ARTICLE IN PRESS

Data & Knowledge Engineering xxx (xxxx) xxx

Contents lists available at ScienceDirect

Data & Knowledge Engineering

journal homepage: www.elsevier.com/locate/datak

An analysis of the collaboration network of the International Conference on Conceptual Modeling at the Age of 40

Lucas Henrique C. Lima [a,*], Alberto H.F. Laender [a], Mirella M. Moro [a], José Palazzo M. de Oliveira [b]

[a] *Universidade Federal de Minas Gerais, Brazil*
[b] *Universidade Federal do Rio Grande do Sul, Brazil*

ARTICLE INFO

Keywords:
Collaboration networks
Coauthorship networks
Databases
Conceptual modeling
ER

ABSTRACT

The International Conference on Conceptual Modeling celebrated 40 years of existence at its 38th edition held in Salvador, Brazil, on 4–7 November 2019. As one of the most traditional and well-known conferences in the database area, it has its origins on the Entity-Relationship Model proposed by Peter P. Chen in 1975. To celebrate such an accomplishment, this article goes over the ER history from distinct perspectives. Overall, we investigate the complete ER collaboration network built on bibliographic data collected from DBLP, comprising its 38 editions held from 1979 to 2019. We analyze several aspects regarding the evolution of its network metrics, such as degree, clustering coefficient and average shortest path, over the four decades. In particular, we analyze the role of the most engaged ER authors, the number of distinct authors, institutions and published papers, and the evolution of some of the most frequent terms presented in the titles of its papers, as well as the influence and impact of the prominent ER authors.

1. Introduction

The International Conference on Conceptual Modeling (or simply the ER conference) celebrated 40 years of existence at its 38th edition held in Salvador, Brazil, on 4–7 November 2019. The ER conference started in Los Angeles, California, in 1979 as a bi-annual event denominated International Conference on the Entity-Relationship Approach, and soon became one of the most traditional and well known conferences in the database area. It has its origins on the Entity-Relationship Model, which was introduced by Peter P. Chen in 1975 at the first edition of the International Conference on Very Large Data Bases[1] [2]. After four biannual editions, the ER conference became very popular and has been uninterruptedly held every year since its fifth edition in Dijon, France, in 1985. The Dijon edition also started a venue rotation scheme between America and Europe, which remained until 1994 when the City of Gold Coast in Australia was chosen as the next venue for the conference. Hence, from 1995 onward, a new rotation scheme started covering three geographical regions, Australasia, Europe and America. Its 1995 edition was also marked by the change of its denomination to International Conference on Object-Oriented and Entity-Relationship Modeling, a clear influence of the object-oriented paradigm so popular in those days. Such new denomination changed again in 1996 to International Conference on Conceptual Modeling, thus reflecting the broader role of its community inside the Computer Science field. Since then, its community has become more mature and explored new important research topics that go beyond traditional conceptual modeling issues, as can be seen not only by the recent volumes of its proceedings, but also by those of its associated workshops. A paper with a summary of the first 30 years of ER the conference was presented at its 2009 edition [3].

* Corresponding author.
E-mail addresses: lucaslima@dcc.ufmg.br (L.H.C. Lima), laender@dcc.ufmg.br (A.H.F. Laender), mirella@dcc.ufmg.br (M.M. Moro), palazzo@inf.ufrgs.br (J.P.M. de Oliveira).

[1] The full article describing the Entity-Relationship Model was published in 1976 in the very first number of the ACM Transactions on Database Systems [1].

https://doi.org/10.1016/j.datak.2020.101866
Received 14 August 2020; Received in revised form 1 October 2020; Accepted 12 October 2020
Available online xxxx

In order to celebrate such an accomplishment and taking advantage of seminal achievements in social network analysis [4–8], this article goes over the ER conference history and provides distinct views of its community. Specifically, we present a deep analysis of the ER collaboration network derived from the peer-reviewed papers presented over its 38 editions and published in its main proceedings. The ER collaboration network was constructed with bibliographic data of the ER conference obtained from the Digital Bibliography & Library Project (DBLP),[2] in which nodes represent authors and edges their coauthorships. The ER collaboration network is represented as a temporal graph, i.e., a graph that changes over time, thus in this article several aspects regarding the evolution of the network metrics are analyzed. As we shall see, the ER conference has evolved steadily over the years, giving rise to a very cohesive community.

