Chaomei Chen, Il-Yeol Song, Weizhong Zhu. (2007) Trends in conceptual modeling: Citation analysis of the ER conference papers (1979-2005). Edited by Daniel Torres-Salinas and Henk F. Moed. Proceedings of the 11th International Conference on the International Society for Scientometrics and Informatrics. CSIC, Madrid, Spain. June 25-27, 2007. pp. 189-200.

Trends in Conceptual Modeling: Citation Analysis of the ER Conference Papers (1979-2005)

Chaomei Chen, Il-Yeol Song, Weizhong Zhu

chaomei.chen@cis.drexel.edu, song@drexel.edu, wz32@ischool.drexel.edu
College of Information Science and Technology, Drexel University, Philadelphia, PA 19104-2875 (USA)

Abstract
We analyze thematic trends and challenging issues in conceptual modeling based on the metadata of 943 research papers published in a series of conferences on conceptual modeling (known as the ER conferences) between 1979 and 2005. We specifically address 1) all-time prominent challenges in conceptual modeling, 2) current challenges and emerging trends, and 3) the structure and dynamics of the conceptual modeling community. We utilize CiteSpace, a progressive domain visualization tool, to identify and visualize the movement of research fronts and intellectual bases, persistent clusters of papers, critical paths connecting these clusters, and the evolution of co-authorship networks as well as citation networks. The work contributes an in-depth analysis of a major forum of conceptual modeling and a practical method that one can use as frequently as needed to keep abreast of the state of the art of conceptual modeling.

Keywords
Progressive knowledge domain visualization, CiteSpace

Introduction

Conceptual modeling represents real-world phenomenon using semantic primitives. It is a basis for understanding the phenomenon and for creating a requirement specification that can be incorporated into an information system. Historically, conceptual modeling served as a framework for database design, database integration, visual query paradigms, and information system development.

The output of conceptual modeling is a conceptual model. The most influential conceptual model in the database community is the entity-relationship (ER) model (Chen 1976). Since the introduction of the ER model, the original ER has been extended to achieve more semantic powers. Many versions of extended ER models have been developed and widely researched such as the Extended ER model (Teorey et al 1986), the E^2R model (Embley & Ling 1989), the hierarchical ER model (Thalheim 2000), Temporal ER models (Gregersen & Jensen 1999), and the starER model (Tryfona et al 1999).

ER conferences refer to a series of conferences on conceptual modeling, namely the International Conference on Conceptual Modeling and its predecessor – the International Conference on Entity-Relationship. This series started in 1979. Up to 1985, the conferences were held every alternative year, and then became annual conferences from 1986. The ER proceedings collect several types of papers: regular research papers, industrial papers, abstracts from keynote speeches or tutorials, a panel statement, and editorials. In this study, our analysis includes all the research papers in the proceedings. Several workshops are also held with the main ER conferences, but we have not included any paper from the ER workshops.

Chaomei Chen, Il-Yeol Song, Weizhong Zhu. (2007) Trends in conceptual modeling: Citation analysis of the ER conference papers (1979-2005). Edited by Daniel Torres-Salinas and Henk F. Moed. Proceedings of the 11th International Conference on the International Society for Scientometrics and Informatrics. CSIC, Madrid, Spain. June 25-27, 2007. pp. 189-200.

The online computer science bibliography DBLP[1] contains more than 700,000 bibliographic records in major computer science subjects, including the ACM SIGMOD Anthology. We retrieved bibliographic records of ER conference papers from DBLP's metadata[2]. In DBLP, citation information is available for some of the records, but not for all of them. Abstracts are not available in the DBLP data. In particular, citation data are available only between 1979 and 1999.

Since DBLP records after 2000 do not contain citation links, we retrieved citation data between 2000 and 2005 from the Web of Science[3] (WoS) instead, including Science Citation Index (SCI), Social Science Citation Index (SSCI), or Art & Humanities (A&H). A bibliographic record in WoS contains fields such as author, title, and abstract. It also has a cited reference field (CR), which contains references cited by the corresponding paper. However, it appears that WoS does not have any records for the ER 2001 conference. As a result, our dataset includes citations made by all ER conference papers except the ones cited by the ER 2001 conference.

For the citation analysis, we used CiteSpace (Chen 2004, Chen 2006). CiteSpace is a Java application for analyzing and visualizing co-citation networks. Its primary goal is to facilitate the analysis of emerging trends in a knowledge domain. It allows the user to take a time series of snapshots of a domain and subsequently merge these snapshots. The initial version of CiteSpace was used to reveal turning points in superstring revolutions in physics. However, several issues remained unresolved when we implemented the first version of CiteSpace. The most distinctive new feature is the combination of computational metrics and visual attributes of pivotal points. The motivation is to substantially reduce the user's cognitive burden as they search for pivotal points in a knowledge structure.

In CiteSpace, users can identify pivotal points by visually scanning a visualized network for nodes that connect different clusters. One of the advantages of this approach is that no additional computing is required. CiteSpace also allows users to identify pivotal points in terms of high betweenness centrality (Freeman 1979). Pivotal points are computationally identified and rendered so that they become preattentative, or pop-out, in the visualized network. Pivotal points are highlighted in the display with a purple ring so that they stand out in a visualized network. Graph-theoretically identifiable pivotal points allow us to reduce network-wide operations to the subset of pivotal nodes only so as to improve the interpretability of the network.

Using CiteSpace, we analyze all the papers of the ER proceedings. We present several citation statistics such as most frequently cited papers in the ER conferences and most frequently cited ER papers as well as co-citation maps and their interpretations.

