The journey of conceptual modeling: Paths from the past to present with trajectories for the future

Jacky Akoka [1], Isabelle Comyn-Wattiau[2], Nicolas Prat[2], and Veda C. Storey[3]

[1] *CEDRIC-CNAM, Paris, France*
[2] *ESSEC Business School, Cergy-Pontoise, France*
[3] *Georgia State University, Atlanta, GA, USA*

Abstract

The field of conceptual modeling has been in existence for over five decades. To understand how the field has evolved and should continue to evolve, it is useful to examine the contributions made over time and the topics that have emerged. In this research, we apply bibliometric analysis to a corpus of 4652 research papers spanning from 1976 to 2022. Co-citation and citation networks are produced that show the different schools of thought, the main topics of the domain, and the relationships among major and influential research papers over time. The co-citation analysis identifies four schools of thought. It elicits the separation between the historical cluster centered around Chen's seminal paper and another cluster proposing grammars and guidelines for representing the real world using conceptual modeling and methods for evaluating these representations. A bibliographic coupling analysis on papers from 2017 to 2022 results in ten clusters that characterize the main themes, including domain-specific conceptual modeling and applications, ontologies and applications, genomics, and datastores and multi-model data. The main path analysis of the citation network identifies several main paths among major and most influential papers. This leads to insights on the lineage of key papers in conceptual modeling research. The primordial nature of the main paths identified encompasses two important aspects. The first revolves around the refinement of the entity-relationship model. The second identifies the contribution of ontologies for conceptual modeling.

Keywords

Conceptual modeling, Bibliographic Topic Analysis, Main Path Analysis, Co-citation Analysis, Bibliographic Coupling Analysis

1. Introduction

Conceptual modeling is an important area of research that has continued to evolve over five decades. It will continue to serve as an important part of future information systems development. To understand the field of conceptual modeling, it would be useful to examine prior work and classify it, in order to appreciate the type of work that has already been carried out and how it is continuing. As in other academic research fields, conceptual modeling has evolved. Its evolution could be informative in understanding its potential for future progression, as well as its importance, relevance, and impact. Furthermore, it might not always be obvious what topics are being investigated and the type of research that takes place [36, 63].

There have been notable attempts to categorize and make sense of the field of conceptual modeling, considering its vast amount of work. Most prior studies, however, have concentrated primarily on

ER2023: Companion Proceedings of the 42nd International Conference on Conceptual Modeling: ER Forum, 7th SCME, Project Exhibitions, Posters and Demos, and Doctoral Consortium, November 06-09, 2023, Lisbon, Portugal
jacky.akoka@lecnam.net (J. Akoka); wattiau@essec.edu (I. Comyn-Wattiau); prat@essec.edu (N. Prat); vstorey@gsu.edu (V. C. Storey)
0000-0002-3228-3266 (J. Akoka); 0000-0002-5257-6189 (I. Comyn-Wattiau); 0000-0002-8064-9357 (N. Prat); 0000-0002-8735-1553 (V. C. Storey)

papers published in the *International Conference on Conceptual Modeling* or related journals, such as *Data and Knowledge Engineering*. There has been less emphasis on studying the entire period of research since the publication of Chen's [27] initial work on the entity-relationship model, thereby establishing the need for conceptual modeling. Furthermore, most studies employ specific keywords. They use a limited set of bibliometric techniques and do not focus on the impact of the research. The latter could be assessed, for example, by observing citation counts. Given the large number of research projects and papers on conceptual modeling, it seems reasonable that an automated approach to analyzing this research is needed.

We aim at exploring the conceptual modeling research domain and analyze its body of literature. More precisely, our objective is to improve our understanding of the conceptual modeling field, gain insights into it, and thus be able to better guide future research. In our bibliographic analysis, we focus on responding to the following general research question: *How can we identify and appreciate the journey of conceptual modeling?* This research question is decomposed into three sub-questions:

- *RQ1:* What is the intellectual structure of conceptual modeling?
- *RQ2:* What are the main themes of conceptual modeling research?
- *RQ3:* What are the paths linking the most influential publications in conceptual modeling?

To address these questions, this paper performs a topic analysis of conceptual modeling research using bibliometric analysis. This research makes several contributions. The first contribution is to extract a timeline that shows how the field of conceptual modeling has progressed. We perform a longitudinal overview of the conceptual modeling (CM) field, allowing us to describe the state of the literature on CM, identify the publication venues. Second, we apply three types of analysis: co-citation analysis (CCA), bibliographic coupling (BCA), and main path analysis (MPA). We identify several important paths that reflect the progression of the field, as well as how the paths or related research topics have evolved and make inferences about what we can learn from this evolution. The analysis is represented by several networks with the identification of "nodes" that signify major topics and how they are related, or linked, to other topics. Specific research papers that have impacted the field are highlighted and discussed for their significance. Third, we provide a summary of the results that might be useful for students and other researchers who wish to participate in this field.

The paper proceeds as follows. Section 2 reviews related research on efforts to understand and structure the field of conceptual modeling. Section 3 provides an overview of how the bibliometric analysis is conducted. Section 4 provides the results obtained. Section 5 discusses the main findings, the limitations, concludes the paper and proposes future research.

2. Related research

There have been prior efforts to study the field of conceptual modeling. First, some authors proposed frameworks to help understand this domain. Delcambre et al. [36], for example, propose a reference framework for identifying and characterizing contributions to conceptual modeling to help researchers position their work. The framework might also help researchers in other fields understand what the field of conceptual modeling is able to offer. Mayr and Thalheim [80] propose a framework for understanding conceptual models, the Triptych paradigm, as an approach that classifies conceptual models linking linguistic terms and encyclopedic notions. Recker et al. [95] reviewed publications in information systems, to create a framework that focused on conceptual modeling in the digital world.

Other authors performed an analysis of conceptual modeling papers. Aguirre-Urreta and Marakas [2] explored the reasons why issues of ontological foundation and modeling practices may result in differences between alternative modeling techniques. Molina et al. [81] surveyed conceptual modeling related to groupware. Fettke [43] investigated how practitioners use conceptual modeling to identify the barriers and success factors when using conceptual modeling. Chen [28] provides an overview of the progression of the conceptual modeling conferences over the first 30 years, identifying continued research topics and collaboration among multiple fields, across international boundaries. Frank et al. [45] conducted a synoptical review of modeling publications in the domain of business information systems engineering, highlighting research that shapes the field of a future research agenda.

Automatic analysis techniques, including bibliometric techniques may be relevant when a domain is growing rapidly [114]. Some authors explain the bibliometric techniques of citation and co-citation graph analysis and apply them to different fields such as desoxyribonucleic acid (DNA) theory or management [65, 127]. Regarding conceptual modeling, C. Chen et al. [26] conducted a citation analysis to identify trends. Cosentino et al. [31] focused on authors and papers in conferences to automate conference analytics. Lima et al. [72] analyzed the collaboration of authors over forty years of participation in the *International Conference on Conceptual Modeling*.

A relevant article that pursues a similar objective to ours is that of Härer and Fill [63] who report on a bibliometric analysis of over 3,000 publications, from nine outlets, using Latent Dirichlet Allocation (LDA). The researchers sought to investigate the different types of conceptual models found in various communities and educational environments to identify topics and visions for future research. They focused on the quantity of papers, authors, major topics, and their evolution. They identified three periods with increasing publications as being: 2005-2009; 2010-2014; and 2015-2019.

These efforts, however, do not always cover the entire period from 1976 to 2022. No prior study exploits bibliographic coupling and main path analyses techniques in the field of conceptual modeling. We intend to fill this research gap by proposing a bibliometric analysis that: 1) covers the entire period from Chen's [27] 1976 seminal paper on the entity-relationship model to 2022; 2) exploits bibliometric techniques to automatically discover the intellectual basis of conceptual modeling, its main themes and main trajectories along citation links; and 3) compares the results obtained by the different techniques.

3. Methodology

In this section, we successively describe the conceptual model that represents the underlying concepts of our approach and their relationships, before providing the steps carried out.

3.1. Underlying conceptual model

Our methodology relies on a conceptual model depicted at Figure 1.

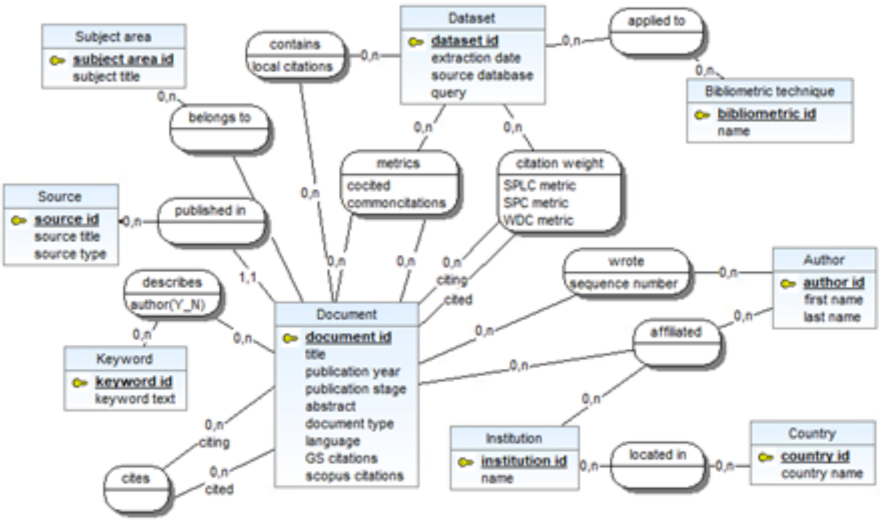

Figure 1: Conceptual model of research reported in this paper

Bibliometrics considers documents published by authors. Each document has a title, a year of publication, a status (article in press or final), an abstract, a type of document (article, review, conference paper, etc.), a language, and a number of citations per bibliometric database (Google Scholar, Scopus, etc.). A document is described by keywords that can be provided by the authors or generated by other means. A document belongs to one or more subject areas (computer science, management, etc.). It is published in a source described by a type (journal, book, conference proceedings, etc.). A document cites other documents and it may, itself, be cited by other documents. A document is written by one or more authors (listed in a specific order). The authors mention, at the time of publication, one or more affiliations in institutions located in different countries. Each bibliometric technique (CCA, BCA, MPA) is applied to a dataset. The latter describes the bibliographic information of a set of documents returned from a bibliographic source database (Scopus, Web of Science, etc.) by a query at a certain date (extraction date). A citing document can be cited in the same dataset by one or more other documents: *local citations* counts the number of times a citing article is cited within a dataset. The different techniques compute metrics and build clusters based on the latter. CCA clusters documents that are cited together (cocited). BCA clusters documents based on their common citations. MPA computes weights attached to edges in a citation graph. In our research, we tested different weights using three different metrics of the literature: Search Path Count (SPC), Search Path Link Count (SPLC) [65], and Weighted Direct Citation (WDC) [90].