We also analyze the involvement of the most engaged and influential ER authors, the number of distinct authors, institutions and published papers, as well as the evolution of some of the most frequent terms presented in the titles of the papers. Finally, we discuss the productivity, influence and impact of the ER authors through centrality measures, show that the ER network follows a phenomenon typical in social networks known as small-world [9] in the first decades of the conference, and present a brief discussion of those institutions with the highest number of affiliated authors in its 38 editions.

The remaining sections of this article are organized as follows. Section 2 covers related work on academic collaboration networks. Section 3 provides background information on data acquisition and network modeling. Then, Sections 4 to 6 present the results of our analyses covering, respectively, basic statistics of the collaboration network, an analysis of its influential authors and a deep discussion of its structure, including connected components, network groups and homophily. Finally, Section 7 summarizes our main conclusions, thus providing some insights for future work.

2. Related work

Social networks have long become subject of serious research, instead of just a means for people to keep in touch with family and friends, or get up-to-date with news and celebrities. Indeed, social network analyses are at the center of many studies with all sorts of goals; for example, check the many workshops and conferences around the subject available in DBLP.[3] The conceptual modeling community has also taken an interest by focusing on data from social media [10–12] or proposing a general conceptual modeling paradigm for both network data and query requirements [13], among others. Nonetheless, our analyses do not require such specific, complex modeling or processing, as we build an academic collaboration network based on bibliographic data.

In such a context of academic collaborations, Newman published some of the first studies in the area [14,15], which show distributions of collaborators and their clusters, and characterize different patterns of collaboration between distinct research areas. Such studies also answer a broad variety of questions about collaboration patterns, presenting several structural and topological features using bibliographic data from biology, physics and mathematics. After Newman's perspective on social network analysis for the academic context, many other studies were published in distinct contexts. For example, there are studies on building and analyzing academic collaboration networks by country [16–20], institution [17,21,22], researchers' features [23,24], research areas [16,18,19,25–27] and scientific venues [24,28–31].

There are also studies that regard Computer Science (CS) as a specific research field. For example, Freire and Figueiredo [32] characterize the structural properties of collaboration networks composed of CS researchers. The authors use such characterization to rank individuals as well. Following a similar perspective, Lima et al. [25] propose a strategy for ranking authors across multiple research areas by characterizing the profile of top Brazilian researchers. In a broader view, Liu et al. [33] analyze productivity, connected components, and authors' impact over the (potentially) largest digital library of CS publications, the DBLP, which has recently achieved five million registered publications.

Similar to the work presented here, Chen et at. [11] carried out a first study of the ER Conference focusing on the period 1979–2007. Other works have addressed specific database-related venues. Nascimento et al. [31] analyze the coauthorship network of the ACM SIGMOD International Conference on Management of Data. Likewise, Ameloot et al. [28] go over the 30 years of history of the ACM Symposium on Principles of Database Systems (PODS). From a country perspective, Lima et al. [34] analyze the Brazilian database research community by evaluating the collaboration network of the Brazilian Symposium on Data Bases (SBBD) on the occasion of its 30th edition. Cosentino et al. [35] present an approach that provides a holistic view of conference-related information, which is able to extract several measures, some of them are adopted in this article. The authors have illustrated these metrics by analyzing the ER conference. In this article, we dive deeper into the analysis of few of these metrics, such as number of papers, number of authors, newcomers. Furthermore, we evaluate some of these metrics over time as an effort to shed light on how the ER conference has been evolving over years.