The rest of this paper is organized as follows. Section 2 discusses data collection and analysis procedure. Section 3 presents several citation statistics of ER papers up to 1999. Section 4 presents high level clustering of research areas addressed in the ER conferences. Section 5 analyzes the conceptual modeling community. Section 6 concludes our paper.

---

[1] http://www.informatik.uni-trier.de/~ley/db/

[2] http://dblp.luni-trier.de/xml/

[3] http://www.isinet.com

Chaomei Chen, Il-Yeol Song, Weizhong Zhu. (2007) Trends in conceptual modeling: Citation analysis of the ER conference papers (1979-2005). Edited by Daniel Torres-Salinas and Henk F. Moed. Proceedings of the 11th International Conference on the International Society for Scientometrics and Informatrics. CSIC, Madrid, Spain. June 25-27, 2007. pp. 189-200.

Data Collection and Analysis Procedure

The trend analysis and visualization based on citation networks consists of nine steps:

1. Identify a knowledge domain. In this study, the knowledge domain of conceptual modeling is defined by full papers published in the ER conference series between 1979 and 2005.
2. Data collection. We collect ER conference papers from two sources, namely, DBLP and the Web of Science (WoS) as follows:
   - All the ER bibliographic data including paper titles and authors were retrieved from DBLP from 1979 to 2005.
   - All the reference data from 1979 to 1999 were retrieved from DBLP.
   - All the reference data including abstracts of the ER papers between 2000 to 2005, except for ER2001, were retrieved from the WoS.
   - DBLP records do not contain abstracts. Neither the WoS nor DBLP contain citation data for the 2001 ER conference.
3. Extract research front terms. Extract phrases, or terms, from titles, abstracts, descriptors, and identifiers of citing articles in the dataset retrieved from the WoS (2000-2005), except 2001. For ER conferences between 1979 and 1999, the extraction is limited to title words only because DBLP does not provide abstracts. Extracted terms are further filtered based on the so-called burst rates, which measure significant increases or decreases of frequencies over a given time interval. Burst terms are used to capture fast-growing interests.
4. Time slicing. Specify the range of the entire time interval and the length of a single time slice.
5. Threshold selection. CiteSpace allows users to specify three sets of threshold levels for citation counts, co-citation counts, and co-citation coefficients. Citation counts are the number of times a publication is cited by the ER conference papers in the combined dataset. Two publications are called co-cited if a paper cites both of them. Co-citation counts for a given pair of publications are the number of ER conference papers in our dataset that cite the pair. Co-citation coefficients are normalized co-citation counts over each time slice. The specified thresholds are applied to three time slices, namely, the earliest slice, the middle one, and the last one. Linear interpolated thresholds are assigned to the rest of slices. In this study, most of our networks contain two types of vertices and three types of edges. Vertices could be authors, papers, journals, and burst terms, whereas edges may represent co-occurrence, co-citation, or referential links.
6. Pruning and Merging. Pathfinder network scaling is the default option in CiteSpace for network pruning (Chen 2004, Schvaneveldt 1990). Users choose whether or not to apply the scaling operation to individual networks. Pathfinder network scaling is an asymptotically expensive algorithm. CiteSpace implements a concurrent version of the algorithm to process multiple networks simultaneously, which substantially reduces the overall waiting time. CiteSpace merges individual networks by taking a set union of all the vertices and selecting links that do not violate a triangle inequality condition in overlapping areas between networks. Users can choose whether or not to prune the merged network as a whole.
7. Layout. CiteSpace supports a standard graph view and a time-zone view.
8. Visual inspection. CiteSpace enables users to interact with the visualization of a knowledge domain in several ways. The user may control the display of visual attributes and labels as well as a variety of parameters used by the underlying layout algorithms.
9. Verify pivotal points. The significance of a marked pivotal point can be verified by asking domain experts, for example, the authors of pivotal-point articles, and/or examining the literature, such as passages containing citations of a pivotal-point article. A particularly interesting direction of research is the development of tools that can automatically summarize the value of a pivotal point.

Chaomei Chen, Il-Yeol Song, Weizhong Zhu. (2007) Trends in conceptual modeling: Citation analysis of the ER conference papers (1979-2005). Edited by Daniel Torres-Salinas and Henk F. Moed. Proceedings of the 11th International Conference on the International Society for Scientometrics and Informatrics. CSIC, Madrid, Spain. June 25-27, 2007. pp. 189-200.

Digital libraries, automated text summarization, machine learning, and several other fields are among the most promising sources of input.

Most Cited Papers

We present most cited papers in several groups. If paper A cites paper B, then paper A is called the source of the citation and paper B is called the target of the citation. DBLP provides a convenient way to locate an ER paper from a paper ID. For example, a paper with ScheuermannSW79 as its ID can be uniquely identified in DBLP as conf/er/ScheuermannSW79.

Table 1 is a list of ER papers that are frequently cited by all conference papers in DBLP, including ER conferences. One column lists citations made by ER papers. Another column lists citations made by papers from conferences other than ER. Table 2 is a list of ER papers that are frequently cited by journal papers indexed in DBLP. Non-ER papers refer to papers that appeared in conferences other than ER conferences. Table 3 lists non-ER papers frequently cited in ER papers up to 1999, whereas Table 4 contains non-ER papers cited in ER papers between 1979 and 2005, except 2001. Source names in Table 4 follow the journal abbreviations used by the Web of Science. See Table 5 for a list of the most popular ones. Table 4 is predominated by journal papers, except the two ER papers by Scheuermann and Santos. Table 5 lists most frequently cited journals in ER papers between 1979 and 1999.

Table 1. Most cited ER papers. Citation source: Papers from all conferences in DBLP up to 1999.