Cosentino et al. [31] also proposed a conceptual model representing bibliographic information. Our work differs, though, because we do not have the same goal. They use the conceptual model as the first step to build a database for analyzing conferences and proceedings, whereas we use our conceptual model as a basis for our methodology. The main concepts of *document, author, institution*, and *country* overlap between the two models. The *dataset* description, the *bibliometric techniques*, and the *metrics* (*local citations, cocited, commoncitations, citation weights, GS citations, scopus citations*), upon which we build our clusters and algorithms, are specific to our conceptual model.

3.2. Steps in bibliometric analysis

Figure 2 summarizes the steps carried out to conduct this bibliometric research. The steps, as described below, rely on the concepts represented in the entity-relationship model of Figure 1.

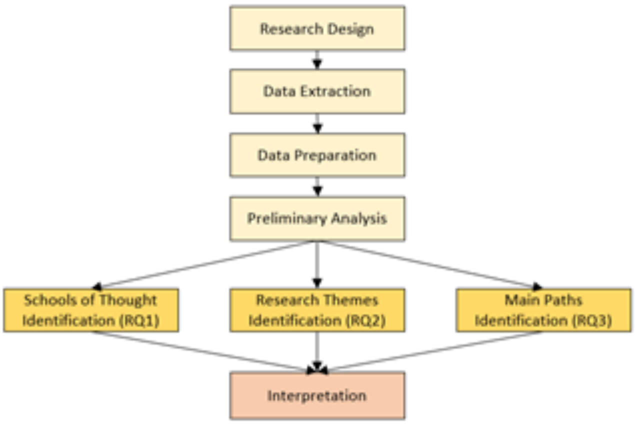

Figure 2: Steps in research process

Research design. This research employs bibliometric techniques to analyze research on conceptual modeling. We extend and update prior co-citation analyses on conceptual modeling. We also investigate two additional, complementary analysis techniques. The first, bibliographic coupling analysis, is used to identify the key themes of research traditions in conceptual modeling. The second, main path analysis, enables us to identify the most influential and central publications in conceptual modeling.

Data extraction. This step identifies relevant bibliographic databases and defines a querying strategy for extracting a suitable dataset. More precisely, we build a common dataset to answer the three research questions. Co-citation analysis is performed using the whole dataset. However, bibliographic coupling is relevant only for a limited, and recent, time span. Therefore, regarding this technique, we use only a subset of the main dataset limited to the most recent publications. Finally, due to its inherent logic, main path analysis is restricted to a specific subset of publications as explained below. The entities *source, document, author, institution, country, keyword, subject area* of the conceptual model of Figure 1 guide the construction of queries and the data extraction.

Data preparation. The information extracted describing the documents from a single bibliographic database presents a good homogeneity, without requiring cleansing operations. On the other hand, to be able to construct the citation and co-citation graphs used by all bibliometric techniques, it is necessary to be able to deduplicate the references of the cited articles sufficiently. The cleansing thus carried out makes it possible to improve the quality of the information collected within the *document* entity and to deduce the citation links of the *cites* relationship of the conceptual model.

Preliminary analysis. This step is a first analysis of the dataset. It consists of the: distribution of papers over time, analysis of the most frequent words, trends in terms of keywords, publication outlets (journals, conferences), authors who had contributed the most publications, and journals and conferences where the papers were published. This step uses the information contained in most of the entities and relationships of the model (*document, source, author, keyword, etc.*) in Figure 1.

Schools of thought identification. In a vast field like conceptual modeling, many contributions coexist. If several articles are often cited by the same articles (*cocited* metric of the conceptual model), they constitute a school of thought. The co-citation analysis relies on this frequency to build clusters whose similarity reveals a common school of thought. The size of the clusters and their cohesion are elements of appreciation of the number of such schools of thought within a research field.

Research themes identification. Inside a research domain, publications that cite many common papers generally address the same research theme (*commoncitations* metric of the conceptual model). Therefore, clustering such papers together is a way to characterize these themes. This is precisely the focus of bibliographic coupling analysis. Scientometric literature justifies bibliographic coupling to analyze more recent articles that do not yet have a significant number of citations. Furthermore, it recommends limiting this bibliographic coupling analysis to the last five or six years to allow a fair analysis of these recent articles.

Main paths identification. Understanding a research field requires eliciting the main research trajectories. The techniques that analyse main paths compute metrics attached to citation links, depending on how many paths traverse these links. Several different metrics (*SPC, SPLC, WDC* metrics of the conceptual model) may be used. Moreover, different algorithms (local, global, key-route) are implemented. Depending on these parameters, it is possible to identify different main paths.

Interpretation. Each bibliometric analysis provides a perspective on exploring the CM research field. Combining them enriches the analysis, and reinforces the findings. It also facilitates the detection of the dynamics of the field. From the main paths that emerge, the researchers can interpret the themes or important topics from the papers identified as being on the paths.

Thus, our approach proposes a conceptual model and a methodology for analyzing conceptual modeling research using scientometric analysis techniques. Its application is described below.

4. Application of our approach

Table 1
Research questions mapped to bibliometric techniques

Research Question	Detailed questions	Bibliometric technique
RQ1: Identify intellectual structure of conceptual modeling	• What are the schools of thought of CM? • What is the structure of the scientific community in CM?	Co-citation Analysis (CCA)
RQ2: Identify main themes of conceptual modeling	• What is the intellectual structure of ***recent/emerging*** literature? • How has the intellectual structure of small niche CM developed through time?	Bibliographic Coupling Analysis (BCA)
RQ3: Identify paths linking the most influential publications in conceptual modeling	• How to trace the development of CM? • How to highlight critical documents in CM?	Main Path Analysis (MPA)

In this section, we describe successively the different steps depicted in Figure 2.

Research design. The objective of this research is to characterize the conceptual modeling research field using bibliometric techniques. These techniques allow us to identify the different schools of thought, the recent research themes being addressed, and the main trajectories of research. We employ three types of quantitative analyses: namely, co-citation, bibliographic coupling, and main path analysis. Guided by our three research questions, we provide a systematic overview of conceptual modeling research with a transverse and a temporal perspective. The comparisons of the different analyses, using mainly citation and co-citation links, enable us to elicit the dynamics of the field. Adapting and extending detailed questions from [127], Table 1 summarizes how these three types of analysis are appropriate for addressing our research questions.

Data extraction. Table 2 summarizes the inclusion criteria we used to define our data extraction strategy. We queried the Scopus bibliographic database,[2] since it offers the largest amount of indexed data in our field. The phrase, conceptual model, is widely accepted and used in multiple fields. Therefore, we prefered to use *conceptual modeling* (with one or two ls) and added *entity-relationship model* as search chains. However, in the *International Conference on Conceptual Modeling (ER*), since conceptual modeling is the universe of discourse, these phrases are often not mentioned. Therefore, in these outlets (ER proceedings), we did not use these search chains, but selected all papers. The search for papers on conceptual modeling included papers from 1976 (Chen's [27] seminal paper) to 2022; papers published in the *International Conference on Conceptual Modeling (ER)* and associated workshop proceedings; papers published in journals, conference proceedings and LNCS book series explicitly addressing conceptual modeling (i.e. containing one of the search chains). The papers were required to be written in English and correspond to subject areas of computer science, management, and decision sciences.

Table 2
Inclusion criteria

Data source	Scopus
Keywords	Conceptual model(l)ing; entity-relationship model
Querying search	Title, author keywords, abstract
Sources	Journals and conference proceedings
Domains	Management, Computer science, Decision sciences
Time span	1976-2022
Language	English

Three queries were used to generate the datasets.

- Dataset 1 (DS1): Extract articles published in the proceedings of the **ER conferences and workshops** present in Scopus (1992-2022).
- Dataset 2 (DS2): Search for the **presence of the exact keywords**: conceptual modeling, conceptual modelling, entity-relationship model in the title, the abstract or the author keywords of the **journal** articles or **conferences** published in English between 1976 and 2022 and referenced in Scopus.
- Dataset 3 (DS3): Search for the presence of the **same keywords** in articles published between 1976 and 2022 in the **LNCS series** to retrieve articles from conferences such as CAiSE, MODELS, PoEMS, RCIS and referenced in Scopus.

Data preparation. Merging the three datasets resulted in 4,652 documents (citing papers). The number of cited references is 102,568. The datasets were cleansed and merged in preparation for the next steps using CRExplorer [111] which supports the merging of datasets extracted from Scopus. The result is the elimination of duplicate cited articles (reduced from 102,568 to 70,747) and the transformation into WoS (Web of Science) format, since this is the best format for most of software employed: Pajek, BibExcel, VOSviewer.

Preliminary Analysis. Figure 3 depicts the distribution of publications over the entire period. Starting our analysis in 1976 is relevant because, before 1979, conceptual modeling research was not significant. Analyzing this distribution curve leads to the identification of three main periods. During the first period that emerged, the total number of publications is 99; an average of seven per year. The first *International Conference on Conceptual Modeling (ER)* in 1979 seems to have triggered the publication of papers in journals and other conferences starting in 1980. Note that the number of publications is minimized in our dataset since the ER proceedings during this period (until 1992) are

[2] https://www.scopus.com/home.uri

not referenced in Scopus (or in WoS). The curve takes a definite turn from 1989 onwards, as the number of publications explodes (111 per year on average), leading to the take-off of the conceptual modelling field. The dip in 1993 and the peak in 1994 have a very simple explanation: the proceedings of the 1993 ER conference were not published until the following year. As illustrated in Figure 3, the largest number of publications (252) is recorded in 2008 and accounts for 5.4% of the total documents. The third period (from 2009) shows a decline in the number of publications. It is possible that authors are now using more precise keywords, as conceptual modeling has matured in both research and professional practice. The figures for 2020 to 2022 may result from the impact of the Covid-19 pandemic on research activity. Also, the figures for 2022 are not definitive, because articles are still being published.

