IFLA WLIC 2019 Athens

Submitted on: 17.07.2019

Data intelligence in libraries: the actual and artificial perspectives
Date: 22-23 August 2019
Location: German National Library – Deutsche Nationalbibliothek (DNB), Frankfurt, Germany

Using Graph Visualization to enhance representation and evaluation of work clusters

Angela Vorndran
Data Management, German National Library (DNB), Frankfurt am Main, Germany.
E-mail address: a.vorndran@dnb.de

Stefan Grund
Data Management, German National Library (DNB), Frankfurt am Main, Germany.
E-mail address: s.grund@dnb.de

Copyright © 2019 by Angela Vorndran and Stefan Grund. This work is made available under the terms of the Creative Commons Attribution 4.0 International License:
http://creativecommons.org/licenses/by/4.0

Abstract:

The German National Library (DNB) uses the platform Culturegraph (www.culturegraph.org) to aggregate metadata of library holdings of German and Austrian library networks. In this data pool of more than 171 million items we perform work clustering. By offering an aggregated view of all publications representing a work, e.g. different editions and translations, the collection appears much more structured and searching is easier.
In this paper we would like to show the use of graph visualization for display, analysis and evaluation of work clusters. Graph visualization enables users to obtain a more transparent view of the connections underlying the structure of a work cluster. Work clustering is achieved by creating and matching keys which combine different metadata elements of a bibliographic record. Applying a breadth-first search, publications with identical matchkeys are grouped together. We use different keys to represent a publication, so each publication can obtain more than one key. Thus, one matchkey establishes a connection between publications. If one of the publications shares a different matchkey with even more publications the network representing a work grows.
Moving beyond visualization we can also gain statistical indicators from the graph to evaluate a work cluster. Degree, average path length or centrality measures can offer information about the internal structure of a work cluster. Particularly with large clusters information about the degree of connections between the members and the existence of more closely related subclusters is important for evaluating clusters.
Using graph visualization thus not only assists in grasping the internal structure of a work cluster more clearly but also helps managing and evaluating large datasets, ultimately leading to better clustering results to support data representation and findability.

Keywords: work cluster, graph visualization

Work clustering

Work clustering strives to aggregate different manifestations and expressions of a work in one work cluster. Work clusters can serve several purposes which appear more and more relevant if a work is published in a high number of different ways, e.g. different editions, translations or issues. The diversity of numerous hits when searching for a publication can be overwhelming and hard to handle, if it is presented in an unstructured way. Because of this fact, several actors have developed models to match publications and group them together in separate subunits, both as works, manifestations or expressions.

Most algorithms developed for this goal concentrate on matching titles and author names such as the "FRBR Work-Set Algorithm"[1] by the Online Computer Library Center (OCLC) or clustering activities by Pfeffer and Wiesenmüller[2]. They vary slightly concerning the metadata fields considered for persons and corporate bodies responsible for or related to the creation of the publication. Also titles can be included from different metadata fields containing title statements, varying forms of titles and uniform titles. A review of other applications of work clustering can be found in Pfeifer/Polak-Bennemann.[3]

A prerequisite necessary to perform work clustering is a clear definition of a work. According to the Library Reference Model (LRM) a work is defined by its specific contents:

> "The essence of the work is the constellation of concepts and ideas that form the shared content of what we define to be expressions of the same work. A work is perceived through the identification of the commonality of content between and among various expressions."[4]

The Functional Requirements for Bibliographic Records (FRBR) offer a more detailed differentiation of what belongs to a work and what constitutes a new work:

> "Similarly, abridgements or enlargements of an existing text, or the addition of parts or an accompaniment to a musical composition are considered to be different *expressions* of the same *work*. Translations from one language to another, musical transcriptions and arrangements, and dubbed or subtitled versions of a film are also considered simply as different *expressions* of the same original *work*. […]
> By contrast, when the modification of a *work* involves a significant degree of independent intellectual or artistic effort, the result is viewed, for the purpose of this study, as a new *work*. Thus paraphrases, rewritings, adaptions for children, parodies,

[1] Thomas B. Hickey and Jenny Toves, „FRBR Work-Set Algorithm. Version 2.0", 2009, https://www.oclc.org/content/dam/research/activities/frbralgorithm/2009-08.pdf.

[2] Cf. Magnus Pfeffer, „Using Clustering Across Union Catalogues to Enrich Entries with Indexing Information", in *Data Analysis, Machine Learning and Knowledge discovery, ed. Myra Spiliopoulou, Lars Schmidt-Thieme, Ruth Janning* (Cham: Springer International Publishing, 2014), 437–45. or Heidrun Wiesenmüller und Magnus Pfeffer, „Abgleichen, anreichern, verknüpfen", *BuB* 35, Nr. 09 (2013): 625–29, http://www.b-u-b.de/pdfarchiv/Heft-BuB_09_2013.pdf.

[3] Barbara Pfeifer and Renate Polak-Bennemann, „Zusammenführen was zusammengehört – Intellektuelle und automatische Erfassung von Werken nach RDA", *o-bib. Das offene Bibliotheksjournal* 3, Nr. 4 (2016): 144–55, https://doi.org/10.5282/o-bib/2016h4s144-155.