Reference ID (conf/er)	**Title**	**Citations By ER papers only**	**Citations by non-ER papers**	**Total Citations**
ScheuermannSW79	Abstraction Capabilities and Invariant Properties Modelling within the Entity-Relationship Approach	41	12	53
SantosNF79	A Data Type Approach to the Entity-Relationship Approach	29	11	40
ElmasriW81	GORDAS: A Formal High-Level Query Language for the Entity-Relationship Model	21	11	32
Klopprogge81	TERM: An Approach to Include Time Dimension in the Entity-Relationship Model	8	16	24
Ling85	A Graphical Query Language for Entity-Relationship Databases	21	2	23
ZhangM83	A Normal Form For Entity-Relationship Diagrams.	18	5	23
DavisA87	Converting A Relational Database Model into an Entity-Relationship Model	18	3	21
AtzeniC81	Completeness of Query Languages for the Entity-Relationship Model	15	6	21
WiederholdE79	The Structural Model for Database Design	6	15	21
NavatheA87	Abstracting Relational and Hierarchical Data with a Semantic Data Model	16	3	19

Chaomei Chen, Il-Yeol Song, Weizhong Zhu. (2007) Trends in conceptual modeling: Citation analysis of the ER conference papers (1979-2005). Edited by Daniel Torres-Salinas and Henk F. Moed. Proceedings of the 11th International Conference on the International Society for Scientometrics and Informatrics. CSIC, Madrid, Spain. June 25-27, 2007. pp. 189-200.

Table 2. Most cited ER papers. Citation source: Journal papers in DBLP up to 1999.

Reference ID (conf/er)	**Title**	**Citations**
WiederholdE79	The Structural Model for Database Design.	7
RoseS91	TOODM - A Temporal Object-Oriented Data Model with Temporal Constraints.	7
ZhangM83	A Graphical Query Language for Entity-Relationship Databases.	7
Klopprogge81	TERM: An Approach to Include Time Dimension in the Entity-Relationship Model.	6
NavatheA87	Abstracting Relational and Hierarchical Data with a Semantic Data Model.	5
LipeckN86	Modelling and Manipulating Objects in Geoscientific Databases.	4
Lien79	On the Semantics of the Entity-Relationship Data Model.	4
SuL79	A Semantic Association Model for Conceptual Design.	4
WongK79	Logical Design and Schema Conversion for Relational and DBTG Databases.	4
RosenthalR87	Theoretically Sound Transformations for Practical Database Design.	3

Table 3. Non-ER papers frequently cited by ER conference papers up to 1999.

DBLP Reference	**Conference Paper Titles**	**# Citations**
afips/Chen77	The Entity-Relationship Model - A basis for the Enterprise View of Data.	26
ds/Abrial74	Data Semantics.	25
sigmod/BanerjeeKKK8	Semantics and Implementation of Schema Evolution in Object-Oriented Databases.	15
dood/AtkinsonBDDMZ8	The Object-Oriented Database System Manifesto.	15
sigmod/HammerM78	The Semantic Data Model: A Modelling Mechanism for Data Base Applications.	15
vldb/BachmanD77	The Role Concept in Data Models.	15
ds/HallOT76	Relations and Entities.	14
sigmod/LuskOP80	A Practical Design Methodology for the Implementation of IMS Databases, Using the Entity-Relationship Model.	14
db-workshops/BrodieR82	On the Design and Specification of Database Transactions.	13
db-workshops/YaoNW78	An Integrated Approach to Database Design.	13

Table 4. Non-ER papers cited by ER conference papers (79-05), except 2001.

Cites	**Burst**	**Centrality**	**Authors**	**Year**	**Source**	**Vol.**	**Page**
308	0.00	0.80	CHEN PP	1976	TODS	1	9
95	8.45	0.22	SMITH JM	1977	TODS	2	105
55	6.78	0.25	TEOREY TJ	1986	CSUR	18	197
49	8.27	0.11	CODD EF	1979	TODS	4	397
44	6.89	0.09	CODD EF	1970	CACM	13	377
42	3.59	0.10	HULL R	1987	CSUR	19	201
40	6.22	0.17	HAMMER M	1981	TODS	6	351
40	7.35	0.03	ELMASRI R	1985	DKE	1	75
38	10.16	0.04	SCHEUERMANN P	1979	ER	0	121
33	14.81	0.27	RUMBAUGH JE	1991	PH	0	0
33	14.60	0.31	BATINI C	1992	BC	0	0
29	8.93	0.05	BATINI C	1986	CSUR	18	323

28	8.40	0.04	SANTOS CSD	1979	ER	0	103
25	3.73	0.03	SHIPMAN DW	1981	TODS	6	140
25	6.81	0.02	ELMASRI R	1989	BC	0	0

Table 5. Most cited Journals in ER papers up to 1999.

Journal Title	# Cites
ACM Transactions on Database Systems	886
Communications of the ACM	278
ACM Computing Surveys	265
IEEE Transactions on Software Engineering	235
Information Systems	225
Data & Knowledge Engineering	162
SIGMOD Record	96
ACM Transactions on Information Systems	93
IEEE Computer	90
IEEE Transactions on Knowledge and Data Engineering	83
Journal of the ACM	49
Computer Journal	44
VLDB Journal	37
IBM Systems Journal	35
Artificial Intelligence	34
IEEE Database Engineering Bulletin	31
Information Science	30
Distributed and Parallel Databases	27
IEEE Software	20
Journal of Intelligent Information Systems	20

All-time Prominent Challenges in Conceptual Modeling

In addition to the citation counts, we are interested in prominent research issues in conceptual modeling and how they have been addressed by the community and, in particular, by papers published in the ER conferences. Co-citation analysis of scientific literature aims to identify emergent patterns in scholarly publications derived from how scientists collectively attribute their work to prior published work. Specifically, the goal of co-citation analysis is to identify clusters of papers that are frequently cited together. Therefore, citations are seen as a filtering mechanism that selects the intellectual work that is valued by peer researchers collectively.