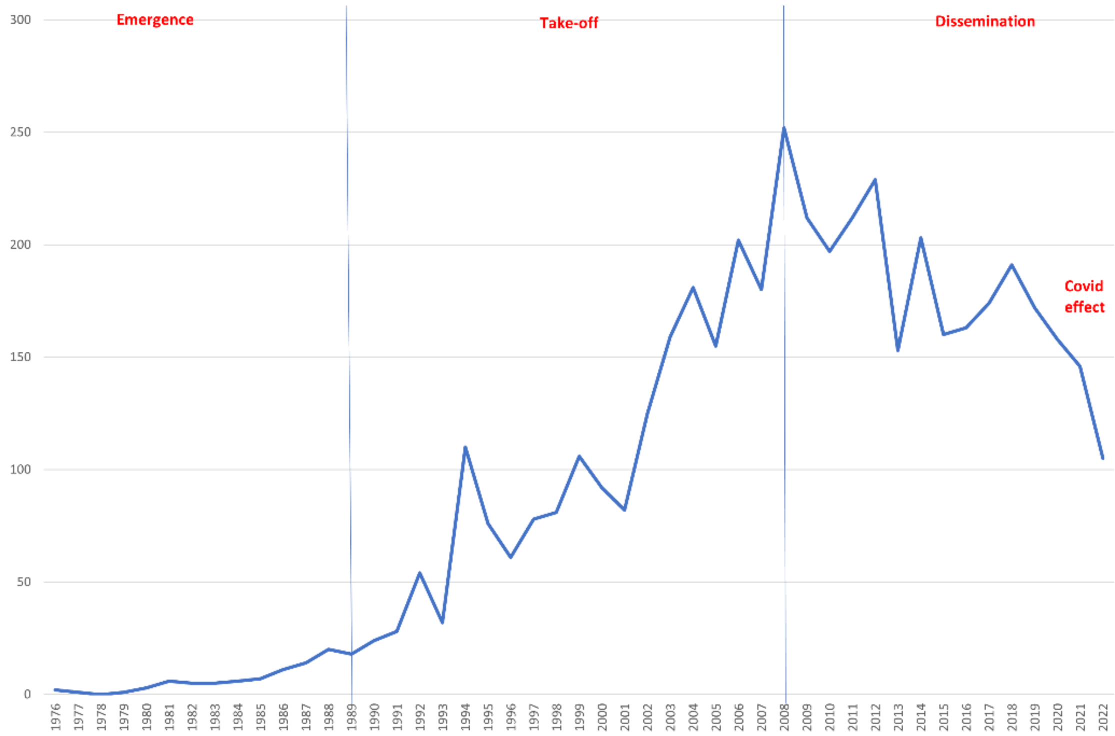

Figure 3: Distribution of conceptual modeling publications over time

The graph showing the average number of citations per year (number of articles*number of citations, normalized by the number of years since) is not provided for space reasons. It is flattened by the 1976 outlier which marks Chen's [27] article, which is by far the most cited, whether in or out of the sample. After removing this outlier, year 1993 marks a slight outlier, which can be explained partly by the seminal paper of Dardenne [33] on goal-directed requirements acquisition. The most relevant source is the book series *Lecture Notes in Computer Science* (LNCS), which represents 56% of the entire publication set. It contains papers from the *International Conference on Conceptual Modeling* and associated workshops that we systematically (without filter with search chain) inserted in the dataset. Also notable is the presence of 58 articles from the Elsevier *Information Systems* journal in the dataset.

In terms of the most cited sources, LNCS is also the first, followed by *ACM Transactions on Database Systems* (due to Chen's [27] paper), Springer (other than LNCS), Addison-Wesley, *Information Systems*, *Communications of the ACM*, *Data & Knowledge Engineering*, *MIS Quarterly*, *IEEE Transactions on Software Engineering* and *Information Systems Research*. In terms of the most local (within the dataset) cited papers, Chen's [27] paper leads by far (488 citations), followed by the PhD thesis of Guizzardi [55] (186 citations), Wand et Weber's [118] *Information Systems Research* commentary (138 citations), and Batini et al.'s [8] conceptual database design book (137 citations). Countries having at least 150 papers in our dataset are, in descending order, USA, Germany, Italy, Spain, Canada, Australia, Brazil, France, Netherlands, China, Belgium, UK, and Austria. The most frequent author keywords in the dataset are (in descending order): conceptual modeling, ontology, conceptual model, entity-relationship model, database design, UML, requirements engineering, and data warehouse. The analysis of frequent keywords over time also shows that conceptual modeling increased until 2008, and somewhat stagnated thereafter.

The longitudinal analysis described above provides a broad overview of the CM field and offers some insights for authors seeking publication venues.

4.1. Schools of thought identification (RQ1)

The structuring of the conceptual modeling field is reflected by grouping researchers around frequently cited articles. These co-citations provide a way to identify commonalities among a set of publications. They thus reflect a *school of thought* that brings together research efforts that are based on the same body of knowledge. Clustering the dataset based on the number of articles cited, in common, leads to the identification of such schools of thought. For the co-citation analysis, we use VOSviewer software, for a variety of reasons including the quality of the visualizations it produced, the performance of the software, and its interoperability with CRExplorer.

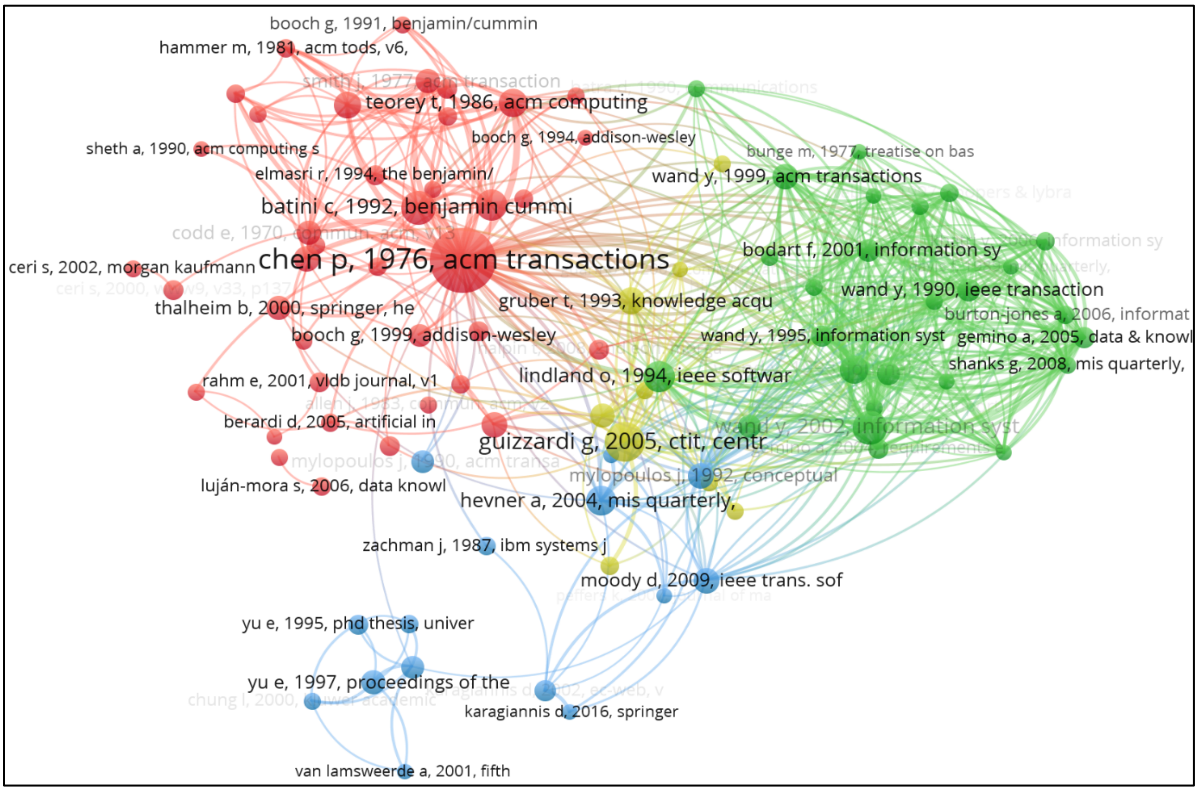

Figure 4: Main schools of thought[3]

[8, 14, 16-18, 23, 24, 25, 27, 38, 46, 52, 55, 62, 64, 69, 73, 76, 83, 92, 103, 105, 108, 109, 115-117, 123, 124, 126]

VOSviewer [112] allows us to first select the cited papers on which it will compute the common citations. There is no guideline regarding the number of papers to consider or the minimum number of citations a paper must have. Therefore, selecting a good setting requires sufficient knowledge of the research domain. We successively tested different settings. Finally, cited references were selected if they were cited at least 30 times (locally in the dataset), resulting in 89 papers. This step also required testing with different values of the VOSviewer resolution parameter leading to a certain level of clustering. The larger the value of the resolution parameter, the larger the number of clusters that are obtained. We opted for a resolution equal to one and obtained the four clusters shown in Figure 4.

In Figure 4, the size of the nodes is related to the number of local citations of the paper, with one color representing one cluster. Note that the constract induced by the size of the nodes had to be reduced to avoid the predominance of highly cited articles (e.g. Chen [27] overpowering the visual importance

[3] There are some observed data quality issues related to the years in the source.

of the others). The thickness of the lines represents the extent to which these references were co-cited. For readability, very thin lines are omitted. We identified four schools of thought that included contributions of: *1) Chen [27], 2) Wand and Weber [118], 3)* Hevner et al. [64], and 4) *Guizzardi [55].* The main cluster (39 papers, in red), emerged, as expected, from Chen's [27] introduction of the entity-relationship model, which provided a simple, usable, and understandable set of constructs that could be used to abstract and model the real world. This paper serves as a well-known defining point in conceptual modeling. An interesting feature of this cluster is the high number of books on databases and object-orientation [1, 25, 38-40, 86, 98, 99, 109]. The second cluster (26 papers, in green) can be described as focusing on the quality of conceptual modeling. Its central paper of Wand and Weber [118] introduced ontology to analyze and understand concepts in conceptual modeling. The third cluster (15 papers, in blue) gathers papers that refer to design science research as well as to requirements engineering, proposing methods such as iStar [125]. Hevner et al.'s paper [64] is central in the graph of Figure 4. It is a seminal paper on design science research, which inspires the work of this cluster and others in the other clusters. Recall that co-citation analysis clusters papers referenced by conceptual modeling papers; these referenced papers may not be directly related to conceptual modeling. Finally, the fourth cluster (9 items in yellow) combines ontologies and conceptual modeling. The most influential paper in this cluster is Guizzardi's work on ontology [55], which defines the general ontological foundations for conceptual modeling. Although diverse, these four schools of thought reflect the composition of work in the field. In short, Cluster 1 corresponds to ***conceptual modeling and databases*** in all their forms. Cluster 2 contains contributions in terms of ***grammars and guidelines for conceptual modeling***. Cluster 3 contains mainly ***requirements engineering models and methods***. Finally, Cluster 4 is composed of ***ontology constructs for conceptual modeling.***

4.2. Research themes identification (RQ2)

Co-citation analysis builds on papers that are most cited, but does not consider recent papers, without an opportunity to be well-cited. Scientometric experts recommend complementing it with other analyses. Thus, we performed a bibliographic coupling analysis (BCA) to identify recent themes of conceptual modeling research. Bibliographic coupling groups together articles sharing common references. More precisely, the more bibliographic references two articles share, the closer they are considered to each other. Clusters are built on this similarity.

