods müssen die statistischen Ergebnisse der LDAs durch qualitatives Labelling in sinnhafte Themen überführt werden. Dieses Verfahren wurde auf über 30.000 Einreichungen der European Conference on Educational Research (ECER) angewendet und liefert die Grundlage für die interaktive Webapp EduTopics (Christ et al. 2025). Sie ermöglicht Nutzenden, eigene qualitative Analyse von Themen und Ko-Variaten sowie explorative Zugänge durch interaktive Visualisierungen durchzuführen. Die Webapp und die darin enthaltenen Themen wurden in mehreren iterativen Schleifen mit Vertretenden der erziehungswissenschaftlichen Community diskutiert, angepasst und weiterentwickelt. Neben der Webapp und ihren Funktionen liegt im Beitrag ein Schwerpunkt auf zwei Granularitätsebenen mit zentralen Superthemen (k = 50) und Subthemen (k = 300). Deren hierarchische Struktur als Über- und Unterkategorien wurde über die semantische Cosinus-Ähnlichkeit der Themen bestimmt. Exemplarisch werden Trends medienpädagogisch relevanter Subthemen wie Media Literacy herangezogen und diskutiert sowie mit den Themen-Trends für das EERA-Netzwerk 06 – Open Learning: Media, Environments and Cultures (i. e. Medienpädagogik) verglichen.*

**Mapping the European Educational Research Landscape Through Topic Modelling. Labelling, Iteration and Analysis of Media Education Term Clusters**

**Abstract**
*Latent Dirichlet Allocation (LDA) is a machine learning method for topic modelling of large text corpora (i. e. for determining latent term and document clusters, so-called topics). LDAs represent a transparent, resource-efficient and intuitively comprehensible alternative to*



Christ, Alexander, Christoph Schindler, und Jens Röschlein. 2026. «Kartierung der europäischen Bildungsforschungslandschaft durch Topic Modelling. Labelling, Iteration und Analyse medienpädagogischer Begriffscluster». *MedienPädagogik* 72 (Mixed Methods): 49–70. https://doi.org/10.21240/mpaed/72/2026.09.20.X.

*LLMs. In the spirit of mixed methods, the statistical results of the LDAs must be converted into meaningful topics by qualitative labelling. This procedure was applied to over 30,000 submissions to the European Conference on Educational Research (ECER) and provides the basis for the interactive webapp EduTopics (Christ et al. 2025). It enables users to conduct their own qualitative analysis of topics and co-variates as well as explorative access through interactive visualisations. The webapp and the topics it contains were discussed, adapted and further developed in several iterative loops with representatives of the educational research community. In addition to the app and its functions, the article focuses on two levels of granularity with central supertopics (k = 50) and subtopics (k = 300). Their hierarchical structure as super- and subcategories was determined by the semantic cosine similarity of the topics. Trends in subtopics relevant to media education such as media literacy are used and discussed as examples and compared with the topic trends for the EERA network 06 – Open Learning: Media, Environments and Cultures (i.e. media education).*

## 1. Einleitung

Parallel zu der annähernd ubiquitären Ausbreitung der medialen und wissenschaftlichen Diskussion zu (Effekten) der sozio-technischen digitalen Transformation der Gesellschaft in den letzten Jahrzehnten – inklusive so unterschiedlicher Phänomene und (medienpädagogischer) Forschungsgegenstände wie künstliche Intelligenz, soziale Medien oder Videospiele – ist auch die Menge und Heterogenität von Forschungsarbeiten zu diesem Themenkomplex substanziell und kontinuierlich gestiegen (Huang et al. 2020). Gleichzeitig hat die verbesserte Rechenkapazität von Computern – und die Notwendigkeit solcher Verfahren für die Analyse heterogener Literaturkorpora – zur (Weiter-)Entwicklung und Etablierung von Verfahren für die automatisierte Textanalyse geführt. Mit diesen maschinellen Lernverfahren der Computerlinguistik lassen sich Informationen über die zentralen Themen, Akteure und ihre Trends selbst aus den grössten und heterogensten Forschungsfeldern extrahieren. Ein prominenter Vertreter solcher Methoden ist die sogenannte *Themenmodellierung* (engl.: Topic Modelling) – insbesondere die Latente Dirichlet Allokation (LDA, siehe Blei, Ng, und Jordan 2003).

Mit der Anwendung von LDAs in Kombination mit weiteren Clusteringverfahren lassen sich Forschungsfelder kartieren, zentrale inhaltliche Trends und Autor:innen bestimmen wie auch (zeitliche) Unterschiede für weitere Kovariaten bestimmen (z. B. Affiliationsländer). Aufgrund der beschriebenen beschleunigten Entwicklung medienpädagogischer Forschungsgegenstände ist gerade die Kartierung erziehungswissenschaftlicher Forschungsarbeiten und -projekte ein interessantes Beispiel für die exemplarische Darstellung der Nützlichkeit von Themenmodellierung, da die Annahme getroffen werden kann, dass die Relevanz medienpädagogischer Forschungsgegenstände über die letzten Jahrzehnte kontinuierlich gestiegen ist und sich auch

die Themen selbst verändert haben (durch beispielsweise Smartphones und soziale Medien). Gleichzeitig kann – wegen der generellen digitalen gesellschaftlichen Transformation – davon ausgegangen werden, dass medienpädagogische Forschungsgegenstände und Inhalte sukzessive auch vermehrt ausserhalb der Medienpädagogik untersucht wurden. Um diese Annahmen auf europäischer Ebene zu überprüfen – und gleichzeitig die verwendeten Methoden exemplarisch darzustellen und zu diskutieren –, werden die Ergebnisse der Webapp EduTopics: ECER (Christ, Röschlein, und Schindler 2025) herangezogen. EduTopics beinhaltet und visualisiert Ergebnisse von über 30.000 Beiträgen zur *European Conference on Educational Research* (ECER) von 1998 bis 2024, die von der European Educational Research Association (EERA) organisiert wurden. Eine genaue Erläuterung der Entwicklung der Webapp und eine vollständige Beschreibung ihrer Funktionen findet sich in Christ et al. (2025). In dem vorliegenden eigenständigen Beitrag wird aufbauend auf Christ et al. (2025) gezeigt, wie die Analysen der Webapp eine Grundlage liefern können, die (thematische) Entwicklung spezifischer paradigmatischer oder inhaltlicher Teilmengen der europäischen Bildungsforschung nachzuzeichnen. In diesem Zusammenhang wird exemplarisch die Entwicklung medienpädagogischer Forschungsgegenstände und -felder aufgezeigt. Dafür verfolg dieser Beitrag zwei Ziele:

1. Es wird dargestellt, wie Themenmodellierung angewandt werden kann, um in grossen, heterogenen Literatursammlungen zentrale Themen zu identifizieren, und wie diese Ergebnisse genutzt werden können, um Forschungsfelder zu beschreiben sowie Schwerpunkte und Trends zu identifizieren.
2. Auf thematischer Ebene wird betrachtet, wie sich medienpädagogisch relevante Themen von 1998 bis 2024 bei der ECER entwickelt haben und ob sich für verschiedene Struktureinheiten wie die EERA-Netzwerke unterschiedliche Trends ergeben haben.

## 2. Themenmodellierung

### *2.1 Annahmen Themenmodellierung*

Themenmodellierung ist ein (in der Computerlinguistik und Metaforschung) weit verbreitetes Bayessches statistisches Verfahren zur Analyse und Kartierung grosser heterogener Literaturkorpora (Bittermann und Fischer 2018; Blei et al. 2003; Blei und Lafferty 2009; Christ et al. 2021; Christ, Smolarczyk, und Kröner 2024; Griffiths und Steyvers 2004; Heidenreich et al. 2019; Jacobs und Tschötschel 2019). Sie gehört zu dem methodologischen Paradigma Computerlinguistik, welches Verfahren zur algorithmischen und (pseudo-)automatisierten Verarbeitung und Analyse unstrukturierter Textdaten umfasst (Gandomi und Haider 2015; O'Mara-Eves et al. 2015; Silge und Robinson 2017). Wie auch bei anderen Verfahren des Clusterings bzw. der Komplexitätsreduktion existieren

verschiedene Ansätze zur Identifikation der Themen in einem Korpus, von denen die latente Dirichlet Allokation (LDA; Blei et al. 2003) die am weitesten verbreitete und intuitiv verständlichste Variante ist. Die Annahmen der LDA sind:

1. *k* verschiedene Themen werden in einem Korpus behandelt: Die Anzahl der Themen in einem Korpus muss a priori festgelegt werden.
2. Dokumente können eine Mischung von Themen behandeln: Es handelt sich um einen *Mixed-Membership*-Ansatz, mit welchem Themen nicht disjunkt klassifiziert werden, sondern deren Verteilung in einem Text prozentual bestimmt wird.
3. Die Themen können basierend auf Wörtern bestimmt werden, die im Vergleich zu anderen Dokumentengruppen häufig (gemeinsam) in Dokumentengruppen vorkommen: Wörter, die im gesamten Korpus häufig vorkommen, sind nicht relevant für die Bestimmung der Themen, sondern nur Wörter, die indikativ für eine spezifische Gruppe an Dokumenten sind.