Prominent Co-Citation Clusters between 1979 and 2005

In this section, we present two co-citation networks. The first is pruned by Pathfinder network scaling algorithm to represent the most salient structure, whereas the second is not pruned so that the network retains more details than the pruned one.

Figure 1 shows a paper co-citation network of 760 citation links of top 487 papers cited by the ER papers between 1979 and 2005, except 2001. As shown in Figure 1 and Table 4, Chen's 1976 paper in *ACM Transactions on Database Systems* (TODS) has the largest citation. Thus, the network is focused on this seminal paper and all other papers linked to the paper.

Chaomei Chen, Il-Yeol Song, Weizhong Zhu. (2007) Trends in conceptual modeling: Citation analysis of the ER conference papers (1979-2005). Edited by Daniel Torres-Salinas and Henk F. Moed. Proceedings of the 11th International Conference on the International Society for Scientometrics and Informatrics. CSIC, Madrid, Spain. June 25-27, 2007. pp. 189-200.

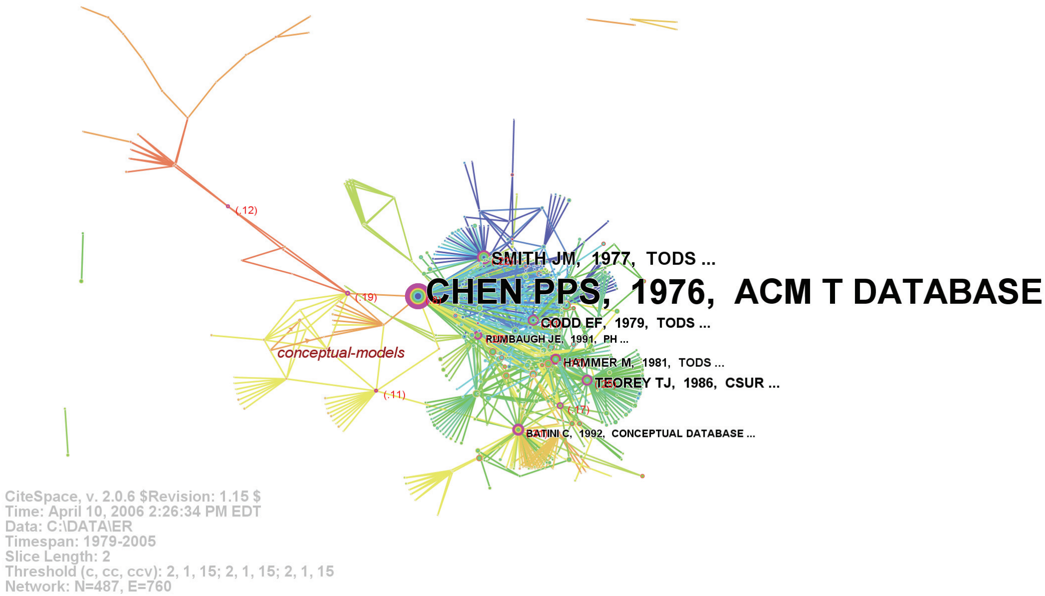

Figure 1: A document co-citation network of conceptual modeling derived from citations made by ER papers (1979-2005), except 2001. This network consists of 487 papers and 760 salient co-citation links. CiteSpace threshold values: c=2, cc=1, ccv=15.

Figure 2 shows four major co-citation clusters generated from top 548 papers and 4697 citation links between 1979 and 2005, except 2001. Again the focal point of the network is Chen's TODS paper. The term 'conceptual models' in the image is a burst term. A burst term means that the term was associated with a sudden increase of popularity. In addition, we showed four rectangles, representing major co-citation clusters of research topics. The four major co-citation clusters of papers emerged from the network are: ER, UML, Design Patterns, and Ontology. We name these clusters based on the most cited members. Table 6 shows the list of top 20 most cited articles in each cluster.

The ER cluster contains 422 papers. It appears to have two subcomponents (blue and green), but their connections are so tight that we regard them as one component. The most prominent papers in this cluster include the 1976 ER paper from Chen PPS, which was cited 308 times, a 1977 paper by Smith (cited 95 times)., followed by Teorey's Paper (cited as 55 times).

The Unified Modeling Language (UML) cluster contains approximately 64 papers. The most cited publication in this cluster is the 1999 book by Rumbaugh on UML, which is cited 9 times. The Design Pattern clusters contain 7 papers, while *the Ontology cluster* contains 20 papers. The 1995 book by Gamma et al. on Design Patterns is also cited 9 times.

The orange colors of the other two clusters indicate that they are more recent than the ER and the UML clusters. UML is linked to a 7-paper cluster on Design Patterns. It is in turn connected to a 20-paper cluster, containing papers on topics such as XML, automatic schema matching, mapping ontologies on the semantic web, and generic schema matching.

Chaomei Chen, Il-Yeol Song, Weizhong Zhu. (2007) Trends in conceptual modeling: Citation analysis of the ER conference papers (1979-2005). Edited by Daniel Torres-Salinas and Henk F. Moed. Proceedings of the 11th International Conference on the International Society for Scientometrics and Informatrics. CSIC, Madrid, Spain. June 25-27, 2007. pp. 189-200.