The BCA is based on 943 papers, a subset of our previous sample limited to papers published over the six last years (2017→2022). We obtained the best result with 10 clusters using VOSviewer, retaining all papers cited at least 5 times, and leading to a connected component of 232 papers (Figure 5). VOSviewer does not allow the specification of a fixed number of clusters. Rather, the user provides a resolution value. The higher the resolution, the larger the number of clusters. Therefore, we used a resolution of 0.70, which generated 10 clusters. Clustering does not necessarily bring together articles that adopt the same approaches, but rather, articles that address the same subjects. In a field such as conceptual modeling, the BCA distinguishes the different application sub-fields. From 2017 to 2022, we obtain 10 such application sub-fields. The size of the cluster represents the importance of the sub-field, as well as the importance of the associated community, since these articles share more references.

The clusters are shown in Table 3. It is not surprising that the first cluster that emerged is related to work on domain-specific approaches, which focus on adapting conceptual modeling to a specific domain and accumulating a body of knowledge for this sub-field. For example, Bork [19] focused on conceptual modeling in smart city domains. Cluster 2 focuses on ontologies, as such, or projected onto application domains. Cluster 6 groups together CM applications to different domains, with possible recourse to ontologies. Cluster 3 is dedicated to requirements engineering, including goal-oriented approaches. Cluster 4 focuses on processes. Cluster 5 brings together contributions on database modeling, mainly relating to the NoSQL world, as these are recent publications. Cluster 7 focuses on the application domain of healthcare, including biology. Cluster 8 contains the most theoretical work on CM. In cluster 9, researchers adopt multi-level modeling. Finally, cluster 10 merges temporal and economic approaches requiring conceptual modeling.

We conducted an additional BCA on previous periods and could identify other application domains of conceptual modeling, for example database design in the early periods or Web design between 2000 and 2010. This proves that conceptual modeling has been around for almost fifty years, and it is still very adaptable. Not only has it kept pace with the evolution of databases (its first field of application and is now present in cluster 5), but it has also broadened its field of application to include, for example, process modeling or genomics.

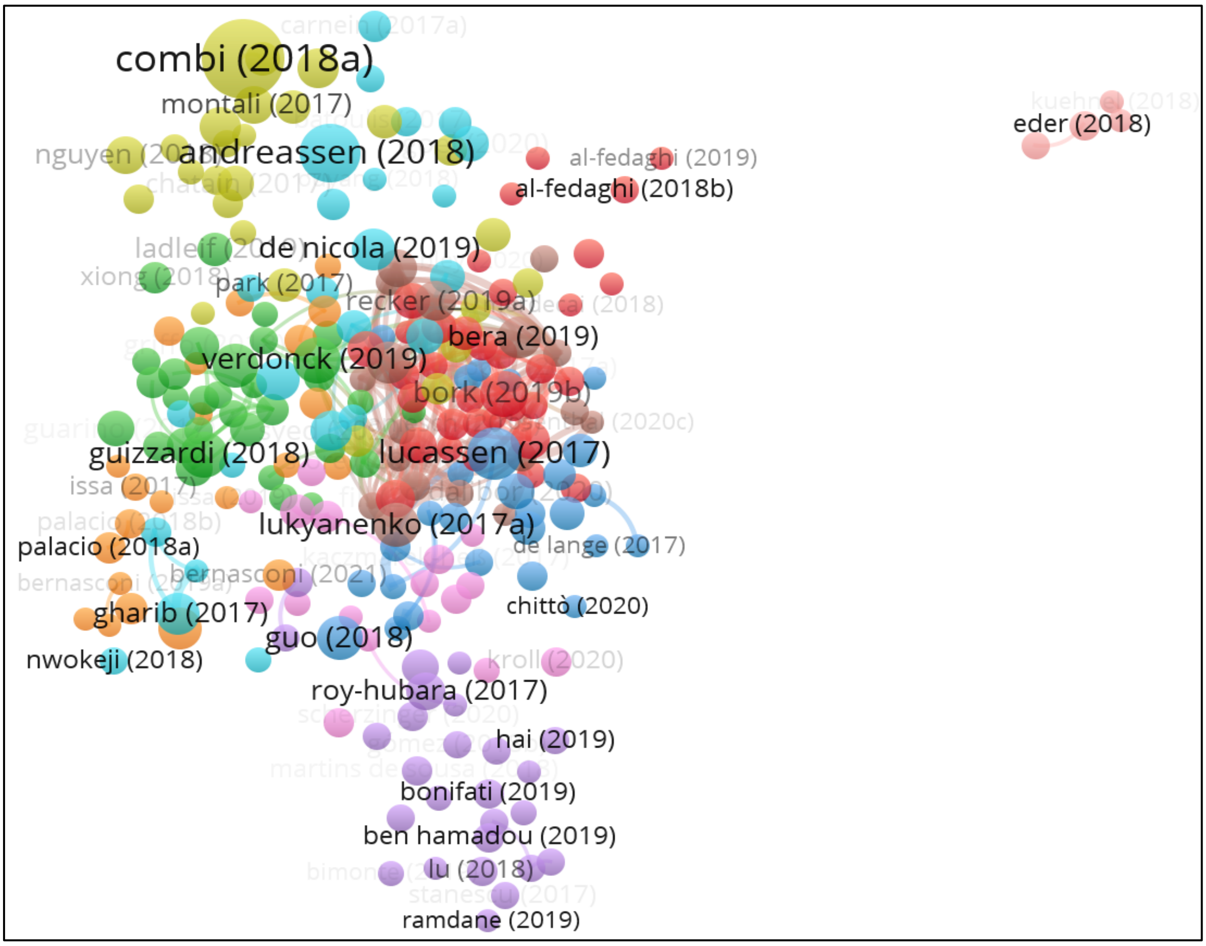

Figure 5: Identification of themes
[3, 7, 11, 12, 15, 20, 29, 30, 35, 37, 59-61, 74, 75, 77, 82, 85, 87, 93, 94, 97, 113]

Table 3
Clusters of conceptual modeling topics

Cluster	Papers	Description	Representative papers
1	35	Domain-specific CM and applications	Bork [19, 21], Buchmann [22], Delcambre [36]; Fill [44], Gonzalez-Perez [49] Johannsen [67]
2	32	Ontologies and applications	Guizzardi [56], Verdonck [113], Sales [100], Griffo [51], Gu [53], Guarino [54]
3	27	Goal models and requirements engineering, incl. digital twins and chatbots	Lucassen [75], Guo [60], Becker [10], Dalibor [32]
4	26	Process model including process modeling, process mining, process behavior including learning and pedagogy	Combi [30], Yeshchenko [122], Nguyen [84]
5	26	Data and databases: datastore, NoSQL, multimodel data	Roy-Hubara [97], De la Vega [34], Bonifati [15]
6	25	Applications with or without ontologies	Andreassen [6], Lukyanenko [68], de Nicola [35]

7	22	Genomics, healthcare, medicine, Covid	Bernasconi [13], Palacio [87], Issa [66]
8	18	Understanding CM; theoretical developments	Kaczmarek-hess [68], Lukyanenko [77], Recker [94], Wand [119]
9	17	Multilevel models	Almeida [5], Giebler [47]
10	4	Temporal or economic view of the field	Eder [37], Kuehnel [70, 71], Pichler [91]

Note: The analysis provides only the last name of the first author.

4.3. Main paths identification (RQ3)

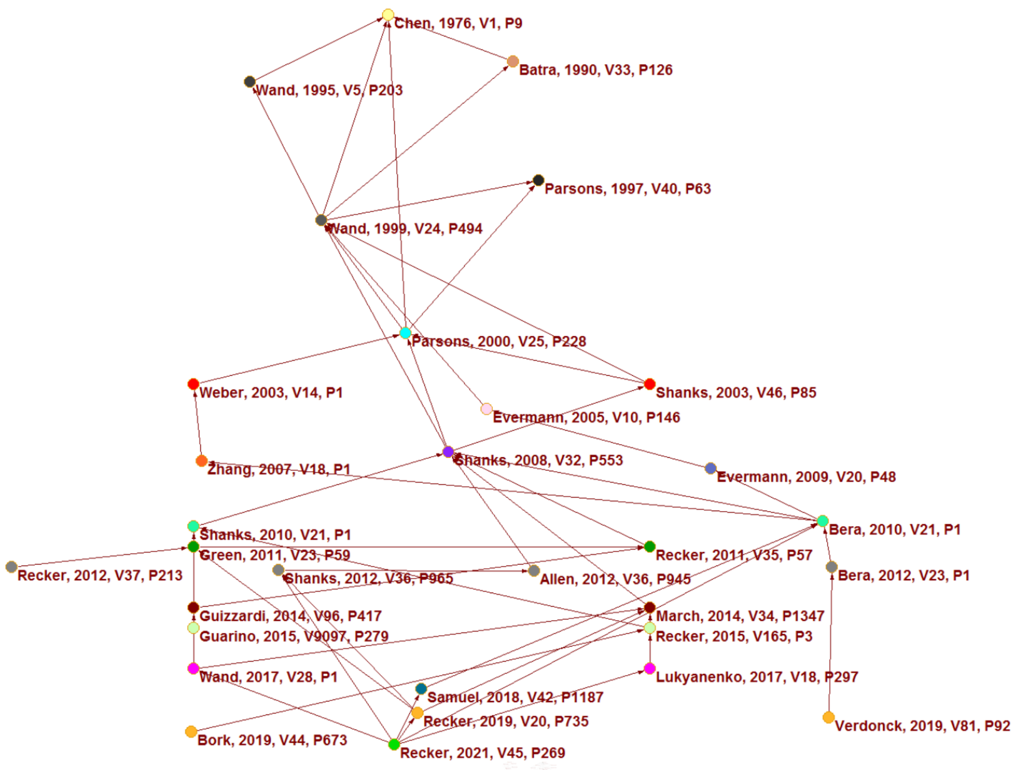

Figure 6: Main path analysis based on 30 key routes and SPC metric[4]
[4, 9, 20, 27, 42, 78, 88, 89, 94, 95, 101, 102, 113, 115, 119, 120]

We used Main Path Analysis (MPA) to identify the most influential publications in the CM field. We, therefore, constructed the citation network using Scopus and mapped it to identify the most important paths of influence among the publications. Three parameters must be adjusted. First, the edges are labeled with centrality measures. The choice of a centrality measure provides a different perspective on the importance of nodes and links in the network. We tested different measures and different path calculation algorithms with different settings. It appears that the choice between SPC, SPLC and SPNP metrics (as offered by the Pajek software) does not clearly impact the results. Thus, we chose SPC, as recommended and employed by other researchers. The SPC metric is the number of times a citation link is traversed when exhausting the search from the source nodes to the sink nodes. A high SPC value characterizes a citation link that helps a great deal in knowledge diffusion. Second, several algorithms (mainly local search and global search) may be used to compute these metrics. In terms of the algorithm, a local search provided a richer graph, as confirmed by several authors. A third parameter is the number of top citation links (called key-routes) that we want to include at least in the graph. We applied the graph calculation on 30 key-routes to express the maximum possible paths, while remaining readable. The result is shown in Figure 6.

