

Quo Vadis Fachbereiche und Schulen der DHd: Netzwerkanalyse der DHd Abstracts 2014-2023

Haider, Thomas Nikolaus

thomas.haider@uni-passau.de
Universität Passau, Deutschland
ORCID: 0000-0003-1522-4026

Gassner, Sebastian

sebastian.gassner@uni-passau.de
Universität Passau, Deutschland

Rehbein, Malte

malte.rehbein@uni-passau.de
Universität Passau, Deutschland
ORCID: 0000-0002-3252-0604

Einleitung

Die Digital Humanities (DH) sind ein wachsendes Forschungsfeld (Tang et al., 2017), und nach 10 Jahren DHd fragt die Konferenz in Passau 2024 nach ‚Quo Vadis DH?‘. Wir möchten mit diesem Beitrag einen Überblick der Fächerlandschaft und des Zitierverhaltens der deutschsprachigen DH Community gewinnen. Dies verfolgen wir mit Methoden der Netzwerkanalyse, wobei wir uns insbesondere Ko-Keywords, Ko-Autorschaft, und Ko-Zitation ansehen. Zum Zitierverhalten werfen wir außerdem einen Blick auf die Länge der Bibliographien und das Alter der Zitationen.

Während bestehende bibliometrische Studien sich mit verwandten Bereichen wie der Computerlinguistik (Bollmann & Elliot, 2020), dem Management (Wieczorek et al., 2021) und den Digital Humanities international (Tang et al., 2017) sowie mit spezifischen Analysen einzelner DHd-Konferenzen (Henny-Kramer & Sahle, 2018; Hoenen, 2019) beschäftigt haben, fehlt ein umfassender Überblick der Zitationslandschaft der DHd-Konferenz. Diese Lücke motiviert unsere Studie, mit dem Ziel eine Bewertung und einen Einblick in die Zitationsdynamik innerhalb dieses wissenschaftlichen Forums zu geben.

Wir verfolgen folgende Analysen:

Ko-Keywords: Durch das Extrahieren und Analysieren von Keywords aus wissenschaftlichen Arbeiten können wir wiederkehrende Themen, gemeinsame Begriffe und Gruppen von verwandten Problemen identifizieren. Die Beziehungen zwischen Keywords sollten Einblicke in die Teilgebiete der DH bieten.

Ko-Autorschaft: Die Zusammenarbeit ist das Herzstück des wissenschaftlichen Fortschritts, und Koautorenschaftsnetzwerke sind ein Instrument zur Untersuchung der Interaktionen zwischen Forscher:innen. Die Analyse von Koautorenschaftsmustern kann helfen, einflussreiche Forschungsgruppen, intellektuelle Zentren und Communities mit gemeinsamen Forschungsinteressen zu identifizieren. Die Aufdeckung dieser Netzwerke ermöglicht es uns, unterschiedliche Denkrichtungen, ihre geografische Verteilung und möglicherweise ihren Einfluss auf den wissenschaftlichen Fortschritt zu ermitteln.

Zitierverhalten und Ko-Zitation: In wissenschaftlichen Arbeiten werden typischerweise frühere Arbeiten zitiert, um die intellektuelle Abstammung von Ideen zu belegen und grundlegende Beiträge zu würdigen. Durch die Erstellung von Kozytationsnetzwerken können wir etwa einflussreiche Arbeiten aufspüren und nachverfolgen, und Einsicht in die Verbindungen zwischen verschiedenen Subdisziplinen bekommen.

Korpus

Als Hauptkorpus verwenden wir die DHd Abstracts von 2016 bis 2023, welche komplett in TEI XML vorliegen (<https://github.com/DHd-Verband>). Die Volltexte der Jahrgänge 2014 und 2015 liegen nur in .pdf vor. Die Information zu Keywords und Bibliographie ist in diesen frühen Jahrgängen inkonsistent und wird daher nicht verwendet. Allerdings verwenden wir für Ko-Autorschaft die Überblicks-XML (die alle Beiträge der jeweiligen Jahrgänge 2014 und 2015 vereint).

Methodik

Die Hauptaspekte dieses Papiers (Ko-Keywords, Ko-Autorschaft, Ko-Zitation) modellieren wir mit Netzwerkanalyse. Die Analysen bewegen sich auf Dokument-Ebene: Pro DHd-Abstract verwenden wir eine Liste mit Keywords, eine Liste mit Autor:innen, und eine Liste mit Literaturreferenzen.