For other areas within Computer Science, Smeaton et al. [36] analyze the collaboration network and research topics at the 25-year celebration of the International ACM SIGIR Conference on Research and Development in Information Retrieval. On a similar line, Liu et al. [37] focus on the first decade of the Digital Libraries community by analyzing the collaboration network of past editions of the ACM/IEEE Joint Conference on Digital Libraries. Finally, Maia et al. [30] study the collaboration network of the Brazilian Symposium on Computer Networks and Distributed Systems (SBRC) over its 30 year editions.

Overall, the contributions of this article can be summarized as an X-ray over the ER community and its history. From a cohesive dataset that includes data on authors and their papers from all ER editions, we build a collaboration network reflecting the 40 years of the conference and analyze it under different perspectives (from topological network features to topic evolution). We also expand the original dataset to include data on each author's country and gender, then broadening our analyses to consider homophily as well. At the end, this is a study that provides knowledge to the community about its members and evolution.

[2] https://dblp.uni-trier.de/.

[3] A query on DBLP returns 22 events: https://dblp.uni-trier.de/search/venue?q=social+networks, May/2020.

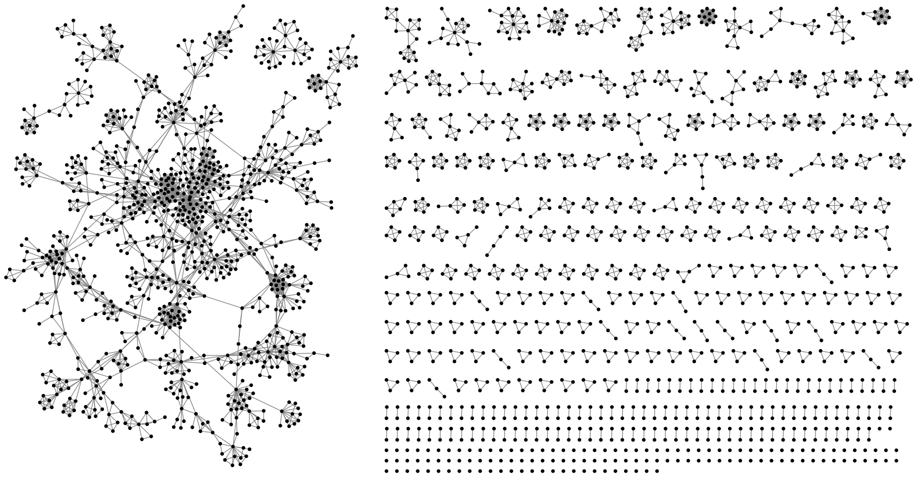


Fig. 1. General view of the complete ER collaboration network: almost 40% of the authors in the largest connected component, 45% within smaller connected components with three to 29 authors, 10.4% in pairs, and 5.6% as sole authors.

3. Data acquisition and network modeling

This section describes background information on data acquisition and network modeling. Since data acquisition is a central point in the present work, we start by describing how our dataset has been created and then we describe the network model we adopt to represent the ER collaboration network.

3.1. Our dataset

Our study considers both publication statistics and a collaboration network built upon all ER authors.[4] Its dataset comprises bibliographic data of the 38 editions of the ER conference held from 1979 to 2019 and collected from DBLP,[5] which includes for each paper: its title, year of publication and list of authors. Such a dataset was also expanded with the authors' affiliation, country and gender (sex) as follows. Altogether, we consider 1396 papers authored by 2311 researchers over the 40 years of the conference.