[4] The output is given in terms of only the last name of the first author. The same labeling is used by us in the text when referencing the nodes. There were some observed quality issues with the year in Scopus.

Analysis of the graph in Figure 6 reveals three main dynamics that follow one another over time. The first dynamic, which we call "**refinement of ER model**", integrates the paths from the Chen 1976 [27] node to the Parsons 2000 [89] node. It includes Batra 1990 [9], who compares the representational capacity of the entity-relationship model and the relational model. Wand 1995 [117] proposes the conceptual modeling requirements that information systems must meet. Parsons 1997 [88] is the first source article for this graph that does not cite Chen 1976 [27]. The article proposes guidelines for facilitating the choice of object classes to be modeled in an effort to represent a domain. Wand 1999 [115] complements Parsons 1997 [88] with guidelines for relationships in entity-relationship conceptual modeling. Finally, Parsons 2000 [89] proposes a two-level modeling of classes and their instances.

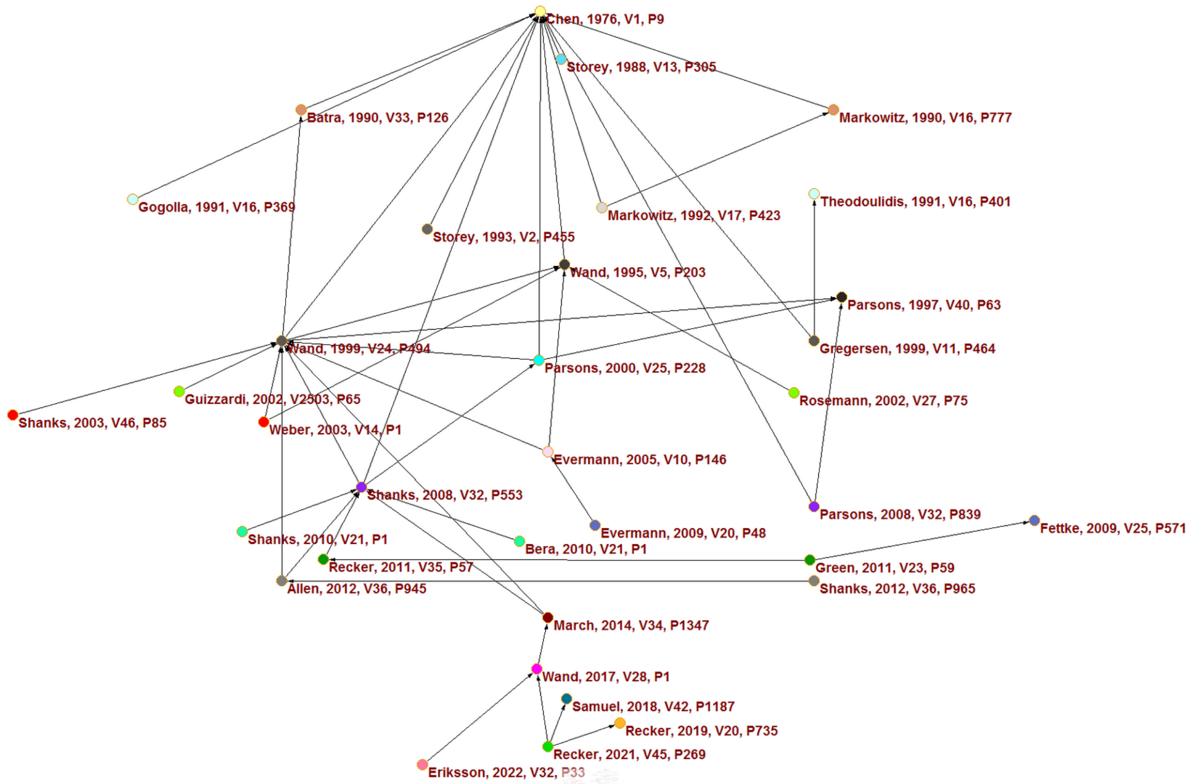

Figure 7: Alternative path analysis
[4, 9, 27, 41-43, 48, 50, 57, 78, 79, 88, 89, 95, 96, 101-104, 106, 107, 110, 115, 117, 119, 120]

The second dynamic is what we call "**improvement of conceptual modeling with ontologies**". This dynamic is much more extensive in terms of the number of contributions and the number of links between them. The crossroads of many paths are Shanks 2003 [104], Shanks 2008 [104], Recker 2011 [95], Guizzardi 2014 [58], and Wand 2017 [119]. The introduction of ontologies to facilitate and make more robust conceptual modeling is an important and perennial phenomenon, initiated gradually from the early 1990s. Shanks 2003 [104] is the first paper in these main paths to propose using ontologies to validate conceptual models. Shanks 2008 [103] focuses on the validation of part-whole relationships using the same ontological approach. Finally, Guizzardi 2014 (in fact, published in 2015 in Applied Ontology [58]) describes the foundational ontology UFO and the associated language OntoUML. A third dynamic, which we call "**conceptual modeling for digital world,**" seems to be emerging from 2017, with focus on more recent concepts such as crowdsourcing (Lukyanenko 2017 [77]), open platforms (Bork 2019 [20]), mixing physical and digital realities (Recker 2021 [95]).

The graph in Figure 6 was built using SPC metrics. In SPC (as well as SPLC or SPNP metrics), citation link weights denote information flows, rather than similarity between nodes. This is a well-known limitation [90]. Therefore, we used another metric namely WDC [90]. The latter stands for "Weighted Directed Citation". It enriches SPC by considering the fact that the source and target nodes of a link share common references or are co-cited. Thus, this WDC metric combines information flow with additional semantics. We thus generated an alternative main path as shown in Figure 7. This main path is interesting because it identifies the same 10-12 influential papers of the other graph (Figure 6).

These are also the papers found in all the tests performed in this research. In addition, it enriches the main path in Figure 6 with paths that are somewhat different. Overall, this additional analysis provides the basis for an expanded interpretation of this body of work.

A comparison of the two graphs (Figure 6 and Figure 7) thus generated shows the specificity of the WDC metric, which tends to show publications that are present because of their impact, but which do not follow in the graph insofar as this influence is more localized in time (e.g. [48, 79, 107]). Apart from these few new nodes that appear at different places in the graph (and therefore in time), another difference is the proportion of nodes per epoch. Thus, in the graph of Figure 6, more than three quarters of the nodes are articles published after 2000, whereas in Figure 7, the year 2000 appears to mark the median year of the graph (in the sense that we find almost as many publications before 2000 as after). A more recent new node is Fettke 2009 [43] (in Figure 7), which investigates the use of conceptual modeling by practitioners. Eriksson [41] shows the continuity of the ontological stream. Thus, the common main paths of both graphs may be insightful.

Identifying a clear path for many papers is difficult and leads to challenges in trying to "disentangle" the graph. All the contributions represented in Figures 6 and 7 play an important role in the chain of diffusion of knowledge in conceptual modelling. If we try to identify common paths that integrate the two points of view expressed respectively by the SPC and WDC metrics, we can superimpose the two previous graphs and extract the longest common paths. This results in the two paths shown in Figure 8.

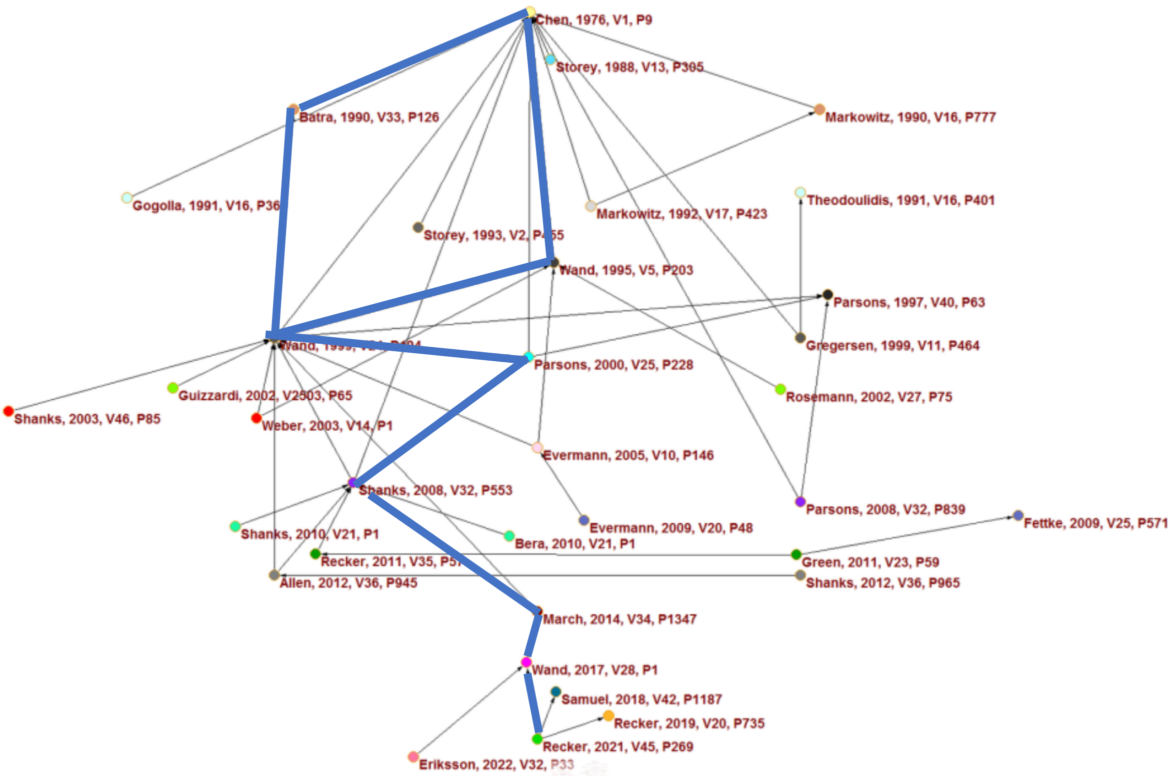

Figure 8: Two possible main paths combining SPC and WDC [9, 27, 78, 89, 95, 103, 115, 117, 119]

4.4. Interpretation

By comparing and confronting the different bibliometric analyses, a periodization of the progress of conceptual modeling research emerges. The longitudinal study highlighted three periods of: *emergence*, *take-off* and *dissemination* of conceptual modeling. The emergence of conceptual modelling (1976-1989) is confirmed by main path analysis, which reveals there are no articles in our dataset between Chen [27] and Storey and Goldstein [107], regardless of the metric used. The take-off period (1989-2008) is prolific in terms of the number of contributions. It includes an initial sub-period of refinement of the entity-relationship model concepts (1989-1995), followed by the appearance of references to ontologies until their full use (1995-2008). The third period of dissemination (starting in 2008) saw the continuation of work on ontologies for conceptual modeling (see main path analysis), as well as the emergence of numerous application themes for conceptual modeling revealed by the bibliographic coupling analysis. Finally, the four schools of thought identified through the co-citation analysis appeared at the start of the take-off phase and continue to this day.