### *2.2 Machine-Learning-Modellierung*

Anhand dieser Annahmen werden *k* Themen in zwei parallelen Clusteringalgorithmen über mehrere hundert bis mehrere tausend Iterationen identifiziert. Dafür wird die sogenannte Dokument-Term-Matrix erstellt, deren Zeilen aus den Dokumenten und Spalten aus den Termen der Dokumente bestehen. In den Zellen der Matrix findet sich die Häufigkeit des Terms im jeweiligen Dokument. Im Clustering selbst wird dann die Matrix in einen sogenannten gerichteten azyklischen Graph (engl.: directed acyclic graph, DAG) überführt, in welchem initial (d. h. in der ersten Iteration) jedes Dokument und jedes Wort einem der *k* Themen zufällig zugewiesen wird. Ab Iteration 2 beginnt dann der maschinelle Lernprozess, indem diejenigen Wörter/Dokumente weiterhin dem gleichen Thema zugewiesen werden, die in gleichen Dokumenten/mit gleichen Wörtern vorkamen, während für diejenigen Wörter/Dokumente, für die dies nicht der Fall war, eine erneute zufällige Zuweisung stattfindet. Über die gesamte Anzahl an Iterationen werden dadurch Cluster an Wörtern identifiziert, die häufig in Dokumenten gemeinsam vorkommen, und Cluster an Dokumenten, die häufig diese Wörter verwenden. Über die Iterationen hinweg wird die Dichte innerhalb der Dokumentencluster minimiert (d. h. die Ähnlichkeit der Dokumente eines Themas maximiert) und der Abstand zwischen den Clustern maximiert (d. h. die Unterschiedlichkeit verschiedener Themen maximiert). Das Ergebnis der Themenmodellierung sind neben statistischen Werten für die Modellgüte (z. B. Log-Likelihood oder $R^2$) zwei Matrizen: die Dokument-Themen-Matrix und die Term-Themen-Matrix.

*Term-Themen-Matrix*: Diese Matrix spannt den semantischen Raum zwischen Wörtern (Zeilen) und Themen (Spalten) auf. In den Zellen steht jeweils das Term-Themen-Gewicht β jedes Terms für jedes Thema. Dieses Gewicht kann nie null sein und gibt an, wie sich die Zuordnung eines Dokuments zu einem Thema ändert, wenn der Term einmal

auftritt; d. h. wenn das Wort *Media* für Thema A ein Gewicht von $\beta_{(Media, A)} = .05$ und für Thema B von $\beta_{(Media, B)} = .001$ hat, dann steigt die Zuordnung eines Textes mit drei Vorkommnissen von *Media* für Thema A um 15 % und für Thema B um .3 %. Alle Gewichte für alle Themen über einen Text hinweg aggregiert ergeben dann das Dokument-Themen -Gewicht ɣ des Texts.

*Dokument-Themen-Matrix*: Die Dokument-Themen-Gewichte ɣ für jeden Text für jedes der *k* Themen werden in dieser Matrix festgehalten. Für einen einzelnen Text über alle Themen hinweg ergibt sich $\Sigma_{(ɣ)} = 1$ mit $ɣ \in ]0; 1[$; d. h. die Zugehörigkeit eines Texts zu einem Thema kann nur im Grenzwertfall 0 oder 1 annehmen, da alle Wörter eines Textes stets eine minimale Restwahrscheinlichkeit nahe 0 haben zu jedem der *k* Themen zu gehören. Ein Text der beispielsweise bei $k = 10$ für das Thema *Videospiele* $ɣ = .6$ und für das Thema *Interview* $ɣ = .3$ aufweist, wird mit sehr hoher Wahrscheinlichkeit jenen Themen zuzuordnen sein; für die restlichen $k = 8$ Themen würde sich somit eine arbiträre durchschnittliche Restwahrscheinlichkeit von jeweils $ɣ = {}^{0.1}/_{8} = .0125$ ergeben, die deutlich niedriger ist als der Erwartungswert bei einer zufälligen Verteilung von $E_{(ɣ)} = .10$ (da ${}^{1}/_{10}$). Für die Betrachtung der Verteilung eines Themas in einem gesamten Korpus oder geclustert anhand weiterer Kovariaten wie Autor:innen, Publikationsjahre oder Länder sollte der Mittelwert $\mu_{(ɣ)}$ anstelle einer diskreten Zuordnung $n_{(Thema)} = \Sigma(Dokument, ɣ_{(max, 1)})$ berechnet werden, da querliegende Themen wie zu wissenschaftlichen Methoden oder breiten Forschungsinteressen (z. B. Sozioökonomischer Status) selten die wichtigsten Themen einzelner Dokumente sind, aber über die für das gesamte Korpus eine breite Relevanz aufweisen. Für jegliche Form der Aggregation von ɣ kann der Bernoulli-Standardfehler mit $se = (\mu_{(ɣ)} * (1 - \mu_{(ɣ)})/n)^{0,5}$ geschätzt werden.

Die Modellierung an sich und die Ergebnisse des iterativen Verfahrens sind abhängig von vier zentralen Aspekten:

1. *Korpus*: Verschiedene Aspekte der Dokumente des Korpus können einen erheblichen Einfluss auf den Prozess der Themenmodellierung und dessen Ergebnisse haben. Generell steigt die Validität und Reliabilität der Ergebnisse durch die Verwendung längerer Texte, da der Algorithmus auf der Häufigkeit von Wörtern basiert – die in kurzen Texten wie Titeln von Forschungsarbeiten meist zwischen 0 und 1 schwankt. Auch die Anzahl an Dokumenten hat einen Einfluss auf die Themenmodellierung da eine zu geringe Anzahl an Dokumenten ein statistisches Power-Problem darstellt. Unabdingbar ist, dass die Texte alle in der gleichen Sprache sind, da andernfalls sprachspezifische Themen erzeugt werden.
2. *Wörter*: Die Verwertbarkeit der Ergebnisse wird bestimmt durch die Grundlage der Themen, d. h. die Wörter, nach denen die Themen identifiziert werden. Eine erste relevante Entscheidung ist die *Art der Wörter*: entweder wie sie im Text stehen, als Wortstämme oder als Lemmata. Bei ersterem wird der Kontext der Wörter besser erhalten – da eben *learning* und *learner* unterschiedliche Bedeutungen tragen. Es besteht aber die Möglichkeit, dass Wörter mit vielen verschiedenen Ausprägungen

dadurch einzeln jeweils so selten werden, dass aus ihnen kein eigenes Thema gebildet wird oder das bzw. die daraus gebildeten Themen bzgl. ihrer Relevanz für das Korpus unterschätzt werden. Daher bieten sich Wortstämme oder Lemmata eher für heterogene Datensätze an, in welchen viele verschiedene Themen behandelt werden, und die exakten Wörter eher für homogene Datensätze, bei welchen die groben Themen a priori klar sind, aber deren Ausdifferenzierung das Ziel der Themenmodellierung ist.

3. *Anzahl der Themen k*: Die Anzahl der Themen *k* wird je nach Ziel der Analyse gewählt. Ist das Anliegen, grössere, grobe Cluster an Themen zu identifizieren, so bieten sich kleinere *k* an und vice versa. Erste Orientierungen für passende *k* können verschiedene Verfahren wie Gibbs-Sampling liefern (Arun et al. 2010; Cao et al. 2009; Deveaud et al. 2014; Steyvers und Griffiths 2007). Diese Verfahren geben für ein gewähltes Maximum an $k_{(max)} \in [2{:}k_{(max)}]$ Modelfitwerte für das jeweilige *k* an. Dabei ist die exakte Anzahl an Themen *k* nur dann besonders relevant, wenn es sich um eine kleine Zahl handelt, da der Unterschied zwischen $k = 5$ und $k = 6$ substanziell ist, der Unterschied zwischen $k = 100$ und $k = 101$ jedoch marginal. In letzterem Fall kann das Gibbs-Sampling genutzt werden, um eine Orientierung der Grössendimension von *k* zu erhalten.
4. *Hyperparameter α und β*: Als Bayessches Verfahren sind die Ergebnisse (Posterior in der Bayesschen Statistik) der Themenmodellierung abhängig von den angenommenen Verteilungen, auch Startwerten oder Hyperparameter genannt (Priors in der Bayesschen Statistik). Für die Themenmodellierung sind es die zwei Hyperparameter α und β.

   α gibt an, ob die Themen im *k*-dimensionalen Raum eher im Zentrum oder eher am Rand liegen sollen, d. h. ob die Dokumente eher wenigen Themen zugewiesen werden (Rand) oder die Dokumente eher aus einer Mischung von vielen Themen bestehen (Zentrum). Für erstere Absicht sollte $\alpha < 1$, für letztere $\alpha > 1$ gewählt werden.

   β gibt an, ob die Themen aus einzelnen Wörtern oder einer Mischung aus Wörtern bestehen sollen. Hohe β-Werte resultieren in Themen, die nur aus wenigen Wörtern bestehen, niedrige β-Werte führen zum Gegenteil.

   Griffiths und Steyvers (2004) empfehlen für die LDA $\alpha = 5/k$ und ein $\beta = 0.01$. Jenseits dieser Empfehlungen können die optimalen α und β-Werte für eine bestimmte Anzahl an Themen *k* auch per Simulationsverfahren bestimmt werden.

   Weiterhin werden die Ergebnisse noch beeinflusst durch die Anzahl an Iterationen und den zufälligen systembedingten Startwert. Bei ersterem kann eine sehr hohe Anzahl an Iterationen keine falsche Entscheidung sein, da – wenn das Lernen des Modells abgeschlossen ist – weitere Iterationen keinen nennenswerten Unterschied mehr auf die Ergebnisse der Modellierung haben. Der Startwert jedoch sollte nicht zufällig vom System festgelegt, sondern auf einen Wert fixiert werden, damit die Reproduzierbarkeit der Ergebnisse vorliegt.