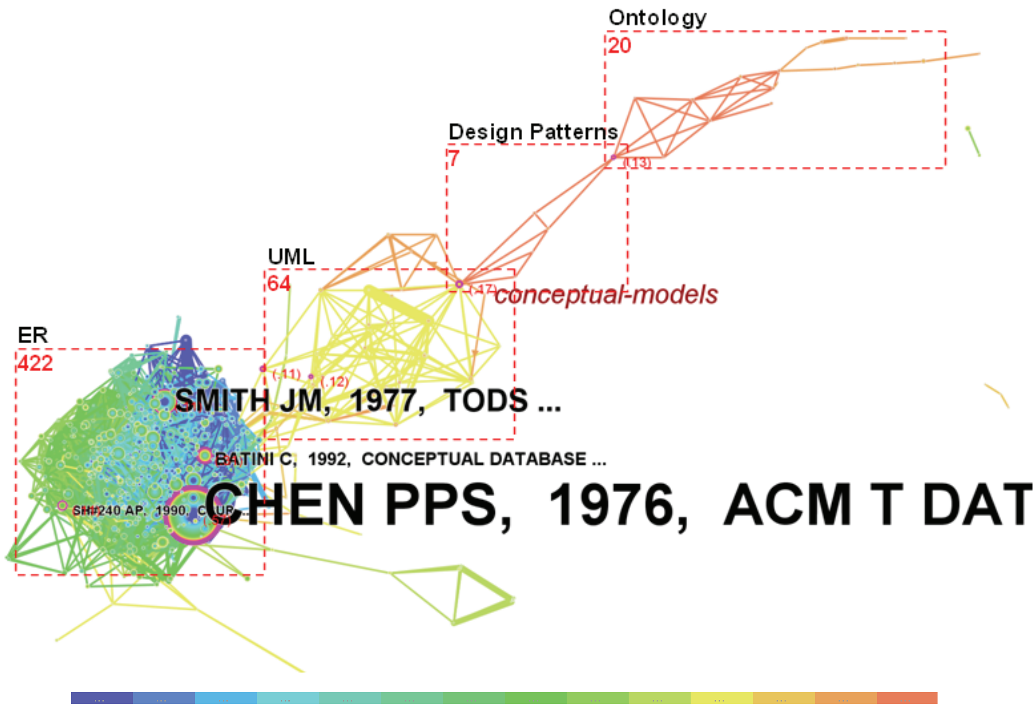

Figure 2: An un-pruned network of documents cited and co-cited by ER papers (1979-2005, two-year slices), containing 548 nodes and 4,697 links. CiteSpace threshold values: c=2, cc=1, ccv=25.

Table 6. Four major co-citation clusters of publications cited by ER papers.

Cluster	Cites	Centrality	Authors	Year	Source	Volume/Page
ER	308	0.57	CHEN PP	1976	TODS	1, 9
	95	0.11	SMITH JM	1977	TODS	2, 105
	55	0.06	TEOREY TJ	1986	CSUR	18, 197
	49	0.08	CODD EF	1979	TODS	4, 397
	44	0.06	CODD EF	1970	CACM	13, 377
UML	9	0.04	RUMBAUGH J	1999	UNIFIED MODELING LAN	Book
	9	0.17	GAMMA E	1995	DESIGN PATTERNS ELEM	Book
	8	0.06	LINDLAND OI	1994	IEEE SOFTWARE	11, 42
	8	0.00	LIEN YE	1979	ER	155
	7	0.02	RUMBAUGH J	1991	OBJECT ORIENTED MODE	Book
Design Patterns	9	0.17	GAMMA E	1995	DESIGN PATTERNS ELEM	Book
	5	0.07	VANDERAALST WMP	2003	DISTRIB PARALLEL DAT	14, 5
	3	0.13	OPDAHL AL	2002	SOFTWARE SYSTEMS MOD	1, 43
	3	0.07	DUMAS M	2001	LNCS	V2185
Ontology	6	0.01	ABITEBOUL S	1995	FDN DATABASES	Book
	5	0.06	GRUBER TR	1993	KNOWL ACQUIS	5, 199

Chaomei Chen, Il-Yeol Song, Weizhong Zhu. (2007) Trends in conceptual modeling: Citation analysis of the ER conference papers (1979-2005). Edited by Daniel Torres-Salinas and Henk F. Moed. Proceedings of the 11th International Conference on the International Society for Scientometrics and Informatrics. CSIC, Madrid, Spain. June 25-27, 2007. pp. 189-200.

	5	0.03	RAHM E	2001	VLDB J	10, 334
	4	0.07	DOAN A	2002	Proc. 11th WWW	662-673

Figure 3 shows a co-citation network of publication sources of all the papers cited by the ER papers published between 1979 and 2005, except 2001. The majority of citations received by TODS of 394 citations are due to Peter Chen's ground-breaking paper in 1976. ER itself is the second largest source, cited 333 times. VLDB in the third place is cited 283 times, followed by SIGMOD 259 times, and by *Communications of the ACM* (CACM) 178 times. Nodes in the top half of the network are more recent than those in the lower half. Some of the nodes are actually the same source with different abbreviations, for example, *Lecture Notes of Computer Science* (LNCS) and the *ACM Transactions on Database Systems* (TODS). CiteSpace supports a function to merge such nodes to a unique node. Figure 3 shows the visualization without merging these nodes.