Table 4
Tentative cross-fertilization of the analyses

<table>
<tr><th></th><th>Periodization</th><th colspan="4">Co-citation</th><th colspan="2">Main path analysis</th><th colspan="3">Bibliographic coupling</th></tr>
<tr><td>1976-1989</td><td>Emergence</td><td rowspan="5">CM and databases</td><td rowspan="5">Grammars & guidelines for CM</td><td rowspan="5">Requirements engineering models & methods</td><td rowspan="5">Ontology constructs for CM</td><td colspan="2">Emergence</td><td rowspan="5">Theoretical developments</td><td rowspan="5">Applications</td><td rowspan="5">Other more specific clusters</td></tr>
<tr><td>1989-1995</td><td rowspan="2">Take-off</td><td colspan="2">Refinement of ER model</td></tr>
<tr><td>1995-2008</td><td rowspan="3">Ontologies</td><td rowspan="2"></td></tr>
<tr><td>2008-2017</td><td rowspan="2">Dissemination</td></tr>
<tr><td>2017-2022</td><td>CM for digital</td></tr>
</table>

Table 4 summarizes the main results of the various bibliometric analyses carried out on the conceptual modeling dataset. The frontier years of the first column are, of course, approximate. The table enables us to compare the different periods or themes revealed by these analyses. Thus, the emergence phase revealed by the longitudinal study finds its equivalent in the main path analysis. As the bibliographic coupling was mainly done over the last 6 years, we have only presented three generic clusters that cover the entire period: the one that brings together theoretical developments around CM; the one that brings together applications; and, finally, other more specific themes. This table provides a cross-fertilization of analyses that merit further development and will be pursued in our future research. It can be seen as the first step of validation of our findings.

5. Discussion

Guided by an objective to understand and appreciate the vast amount of research that has been carried out on conceptual modeling, since the beginning of its history, this research conducted a bibliometric analysis of papers to identify their impact and how they are positioned with respect to the field and with respect to each other. We were particularly interested in identifying: the different periods of research as the field progressed to identify what we could learn from them; and a trajectory for continued research in conceptual modeling. Three types of analysis were applied: co-citation analysis, bibliographic coupling, and main path analysis. The latter was central to our understanding of the main topics, around which others can be organized.

With respect to our research questions, for RQ1 (intellectual structure of conceptual modeling), four interesting schools of thought emerged, each of which can be traced to specific articles related respectively to the entity-relationship model formulation [27] in Cluster 1, ontology in information systems [116] in Cluster 2, the use of conceptual modeling in design science research [64] and conceptual modeling for requirements engineering [124] in Cluster 3, and ontology in computer science [55] in Cluster 4. The four clusters contain many solution-oriented articles (proposing a novel solution for a problem), some books, and vision-oriented publications (e.g., proposing research agendas). Few focus on evaluation or experimentation.

For RQ2 (main themes of conceptual modeling research), ten clusters of research topics emerged from our analysis, which were supported by multiple papers. These were based on more current research, and thereby provide insights into ongoing efforts, which would be of particular interest to students or researchers new to the field.

For RQ3 (paths linking the most influential publications), our path analysis showed that, not surprisingly, there are main paths of research on conceptual modeling that stemmed from the seminal work of Chen [27], which established the field. Less obvious is that, as shown through additional analysis, there is not a unique path, but rather several paths providing a historical timeline.

The overall results show that, despite the current trend of less publications in the field of conceptual modeling, it is still very viable and adaptable to many different application domains. The topics that have been identified reveal a good, but diverse, set of interests. This adaptability is important for both teaching conceptual modeling (and conceptual modeling research seminars) and initiating further research, by either new or experienced students and researchers.

5.1. Limitations

There are several limitations. There is no universal definition of conceptual modeling, making it difficult Conceptual modeling has not a universal definition, making the choice of appropriate keywords questionable. Moreover, we used Scopus to identify a dataset. Of course, it might have provided more, or less, papers than required, due to the difficulty of identifying exact keywords. For example, ER conference proceedings are not indexed by Scopus before 1992. Many pieces of software were used that required many parameters to be set correctly. Since there are very few guidelines available to help select the metrics, the algorithms, and the values of the parameters, we adopted an empirical approach and relied on our personal knowledge and understanding of the conceptual modeling field. Given the sample size of the papers, it would have been impossible to obtain any meaningful results manually. Therefore, we relied on different types of software that might have introduced biases into their algorithms. Finally, since this is a bibliometric study, we focused on past research categorization to help understand the structure and dynamics of a field and its possible evolutions.

Co-citation analysis, bibliographic coupling, and main path analysis have certain limitations. As noted, the data quality may not always be satisfactory. Missing or incorrectly cited documents may affect the accuracy and the completeness of the citation and co-citation graphs. Bibliographic coupling is performed on a limited time window (here six years as generally recommended). The inclusion criteria used to select the data impact the final dataset. For example, conceptual modeling may have synonyms that were omitted. The bibliometric techniques focus on citation and co-occurrence relationships between documents, while ignoring the semantic of these relationships. It is well-known that sometimes the use of self-citations, although relevant, may accentuate the importance of certain links in the calculation of main paths. Finally, the analyses and the visualizations highlight only the first authors, due to the limitations of the software we used.

As far as threats to validity are concerned, we refer to Wohlin et al. [121] who proposed a set of validity threats; namely construct validity, external validity, internal validity, and conclusion validity. *Construct validity* here relates to the choice of the bibliometric techniques to address the research questions. The results obtained with these quantitative techniques depend partly on the choice of the associated parameters. We do not consider *external validity* at this stage, since we did not check the choices (in terms of algorithms, thresholds, metrics, etc.) and the results with experts in the field. With respect to *internal validity*, as already mentioned, there may be some bias introduced into the inclusion criteria. Finally threats to *conclusion validity* are limited by the combination of several techniques.

5.2. Conclusion and future research

This paper has proposed that research on conceptual modeling has matured to a stage where it is possible to identify its themes and transformative research. We conducted a bibliometric analysis composed of six steps: research design, data extraction, data preparation, preliminary analysis, three bibliometric analyses with respect to the three research questions, and interpretation. This study considers a span from 1976 to 2022 to include many papers encompassing conceptual modeling from journals and conference proceedings. The results show the progression of topics that emerged and could be useful to educate the next generation of researchers and to guide future research efforts.

A similar, future bibliometric analysis could be useful to researchers to identify additional publication topics and target outlets. This research is intended to be an initial start. Further research could help to refine the main periods identified, as well as the clustering (and possible sub-clustering) of the literature in the conceptual modeling field. Each bibliographic analysis result would benefit from being enriched using different research taxonomies to detect the possible specificity of certain clusters as well as the absence of certain types of research (for example evaluation or experimentation, which are sometimes not given a high priority in conceptual modeling research). Our efforts should be replicable since all the relevant steps and the details of their composition are explained.

References

[1] Abiteboul, S., R. Hull, and V. Vianu, Foundations of databases. Vol. 8. 1995: Addison-Wesley Reading.

[2] Aguirre-Urreta, M.I. and G.M. Marakas, Comparing conceptual modeling techniques: a critical review of the EER vs. OO empirical literature. ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 2008. **39**(2): p. 9-32.

[3] Al-Fedaghi, S. and Y. Alduwaisan. Modeling of an enterprise and information system: process specification based on the flow of things. in Proceedings of the International Conference on Geoinformatics and Data Analysis. 2018.

[4] Allen, G.N. and S.T. March, A research note on representing part-whole relations in conceptual modeling. MIS Quarterly, 2012: p. 945-964.

[5] Almeida, J.P.A., C.M. Fonseca, and V.A. Carvalho. A comprehensive formal theory for multi-level conceptual modeling. in Conceptual Modeling: 36th International Conference, ER 2017, Valencia, Spain, November 6–9, 2017, Proceedings 36. 2017. Springer.

[6] Andersson, A. and J. Krogstie. Implementation and first evaluation of a molecular modeling language. in Enterprise, Business-Process and Information Systems Modeling: 16th International Conference, BPMDS 2015, 20th International Conference, EMMSAD 2015, Held at CAiSE 2015, Stockholm, Sweden, June 8-9, 2015, Proceedings. 2015. Springer.

[7] Andreassen, T.W., et al., Business model innovation and value-creation: the triadic way. Journal of Service Management, 2018.

[8] Batini, C., S. Ceri, and S. Navathe, Conceptual design–an entity-relationship approach. Redwood: Benjamin Cummings, 1992.

[9] Batra, D., J. Hoffler, and R.P. Bostrom, Comparing representations with relational and EER models. Communications of the ACM, 1990. **33**(2): p. 126-139.

[10] Becker, F., et al. A conceptual model for digital shadows in industry and its application. in Conceptual Modeling: 40th International Conference, ER 2021, Virtual Event, October 18–21, 2021, Proceedings 40. 2021. Springer.

[11] Ben Hamadou, H., E. Gallinucci, and M. Golfarelli. Answering GPSJ queries in a polystore: A dataspace-based approach. in Conceptual Modeling: 38th International Conference, ER 2019, Salvador, Brazil, November 4–7, 2019, Proceedings 38. 2019. Springer.

[12] Bera, P., A. Burton-Jones, and Y. Wand, Improving the representation of roles in conceptual modeling: theory, method, and evidence. Requirements Engineering, 2018. **23**: p. 465-491.

[13] Bernasconi, A., et al. Conceptual modeling for genomics: building an integrated repository of open data. in Conceptual Modeling: 36th International Conference, ER 2017, Valencia, Spain, November 6–9, 2017, Proceedings 36. 2017. Springer.

[14] Bodart, F., et al., Should optional properties be used in conceptual modelling? A theory and three empirical tests. Information Systems Research, 2001. **12**(4): p. 384-405.