### *2.3 Qualitatives Labelling*

Sowohl bei der eventuell vorhandenen Bestimmung der Hyperparameter und der Anzahl der Themen *k* – sollte dies bei dem konkreten Anwendungsfall notwendig sein – als auch für die Beurteilung des finalen Modells ist die qualitative Analyse der Ergebnisse unabdingbar, da die Ergebnisse der Themenmodellierung eine reine Auflistung von Termen und Dokumenten mit hohen Zuordnungsgewichten zu arbiträren numerischen Ziffern sind. Das zentrale Ziel dieser Analyse ist die Bestimmung des Inhalts der Themen durch das qualitative Labelling anhand der jeweiligen Terme mit hohen Term-Themen-Gewichten β und Dokumenten mit hohen Dokument-Themen-Gewichten ɣ. Dabei sollten nicht nur die repräsentativsten Arbeiten betrachtet werden, sondern auch im Sinne der Grounded Theory kontrastierende Fälle. Je höher die Anzahl an Themen *k* a priori festgelegt wurde, desto kleinschrittiger ist auch der Charakter dieses Prozesses da grössere thematische Foki in differenzierte Themen zerlegt werden und dadurch beim Labelling die verschiedenen – aber ähnlichen – Themen häufiger miteinander verglichen werden müssen, damit ihr inhärenter Unterschied sich in ihren Bezeichnungen widerspiegelt.

### *2.4 Ähnlichkeit zwischen verschiedenen Themenmodellierungen*

Je nach Zielstellung des Vorhabens kann es interessant sein, den Zusammenhang zwischen den Themen innerhalb eines einzelnen Modells oder zwischen verschiedenen Modellen zu bestimmen. Wenn ersteres untersucht werden soll, stossen klassische statistische Verfahren wie Korrelationen an ihre Grenzen, da der Zusammenhang zwischen den Dokument-Themen-Gewichten ɣ innerhalb eines Modells stets indirekt proportional zueinander ist, da aus einem hohen ɣ zu einem Thema automatisch nur ein niedriges ɣ zu anderen Themen resultieren kann. Stattdessen wird für die Bestimmung der Ähnlichkeit von Themen nicht die Verteilung der Dokumente über die Themen, sondern die semantische Ähnlichkeit der Themen betrachtet, da dies auch den Vergleich über verschiedene Korpora hinweg erlaubt, nicht nur die Betrachtung der Ähnlichkeit innerhalb eines Korpus. Für die Bestimmung der Ähnlichkeit bestehen verschiedene Möglichkeiten, die von Aletras und Stevenson (2014) verglichen wurden (für deren Anwendung bei Themenmodellierungen: Kullback-Leibler-Divergenz: Wang et al. 2009; Newman et al. 2009; Cosinus-Ähnlichkeit: He et al. 2009; Ramage et al. 2009). Von diesen erreichte das (zeit- und ressourcenschonende) intuitiv-verständliche Verfahren der Bestimmung der Cosinus-Ähnlichkeit der wichtigsten 20 Wörter pro Thema sehr gute Perfomanzwerte. Für die Bestimmung der semantischen Cosinus-Ähnlichkeit $\cos_{(\theta)}$ wird der semantische Raum der Themen in einer Matrix (Themen × Terme) aufgespannt, wodurch jedem Thema ein Wort-Vektor zugewiesen wird. Zwischen jeweils zwei Vektoren (d. h. Themen) im n-dimensionalen Raum (*n* = Anzahl der Terme) kann somit ein normierter Winkel- und Längenunterschied bestimmt werden, wobei ähnliche Vektoren (d. h. Themen) Werte

nahe 1 und unähnliche Themen Werte nahe 0 erhalten. Dadurch können die jeweils nächsten Nachbarn von Themen bestimmt werden, sowohl innerhalb eines Modells als auch zwischen Modellen.

## 3. Methode

### *3.1 Korpus*

Das zugrundeliegende Korpus wurde bereitgestellt von der EERA und enthielt alle Informationen zu den Beiträgen zur ECER von 1998 bis 2024. Die Namen und Affiliationen der Autor:innen waren während der Beitragseinreichung selbst getätigte Informationen. Im Ausgangszustand bestand das Korpus aus $N = 43.221$ mit den Variablen *Dokumenten-ID*, *Titel*, *Jahr der ECER*, *Beitragsformat*, *EERA-Netzwerk* und den drei inhaltlichen Feldern *detaillierte Informationen*, *Methoden* und *Ergebnisse*. Die Texte der drei inhaltlichen Felder wurden in *einer* Textvariablen zusammengefasst, da erst ab 2008 eine Differenzierung der Zusammenfassungen der Beiträge erfolgte.

Nach der Bereinigung – die im Detail in Christ, Röschlein und Schindler (2025) nachgelesen werden kann – ergaben sich insgesamt $n = 34.605$ relevante Beiträge von $n = 29.847$ verschiedenen Autor:innen, verteilt auf $n = 129$ verschiedene Affiliationsländer (mit einem Missing-Anteil von 14.7 %) und $n = 38$ EERA-Netzwerke. Aus der inhaltlichen Textvariablen wurden Exportprobleme wie HTML-Fragmente (z. B. <p>, <li>, <br>), Bindewörter (wie *and, the*, *be*), alle Ländernamen (z. B. Germany, United Kingdom, USA) und Sprachen (z. B. English, German, Finnish) entfernt und die übrigen Wörter auf ihre Wortstämme reduziert.

### *3.2 Bibliografische Analysen*

Für die bibliografischen Analysen wurden für jeden Beitrag jeweils die Autor:innen, die Affiliationsländer der Autor:innen, die Zugehörigkeit des Beitrags zu einem oder mehreren EERA-Netzwerken und das Jahr der ECER bestimmt. Für die Affiliationsländer der Beiträge wurden die gleichen Länder jeweils nur einmal gezählt, d. h. ein Beitrag von fünf Autor:innen aus Schweden zählt in den Analysen nur einmal als schwedischer Beitrag.

### *3.3 Themenmodellierung mit zwei Granularitätsebenen*

Für die Bestimmung der Inhalte des Korpus und der einzelnen Beiträge wurde Themenmodellierung mit zwei Granularitätsebenen gewählt, um sowohl gröbere Schwerpunkte und Trends in den Superthemen als auch differenziertere Betrachtungen über die Subthemen zu ermöglichen. Die Grundlagen für beide Modelle waren das gleiche

Korpus und die gleichen Wortstämme, und zwar alle Wortstämme, die in mindestens $n = 50$ Beiträgen vorkamen.

*Bestimmung der Anzahl der Themen k:* Für die Bestimmung der Anzahl an Super- und Subthemen wurde ein Gibbs-Sampling nach Steyvers und Griffiths (2007) mit den Massen von Cao et al. (2009), Arun et al. (2010), Deveaud et al. (2014) sowie Griffiths und Steyvers (2004) inklusive Perplexitäts- (De Waal und Barnard 2008) und Kohärenzanalysen (Mimno et al. 2011; Newman et al. 2009) durchgeführt. Für die Bestimmung der Anzahl der Superthemen wurde a priori das Intervall $k \in [40; 80]$ in $k = 5$ Schritten festgelegt, für die Subthemen $k \in [200; 400]$ in $k = 10$ Schritten. Bei beiden Simulationen wurden diejenigen Themenmodellierungen mit lokalen Suprema, d. h. guten Modell-Fit-Werten, in den dazugehörigen Massen (bzw. mit geometrischen Wendepunkten bei der Perplexität und Kohärenz) separat betrachtet und der Inhalt der einzelnen Themen qualitativ bestimmt. Dafür wurden für jedes Thema jeweils die $n = 25$ Wörter mit den höchsten Term-Themen-Gewichten $\beta$ und die $n = 50$ Beiträge mit den höchsten Dokument-Themen-Gewichten $\gamma$ betrachtet, um das Thema zu labeln. Dabei wurden aus den Modellen mit gutem Fit diejenigen mit dem höchsten $k$-Wert als erstes analysiert, da dadurch in den qualitativen Vergleichen mit den Modellen mit niedrigerem $k$ die Cosinus-Ähnlichkeit genutzt werden konnte um zu bestimmen, ob ein niedrigeres $k$ zum Verschwinden inhaltlich-distinkter Themen geführt hat oder ob das kleinere Modell nur dazu geführt hat, dass mehrere sehr ähnliche Themen in ein Thema überführt wurden. Dieser Prozess hat dazu geführt, dass $k = 50$ für die Superthemen und $k = 300$ für die Subthemen als beste Lösung identifiziert wurde.

*Labelling*: Die Label der Themen beider Modelle wurden – wie bereits beschrieben – qualitativ bestimmt. Bei mehreren verschiedenen Vorträgen zu dem Projekt und dessen Ergebnissen wurden die Themenbezeichnungen mit Vertreter:innen der erziehungswissenschaftlichen Community (z. B. mit den Organisatoren der EERA-Netzwerke oder Vertreter:innen nationaler erziehungswissenschaftlicher Gesellschaften) diskutiert und entsprechend überarbeitet. Dabei musste bei der Bestimmung der Labels stets ein Kompromiss zwischen Präzision des Labels und seiner Textlänge gefunden werden, da zu lange Labels bei der Darstellung der interaktiven Visualisierungen in der Webapp Probleme bereitet hätten.