Authors' affiliations and countries were retrieved from DBLP Person Information. Then, a manual collection over ER proceedings was necessary for authors without such DBLP data. This method allowed us to get nearly 88% of the authors' affiliations and respective countries. Lastly, we also used two simple steps to infer gender: first, we run common gender predictions based on the name of the authors; then we manually identified the gender of authors without results in the previous step, which includes the investigation of the Google Scholar or institutional pages of authors. We were not able to infer gender for the authors who have gender-neutral names (and no other information, such as a Google Scholar page) and for those who have abbreviated their names (e.g., R. R. Brown). At the end, about 95% of the authors are associated with a gender, which is a good coverage for statistical purposes, as we shall see later.

3.2. Network model

Following usual modeling (e.g., [30]), we represent the ER collaboration network as a temporal graph $G_y = (V_y, E_y)$, where V_y is the set of vertices, E_y is the set of edges, and y is the year a network refers to. The graph $G_y = (V_y, E_y)$ is an undirected weighted graph, where the vertices represent authors, and the edges indicate that two authors have published together in or before the year y. Fig. 1 shows the complete ER collaboration network, i.e., with $y = 2019$. Built from all peer-reviewed papers in the main proceedings of the 38 ER editions, it contains a total of 2311 authors (vertices) and 3825 collaborations (edges), corresponding to 1396 papers. Such network modeling is helpful on the investigation of influential authors and highly connected authors, as well as features and properties of the network, for instance, which can provide valuable insights on the analysis of the ER conference over years.

As for topological features, the complete ER collaboration network has 466 distinct connected components, where the largest one contains 900 nodes, representing almost 40% of the whole community. In addition, 1042 nodes (45%) compose smaller components with three to 29 authors. There are also 240 nodes (10.4%) that form pairs of authors, and 129 nodes (5.6%) that correspond to sole authors. Finally, considering each ER edition, the average number of papers is 34.07 (with a standard deviation of 13.01), and the average number of authors is 85.68 (with a standard deviation of 39.13), while the average number of papers per author is 1.61 (with a standard deviation of 1.90), and the average number of authors per paper is 2.68 (with a standard deviation of 1.36).

[4] We consider *ER authors* as those who have published peer-reviewed papers in the main track of the ER conference (i.e., tutorials, keynotes, posters, demos and workshops are not considered here).

[5] DBLP ER webpage: https://dblp.uni-trier.de/db/conf/er/index.html.