[4] Pat Riva, Patrick Le Bœuf, and Maja Žumer, „IFLA Library Reference Model. A Conceptual Model for Bibliographic Information", 2017, 21f., https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf.

musical variations on a theme and free transcriptions of a musical composition are considered to represent new *works*."[5]

Work clustering in our case aims to include all expressions of a work including translations and minor changes. The creation of a new work strongly depends on the degree of intellectual effort necessary, e.g. when adapting a work for a different target audience or genre. The amount and extent of changes can hardly be traced exactly on the basis of metadata so this is one of the challenges automatic work clustering is facing.

Work clustering in this context is applied on a data basis of more than 170 million metadata records of German and Austrian library networks available in Culturegraph. Work clusters are formed by using an algorithm which creates keys for each publication combining specific metadata elements. The algorithm is based on earlier works in the context of Culturegraph and the literature cited above. On the one hand a combination of unique identifiers (such as International Standard Book Number (ISBN) or library control numbers), Number of part/section of a work, Title and publication type forms the key. On the other hand Title and Remainder of title, Creator, Number of part/section of a work and publication type are used (cf. Fig. 1). Library control numbers such as the OCLC Control Number or, in a German context, the Erstkatalogisierungs-ID (EKI, ID of the first catalogue record of a publication) serve as a common identifier for catalogue records of the same publication, e.g. when adapting records from other library catalogues.

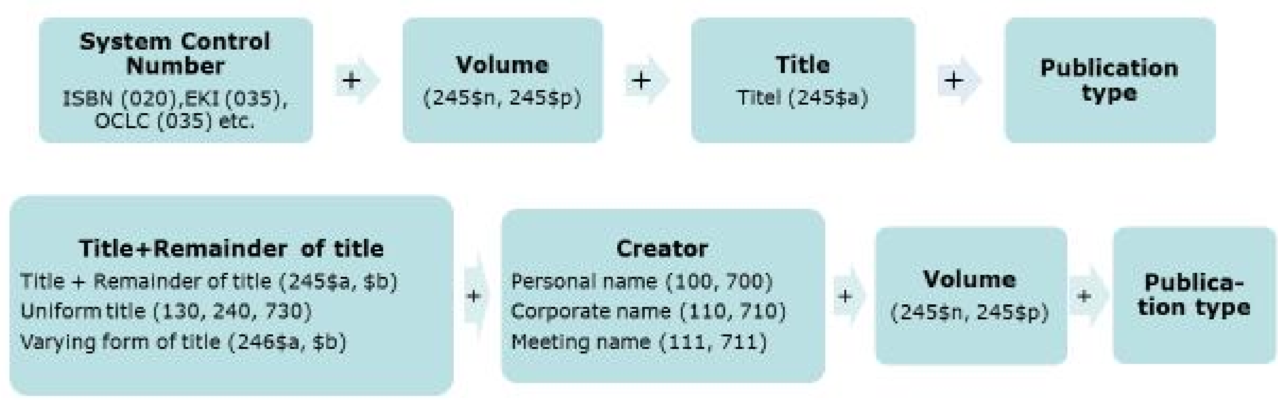

Fig. 1: Keys used in work clustering

Publications with identical keys are grouped together in clusters. As a breadth-first search is applied one identical key is sufficient to connect publications. Publications usually obtain more than one key thus the different keys can establish connections between several publications which by themselves are connected to other publications. In this way a network of connected publications emerges and forms a work cluster.

The handling of metadata records in Culturegraph involves several data analysis tools. Data is provided in the formats MARC 21 or MARCXML. To create keys and an intermediate working format the Java library Metafacture is used (https://github.com/metafacture) [6] as well as XSL Transformations and Python and Shell scripts.

[5] Functional Requirements for Bibliographic Records, Final Report, IFLA, 1998, p.16f., https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf.

[6] Markus Michael Geipel, Christoph Böhme, and Jan Hannemann, „Metamorph: A Transformation Language for Semi-Structured Data“, *D-Lib Magazine* 21, Nr. 5/6, 2015, https://doi.org/10.1045/may2015-boehme.

Link zu diesem Datensatz	http://d-nb.info/94284601X
Titel	Strukturiertes Systemdesign : ein praktischer Leitfaden / Meilir Page-Jones.
Person(en)	Page-Jones, Meilir (Verfasser)
Werk(e)	The practical guide to structured systems design (dt.)
Verlag	München ; Wien : Hanser - London : Prentice Hall Internat.
Zeitliche Einordnung	Erscheinungsdatum: 1995
Umfang/Format	XIX, 378 S. : graph. Darst. ; 23 cm
ISBN/Einband/Preis	978-3-446-16302-7 (Hanser) kart. : DM 78.00 3-446-16302-6 (Hanser) kart. : DM 78.00 978-0-13-678020-5 (Prentice Hall Internat.) kart. 0-13-678020-2 (Prentice Hall Internat.) kart.
Anmerkungen	Literaturverz. S. 371 - 373
Schlagwörter	Strukturierte Systementwicklung
Sachgruppe(n)	28 Informatik, Datenverarbeitung

Fig. 2 Bibliographic record

As an example of how keys for work clustering are created the keys for the catalogue record displayed in Fig. 2 are listed below:

- ISBN-Volume-Title-Publication Type:
  9780136780205-X-strukturiertessystemdesign-text
  9783446163027-X-strukturiertessystemdesign-text
- System Control Number-Volume-Title-Publication Type:
  oclc75512737-X-strukturiertessystemdesign-text
- ID-Volume-Title-Publication Type:
  DNB94284601X-X-strukturiertessystemdesign-text
- Title+Remainder of Title-Creator-Volume-Publication Type:
  strukturiertessystemdesignpraktleitfaden-pagejonesmeilir-X-text
- Uniform title-Creator-Volume-Publication type:
  practicalguidetostructuredsystemsdesign-pagejonesmeilir-X-text

As of May 2019 107,218,433 records were processed. Because work clustering is not executed for serials, serial component parts and parts of sets with dependent titles some records are not considered for key creation. 14,972,287 clusters with a size larger than one emerged containing 72,024,484 records. The average cluster size ist five.