Die Größe der Knoten ist durchgängig durch Betweenness Centrality bestimmt (wie wichtig ist ein Knoten für das gesamte Netzwerk), und die Kantenstärke ist durch die absolute Frequenz des gemeinsamen Vorkommens von zwei Knoten bestimmt (zum Beispiel zeigt eine Kante im Ko-Autorenschaftsnetzwerk an wie oft zwei Autor:innen auf der DHd zusammen publiziert haben).

Für das Layout der Netzwerke verwenden wir Force Atlas 2, und die (Cluster)farben werden durch Modularity bestimmt.

Netzwerke

Keywords-Netzwerk

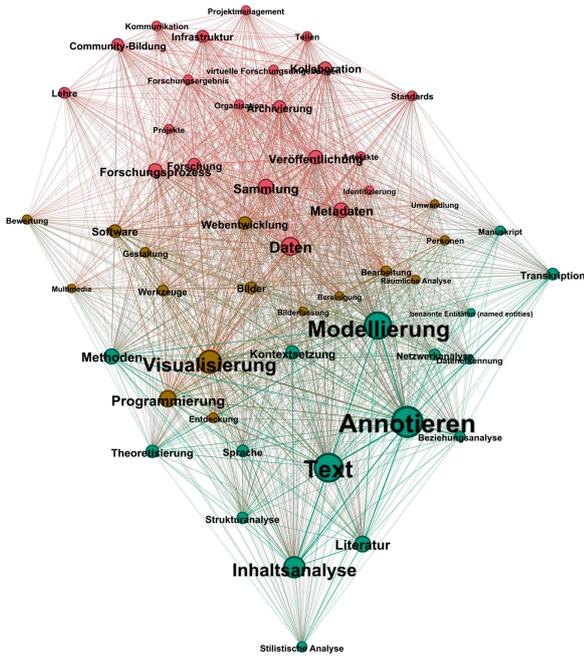


Abbildung 1: Ko-Keywords Netzwerk

Abbildung 1 zeigt das Ko-Keywords Netzwerk. Die DHD-Abstracts verfügen über zwei separate Keyword Auszeichnungen, sog. 'Keywords' und 'Topics'. Es zeigt sich, dass die 'Keywords' eher sparse sind (da diese durch ein Freifeld eingegeben werden), und die 'Topics' sehr dicht sind. Daher zeigen die Netzwerke der einzelnen Felder wenig aufschlussreiches. Wenn allerdings beide Annotationen auf Dokumentenebene zusammengeführt werden, so zeigt sich, dass die Topics durch die Keywords sinnvoll miteinander verbunden werden. Der Übersicht halber wurden seltene Keywords gefiltert.

Wir finden einige Cluster: Rechts unten die Annotation von Text, insbesondere für die Inhalts- und Stilanalyse in der Literaturwissenschaft, mittig Modellierung und Visualisierung, dazu Methoden und Metadaten. Zwischen den ersten beiden Clustern die Bildwissenschaft. Mittig und Oben finden sich Daten, deren Sammlung, Veröffentlichung, und Archivierung. Oben die Forschung selbst, mit der dazugehörigen Community, Infrastruktur und Lehre.

Ko-Autorschaft Netzwerk

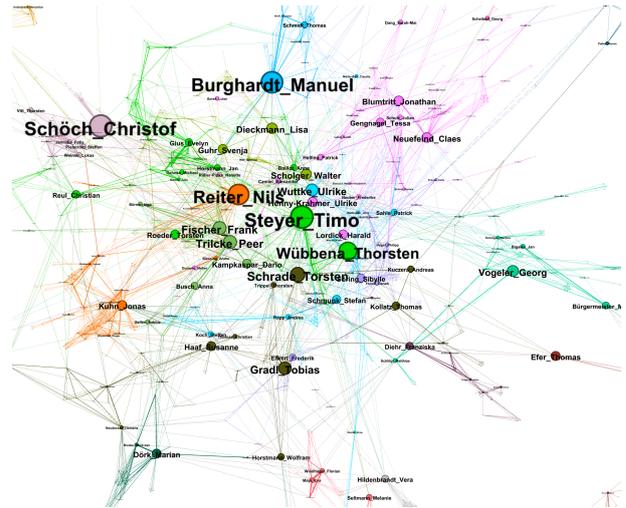


Abbildung 2: Ko-Autorschaft Netzwerk

In Abbildung 2, dem Ko-Autorschafts-Netzwerk zeigen sich einige regionale Zentren. In grün mittig links die Verbindung Potsdam-Berlin (Trilcke, Fischer). Rechts in türkis Graz. In orange und grün links und zentral Stuttgart-Darmstadt-Hamburg. Links oben Würzburg-Trier (Jannidis, Schöch). Oben in blau Leipzig-Regensburg (Burghardt). Zentral hellgrün findet sich ein dichtes Feld mit gemischten Affiliationen, darunter einige Kölner Namen.