*Ähnlichkeit zwischen den Super- und Subthemen*: Für die Relationen innerhalb und zwischen den Themen beider Modelle wurden wie oben beschrieben die semantischen paarweisen Cosinus-Ähnlichkeiten zwischen allen $k = 300 + 50$ Themen bestimmt. Daraus resultierten insgesamt $n = 61.425$ paarweise Ähnlichkeiten. Bei insgesamt $n = 32.538$ paarweisen Ähnlichkeiten war $\cos_{(\theta)} < .01$, was einer annähernden Orthogonalität der Vektoren entspricht und daher die jeweilige Paarung für alle Analysen vernachlässigbar gemacht hat.

### *3.4 Clustering und Visualisierungen der Ergebnisse*

*Clustering*: Um nicht nur die Ergebnisse der Themenmodellierung per se darzustellen, sondern auch Trends und Relationen zwischen den bibliografischen Variablen (d. h. Autor:innen, Affiliationsländern, EERA-Netzwerken) und den Themen darzustellen, wurden jeweils die Mittelwerte der Dokument-Themen-Gewichte $\mu_{(\gamma)}$ und die dazugehörigen Bernoulli-Standardfehler $se_{(\gamma)}$ berechnet.

*Visualisierungen*: Die meisten genutzten Visualisierungen erfordern keine Erklärung, da es sich um gängige Verfahren wie Streudiagramme, Balkendiagramme oder Heatmaps handelt. Graphendarstellungen sind hingegen nicht gängig und werden daher im Folgenden kurz erläutert: Graphen können genutzt werden, um Beziehungen zwischen Datenpunkten und makroskopische Strukturen in Datensätzen darzustellen. Dafür werden die Ausprägungen kategorialer Variablen als sogenannte Nodes und die Beziehung zwischen ihnen als Edges repräsentiert. Als Nodes eignen sich zählbare Variablen wie EERA-Netzwerke oder Autor:innen. Der assoziierte (numerische) Wert der Nodes wird bestimmt durch beispielsweise die Anzahl der Beiträge eines EERA-Netzwerks/einer Autor:in. Die Werte der Edges ergeben sich aus dem Zusammenhang zwischen zwei Nodes, beispielsweise der Häufigkeit der Kooperation zwischen zwei EERA-Netzwerken oder zwei Autor:innen. Die makroskopische Struktur des Graphen wird durch einen Algorithmus bestimmt. Der weitverbreitete und hier verwendete kräftebasierte Fruchtermann-Reingold-Algorithmus (FR, Fruchtermann und Reingold 1991) ist intuitiv verständlich, robust und erlaubt eine klare und präzise Visualisierung von Datensätzen (Huang et al. 2024; Kobourov 2012; Stone 2022). Bei der FR-Methode wird jeder Node basierend auf ihrem Wert eine negative elektrische Ladung *U* zugewiesen. Jede Edge wird als Feder mit einer wertbasierten Federlänge und konstanten Federhärte *D* basierend auf dem Hookeschen Gesetz betrachtet. Die Nodes stossen sich durch ihre gleiche Ladung gegenseitig ab, werden aber durch die Federn zwischen ihnen zusammengezogen. Auf einer zweidimensionalen Ebene werden dadurch Nodes mit vielen starken Verbindungen (d. h. Edges) in die Mitte der Ebene gedrückt und Nodes mit wenigen (oder keinen) Verbindungen an den Rand der Ebene. Auf makroskopischer Ebene lassen sich dadurch die zentralen und peripheren Nodes bzw. Node-Cluster (d. h. beispielsweise EERA-Netzwerke) in einem Datensatz bestimmen. Auf bivariater Ebene lassen sich die nächsten Nachbarn einer Node identifizieren.

### *3.5 Entwicklung der Webapp EduTopics*

Die Ergebnisse der Analysen wurden visualisiert, mit dem R-Paket Shiny als interaktive Webapp programmiert und anschliessend zur Verfügung gestellt (url: https://dipf-lis.shinyapps.io/EduTopicsECER/. Zur genauen Beschreibung der Webapp siehe Christ et al., 2025). Eine App besteht aus insgesamt zehn Abschnitten: Eine Startseite, ein Abschnitt zur Nutzung der Webapp, einer zur Erläuterung der angewandten Methoden und jeweils

ein Abschnitt für die bibliografischen Variablen EERA-Netzwerk, Affiliationsland und Autor:innen. Für die Ergebnisse der Themenmodellierung ist ein Abschnitt zum Clustering der Super- und Subthemen mit manipulierbaren Graphen vorhanden sowie ein weiterer für die Ergebnisse der einzelnen Super- und Subthemen. Abgeschlossen wird die Webapp mit dem Abschnitt «Toolbox»: Diese ermöglicht den Nutzenden eine eigenständige Exploration des Korpus und der assoziierten Ergebnisse anhand der Variablen Jahr, EERA-Netzwerke, Affiliationsländer und Themen. Zudem wird ermöglicht, eigene Text(sammlungen) einzugeben und anschliessend deren Verteilung über die Super- und Subthemen anhand des ECER-Modells ausgeben zu lassen.

Jeder der einzelnen Abschnitte der Webapp beinhaltet eine Beschreibung der darstellbaren Ergebnisse und der Optionen für die interaktive Manipulation der Abbildungen und der Tabellen. So sind für alle Tabellen Filter- und Sortieroptionen integriert und für alle Abbildungen Zoom- und Exportmöglichkeiten vorhanden, wie sie zur Visualisierung in diesem Beitrag genutzt wurden.

## 4. Ergebnisse

Die Ergebnisse der Themenmodellierungen werden im Folgenden exemplarisch anhand medienpädagogisch relevanter Themen aus der Webapp EduTopics dargestellt. Hierfür werden die englischen Originallabels verwendet, damit auch eine direkte Nachvollziehbarkeit zur Webapp vorhanden ist. Bei der jeweils ersten Erwähnung eines Themas werden in Klammern deutsche Übersetzungen geliefert.

### *4.1 Das Superthema «Digital Technology, Media and ICT»*

Bei der Modellierung der Superthemen ergab sich ein dediziertes medienpädagogisches Thema zu *Digital Technology, Media and ICT* (*Digitale Technologien, Medien sowie Informations- und Kommunikationstechnik*). Für insgesamt $n$ = 863 Dokumente war dies das relevanteste Superthema, wodurch es auf Rang 14 von 50 landete. Eine Sortierung der Superthemen nach ihren durchschnittlichen Dokument-Themen-Gewichten $\mu_{(\gamma)}$ ergab Rang 26, woraus sich schliessen lässt, dass es eher kein querliegendes Thema ist, wie es die eher methodisch oder theoretisch paradigmatischen Themen *Theories, Concepts and Knowledge* (*Theorien, Konzepte und Wissen*; Rang 1), *Qualitative Research and Interviews* (*Qualitative Forschung und Interviews*; Rang 6) oder *Discourse Analysis and Power* (*Diskursanalyse und Macht(-verhältnisse)*; Rang 10) sind. Dies legt nahe, dass zwar medienpädagogische Forschungsthemen eine zentrale Rolle in der europäischen Bildungsforschung einnehmen, sie aber – zumindest für den gesamten Zeitraum von 1998 bis 2024 – stärker als explizites Thema und nicht als flächendeckendes Querschnittsthema wahrgenommen werden.

Zu den Begriffen mit hohen β-Gewichten für dieses Thema zählten *Digital, Technology, Media, ICT, Online, Education, Information, Tool* etc. (siehe Abbildung 1).

Show 10 entries Search:

Top 25 Terms by Term-Topic-Weight β

| | Word | β |
|---|---|---|
| 1 | Digital | 0.068 |
| 2 | Technology | 0.064 |
| 3 | Media | 0.034 |
| 4 | ICT | 0.033 |
| 5 | Online | 0.03 |
| 6 | Education | 0.022 |
| 7 | Information | 0.02 |
| 8 | Tool | 0.017 |
| 9 | Computer | 0.016 |
| 10 | Learning/Learner | 0.016 |

Showing 1 to 10 of 25 entries Previous 1 2 3 Next

**Abb. 1:** Top 10 Begriffe nach Term-Themen-Gewicht β für das Superthema Digital Technology, Media and ICT (Digitale Technologien, Medien und Informations- und Kommunikationstechnik).

Für das Superthema zeigte sich eine zeitliche Stabilität mit $\mu(\gamma) \in [.014; .021]$ mit Ausreissern für die Jahre 1998 ($\mu_{(\gamma)} = .005$), 2000 ($\mu_{(\gamma)} = .026$) und 2021/2022 ($\mu_{(\gamma)} = .023/.028$; siehe Abbildung 2). Die Anstiege für die Jahre 2021/2022 deuten auf die erhöhte Relevanz des Themas im Zuge der Corona-Pandemie hin, wobei auch die nachfolgenden Jahre einen überdurchschnittlichen Wert aufweisen. Dies könnten Anzeichen für eine generell steigende Relevanz des Superthemas auf europäischer Ebene sein, jedoch sind für eine diesbezügliche Einschätzung aktuell zu wenige Datenpunkte vorhanden.