[15] Bonifati, A., et al. Schema validation and evolution for graph databases. in Conceptual Modeling: 38th International Conference, ER 2019, Salvador, Brazil, November 4–7, 2019, Proceedings 38. 2019. Springer.

[16] Booch, G., Object oriented design with applications. 1990: Benjamin-Cummings Publishing Co., Inc.

[17] Booch, G., D.L. Bryan, and C.G. Petersen, Software engineering with Ada. Vol. 30608. 1994: Addison-Wesley Professional.

[18] Booch, G., J. Rumbaugh, and I. Jacobson, The unified modeling language reference manual. 1999.

[19] Bork, D., et al. Using conceptual modeling to support innovation challenges in smart cities. in 2016 IEEE 18th International Conference on High Performance Computing and

Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS). 2016. IEEE.
[20] Bork, D., et al. An open platform for modeling method conceptualization: the OMiLAB digital ecosystem. 2019. Association for Information Systems.
[21] Bork, D., D. Karagiannis, and B. Pittl, A survey of modeling language specification techniques. Information Systems, 2020. **87**: p. 101425.
[22] Buchmann, R.A. and A.-M. Ghiran. Serviceology-as-a-service: a knowledge-centric interpretation. in Serviceology for Services: 5th International Conference, ICServ 2017, Vienna, Austria, July 12-14, 2017, Proceedings 5. 2017. Springer.
[23] Bunge, M., Treatise on basic philosophy: Ontology I: the furniture of the world. Vol. 3. 1977: Springer Science & Business Media.
[24] Burton-Jones, A. and P.N. Meso, Conceptualizing systems for understanding: An empirical test of decomposition principles in object-oriented analysis. Information Systems Research, 2006. **17**(1): p. 38-60.
[25] Ceri, S., P. Fraternali, and M. Matera, Conceptual modeling of data-intensive Web applications. IEEE Internet computing, 2002. **6**(4): p. 20-30.
[26] Chen, C., I.-Y. Song, and W. Zhu. Trends in conceptual modeling: Citation analysis of the ER conference papers (1979-2005). in Proceedings of the 11th International Conference on the International Society for Scientometrics and Informatrics. 2007.
[27] Chen, P.P.-S., The entity-relationship model—toward a unified view of data. ACM transactions on database systems (TODS), 1976. **1**(1): p. 9-36.
[28] Chen, P.P. Thirty years of ER conferences: Milestones, achievements, and future directions. in Conceptual Modeling-ER 2009: 28th International Conference on Conceptual Modeling, Gramado, Brazil, November 9-12, 2009. Proceedings 28. 2009. Springer.
[29] Chittò, P., et al. Automatic generation of chatbots for conversational web browsing. in Conceptual Modeling: 39th International Conference, ER 2020, Vienna, Austria, November 3–6, 2020, Proceedings 39. 2020. Springer.
[30] Combi, C., et al. Conceptual modeling of processes and data: Connecting different perspectives. in Conceptual Modeling: 37th International Conference, ER 2018, Xi'an, China, October 22–25, 2018, Proceedings 37. 2018. Springer.
[31] Cosentino, V., J.L. Cánovas Izquierdo, and J. Cabot. Metascience: An holistic approach for research modeling. in Conceptual Modeling: 35th International Conference, ER 2016, Gifu, Japan, November 14-17, 2016, Proceedings 35. 2016. Springer.
[32] Dalibor, M., et al. Towards a model-driven architecture for interactive digital twin cockpits. in Conceptual Modeling: 39th International Conference, ER 2020, Vienna, Austria, November 3–6, 2020, Proceedings. 2020. Springer.
[33] Dardenne, A., A. Van Lamsweerde, and S. Fickas, Goal-directed requirements acquisition. Science of computer programming, 1993. **20**(1-2): p. 3-50.
[34] de la Vega, A., et al., Mortadelo: Automatic generation of NoSQL stores from platform-independent data models. Future Generation Computer Systems, 2020. **105**: p. 455-474.
[35] De Nicola, A., M. Melchiori, and M.L. Villani, Creative design of emergency management scenarios driven by semantics: An application to smart cities. Information Systems, 2019. **81**: p. 21-48.
[36] Delcambre, L.M., et al. A reference framework for conceptual modeling. in Conceptual Modeling: 37th International Conference, ER 2018, Xi'an, China, October 22–25, 2018, Proceedings 37. 2018. Springer.
[37] Eder, J., et al. Expressiveness of temporal constraints for process models. in Advances in Conceptual Modeling: ER 2018 Workshops Emp-ER, MoBiD, MREBA, QMMQ, SCME, Xi'an, China, October 22-25, 2018, Proceedings 37. 2018. Springer.
[38] Elmasri, R. and S. Navathe, Fundamentals of database systems. 1994, Benjamin/Cummings Publishing Company.
[39] Embley, D.W., B.D. Kurtz, and S.N. Woodfield, Object-oriented systems analysis: a model-driven approach. 1991: Yourdon Press.
[40] Embley, D.W. and B. Thalheim, Handbook of conceptual modeling: theory, practice, and research challenges. 2012: Springer.

[41] Eriksson, O. and P.J. Ågerfalk, Speaking things into existence: Ontological foundations of identity representation and management. Information Systems Journal, 2022. **32**(1): p. 33-60.
[42] Evermann, J. and Y. Wand, Ontology based object-oriented domain modelling: fundamental concepts. Requirements engineering, 2005. **10**: p. 146-160.
[43] Fettke, P., How conceptual modeling is used. Communications of the Association for Information Systems, 2009. **25**(1): p. 43.
[44] Fill, H.-G., B. Pittl, and G. Honegger. A modeling environment for visual SWRL rules based on the SeMFIS platform. in Designing the Digital Transformation: 12th International Conference, DESRIST 2017, Karlsruhe, Germany, May 30–June 1, 2017, Proceedings 12. 2017. Springer.
[45] Frank, U., et al., The Research Field "Modeling Business Information Systems" Current Challenges and Elements of a Future Research Agenda. WIRTSCHAFTSINFORMATIK, 2014. **56**: p. 49-54.
[46] Gemino, A. and Y. Wand, Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties. Data & Knowledge Engineering, 2005. **55**(3): p. 301-326.
[47] Giebler, C., et al. Leveraging the data lake: Current state and challenges. in Big Data Analytics and Knowledge Discovery: 21st International Conference, DaWaK 2019, Linz, Austria, August 26–29, 2019, Proceedings 21. 2019. Springer.
[48] Gogolla, M. and U. Hohenstein, Towards a semantic view of an extended entity-relationship model. ACM Transactions on Database Systems (TODS), 1991. **16**(3): p. 369-416.
[49] Gonzalez-Perez, C. and P. Martin-Rodilla. An alternative approach to metainformation conceptualisation and use. in Conceptual Modeling: 36th International Conference, ER 2017, Valencia, Spain, November 6–9, 2017, Proceedings 36. 2017. Springer.
[50] Gregersen, H. and C.S. Jensen, Temporal entity-relationship models-a survey. IEEE Transactions on knowledge and data engineering, 1999. **11**(3): p. 464-497.
[51] Griffo, C., et al. Legal Power-Subjection Relations: Ontological Analysis and Modeling Pattern. in Conceptual Modeling: 41st International Conference, ER 2022, Hyderabad, India, October 17–20, 2022, Proceedings. 2022. Springer.
[52] Gruber, T.R., A translation approach to portable ontology specifications. Knowledge acquisition, 1993. **5**(2): p. 199-220.
[53] Gu, Y., V.C. Storey, and C.C. Woo. Conceptual modeling for financial investment with text mining. in Conceptual Modeling: 34th International Conference, ER 2015, Stockholm, Sweden, October 19-22, 2015, Proceedings 34. 2015. Springer.
[54] Guarino, N. On the semantics of ongoing and future occurrence identifiers. in Conceptual Modeling: 36th International Conference, ER 2017, Valencia, Spain, November 6–9, 2017, Proceedings 36. 2017. Springer.
[55] Guizzardi, G., Ontological foundations for structural conceptual models. 2005.
[56] Guizzardi, G., et al. Endurant types in ontology-driven conceptual modeling: Towards OntoUML 2.0. in Conceptual Modeling: 37th International Conference, ER 2018, Xi'an, China, October 22–25, 2018, Proceedings 37. 2018. Springer.
[57] Guizzardi, G., H. Herre, and G. Wagner. Towards ontological foundations for UML conceptual models. in On the move to meaningful internet systems 2002: CoopIS, DOA, and ODBASE: Confederated International Conferences CoopIS, DOA, and ODBASE 2002 Proceedings. 2002. Springer.
[58] Guizzardi, G., et al., Towards ontological foundations for conceptual modeling: The unified foundational ontology (UFO) story. Applied Ontology, 2015. **10**(3-4): p. 259-271.
[59] Guizzardi, R.S. and A. Perini. Goal-oriented Decision Modeling: A Position Paper. in iSTAR@ CAiSE. 2018.
[60] Guo, G. An active workflow method for entity-oriented data collection. in Advances in Conceptual Modeling: ER 2018 Workshops Emp-ER, MoBiD, MREBA, QMMQ, SCME, Xi'an, China, October 22-25, 2018, Proceedings 37. 2018. Springer.
[61] Hai, R., C. Quix, and D. Wang. Relaxed functional dependency discovery in heterogeneous data lakes. in Conceptual Modeling: 38th International Conference, ER 2019, Salvador, Brazil, November 4–7, 2019, Proceedings 38. 2019. Springer.
[62] Hammer, M. and D. Mc Leod, Database description with SDM: A semantic database model. ACM Transactions on Database Systems (TODS), 1981. **6**(3): p. 351-386.