Zitierverhalten & Ko-Zitation

Das Parsing von Referenzen/Zitationen fokussiert auf first-level Referenzen (also auf Artikel-Ebene, und nicht z.B. in welchem Sammelband etwas erschienen ist). Im Folgenden (Tabelle 1) sind die 20 am häufigsten zitierten Arbeiten aufgeführt. Darunter finden sich unter den Hauptautor:innen drei Frauen, Johanna Drucker, Birgit Hamp, und Verena Henrich. Die Themenwahl reicht von den FAIR Principles über Topic Modelling/Embeddings und Distant Reading über Editionswissenschaft und Korpora bis in die Bildwissenschaft. Es zeigt sich ein ähnliches Bild wie bei der Keywordanalyse.

Title	Year	Authors	Citations
The Fair Guiding Principles for Scientific Data Management and Stewardship	2016	Wilkinson et al.	23
Latent Dirichlet Allocation	2003	Blei et al.	18
Digitale Editionsformen	2013	Sahle	18
Distant Reading	2013	Moretti	14
TEI P5 Guidelines for Electronic Text Encoding and Interchange	(2007)	TEI Consortium	12
Probabilistic Topic Models	2012	Blei	12
Germanet: A Lexicalsemantic Net for German	1997	Hamp & Feldweg	11
Stylometry with R: A Package for Computational Text Analysis	2016	Eder et al.	11
GernEdit: The Germanet Editing Tool	2010	Henrich & Hinrichs	10
Generous Interfaces for Digital Cultural Collections	2015	Whitelaw	10
Digitales Publizieren	2017	Kohle	10
BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding	2018	Devlin et al.	10
Humanities Approaches to Graphical Display	2011	Drucker	9
Efficient Estimation of Word Representations in Vector Space	2013	Mikolov et al.	9
MALLET: A Machine Learning for Language Toolkit	2002	McCallum	9
Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges	2018	Windhager et al.	9
Literaturwissenschaft als Hackathon: Zur Praxeologie der Digital Literary Studies und ihren Epistemischen Dingen	2018	Trilcke & Fischer	9
Digital Humanities: Eine Einführung	2017	Jannidis & Kohle & Rehbein	8
Leistung aus Vielfalt	2016	Rat für Informationsinfrastrukturen (RfII)	8
Das Drama	2001	Pfister	8

Tabelle 1: Top 20 am häufigsten zitierte Arbeiten

Um zu untersuchen wie die Bibliographien beschaffen sind, untersuchen wir deren Länge und das Alter der Referenzen, wie von Bollmann und Elliot (2020) für die ACL gezeigt, um zu untersuchen ob ältere Publikationen 'vergessen' werden. Es folgen zwei 'letter value' plots, die die Länge der Bibliographie der Abstracts und das Alter der Referenzen zeigen.