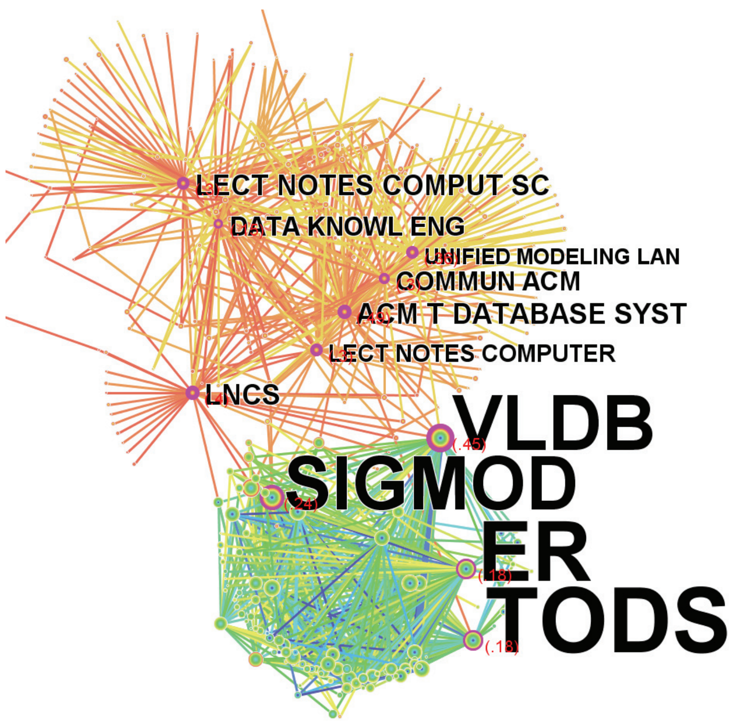

Figure 3: A Pathfinder pruned co-citation network of sources, i.e. journals and books as well as conferences derived from citations made by ER papers (1979-2005). CiteSpace (1-year slices) threshold values: c= 2,cc=1, and ccv=5. The network contains 311 sources and 791 co-citation links.

The Conceptual Modeling Community

In this section, we present the co-authorship map of 1,349 authors and 2,125 co-authoring links (1979-2005, slice length=3 years) in Figure 4. The ER conference co-authorship map depicts a social network of authors who have joint publications in the ER conferences. The map contains two types of vertices: authors who have published in the ER conferences and key phrases that appeared in the metadata of ER conference papers such as titles and abstracts. The size of a vertex represents the number of papers an author has published in the ER conferences. The larger the rings are, the more

Chaomei Chen, Il-Yeol Song, Weizhong Zhu. (2007) Trends in conceptual modeling: Citation analysis of the ER conference papers (1979-2005). Edited by Daniel Torres-Salinas and Henk F. Moed. Proceedings of the 11th International Conference on the International Society for Scientometrics and Informatrics. CSIC, Madrid, Spain. June 25-27, 2007. pp. 189-200.

papers they represent. The color of each ring corresponds to the year of an ER conference in which their papers are published. The network is a hybrid network of directed and undirected graphs. Links between authors are co-authorship, which is undirected, whereas links between key phrases and authors are directed, meaning the authors used key phrases in their papers' titles and/or abstracts.

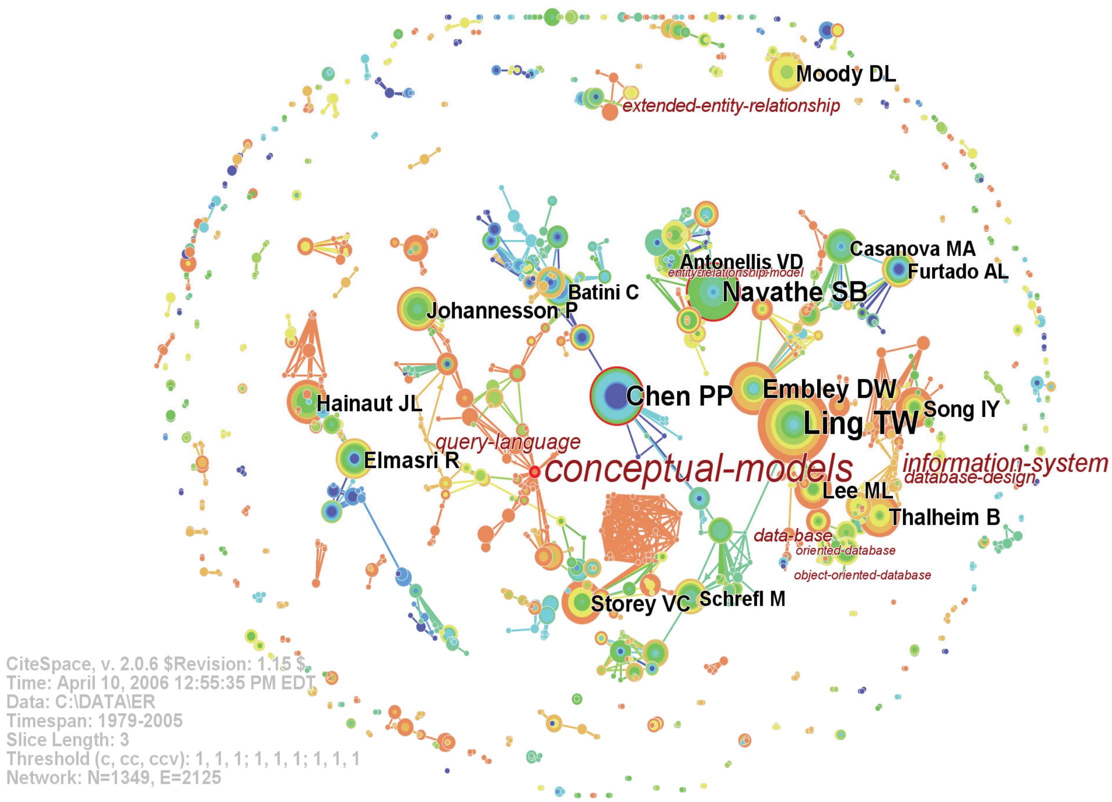

Figure 4: An ER conference co-authorship map of 1,349 authors and 2,125 co-authoring links (1979-2005, slice length=3 years). Red circles indicate burst of productivity during the entire interval.