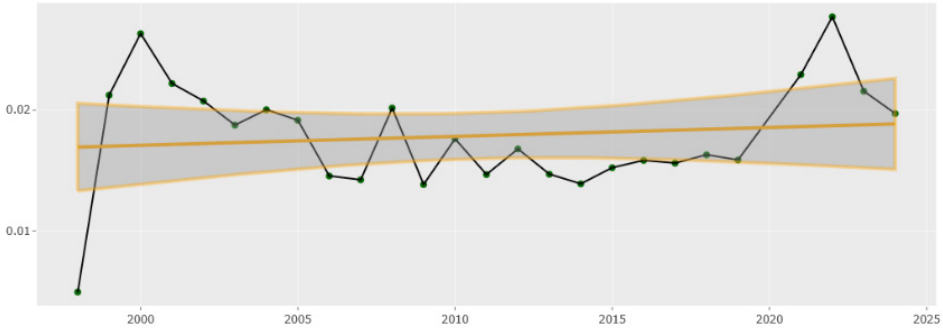


**Abb. 2:** Durchschnittliche Dokument-Themen-Gewichte $\mu_{(\gamma)}$ des Superthemas Digital Technology, Media and ICT (Digitale Technologien, Medien und Informations- und Kommunikationstechnik) von 1998 bis 2024.

Für insgesamt *n* = 4 EERA-Netzwerke ergaben sich durchschnittliche Dokument-Themen-Gewichte $\mu_{(\gamma)}$ oberhalb des Erwartungswerts von $E_{(\gamma)} = .02$ ($\mu_{(\gamma)} = 1/50$): *NW 16 – ICT*, *NW 06 – Open Learning*, *NW 12 – Open Research* in Education und *NW 28 – Sociologies of Education*, wobei der Abfall zwischen NW 12 und NW 28 von $\mu_{(\gamma)} = .11$ auf $\mu_{(\gamma)} = .03$ substanziell und signifikant war. Dies wird insbesondere dann ersichtlich, wenn die durchschnittlichen Netzwerk-Themen-Gewichte $\mu_{(\gamma)}$ im Vergleich zum Erwartungswert $E_{(\gamma)} = .02$ betrachtet werden. Sie liegen für NW 16 bei ca. 8 x $E_{(\gamma)}$, NW 06 ca. 7 x $E_{(\gamma)}$, NW 12 ca. 5.5 x $E_{(\gamma)}$ und NW 28 ca. 1.5 x $E_{(\gamma)}$. Für alle weiteren Netzwerke ergaben sich $\mu_{(\gamma)} < .02$.

### *4.2 Verortung des Superthemas bei der ECER*

Die Betrachtung der Verortung des Superthemas in der Makrostruktur der Themen der ECER seit 1998 anhand der semantischen paarweisen Ähnlichkeiten der Super- und Subthemen zeigt, dass *Digital Technologies, Media and ICT* eher an der Peripherie der thematischen zweidimensionalen Karte verortet ist und lediglich über das Subthema *Online and Distance (Blended) Learning (Online, Distanz und integriertes (blended) Lernen)* mit dem breiteren Themencluster der ECER verbunden ist (siehe Abbildung 3).

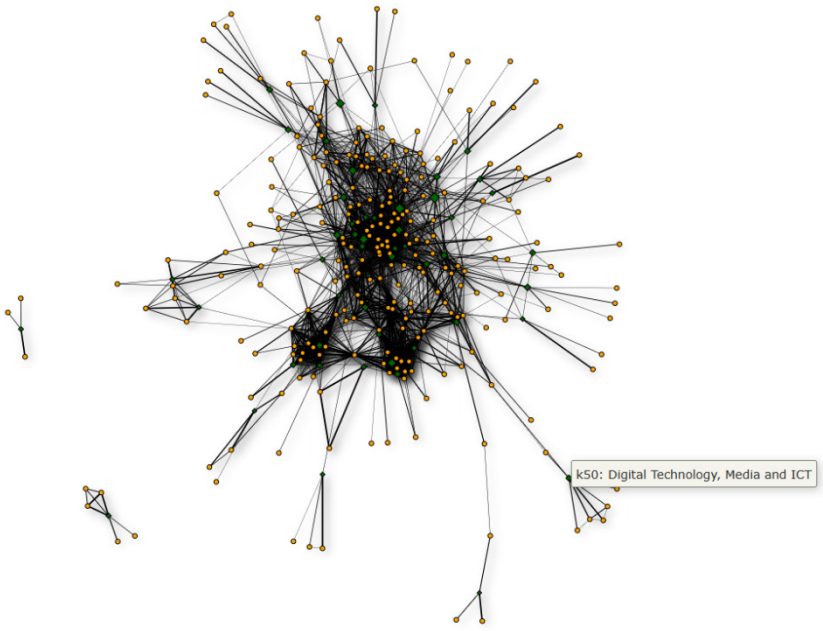


**Abb. 3:** Zweidimensionale Graphenkartierung der Super- und Subthemen der Beiträge der ECER seit 1998. Grüne Diamanten stellen Superthemen, orange Kreise Subthemen dar. Die Dicke der Verbindungen der Nodes repräsentiert die paarweise semantische Ähnlichkeit $\cos_{(\theta)} > .7$. Hervorgehoben ist das Superthema Digital Technologies, Media and ICT (Digitale Technologien, Medien und Informations- und Kommunikationstechnik).

### *4.3 Medienpädagogische Subthemen*

Die *n* = 7 semantisch ähnlichsten Subthemen zu diesem Superthemen wiesen alle medienpädagogische Bezüge auf: mit dem eher allgemeinen Thema *Digital Technologies, Tools and Platforms* (*Digitale Technologien, Werkzeuge und Plattformen*; $\cos_{(\theta)} = .922$) auf Rang 1 bis *Media Literacy* (*Medienkompetenz und Literacy*; $\cos_{(\theta)} = .745$) auf Rang 7. Von Rang 7 auf Rang 8 mit *Information, Libraries and Internet* (*Informationen, Bibliotheken und Internet*; $\cos_{(\theta)} = .643$) ergab sich sowohl ein substanzieller und signifikanter Abfall der semantischen Ähnlichkeit ($\Delta\cos_{(\theta)} = .1$) als auch eine Verschiebung der Inhalte des Subthemas von spezifisch auf medienpädagogische Fragestellungen bezogenen Themen hin zu querliegenden Themen, die für viele verschiedene Forschungsbereiche relevant sind, wie das Subthema *Proposals for Educational Research* (*Forschungsvorhaben und -anträge*; $\cos_{(\theta)} = .591$) oder *Designing Learning Environments* (*Gestaltung von Lernumgebungen*; $\cos_{(\theta)} = .574$; siehe Abbildung 4).

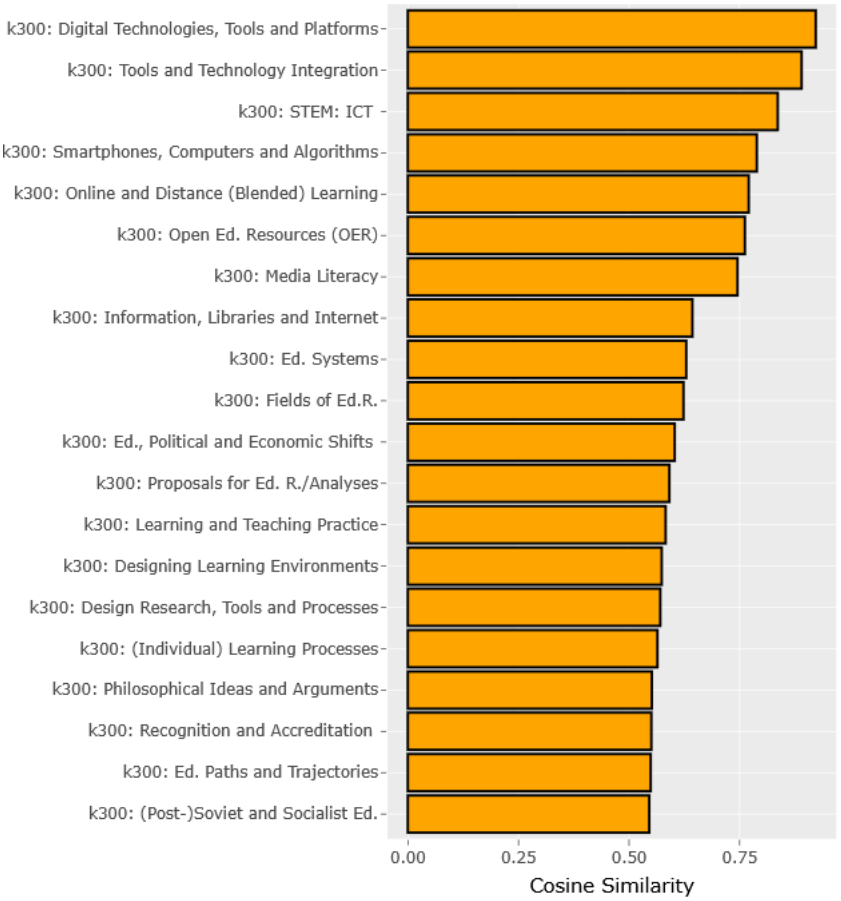


**Abb. 4:** Balkendiagramm der Top 20 Subthemen des Superthemas Digital Technology, Media and ICT (Digitale Technologien, Medien und Informations- und Kommunikationstechnik) basierend auf der semantischen Ähnlichkeit $\cos_{(\theta)}$.