Graph visualization

Work clusters represented as graphs

In this paper we would like to explore the use of graph visualization techniques to represent as well as to evaluate the work clusters created in Culturegraph.
On the Culturegraph website an accumulated view of all records belonging to a cluster on one result page is displayed. This has the advantage of offering an easy comparison of the bibliographic metadata elements of each entry. It makes visible how different editions, publication years or translations are summarized in a cluster (cf. Fig. 3). But this way of presentation cannot provide information on how the cluster is connected, which keys are used and how the cluster is structured.

(DE-101)94284601X EINZEL-ANSICHT »

Überblick

Hauptsachtitel	Strukturiertes Systemdesign
Zusatz	Ein Praktischer Leitfaden
Person	aut Page-Jones, Meilir 112907873
Körperschaft	-
Umfang	XIX, 378 S.
Erscheinungsjahr	1995
Ländercode	XA-DE
Material	book
Erscheinungsort	London
Herausgeber	Prentice Hall Internat.
Schlagwort	650 Strukturierte Systementwicklung 4259060-7
Sachgruppe	DNB 004 DNB 28 DNB 30
Standard-Identifier	OCLC 75512737 DNB 94284601X EKI DNB94284601X
Verlags-Identifier	ISBN13 9780136780205 ISBN13 9783446163027
Export	MARCXML RDF

(DE-603)026460920 EINZEL-ANSICHT »

Überblick

Hauptsachtitel	The Practical Guide To Structured Systems Design
Inhalt	Inhaltsverzeichnis
Person	Page-Jones, Meilir 112907873
Körperschaft	-
Umfang	XXI, 368 S
Erscheinungsjahr	1988
Material	book
Erscheinungsort	London [u.a.]
Herausgeber	Prentice-Hall Internat.
Schlagwort	650 Strukturierte Systementwicklung 4259060-7
Klassifikation	DDC 004.21 RVK ST 230 RVK ST 237
Standard-Identifier	OCLC 59143155 EKI HEB026460920
Verlags-Identifier	ISBN13 9780136907695 ISBN13 9780136907770
Export	MARCXML RDF

Fig. 3: Presentation of work cluster on Culturegraph website

In contrast, graph visualization is able to illustrate the internal structure of a work cluster much clearer than any text-based way. The software Gephi[7] is used for the visualization of graphs. Here, it is easy to visually distinguish between different subclusters in a cluster and to identify single keys connecting different areas of a cluster. A work cluster contains vertices representing publications and vertices representing keys. The edges show which keys belong to a publication and thereby also establish the connections between several publications via the same keys.

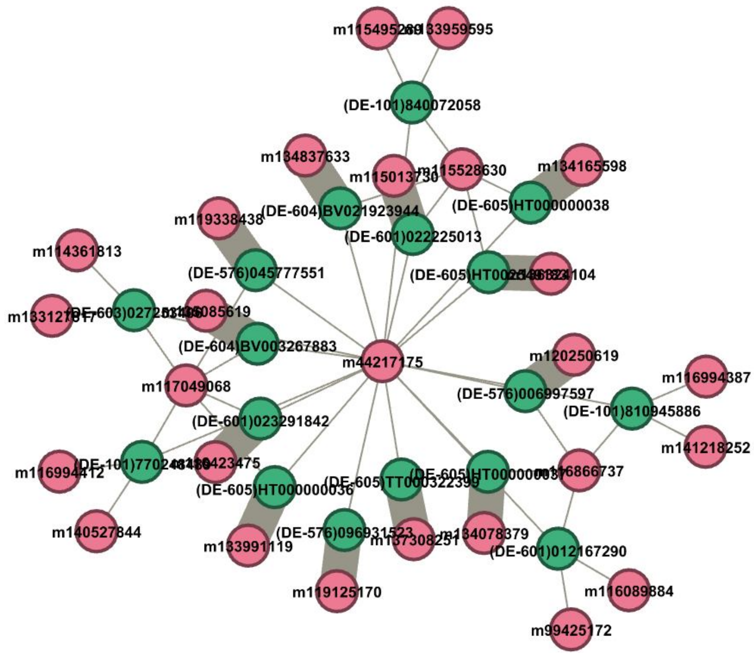

Fig. 4 Cluster 22

[7] https://gephi.org/, Cf. M. Bastian, M., Heymann, S. and Jacomy, M., Gephi: an open source software for exploring and manipulating network, International AAAI Conference on Weblogs and Social Media, 2009.