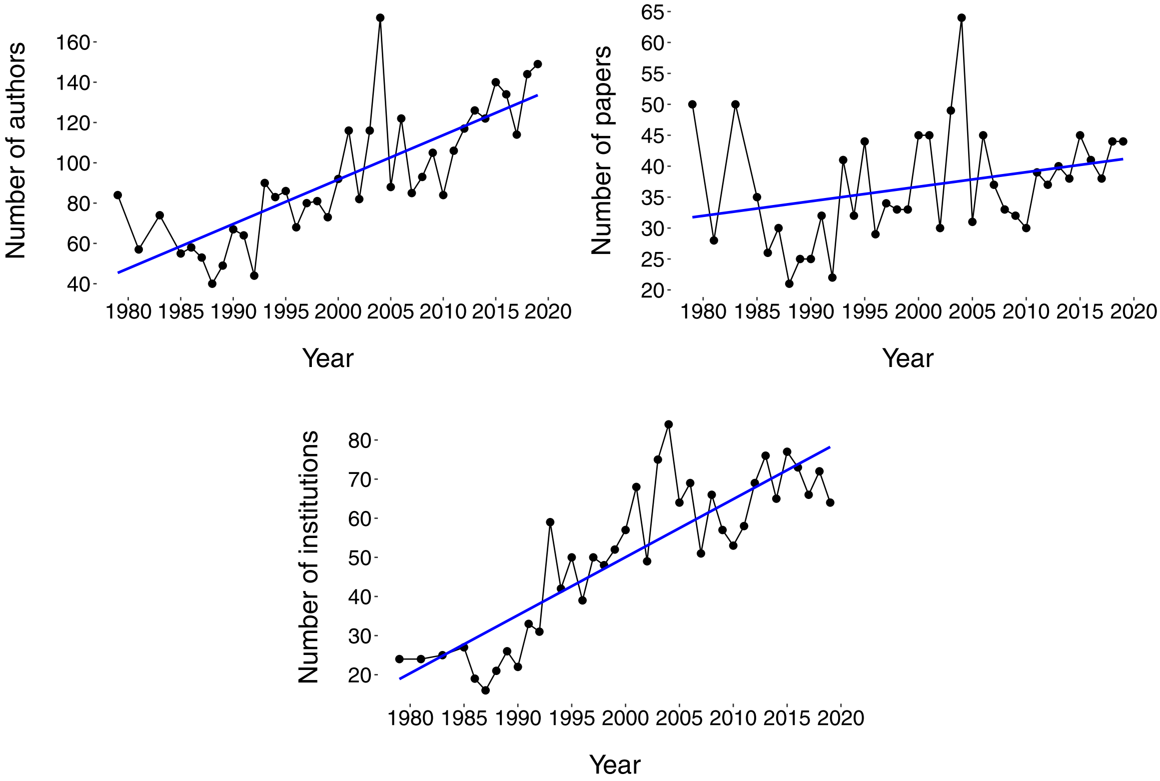


Fig. 2. Annual evolution of the number of distinct authors, papers and institutions over the 40 years (38 editions) of the ER conference.

Overall, in this article we investigate how the ER community evolves over time by analyzing different snapshots of its collaboration network. Given the bibliographic data collected from the DBLP for the 38 ER editions, we apply distinct metrics to characterize the ER community, its collaborations, and main topics approached in 40 years of history. Our analyses rely on bibliographic data, network metrics, and external data to provide an in-depth overview of the International Conference on Conceptual Modeling at the Age of 40. The next sections cover such analyses, starting with basic statistics and topic analyses.

4. Basic statistics and topic analyses

This section goes over basic statistics and addresses two specific topic analyses. We start by presenting some numbers on authorships, papers and affiliations, which is followed by a demographic view of the authors based on their countries (given by affiliation) and gender. Then, we address a topic analysis based on the paper titles and provide a view of the main topics addressed by the authors over time.

4.1. Basic statistics

Fig. 2 shows the evolution of the number of distinct authors (left), published papers (middle) and authors' institutions (right) per year. The regression lines on the graphs show the number of authors, papers and institutions has increased over the years, which suggests the ER conference has been attracting more contributions, with a part from new authors associated with new institutions, i.e., its community has been increasing steadily over the years. Also, the number of authors increases faster than the number of institutions, which increases faster than the number of papers. Such trends may indicate more collaboration on papers and ER spreading its reach towards new places as well.

Fig. 3 presents the ER network density over the years (left) and the node degree distribution (right) as viewed in 2019. The network density is calculated by dividing the number of edges by the number of nodes present in the network. Here, ER density oscillates in the first decade, but has a clear increasing trend later. Such result is possibly explained by the arrival of new authors that collaborate with other ones that are already in the network. In addition, the increasing density could be explained by authors who already belong to the network starting to collaborate together. We go deeper into the analysis of the collaborations in the next sections.

The degree of a vertex is given by the number of its adjacent edges. Hence, authors who have a high number of papers with different coauthors also have a high degree. The degree distribution is the probability distribution of these degrees over the entire