The most productive authors who have published 8 or more ER full papers are listed in the following Table 7. For example, Tok Wang Ling published 18 papers in the ER conferences. Peter P. Chen and Shamkant B. Navathe both published 14 papers; they are also associated with a high burst rate of 6.01 and 6.71 respectively. We examined the history of each of the two authors and found that the high burst rate for Chen was due to an early peak of the number of papers in the ER conference, whereas Navathe's burst rate was due to an episode of an increasing number of papers, including 6 papers in a 3-year period starting 1991.

Table 7. Authors who have published 8 or more ER research papers (1979-2005).

# Papers	Burst Rate	Centrality	Authors
18	0.00	0.02	Tok Wang Ling
14	6.01	0.01	Peter P. Chen
14	6.71	0.00	Shamkant B. Navathe
12	0.00	0.01	David W. Embley

10	0.00	0.00	Bernhard Thalheim
10	0.00	0.00	Jean-Luc Hainaut
10	0.00	0.00	Paul Johannesson
10	0.00	0.00	Veda C. Storey
9	0.00	0.00	Daniel L. Moody
9	0.00	0.00	Il-Yeol Song
9	0.00	0.00	Mong-Li Lee
9	0.00	0.00	Ramez Elmasri
8	0.00	0.00	Antonio L. Furtado
8	0.00	0.01	Carlo Batini
8	0.00	0.00	Marco A. Casanova
8	0.00	0.01	Michael Schrefl
8	0.00	0.00	Valeria De Antonellis

Figure 5 shows a hybrid Pathfinder network of co-cited (first) authors and burst terms found in citing papers between 1979 and 1999. Most cited authors are listed in Table 8 along with their citation counts and centrality scores. For example, Chen PP is the most cited and has the highest centrality, indicating his predominant influence to the ER community. The only visible burst term in this period is the term *database*.

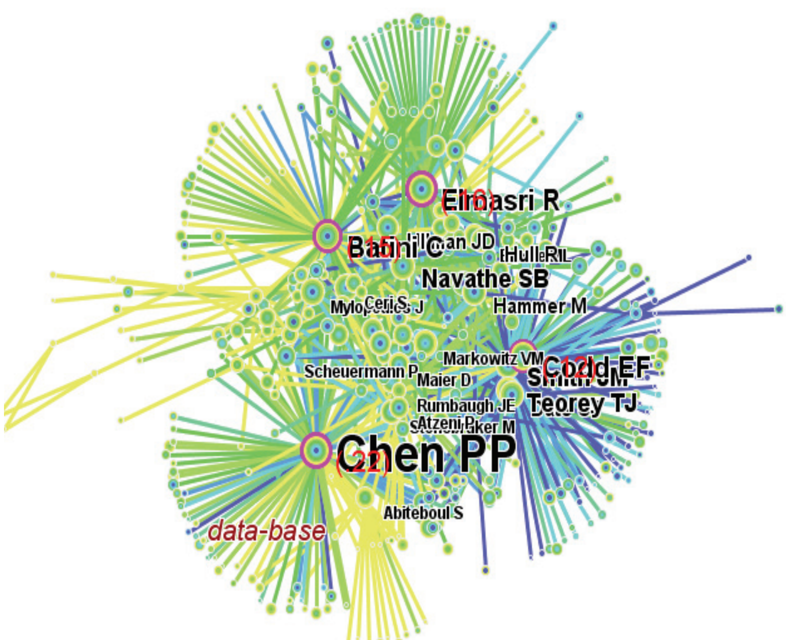

Figure 5: A hybrid Pathfinder network of author co-citations and burst term co-occurrences representing ER papers' citing behavior between 1979 and 1999 (first authors only).

Table 8. Most cited authors in the author co-citation network shown in Figure 5 (first authors only).

Cites	**Centrality**	**Authors**
308	0.22	Chen PP
109	0.16	Elmasri R
108	0.15	Batini C

99	0.12	Codd EF
97	0.08	Smith JM
73	0.08	Teorey TJ
73	0.03	Navathe SB
64	0.06	Date CJ
61	0.02	Hammer M
55	0.02	Hull R

Figure 6 shows a hybrid network between 2000 and 2005, except 2001. Most cited authors are listed in Table 9. During this period, Abiteboul and Booch are most cited authors. Abiteboul is also significant in terms of its rate of citations. Several burst terms appear in this period, including terms such as *conceptual models*, *information system*, and *query language*. These burst terms identify potentially fast-growing research topics.

According to a summary[4], the Booch methodology introduced by Booch in 1991 was a widely used method in object-oriented analysis and design. The third place is James Rumbaugh, who developed the object-modeling technique (OMT), which is one of the precursors to UML. In 1994 Grady Booch and Jim Rumbaugh worked together to unify the Booch and OMT methods, which are regarded as the predecessor of UML. Carlo Batini's name appears as the fourth most cited author. Carlo Batini and Maurizio Lenzerini in their 1983 ER paper introduced a methodology for data schema integration in the ER model. In 1989, David W. Embley and Tok Wang Ling proposed what is known as the E^2R model in their paper "Synergistic Database Design with an Extended Entity-Relationship Model" that solved two major problems of ER models. In their E^2R model, designers no longer have to distinguish between attributes and entities and it also supports the normalization at the model level.

[4] http://cs-exhibitions.uni-klu.ac.at/index.php?id=447

Chaomei Chen, Il-Yeol Song, Weizhong Zhu. (2007) Trends in conceptual modeling: Citation analysis of the ER conference papers (1979-2005). Edited by Daniel Torres-Salinas and Henk F. Moed. Proceedings of the 11th International Conference on the International Society for Scientometrics and Informatrics. CSIC, Madrid, Spain. June 25-27, 2007. pp. 189-200.