Figure 4 demonstrates the structure of a work cluster. Datasets representing publications are manifested as green nodes, whereas red nodes stand for keys assigned to the publications. As we can see, most publications receive several keys, which they can share with different other publications. The central key m44217175 occurs in all publications, other keys (e.g. m115528630 at the top) are common to a smaller number of publications. And still more keys, which are located at the outer edge, occur in only one publication and do not serve as connectors to other members of the cluster. Some publications possess two identical keys because of identical control numbers in different metadata fields. This is represented by edges depicted as grey strips instead of lines.
Graph visualization is most beneficial for understanding structures of large clusters. Fig. 5 shows a cluster, where graph visualization enables the user to discern between different subclusters representing different forms of a work. The subcluster on the left is grouped around a manifestation with a specific ISBN. The other subclusters divide the cluster up into different expressions, i.e. the French original text and the German translation as well as a few datasets of the English translation in the top right corner.
This way of depicting a work cluster conveniently gives an insight into how homogeneous a cluster is or how many different expressions and manifestations of a work it encompasses. It shows how sometimes a single key can establish a connection between two different subclusters.

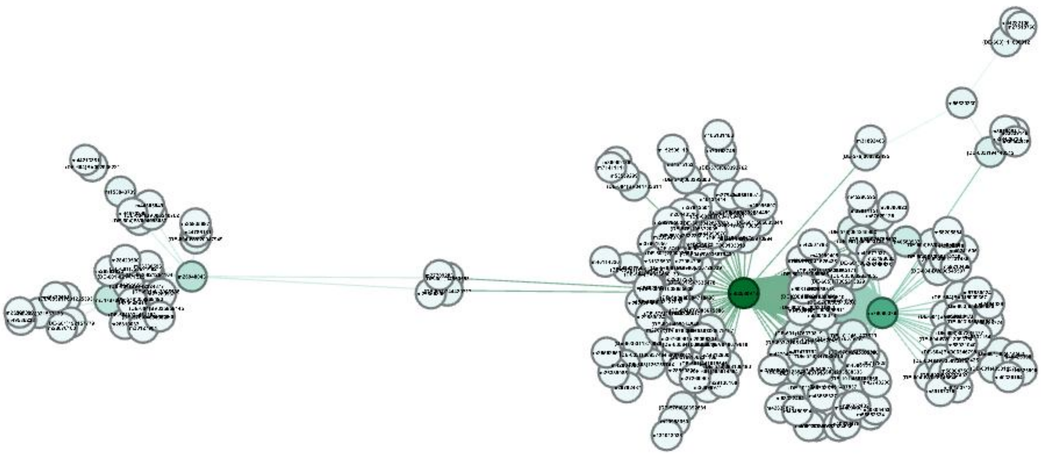

Fig. 5 Cluster 190701

Graph visualization to analyze work clusters

Apart from visual impressions, metrics, which can be calculated from graphs, can be used to explore the connections between the members of a cluster and describe the composition of a cluster.
As a first example of how these metrics can be indicative of the structure, representation and evaluation of work clusters, Cluster 1186856 is discussed here in greater detail. It contains 68 vertices and 97 edges. This network describes the connections between 29 publications establishing a work cluster. Publications can be identified by their IDs starting with „(DE-“ or „(AT-“. The work cluster contains publications from five different library networks and the German National Library. Seven are in German language, whereas 22 are in the original language English. As can be deferred from Fig. 6, most publications are connected to the central node m143553364 representing the key „practicalguidetostructuredsystemsdesign-pagejonesmeilir-X-text“. They either construct this key from the title, in case of English

publications, or the uniform title added to German translations, the author name, „X“ representing a missing volume number and the publication type.

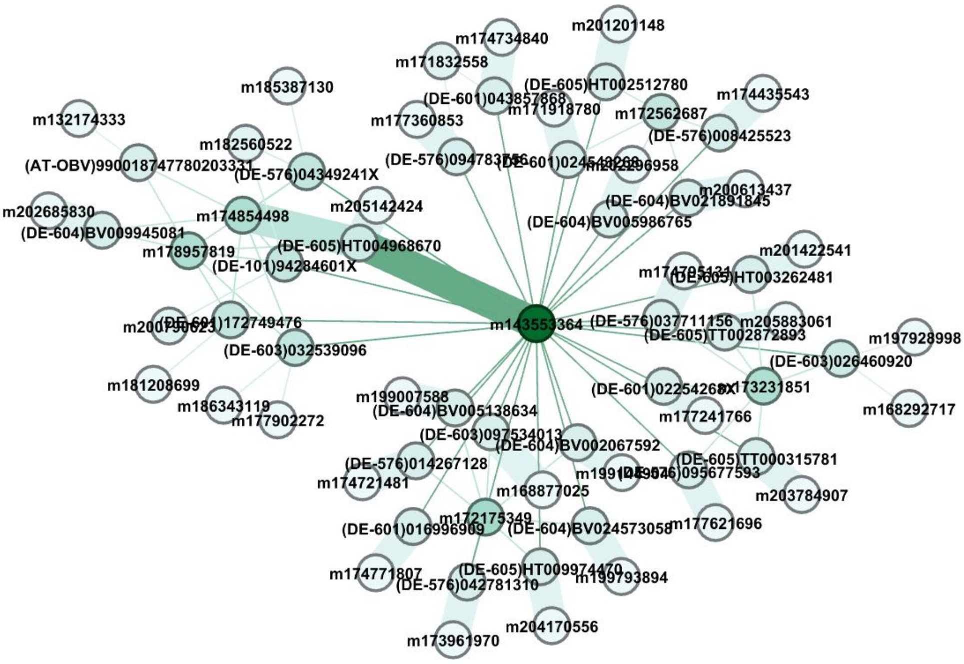

Fig. 6: Cluster 1186856, Degree

A simple but meaningful metric indicator describing the structure of a graph is the degree, i.e. the number of edges adjacent to a vertex [8]. It shows which vertices are placed in the center of a cluster and are well connected with other vertices. Also the average degree of a vertex can lead to meaningful insights.