Für die so identifizierten $n$ = 7 medienpädagogisch relevanten Subthemen ergaben sich seit 1998 gänzlich unterschiedliche zeitliche Entwicklungen: Beispielsweise zeigte sich, dass die Relevanz von *STEM: ICT* (*MINT: Informatikunterricht*) seit Beginn der ECER kontinuierlich gesunken ist von fast 2–3-mal dem Erwartungswert $E_{(\gamma)} = 1/300$ von 1999–2005 unter den Erwartungswert ab 2017. Gleichzeitig ist die Relevanz von *Digital Technologies, Tools and Platforms* kontinuierlich gestiegen und war ab 2017 höher als diejenige von *STEM: ICT* – abermals mit Corona- assoziiertem Maximum (siehe Abbildung 5). Für 2023 und 2024 lag der Wert für *Digital Technologies, Tools and Platforms* weiterhin deutlich über dem Erwartungswert $E_{(\gamma)} = 1/300$ mit $\mu_{(\gamma)} > .005$, wodurch das Subthema für diese Jahre unter allen 300 Subthemen auf Rang 37 kommt. Sollte sich in Zukunft erweisen, dass diese Werte weiter ansteigen oder zumindest stabil bleiben, zeigt dies den Einzug digitalisierungsbezogener (und dadurch medienpädagogischer) Fragestellungen in verschiedene weitere Bereiche der Bildungsforschung.

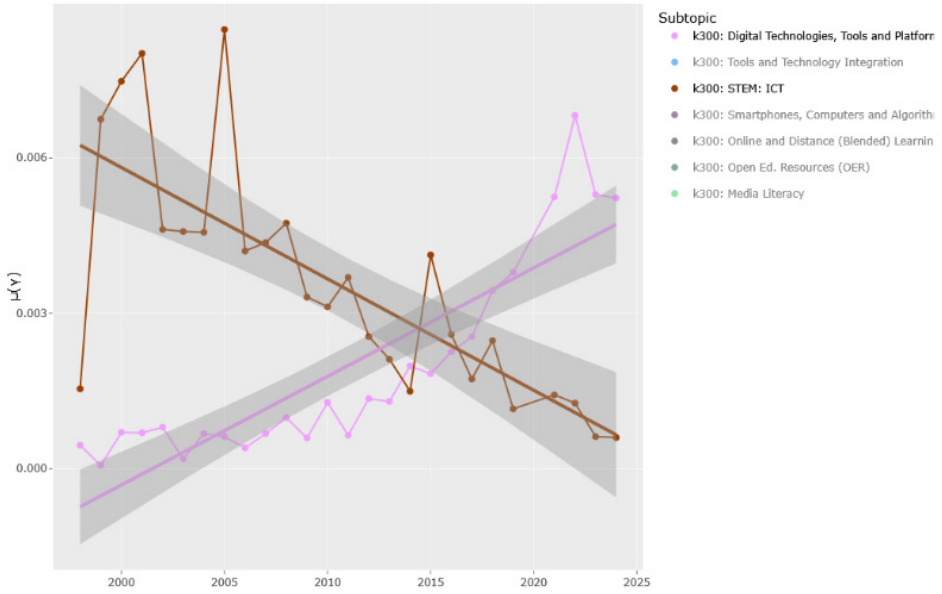


**Abb. 5:** Zeitliche Verläufe für die Subthemen Digital Technologies, Tools and Platforms (Digitale Technologien, Werkzeuge und Plattformen) und STEM: ICT (MINT: Informatikunterricht) von 1998 bis 2024.

Für die Subthemen Tools and Technology Integration ((Digitale) Werkzeuge und Technologieintegration; $\mu_{(\gamma)} \in [.001; .004]$), Smartphones, Computers and Algorithms (Smartphones, Computer und Algorithmen; $\mu_{(\gamma)} \in [.00; .002]$), Open Educational Resources (Freie Lern- und Lehrmaterialien; $\mu_{(\gamma)} \in [.001; .004]$) und Media Literacy (Medienkompetenz und Literacy; $\mu_{(\gamma)} \in [.00; .002]$) konnten keine nennenswerten zeitlichen Trends bezüglich der Relevanz festgestellt werden. Hier lagen auch die durchschnittlichen Dokument-Themen-Gewichte pro Jahr meist unter dem Erwartungswert $E_{(\gamma)} = 1/300$.

Anders hingegen verhielt es sich für das Subthema *Online and Distance (Blended) Learning*, welches von 1999 bis 2002 einen hohen Wert aufwies, welcher danach deutlich fiel. Für 2021 und 2022 ergab sich abermals eine substanzielle Steigerung von $\mu_{(\gamma, 2019)} = .002$ auf $\mu_{(\gamma, 2021)} = .004$ und $\mu_{(\gamma, 2022)} = .005$ (siehe Abbildung 6).

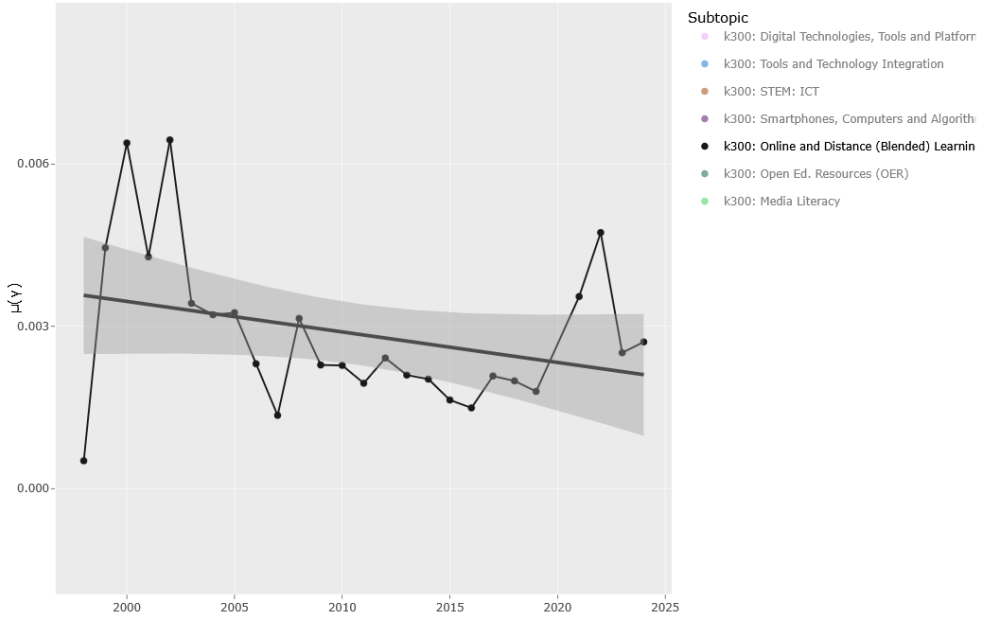


**Abb. 6:** Zeitlicher Verlauf des Subthemas Online and Distance (Blended) Learning (Online, Distanz und integriertes (blended) Lernen) von 1998 bis 2024.

Final werden die Entwicklungen der $n = 7$ medienpädagogisch relevanten Subthemen seit 1998 betrachtet. Zudem wurden – als weiteres Netzwerk – die Beiträge der Emerging Researcher Conference (ERC) hinzugefügt, um auch die Entwicklung medienpädagogischer Inhalte in der Forschung der Wissenschaftler:innen in Qualifizierungsphasen genauer zu betrachten. Um die Vergleichbarkeit sicherzustellen, wurden ausserdem die Entwicklungen der o. g. Subthemen auch bei allen Beiträgen aller verbliebenen Netzwerke betrachtet (siehe Abbildung 7). Es zeigte sich, dass die meisten Subthemen über die Jahre hinweg (wenn auch nur minimal) gestiegen sind (siehe z. B. *Media Literacy; Smartphones, Computers and Algorithms; Tools and Technology Integration*). Insbesondere durch den Anstieg bei der ERC und bei den Beiträgen der restlichen Netzwerke (Zeile 1 und Zeile 5 in Abb. 7) deutet dies darauf hin, dass Aspekte der Digitalisierung und – daraus resultierend – auch medienpädagogische Subthemen nicht nur generell an gesellschaftlicher Relevanz gewonnen haben, sondern auch in den letzten Jahren sukzessive mehr Einzug in die Forschung jenseits der dediziert medienpädagogischen Netzwerke gefunden haben. Somit konnten sie sich von einem – wie oben beschrieben – expliziten Thema zunehmend zu einem Querschnittsthema entwickeln.

Für *NW 06 – Open Learning* ergaben sich Anstiege für *Digital Technologies, Tools and Platforms* und *Media Literacy*, während *Online and Distance (Blended)* abfiel und die anderen 4 Subthemen tendenziell konstant blieben.

Für *NW 16 – ICT* hingegen stieg nur *Digital Technology, Tools and Platforms* über die Jahre an, *STEM: ICT* fiel ab. Für alle weiteren Subthemen ergaben sich keine substanziellen Veränderungen.

Die Zugehörigkeit von *NW 12 – Open Research* in Education zu dem medienpädagogischen Superthema basierte nahezu ausschliesslich auf dem Subthema *Open Educational Resources (OER)*, da für die anderen $n = 6$ Subthemen $\mu_{(\gamma, \text{Jahr})} \approx 0$ war (siehe Abbildung 7).