[63] Härer, F. and H.-G. Fill. Past trends and future prospects in conceptual modeling-a bibliometric analysis. in Conceptual Modeling: 39th International Conference, ER 2020, Vienna, Austria, November 3–6, 2020, Proceedings 39. 2020. Springer.
[64] Hevner, A., et al., Design Science in Information Systems Research. MIS Quarterly, 2004. **28**(1): p. 75-105.
[65] Hummon, N.P. and P. Dereian, Connectivity in a citation network: The development of DNA theory. Social networks, 1989. **11**(1): p. 39-63.
[66] Issa, S., et al. Revealing the conceptual schemas of RDF datasets. in Advanced Information Systems Engineering: 31st International Conference, CAiSE 2019, Rome, Italy, June 3–7, 2019, Proceedings 31. 2019. Springer.
[67] Johannesson, P., M. Bergholtz, and O. Eriksson. Institution Aware Conceptual Modelling. in The ER Forum 2017 and the ER 2017 Demo Track, Valencia, Spain, November 6-9, 2017. 2017. CEUR-WS. org.
[68] Kaczmarek-Heß, M. and S. De Kinderen, A multilevel model of IT platforms for the needs of enterprise IT landscape analyses. Business & Information Systems Engineering, 2017. **59**: p. 315-329.
[69] Karagiannis, D. and H. Kühn. Metamodelling platforms. in EC-web. 2002. Citeseer.
[70] Kuehnel, S., S.T.-N. Trang, and S. Lindner. Conceptualization, design, and implementation of EconBPC–A software artifact for the economic analysis of business process compliance. in Conceptual Modeling: 38th International Conference, ER 2019, Salvador, Brazil, November 4–7, 2019, Proceedings 38. 2019. Springer.
[71] Kuehnel, S. and A. Zasada. An approach toward the economic assessment of business process compliance. in Advances in Conceptual Modeling: ER 2018 Workshops Emp-ER, MoBiD, MREBA, QMMQ, SCME, Xi’an, China, October 22-25, 2018, Proceedings 37. 2018. Springer.
[72] Lima, L.H.C., et al., An analysis of the collaboration network of the international conference on conceptual modeling at the age of 40. Data & Knowledge Engineering, 2020. **130**: p. 101866.
[73] Lindland, O.I., G. Sindre, and A. Solvberg, Understanding quality in conceptual modeling. IEEE software, 1994. **11**(2): p. 42-49.
[74] Lu, J., et al. UDBMS: road to unification for multi-model data management. in Advances in Conceptual Modeling: ER 2018 Workshops Emp-ER, MoBiD, MREBA, QMMQ, SCME, Xi’an, China, October 22-25, 2018, Proceedings 37. 2018. Springer.
[75] Lucassen, G., et al. Visualizing user story requirements at multiple granularity levels via semantic relatedness. in Conceptual Modeling: 35th International Conference, ER 2016, Gifu, Japan, November 14-17, 2016, Proceedings 35. 2016. Springer.
[76] Luján-Mora, S., J. Trujillo, and I.-Y. Song, A UML profile for multidimensional modeling in data warehouses. Data & knowledge engineering, 2006. **59**(3): p. 725-769.
[77] Lukyanenko, R., et al., Representing crowd knowledge: Guidelines for conceptual modeling of user-generated content. Journal of the Association for Information Systems, 2017. **18**(4): p. 2.
[78] March, S.T. and G.N. Allen, Toward a social ontology for conceptual modeling. Communications of the AIS, 2014. **34**.
[79] Markowitz, V.M. and A. Shoshani, Representing extended entity-relationship structures in relational databases: A modular approach. ACM Transactions on Database Systems (TODS), 1992. **17**(3): p. 423-464.
[80] Mayr, H.C. and B. Thalheim, The triptych of conceptual modeling: A framework for a better understanding of conceptual modeling. Software and Systems Modeling, 2021. **20**(1): p. 7-24.
[81] Molina, A.I., M.A. Redondo, and M. Ortega, A review of notations for conceptual modeling of groupware systems. New trends on human–computer interaction: research, development, new tools and methods, 2009: p. 75-86.
[82] Montali, M. and A. Rivkin, Db-nets: On the marriage of colored petri nets and relational databases. Transactions on Petri Nets and Other Models of Concurrency XII, 2017: p. 91-118.
[83] Moody, D., The “physics” of notations: toward a scientific basis for constructing visual notations in software engineering. IEEE Transactions on software engineering, 2009. **35**(6): p. 756-779.
[84] Nguyen, H., et al. Multi-perspective comparison of business process variants based on event logs. in Conceptual Modeling: 37th International Conference, ER 2018, Xi'an, China, October 22–25, 2018, Proceedings 37. 2018. Springer.

[85] Nwokeji, J.C., et al. Big Data ETL Implementation Approaches: A Systematic Literature Review (P). in SEKE. 2018.
[86] Olivé, A., Conceptual Modeling of Information Systems. 2007: Springer Science & Business Media.
[87] Palacio, A.L. and Ó.P. López. Towards an effective medicine of precision by using conceptual modelling of the genome: short paper. in Proceedings of the International Workshop on Software Engineering in Healthcare Systems. 2018.
[88] Parsons, J. and Y. Wand, Choosing classes in conceptual modeling. Communications of the ACM, 1997. **40**(6): p. 63-69.
[89] Parsons, J. and Y. Wand, Emancipating instances from the tyranny of classes in information modeling. ACM Transactions on Database Systems (TODS), 2000. **25**(2): p. 228-268.
[90] Persson, O., Identifying research themes with weighted direct citation links. Journal of Informetrics, 2010. **4**(3): p. 415-422.
[91] Pichler, H., J. Eder, and M. Ciglic. Modelling processes with time-dependent control structures. in Conceptual Modeling: 36th International Conference, ER 2017, Valencia, Spain, November 6–9, 2017, Proceedings 36. 2017. Springer.
[92] Rahm, E. and P.A. Bernstein, A survey of approaches to automatic schema matching. the VLDB Journal, 2001. **10**: p. 334-350.
[93] Ramdane, Y., et al. SkipSJoin: A new physical design for distributed big data warehouses in hadoop. in Conceptual Modeling: 38th International Conference, ER 2019, Salvador, Brazil, November 4–7, 2019, Proceedings 38. 2019. Springer.
[94] Recker, J., et al., Information systems as representations: A review of the theory and evidence. Journal of the Association for Information Systems, 2019. **20**(6): p. 5.
[95] Recker, J.C., et al., From representation to mediation: a new agenda for conceptual modeling research in a digital world. MIS Quarterly, 2021. **45**(1): p. 269-300.
[96] Rosemann, M. and P. Green, Developing a meta model for the Bunge–Wand–Weber ontological constructs. Information Systems, 2002. **27**(2): p. 75-91.
[97] Roy-Hubara, N., et al., Modeling graph database schema. IT Professional, 2017. **19**(6): p. 34-43.
[98] Rumbaugh, J., et al., Object-oriented modeling and design. Vol. 199. 1991: Prentice-hall Englewood Cliffs, NJ.
[99] Rumbaugh, J., G. Booch, and I. Jacobson, The unified modeling language user guide. 1999: Addison-wesley.
[100] Sales, T.P. and G. Guizzardi. Ontological anti-patterns in taxonomic structures. in Proceedings of the XII Seminar on Ontology Research in Brazil and III Doctoral and Masters Consortium on Ontologies. 2019. CEUR-WS.
[101] Samuel, B.M., V. Khatri, and V. Ramesh, Exploring the effects of extensional versus intensional representations on domain understanding. MIS Quarterly, 2018. **42**(4): p. 1187-1210.
[102] Shanks, G., et al., Representing classes of things and properties in general in conceptual modelling: An empirical evaluation. Journal of Database Management (JDM), 2010. **21**(2): p. 1-25.
[103] Shanks, G., et al., Representing part-whole relations in conceptual modeling: An empirical evaluation. MIS Quarterly, 2008: p. 553-573.
[104] Shanks, G., E. Tansley, and R. Weber, Using ontology to validate conceptual models. Communications of the ACM, 2003. **46**(10): p. 85-89.
[105] Sheth, A.P. and J.A. Larson, Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Computing Surveys (CSUR), 1990. **22**(3): p. 183-236.
[106] Storey, V.C., Understanding semantic relationships. The VLDB Journal, 1993. **2**: p. 455-488.
[107] Storey, V.C. and R.C. Goldstein, A methodology for creating user views in database design. ACM Transactions on Database Systems (TODS), 1988. **13**(3): p. 305-338.
[108] Teorey, T.J., D. Yang, and J.P. Fry, A logical design methodology for relational databases using the extended entity-relationship model. ACM Computing Surveys (CSUR), 1986. **18**(2): p. 197-222.
[109] Thalheim, B., Entity-relationship modeling: foundations of database technology. 2000: Springer Science & Business Media.

[110] Theodoulidis, C., P. Loucopoulos, and B. Wangler, A conceptual modelling formalism for temporal database applications. Information Systems, 1991. **16**(4): p. 401-416.
[111] Thor, A., et al., Introducing CitedReferencesExplorer (CRExplorer): A program for reference publication year spectroscopy with cited references standardization. Journal of Informetrics, 2016. **10**(2): p. 503-515.
[112] Van Eck, N. and L. Waltman, Software survey: VOSviewer, a computer program for bibliometric mapping. scientometrics, 2010. **84**(2): p. 523-538.
[113] Verdonck, M., et al., Comparing traditional conceptual modeling with ontology-driven conceptual modeling: An empirical study. Information Systems, 2019. **81**: p. 92-103.
[114] Wagner, G., R. Lukyanenko, and G. Paré, Artificial intelligence and the conduct of literature reviews. Journal of Information Technology, 2022. **37**(2): p. 209-226.
[115] Wand, Y., V.C. Storey, and R. Weber, An ontological analysis of the relationship construct in conceptual modeling. ACM Transactions on Database Systems (TODS), 1999. **24**(4): p. 494-528.
[116] Wand, Y. and R. Weber, An ontological model of an information system. IEEE Transactions on Software Engineering, 1990. **16**(11): p. 1282-1292.
[117] Wand, Y. and R. Weber, On the deep structure of information systems. Information Systems Journal, 1995. **5**(3): p. 203-223.
[118] Wand, Y. and R. Weber, Research commentary: information systems and conceptual modeling—a research agenda. Information Systems Research, 2002. **13**(4): p. 363-376.
[119] Wand, Y. and R. Weber, Thirty years later: some reflections on ontological analysis in conceptual modeling. Journal of Database Management (JDM), 2017. **28**(1): p. 1-17.
[120] Weber, R., Conceptual modelling and ontology: Possibilities and pitfalls. Journal of Database Management (JDM), 2003. **14**(3): p. 1-20.
[121] Wohlin, C., et al., Experimentation in software engineering. 2012: Springer Science & Business Media.
[122] Yeshchenko, A., et al. Comprehensive process drift detection with visual analytics. in Conceptual Modeling: 38th International Conference, ER 2019, Salvador, Brazil, November 4–7, 2019, Proceedings 38. 2019. Springer.
[123] Yu, E.S. Models for supporting the redesign of organizational work. in Proceedings of conference on Organizational computing systems. 1995.
[124] Yu, E.S. Towards modelling and reasoning support for early-phase requirements engineering. in Proceedings of ISRE'97: 3rd IEEE International Symposium on Requirements Engineering. 1997. IEEE.
[125] Yu, E.S., Social Modeling and i. Conceptual modeling: Foundations and applications: Essays in honor of John Mylopoulos, 2009: p. 99-121.
[126] Zachman, J.A., A framework for information systems architecture. IBM systems journal, 1987. **26**(3): p. 276-292.
[127] Zupic, I. and T. Čater, Bibliometric methods in management and organization. Organizational research methods, 2015. **18**(3): p. 429-472.