Fig. 6 shows that when looking at the highest degree, vertex m143553364 exhibits the maximum represented by the dark green color. On the left side a smaller set of publications is more closely connected. These are German translations of the publication all receiving the key “strukturiertessystemdesignpraktleitfaden-pagejonesmeilir-X-text”. Whereas several publications possess the English uniform title two publications are only identified by the German title ((AT-OBV)990018747780203333), (DE-604)BV069945081) and are connected to the graph via those publications with the German and the English uniform title. Graph visualisation enables the user to realize this fact instantly whereas when looking at a list of the keys of these publications it is difficult and time-consuming to understand the connections between the cluster members.

The degree of those publications with both titles accordingly is higher, indicated by a slightly darker green. Other subgroups of publications with identical keys mostly result from identical ISBNs, such as the vertex m173231851 on the right side representing the key “9780136907695-X-practicalguidetostructuredsystemsdesign-text.”

In contrast to the very centralized graph of Cluster 1186856 where almost all publication share one common key, Cluster 2748844 (cf. Fig. 7) is more widely branched. The analysis of degrees shows that one key exists which prevails in numbers of connected edges (m201915181) but several other well connected vertices exist in the highly complex network.

[8] Cf. Krishna Raj P.M. and Ankith Mohan K. G. Srinivasa, Practical Social Network Analysis with Python, Springer International Publishing, 2018, p. 2, see also e.g. Newman, M.E.J. Networks. An Introduction, 2010, p. 168.

Cluster 2748844 consists of 41 vertices and 74 edges. The most common key is „dreamsportaltosource-whitmontedward-X-text“ (m201915181). Some other keys assembling the second author and the English title or both authors with the German title of the translation of the work also appear in many datasets. On the left we can see one branch with a publication where an additional remainder of title is part of the metadata and thus the author-title key does not match. This publication becomes part of the cluster when aligning the ISBN-title keys.

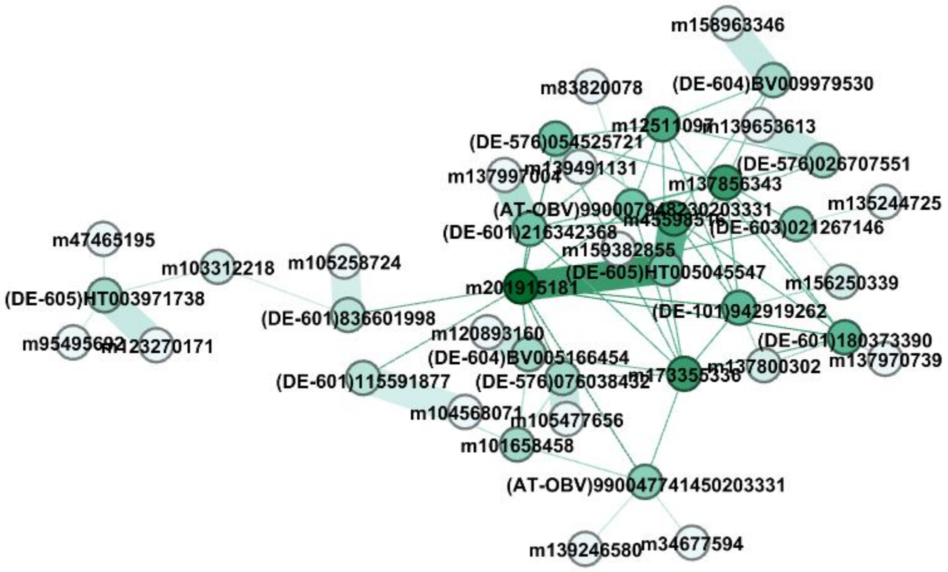

Fig. 7 Cluster 2748844, Degree

In addition to degree, also the path length in a graph can be of significance. The length of a path is calculated by counting the edges forming the path between two vertices. The average path length describes the average of the distances between all vertices in the network.[9] The average path length of Cluster 1186856 is 3.2 and of Cluster 2748844 3.4. The average path length also can depend on the size of the cluster. Another example of cluster with 1978 vertices and 3718 edges has an average path length of 8.6.

Fig. 8 pictures Cluster 2748844 as well but the values of betweenness centrality are indicated by the purple colour. Betweenness centrality is calculated by counting the times a vertex is situated on a shortest path between other vertices[10]. Vertices which serve as a connection for many vertices in a network possess a central position, are important for the network and influence the network strongly. In Cluster 1186856 the one central key also scores the highest value in betweenness centrality. Fig. 8 depicts Cluster 2748844 which also shows a vertex of maximum betweenness centrality representing the key „dreamsportaltosource-whitmontedward-X-text“. But here, some other vertices such as the publication (DE-601)836601998 are relevant intersections because they connect publications which do not possess the most common key with the cluster.