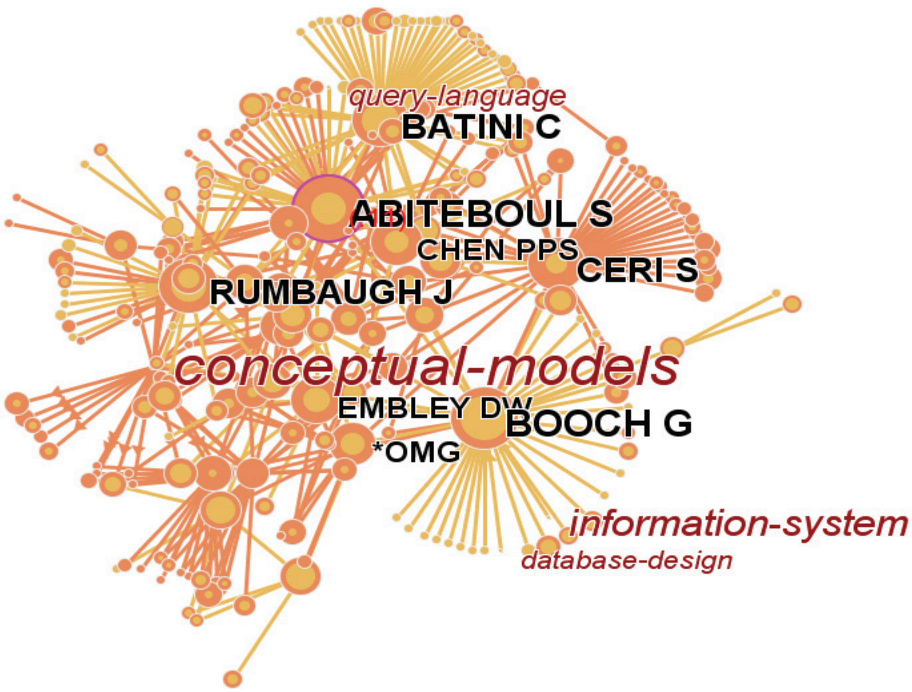

Figure 6: A hybrid Pathfinder network of author co-citations and burst term co-occurrences representing ER papers' citing behavior between 2000 and 2005, except 2001 (first authors only).

Table 9. Most cited authors from the author co-citation network shown in Figure 6 (first authors only).

Cites	Centrality	Authors
20	0.11	ABITEBOUL S
20	0.09	BOOCH G
17	0.03	RUMBAUGH J
16	0.09	BATINI C
16	0.08	CERI S
14	0.03	CHEN PPS
14	0.02	EMBLEY DW
13	0.02	*OMG
11	0.01	AGRAWAL R
11	0.00	FOWLER M
11	0.00	JACOBSON I
11	0.01	THALHEIM B

Conclusion

In this paper, we have presented citation analysis of all the papers published in regular ER conferences from 1979 to 2005. In some of statistics, 2001 citation data are missing as it was available neither in DBLP nor the Web of Science. We presented several citation statistics and visualizations of co-citation networks and social networks of collaborating authors. Our analysis indicates that bibliographic data can be used to identify key research focuses of conceptual modeling at various periods of time. The four identified co-citation clusters represent a trend of progression from the pioneering ER modeling, to object-oriented UML, to Design patterns, and to the more recent ontology. Finally, we showed a community of conceptual modeling in terms of the most publishing authors and most cited authors. On the other hand, we study has also revealed that the available bibliographic resources are still not readily available, for example the absence of abstracts in DBLP and missing earlier ER conferences in the ISI's Web of Science.

We believe our work contributed an in-depth analysis of a major forum of conceptual modeling and a systematic and streamlined method that one can use as frequently as needed to keep abreast of the history and the state of the art of conceptual modeling.

Notes

CiteSpace is available at http://cluster.cis.drexel.edu/~cchen/citespace. Color versions of the figures are available at http://cluster.cis.drexel.edu/~cchen/papers/issi2007/.

References

Chen C. 2004. Searching for intellectual turning points: Progressive Knowledge Domain Visualization. *Proc. Natl. Acad. Sci. USA* 101: 5303-10

Chen C. 2006. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology* 57: 359-77

Chen PP. 1976. The Entity-Relationship Model: Toward a Unified View of Data. *ACM Transactions on Database Systems* 1: 9-36

Embley DW, Ling TW. 1989. *Synergistic Database Design with an Extended Entity Relationship Model*. Presented at Proc. of the 8th International Conf. on Entity-Relationship Approach, Toronto, Canada

Freeman LC. 1979. Centrality in social networks: Conceptual clarification. *Social Networks* 1: 215-39

Gregersen H, Jensen CS. 1999. Temporal Entity-Relationship Models: A Survey. *IEEE Transaction on Knowledge and Data Engineering* 11: 464-97

Schvaneveldt RW, ed. 1990. *Pathfinder Associative Networks: Studies in Knowledge Organization*. Norwood, New Jersey: Ablex Publishing Corporations

Teorey TJ, Yang D, Fry JP. 1986. A Logical Design Methodology for Relational Databases Using the Ex-tended Entity-Relationship Model. *ACM Computing Surveys* 18: 197-222

Thalheim B. 2000. *Entity-Relationship Modeling: Foundations of Database Technology*: Springer

Tryfona N, Busborg F, Christiansen JGB. 1999. *StarER: A Conceptual Model for Data Warehouse Design*. Presented at Proc. of ACM Second Int'l Workshop on Data Warehousing and OLAP (DOLAP '99), Kansas City, MO