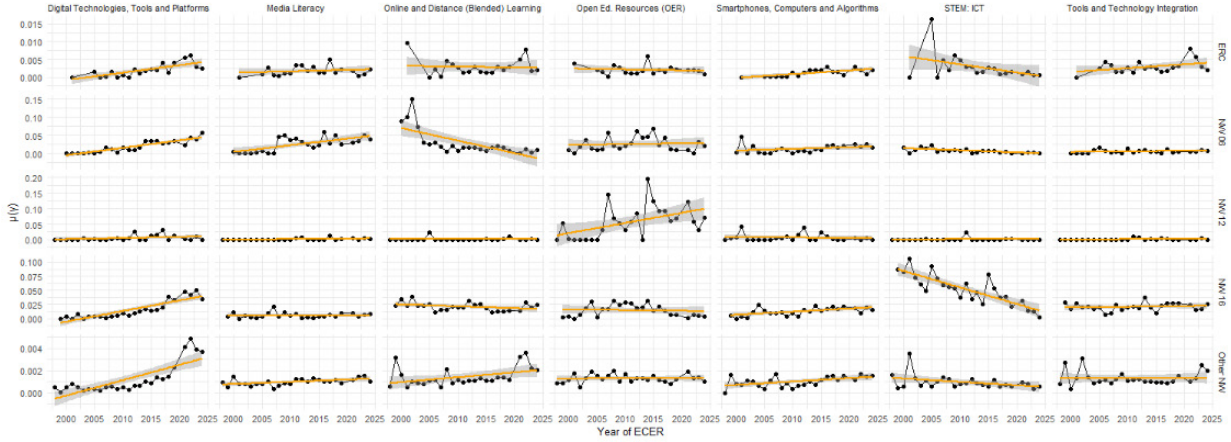


**Abb. 7:** Zeitliche Verläufe von 1998 bis 2024 medienpädagogischer Subthemen (Spalten) für die drei EERA-Netzwerke NW 06 – Open Learning, NW 12 – Open Research in Education, NW 16 – ICT in Education and Training, die Emerging Researcher Conference (ERC) und die Beiträge aller weiteren Netzwerke. Relative Skalierung der y-Achsen anhand des Maximums der mittleren Dokument-Themen-Gewichts $\mu_{(\gamma)}$ der jeweiligen Zeile.

## 5. Diskussion

Dieser Artikel verfolgte zwei Ziele: (a) die Darstellung des Konzepts der Themenmodellierung und der Webapp EduTopics und (b) die exemplarische Darstellung der Entwicklung medienpädagogisch relevanter Themen anhand der Beiträge zur ECER seit 1998. Es wurde gezeigt, wie die Methoden aus (a) die Erreichung des zweiten Ziels ermöglichen konnten und wie dabei auf qualitative Methoden zur Beurteilung der Ergebnisse zurückgegriffen wird. Dabei sind qualitative Verfahren nicht nur für die Beurteilung der Ergebnisse von Themenmodellierung relevant, sondern insbesondere auch in Bezug auf die Nachnutzbarkeit der Ergebnisse der Webapp. So wurde durch die Datenbereinigung, die verschiedenen Clustering-Verfahren und die interaktive Bereitstellung der Ergebnisse Nutzenden ein interaktiver und niedrigschwelliger Zugang zu den aggregierten und disaggregierten Daten ermöglicht, der so vorher nicht vorhanden war und auch nicht ohne Weiteres möglich gewesen wäre. So können die Ergebnisse in EduTopics für vertiefende

qualitative Analysen genutzt werden, auch von Personen, die keinerlei Kompetenzen in der Bereinigung, Modellierung und Visualisierung komplexer und grosser Datensätze aufweisen. Dazu zählen nicht nur Analysen thematischer Schwerpunkte, Verläufe oder Themen-Hierarchien, sondern auch Analysen zu Länder- und Autor:innen-Kooperationen.

Bezüglich der Entwicklung medienpädagogischer Inhalte hat sich für die ECER gezeigt, dass die medienpädagogischen Inhalte und Fragestellungen generell mehr Einzug in alle Forschungsbereiche der ECER gefunden haben. Zurückzuführen ist diese Entwicklung auf die generell zunehmende Digitalisierung der Gesellschaft, da gerade in den letzten Jahren Digitale Technologien, Lernen, Mobiltelefone und Algorithmen kein Nischenthema mehr sind wie zu Beginn des Jahrhunderts, sondern zum «cultural modus operandi of modern civilization» (Aguaded et al. 2022, 9) geworden sind. Insbesondere zeigt sich auch der Effekt der Covid-Pandemie eindeutig durch lokale Maxima für Themen um Distanzunterricht oder generell zu digitalen Technologien und deren Integration in Bildungsprozesse. Weiterhin hat das Thema *ICT* über die Jahre hinweg kontinuierlich abgenommen – selbst für das dedizierte ICT-Netzwerk. Stattdessen stieg das Subthema Medienkompetenz und Literacy sukzessive an. Hier lässt sich vermuten, dass der Fokus medienpädagogischer Forschung sich in den letzten Jahrzehnten verschoben hat – von handlungsbezogenen auf kritische Medienkompetenzen, wie es auch bereits schon 2009 von Hobbs und Jensen antizipiert wurde (2009). Für zukünftige Modellierungen sollte besonders interessant werden, welche Rolle Themen zu KI und LLMs einnehmen (ein Thema, welches sich bis 2024 noch nicht herausgebildet hat), wie diese sich auf die verschiedenen Netzwerke verteilen und in Kombination mit welchen weiteren Themen sie besonders prävalent auftreten.

### *5.1 Limitationen*

Eine der zentralen Limitationen der Ergebnisse der Modellierungen stellt die Datengrundlage dar. Die Ergebnisse basieren auf dem Clustering und der Modellierung einer einzelnen europäischen Konferenz, wodurch die Ergebnisse stark beeinflusst werden durch die Themensetzung der Konferenz und die theoretischen, methodologischen und ggf. auch nationalen Paradigmen der Teilnehmenden. Dem lässt sich jedoch entgegenhalten, dass trotz der potenziell eingeschränkten Generalisierbarkeit der Ergebnisse EduTopics einen bisher einzigartigen Einblick in die Themenschwerpunkte und Themenentwicklung der europäischen Bildungsforschung ermöglicht.

Weiterhin basieren die Ergebnisse von EduTopics lediglich auf den Zusammenfassungen der Beiträge. Zwar verfügen die Zusammenfassungen über eine ausreichende Textlänge für reliable und valide Ergebnisse der Themenmodellierung. Jedoch würde eine Modellierung von Volltexten einen tieferen Einblick in die thematische Entwicklung ermöglichen – auch wenn dies bei Konferenzbeiträgen aufgrund ihres Formats nicht möglich ist, sondern sich eher für die Modellierung von Forschungsarbeiten anbietet.

Ebenso haben sich strukturelle Änderungen bei der Einreichung zur ECER ergeben, die geeignet sind, die Modellierung zu erschweren: So wurden erst ab 2008 bei der Einreichung dedizierte Methoden- und Ergebnissteile gefordert, wodurch eine separate Analyse der theoretischen und methodologischen Paradigmen daran scheitert, dass diese Textteile vor 2008 nicht ohne Weiteres in den Beiträgen identifizierbar sind. Ebenso haben sich Erfordernisse für die Angabe von Referenzen über die Jahrzehnte verändert, wodurch diese nicht genutzt werden können, um weitere Informationen zu extrahieren. Dies liegt insbesondere auch an der Entscheidung, die historische Dimension der ECER seit 1998 abzubilden. Hätte der Fokus stattdessen auf der Entwicklung aktueller Trends etwa der letzten 10 Jahre gelegen, so wäre eine separate Modellierung von Inhalten, Methoden und Referenzen gegebenenfalls möglich gewesen.

Je nach Zielsetzung der Analyse könnte auch die Modellierung der ECER-Beiträge über alle EERA-Netzwerke und Affiliationsländer hinweg als Limitation betrachtet werden, da aufgrund der geografischen Streuung eher Themen extrahiert werden, die eine grössere Verbreitung innerhalb der Beiträge aufweisen, wodurch thematische Verschiebungen innerhalb einzelner Netzwerke oder Länder eventuell nicht identifiziert werden können. Eine solche Analyse würde jedoch die Extraktion netzwerk- und länderübergreifender Themen erschweren und gleichzeitig mit n = 38 EERA-Netzwerken in mehr als 70 Ländern einen erheblichen Mehraufwand bedeuten, ohne dass der tatsächliche Mehrwert solcher Modellierungen absehbar wäre. Jenseits dessen würde eine länderspezifische Modellierung aufgrund der Substichprobengrössen für viele Länder weder reliable noch valide Ergebnisse liefern, sofern sie wegen kleiner Fallzahlen überhaupt möglich wäre.

Das qualitative Labelling muss, wie alle anderen qualitativen Prozesse, stets unter der Prämisse betrachtet werden, dass bei diesem Prozess gegebenenfalls Biases oder fehlendes Wissen über bestimmte Forschungsbereiche eine Rolle gespielt haben können. Es wurde versucht, diese Biases zu reduzieren, indem mit Vertreter:innen der Communities – wie oben beschrieben – über die Labels diskutiert wurde. Im Gegensatz zu klassischen Publikationsformaten ermöglicht jedoch die Webapp allen Nutzer:innen, eigene Interpretationen basierend auf den indikativen Termen, Dokumenten, Ähnlichkeiten zu anderen Themen und Autor:innen-Gewichten durchzuführen und gegebenenfalls zu anderen Labels zu kommen. Auch war die Bestimmung der Labels eingeschränkt durch die Textlänge, da zu lange Labels zu Darstellungsproblemen in den interaktiven Visualisierungen der Ergebnisse in EduTopics geführt hätten.

### *5.2 Implikationen für die Weiterentwicklung der App*

An den beiden soeben beschriebenen Limitationen der Ergebnisse wird aktuell gearbeitet, indem Treffen mit weiteren Vertreter:innen von Fachgesellschaften stattfinden, um das Korpus mit Daten weiterer internationaler und nationaler Konferenzen anzureichern. Dabei werden neue Herausforderungen bezüglich der Modellierung und des Labellings

entstehen wie der Umgang mit verschiedenen Sprachen, verschiedenen Substichprobengrössen und nationalen Forschungsparadigmen. Ebenso stellt sich bei den jährlichen Updates der Webapp die Frage, in welchem Turnus die Modelle überprüft und ggf. auch neu durchgeführt werden müssen, damit zwar eine Stabilität der Modelle über die Jahre hinweg vorhanden ist, jedoch neue, bisher noch nicht vorhandene Themen wie zu Large Language Models und KI auch dann identifiziert und in die Ergebnisse der App integriert werden, sobald sie eine gewisse Prävalenz erreicht haben.