[9] Krishna Raj P.M. and Ankith Mohan K. G. Srinivasa, Practical Social Network Analysis with Python, Springer International Publishing, 2018, p. 16f.

[10] Cf. Newman, M.E.J. Networks. An Introduction, 2010, p. 187.

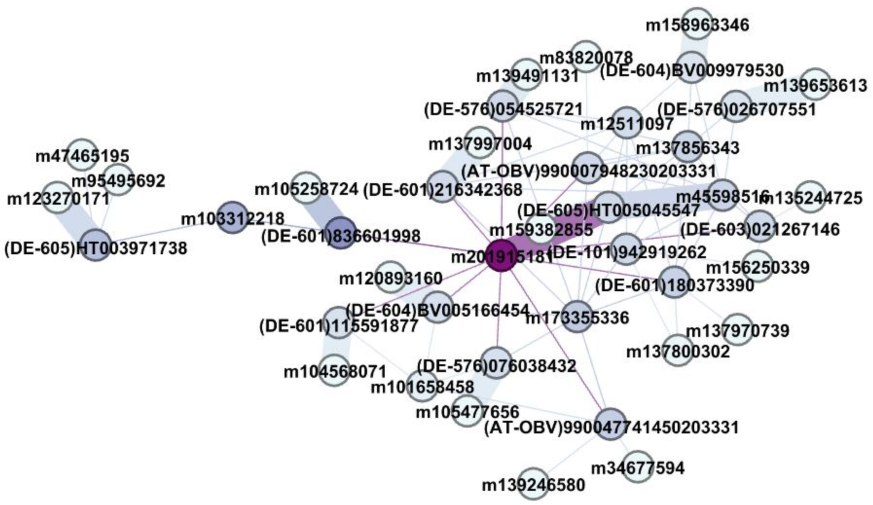

Fig. 8 Cluster 2748844, Betweenness Centrality

Another measure commonly used in network analysis is closeness centrality. It measures the mean distance between vertices. If a vertex is connected to many vertices in a short distance it is well connected and easily reached[11]. In Fig. 9 Cluster 2748844 again shows that there are closer connections on the right side of the cluster between several publications sharing more than one common key. Apart from the most common key combining title and first author they also share keys for the German title and first author (m45598516), German title and second author (m12511097) or English title and second author (m173355336). So the measure of closeness centrality in this cluster gives an indication of a subcluster consisting of the German translations of the publication. For this more closely connected subset of publications in a cluster as dense as this one, closeness centrality is an appropriate indicator, because it appraises the fact that path lengths are short because of numerous connections between the vertices.

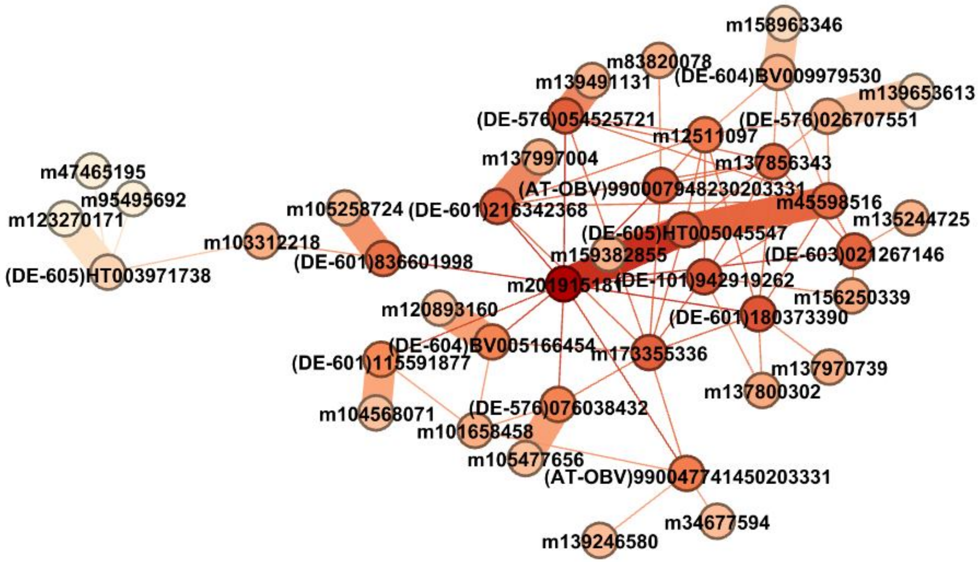

Fig. 9 Cluster 2748844, Closeness Centrality

[11] Ibd., p. 183.

Furthermore, the measure of eccentricity describes the contrary situation highlighting those vertices which are not closely connected and lie at the peripheral areas of a cluster (cf. Fig. 10). High eccentricity values thus manifest the maximum distance from one vertex to any other vertex in the network.[12] The vertices which are closely connected to many other vertices have access to more paths through the network and can reach many vertices on a short path. However, vertices at the margins must possibly traverse several nodes to reach the centre to move on to vertices located in the opposite direction of the cluster such as vertices (DE-605)HT003971738 or (DE-604)BV009979530 in Cluster 2748844.