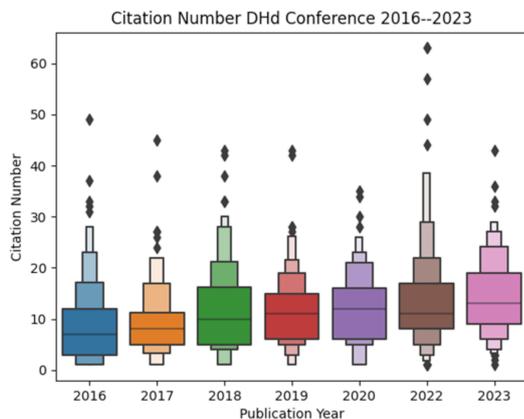


Abbildung 3: Letter Value Plot für Länge der Bibliographien

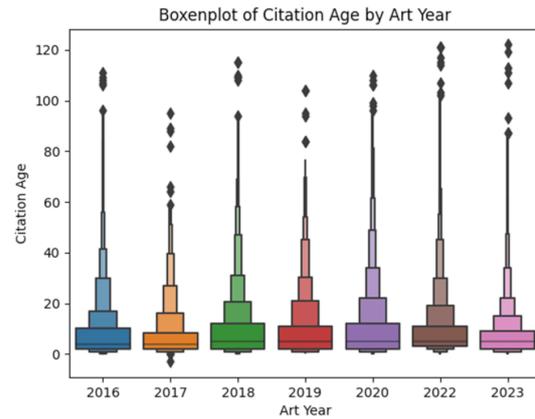


Abbildung 4: Letter Value Plot für Alter der Referenzen

Ingesamt ist zu sagen, dass die Bibliographien über die Jahre immer länger geworden sind (Median 7 bis Median 13), was für die Wissenschaftlichkeit und die Anerkennung des Fachs sicherlich fördernd ist. Zu Abbildung 4: Im Median sind Referenzen vier oder fünf Jahre alt, wenn sie zitiert werden. Es werden auch sehr alte Arbeiten zitiert, aber hier wurde erst ab 1900 berücksichtigt um etwaige Parsingerror zu minimieren. Das arithmetische Mittel scheint die letzten Jahre etwas zu fallen, was darauf hinweisen könnte, dass weniger alte Arbeiten zitiert werden.

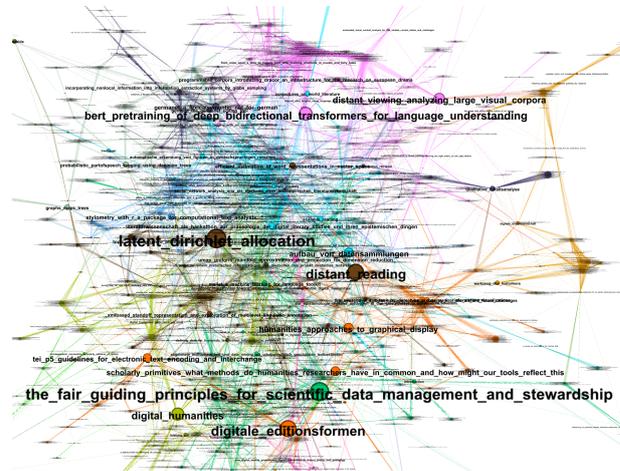


Abbildung 5: Ko-Zitationsnetzwerk

Abbildung 5 zeigt das Ko-Zitationsnetzwerk. Eine Kante wird gebildet, wenn zwei Referenzen zusammen in einer Bibliographie auftauchen, und das Gewicht der Kante bestimmt in wievielen Abstracts dies der Fall ist. Wir sehen wieder ein gewohntes Bild: Oben beginnt das Netzwerk mit Sentiment Analysis und Transformer Modellen los und bewegt sich dann über Computational Literary Studies (mit z.B. Netzwerkanalyse für Dramen) über Topic Analyse (Latent Dirichlet Allocation) und Distant Reading

hin zu Forschungsdatenmanagement (FAIR) bis zur Editionswissenschaft.

Zusammenfassung

Wir zeigen auf diesem Poster einen Querschnitt der Themen (Ko-Keywords), der Ko-Autorschaft und das Zitierverhalten der DHd Community. In Zukunft wollen wir einzelne Variablen miteinander verbinden (etwa wer für welche Keywords zitiert wird).

Bibliographie

Bollmann, Marcel, and Desmond Elliott. 2020. "On forgetting to cite older papers: An analysis of the acl anthology." In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics , pp. 7819-7827. 2020.

Henny-Krahmer, Ulrike und Patrick Sahle. 2018. "Einreichungen zur DHd 2018." Zugriff 19. Juli 2023. <https://dhd-blog.org/?p=9001>.

Hoenen, Armin. 2019. "Einreichungen zur DHd 2019 - II." Zugriff 19. Juli 2023. <https://dhd-blog.org/?p=11418>.

Tang, Muh-Chyun, Yun Jen Cheng, and Kuang Hua Chen. 2017. "A longitudinal study of intellectual cohesion in digital humanities using bibliometric analyses." *Scientometrics* 113 (2017): 985-1008.

Wieczorek, Oliver, Markus Eckl, Madeleine Bausch, Erik Radisch, Christoph Barmeyer, and Malte Rehbein. 2021. "Better, Faster, Stronger: The Evolution of Co-authorship in International Management Research Between 1990 and 2016." *Sage open* 11, no. 4 (2021): 21582440211061561.