### *5.3 Fazit*

Der vorliegende Beitrag gibt einen Einblick in das computerlinguistische Verfahren der Themenmodellierung. Er verdeutlicht das enthaltene Potenzial für die Analyse und Kartierung grosser Literatursammlungen und unterstreicht sowohl die eminente Rolle qualitativer Analysen als auch, welche neuen Zugänge durch solche Verfahren für qualitative Forschung ermöglicht werden. Die Charakteristika des Korpus erlauben eine Betrachtung unterschiedlicher medienpädagogischer Forschungsgegenstände auf europäischer Ebene, sind in ihrer Granularität angesichts der Grösse und Heterogenität des Korpus jedoch limitiert. Ein präzisere Anwendung der Themenmodellierung auf medienpädagogische Korpora würde zwar eine differenziertere Betrachtung ermöglichen, allerdings die Analyse der Möglichkeit berauben, die zunehmende Durchsetzung aller Bereiche der Bildungsforschung mit medienpädagogischen Inhalten nachzuvollziehen. Dennoch wäre ein solches Unterfangen gewinnbringend, um detailliertere und differenzierte Informationen über die paradigmatische Entwicklung der Medienpädagogik und ihres Diskurses zu erhalten. Der Umfang eines Korpus für eine solche Fragestellung würde auch in einem solchen Fall die Anwendung von computerlinguistischen maschinellen Lernverfahren wie der Themenmodellierung benötigen.

**Literaturverzeichnis**

Aguaded, Igacio, Sabina Civila, und Arantxa Vizcaíno-Verdú. (2022). «Paradigm changes and new challenges for media education: Review and science mapping (2000–2021)». *Profesional de la Información, 31*(6). https://doi.org/10.3145/epi.2022.nov.06.

Aletras, Nikolaos, und Mark Stevenson. 2014. «Measuring the similarity between automatically generated topics». In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers:* 22–27. Göteborg. https://aclanthology.org/E14-4005.pdf.

Arun, Rajkumar, Vishnu Suresh, Veni Madhavan, und Murthy Narasimha. 2010. «On finding the natural number of topics with latent Dirichlet allocation: Some observations». In *Advances in Knowledge Discovery and Data Mining*. PAKDD 2010. Lecture Notes in Computer Science, vol. 6118. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-13657-3_43.

Bittermann, André, und Andreas Fischer. 2018. «How to identify hot topics in psychology using topic modeling». *Zeitschrift für Psychologie* 226 (1): 3–13. https://doi.org/10.1027/2151-2604/a000318.

Blei, David. M., und John D. Lafferty. 2009. «Topic Models». In *Text Mining*, herausgegeben von Ashok Srivastava und Sahami Mehran, 101–24. New York: Chapman and Hall/CRC. https://doi.org/10.1201/9781420059458-12.

Blei, David. M., Andrew. Y. Ng, und Michael I. Jordan. 2003. «Latent Dirichlet allocation». *Journal of machine Learning research* 3 (Jan): 993–1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

Cao, Juan, Tian Xia, Jintao Li, Yongdong Zhang, und Sheng Tang. 2009. «A density-based method for adaptive LDA model selection». *Neurocomputing* 72 (7–9): 1775–81. https://doi.org/10.1016/j.neucom.2008.06.011.

Christ, Alexander, Jens Röschlein, und Christoph Schindler. 2025. «EduTopics: ECER – A webapp for interactive visualisation and exploration of ECER contributions since for 1998». *European Educational Research Journal*. https://doi.org/10.1177/14749041251405570.

Christ, Alexander, Kathrin Smolarczyk, und Stephan Kröner. 2024. «Schwerpunktthemen der quantitativ-empirischen Forschung mit Bezug zur Digitalisierung in der kulturellen Bildung: Eine kartierende Forschungssynthese». *Zeitschrift für Erziehungswissenschaft* 27 (2): 351–92. https://doi.org/10.1007/s11618-023-01210-7.

Christ, Alexander, Marcus Penthin, und Stephan Kröner. 2021. «Big data and digital aesthetic, arts, and cultural education: Hot spots of current quantitative research». *Social Science Computer Review* 39 (5): 821–43. https://doi.org/10.1177/0894439319888455.

De Waal, Alta, und Etienne Barnard. 2008. «Evaluating topic models with stability». Cape Town: PRASA 2008. http://hdl.handle.net/10204/3016.

Deveaud, Romain, Eric SanJuan, und Patrice Bellot. 2014. «Accurate and effective latent concept modeling for ad hoc information retrieval». *Document numérique* 17 (1): 61–84. https://doi.org/10.3166/DN.17.1.61-–84.

Fruchterman, Thomas M., und Reingold, Edward. M. 1991. «Graph drawing by force-directed placement». *Software: Practice and experience* 21 (11): 1129-64. https://doi.org/10.1002/spe.4380211102.

Gandomi, Amir, und Murtaza Haider. 2015. «Beyond the hype: Big data concepts, methods, and analytics». *International journal of information management* 35 (2): 137–44. https://doi.org/10.1016/j.ijinfomgt.2014.10.007.

Griffiths, Thomas. L., und Mark Steyvers. 2004. «Finding Scientific Topics». *Proceedings of the National academy of Sciences*, 101(suppl_1): 5228–35. https://doi.org/10.1073/pnas.0307752101.

He, Qj, Bi Chen, Jian Pei, Baojun Qiu, Prasenjit Mitra, und Lee Giles. 2009. «Detecting topic evolution in scientific literature: how can citations help?». In *Proceedings of the 18th ACM conference on Information and knowledge management*: 957–66. Hong Kong: ACM. https://doi.org/10.1145/1645953.1646076.

Heidenreich, Tobias, Fabienne Lind, Jakob-Moritz Eberl, und Hajo G. Boomgaarden. 2019. «Media framing dynamics of the ‹European refugee crisis›: A comparative topic modelling approach». *Journal of Refugee Studies* 32 (Special_Issue_1): i172-i182. https://doi.org/10.1093/jrs/fez025.

Hobbs, Renee, und Amy Petersen Jensen. (2009). «The past, present, and future of media literacy education». *Journal of media literacy education* 1 (1). https://doi.org/10.23860/jmle-1-1-1.

Huang, Rui, Hailong Gai, Rong Jing, und Wan Fucheng. 2024. «Word Spectral Visualization Base on Fruchterman-Reingold Optimized ForceAtlas 2». *Journal of Electrical Systems* 20: 7–15. https://doi.org/10.52783/jes.1086.

Huang, Cui, Chao Yang, Shutao Wang, Wei Wu, Jun Su, und Chuying Liang. 2020. «Evolution of topics in education research: A systematic review using bibliometric analysis». *Educational Review* 72 (3): 281–97. https://doi.org/10.1080/00131911.2019.1566212.

Jacobs, Thomas, und Robin Tschötschel. 2019. «Topic models meet discourse analysis: a quantitative tool for a qualitative approach». *International Journal of Social Research Methodology* 22 (5): 469–85. https://doi.org/10.1080/13645579.2019.1576317.

Kobourov, Steohen G. 2012. «Spring embedders and force directed graph drawing algorithms». *arXiv preprint arXiv:1201.3011*. https://doi.org/10.48550/arXiv.1201.3011.

Mimno, David, Hanna M. Wallach, Edmund Talley, Miriam Leenders, und Andrew McCallum. 2011. «Optimizing semantic coherence in topic models». In *Proceedings of the 2011 conference on empirical methods in natural language processing*: 262–72. Edinburgh. https://aclanthology.org/D11-1024.Pdf.

Newman, David, Arthur Asuncion, Padhraic Smyth, und Max Welling. 2009. «Distributed algorithms for topic models». *Journal of Machine Learning Research* 10 (8): 1801–28. https://www.jmlr.org/papers/volume10/newman09a/newman09a.pdf.

O'Mara-Eves, Alison, James Thomas, John McNaught, Makoto Miwa, und Sophia Ananiadou. 2015. «Using text mining for study identification in systematic reviews: a systematic review of current approaches». *Systematic reviews* 4: 1–22. https://doi.org/10.1186/2046-4053-4-5.

Ramage, Daniel, David Hall, Ramesh Nallapati, und Christopher D. Manning. 2009. «Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora». In *Proceedings of the 2009 conference on empirical methods in natural language processing*: 248–56. https://aclanthology.org/D09-1026.pdf.

Silge, Julia, und David Robinson. 2017. *Text mining with R: A tidy approach*. Boston (MA): O'Reilly Media. https://www.tidytextmining.com/.

Steyvers, Mark, und Thomas L. Griffiths. 2007. «Probabilistic topic models». *Handbook of Latent Semantic Analysis*, 427: 424–40.

Stone, Matthew R. 2022. «Analysis of graph layout algorithms for use in command and control network graphs». *Theses and Dissertations*. 5546. https://scholar.afit.edu/etd/5546.

Wang, Xiang, Kai Zhang, Xiaoming Jin, und Dou Shen. 2009. «Mining common topics from multiple asynchronous text streams». In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*: 192–201. https://doi.org/10.1145/1498759.1498826.