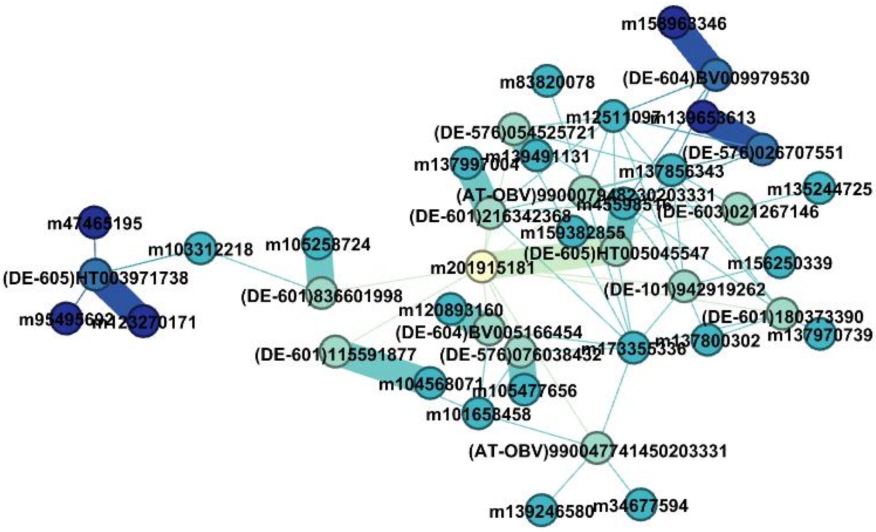

Fig. 10 Cluster 2748844, Eccentricity

The metrics which can be gained from analyzing the graphs of work clusters can give meaningful insights into the underlying structure and the significance of each vertex. In the following we would like to explore further if this information can be used for the evaluation of work clusters.

Using graph visualization as a means to support evaluation of work clusters

The structure of a work cluster, which can be grasped much easier through graph visualization, can give valuable indications as to which are the crucial structures to be reviewed first when evaluating a cluster. The statistic values describing the internal composition of a cluster described above can serve as indicators which assist in evaluating clusters or crucial connections within a cluster. Some observations concerning characteristic features of different variants of clusters can be helpful in this process.

A key with a high degree and high values of betweenness centrality and/or closeness centrality seems to be relevant to a large number of vertices in the cluster. If a key is connected to a high share of publications in the cluster, that seems to be an indicator that this portion of a cluster is very homogeneous. All publications share a common key and that means, for example, that they all share the same title and author or the same ISBN. In such a structure the probability of the (sub)cluster being correct is very high.

[12] Cf. Brandes, Ulrik and Thomas Erlebach. Network Analysis: Methodological Foundations, Springer-Verlag Berlin, Heidelberg, 2005, p. 21.

A different situation can be found when a vertex is characterized by high values of betweenness centrality and/or closeness centrality but a low degree. As in Fig. 11 only one publication connects two otherwise independent subclusters, which are in themselves closely intertwined. This publication is a decisive actor in the cluster and could, if faulty, be the cause of a large incorrect cluster. So the evaluation of the keys created for this publication has a high priority.

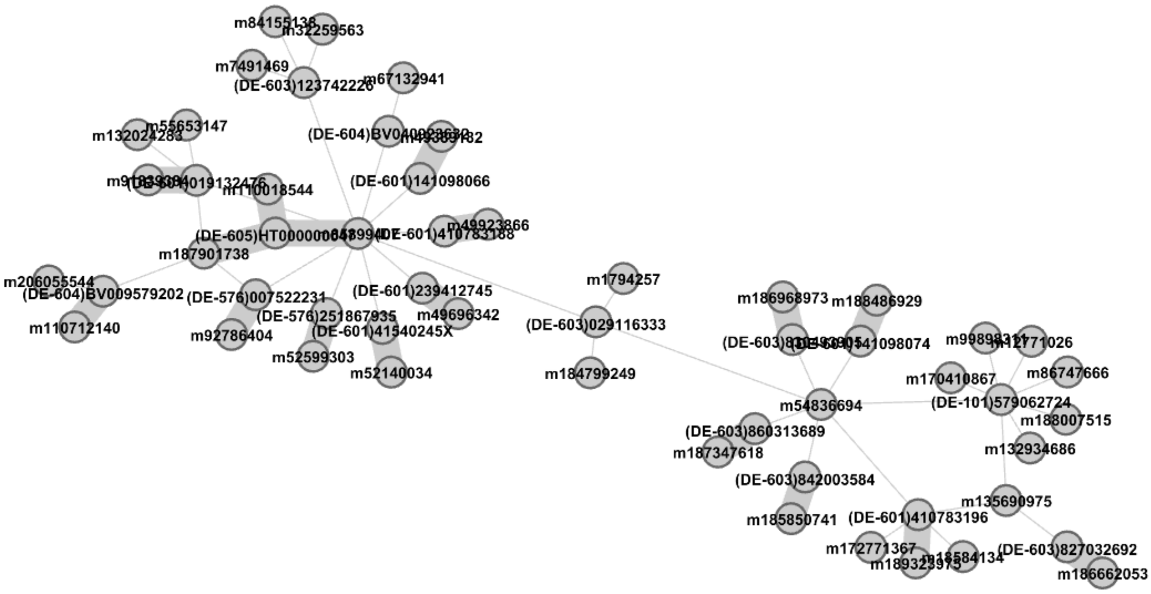

Fig. 11 Cluster 29

Publications located at the margins of a cluster and showing high values of eccentricity also often are connected to the cluster via a single key. Here also, the possibility of an erroneous connection should be investigated. On the contrary, publications sharing several keys with other publications in the cluster bear a lower risk of false assignment to the cluster. Again, higher values of betweenness centrality stand for a higher probability of correct clusters. If several publications share more than one key e.g. keys with titles and authors, uniform titles and authors and/or ISBN and title they most likely belong to the same work.

Degree only is not a very indicative factor. As can be seen in vertex (DE-101)579062724 in Fig. 11 on the right side, which represents a publication, it can be the case that a vertex is connected to many edges but they lead to keys which are not further connected. These dead ends can aid in characterizing the publication linked to it but they do not provide any further indication as to whether the grouping together with the other members of the cluster is correct.

However, if a vertex shows a high degree, and represents a key which is connected to most publications in a cluster, this means that all the publications in the cluster are likely to be fairly homogeneous in content and keys so that the cluster is most probably correct (cf. Fig. 6).

Conclusion

Graph visualization is a valuable addition to traditional result lists displaying bibliographical information. Work clustering, as it was developed in Culturegraph, initially is aimed at exchanging classification numbers and subject headings amongst cluster members. As clusters can reach sizes of several thousand members it is a complex task to understand, review and evaluate them on a textual basis. Graph visualization enables a user to understand the structures in a cluster much easier. Subclusters, which are characterized by a number of edges, connecting vertices several times in a closer circle within a larger cluster can give indications of different manifestations or expressions within a work. Particularly for evaluation purposes the visual impression of a work graph is very helpful because it steers the attention to vertices, which are crucial to the structure of a work cluster. Often these are single publications connecting two otherwise independent clusters. They should be reviewed with high priority as their position causes them to be the source of a large erroneous cluster, if faulty.

Of course, the quality of metadata is essential. Work clusters comprising different translations of a work rely on uniform titles being included in the record. A clear separation of titles and remainders of titles is important to form distinct keys for each publication which can be aligned with the keys of other publications. Likewise, the use of abbreviations can be harmful for the matching of titles. Furthermore, the inclusion of all persons responsible for creating the work and preferably also relator codes for these persons is useful. In the creation of keys considerations of which and how many persons to include can have a strong influence the size of clusters, as some persons with specific relations tend to be part of the creation of numerous publications (e.g. editors of classic works).

However, the technique applied here, of creating a number of different keys encompassing different characteristics of a publication proved fairly robust concerning ambiguous or erroneous metadata content.

As a topic to elaborate further on it could be promising to analyze the vertices and their contents event more thoroughly e.g. trying to find out whether numeric keys making use of control numbers or textual keys constructed from titles and authors are more likely to lead to erroneous clusters.

References

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. ‘Gephi: An Open Source Software for Exploring and Manipulating Networks.’ International AAAI Conference on Weblogs and Social Media, 2009. https://gephi.org/publications/gephi-bastian-feb09.pdf.

Brandes, Ulrik, and Thomas Erlebach, eds. *Network Analysis*. Vol. 3418. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. https://doi.org/10.1007/b106453.

IFLA. ‘Functional Requirements for Bibliographic Records. Final Report’, 1998. https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf.

Geipel, Markus Michael, Christoph Böhme, and Jan Hannemann. ‘Metamorph: A Transformation Language for Semi-Structured Data’. *D-Lib Magazine* 21, no. 5/6 (May 2015). https://doi.org/10.1045/may2015-boehme.

Hickey, Thomas B., and Jenny Toves. ‘FRBR Work-Set Algorithm. Version 2.0’, 2009. https://www.oclc.org/content/dam/research/activities/frbralgorithm/2009-08.pdf.

Newman, M. E. J. *Networks: An Introduction*. Oxford ; New York: Oxford University Press, 2010.

Pfeffer, Magnus. ‘Using Clustering Across Union Catalogues to Enrich Entries with Indexing Information’. In *Data Analysis, Machine Learning and Knowledge Discovery, Hrsg. Myra Spiliopoulou, Lars Schmidt-Thieme, Ruth Janning*, 437–45. Cham: Springer International Publishing, 2014.

Pfeifer, Barbara, and Renate Polak-Bennemann. ‘Zusammenführen was zusammengehört – Intellektuelle und automatische Erfassung von Werken nach RDA’. *o-bib. Das offene Bibliotheksjournal* 3, no. 4 (2016): 144–55. https://doi.org/10.5282/o-bib/2016h4s144-155.

Raj P.M., Krishna, Ankith Mohan, and K.G. Srinivasa. *Practical Social Network Analysis with Python*. Computer Communications and Networks. Cham: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-96746-2.

Riva, Pat, Patrick Le Bœuf, and Maja Žumer. ‘IFLA Library Reference Model. A Conceptual Model for Bibliographic Information’, 2017. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017.pdf.

Wiesenmüller, Heidrun, and Magnus Pfeffer. ‘Abgleichen, Anreichern, Verknüpfen’. *BuB* 35, no. 09 (2013): 625–29.