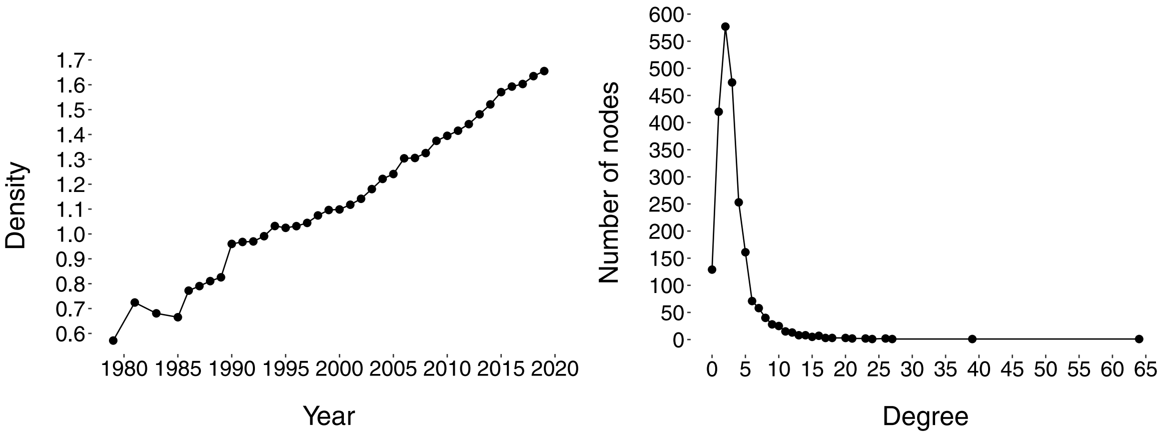


Fig. 3. Density evolution over the years (left) and degree distribution of the ER collaboration network (right).

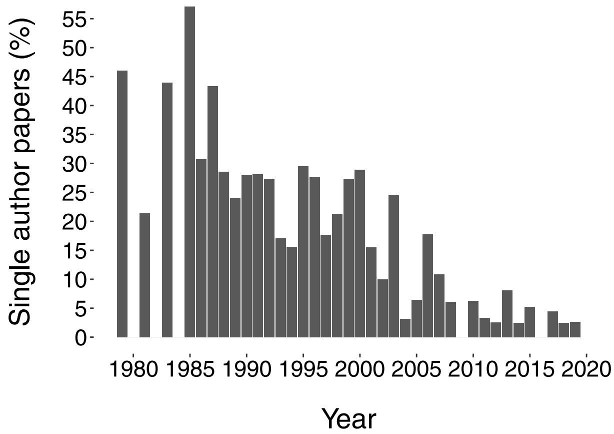


Fig. 4. Percentage of single-author papers per year.

Table 1
Distributions of authors per paper and of papers per author.

(a) Authors by published papers.		(b) Papers by number of authors.	
# papers	# authors (%)	# authors	# papers (%)
1	1764 (76.33)	1	248 (17.77)
2	288 (12.46)	2	463 (33.16)
3	97 (4.20)	3	375 (26.86)
4	58 (2.51)	4	202 (14.47)
5 or more	104 (4.50)	5 or more	108 (7.74)

network. According to Fig. 3 (right), the degree distribution of the ER network follows a power-law (disregarding authors with none or only one coauthor). In other words, there are few authors with high degree, and most of the authors have only few collaborations. Particularly, the ER collaboration network includes 129 single authors (i.e., authors with no collaboration at all), 420 authors with one coauthor (i.e., authors with degree equals 1), 577 with two coauthors (i.e., authors with degree equals 2) and 1314 with three or more coauthors (i.e., authors with degree at least 3).

Such network metrics are also explained by authorship distributions. Therefore, Table 1a and b show the distributions of authors per paper and papers per author, respectively. Most ER authors (76.33%) have only one paper published throughout the years, and about 24% have two or more papers. Pair-authored papers form one third of the ER publications (nearly 33%), which might comprise those papers written by students with their supervisors, whereas three-author papers account for about one fourth (26.86%) of the publications. This distribution clearly shows the conference has been attracting the participation of new researchers, a very important feature to keep the community strong.

As the number of single-author papers is relatively high, Fig. 4 shows its evolution over the years in percentage. Notice that the conference was not held in the years where there were no single author papers (1980, 1982, 1984), as ER started as a bi-annual event. There are 207 (8.96%) authors who have published single-author papers over four decades. From those, 129 are in fact

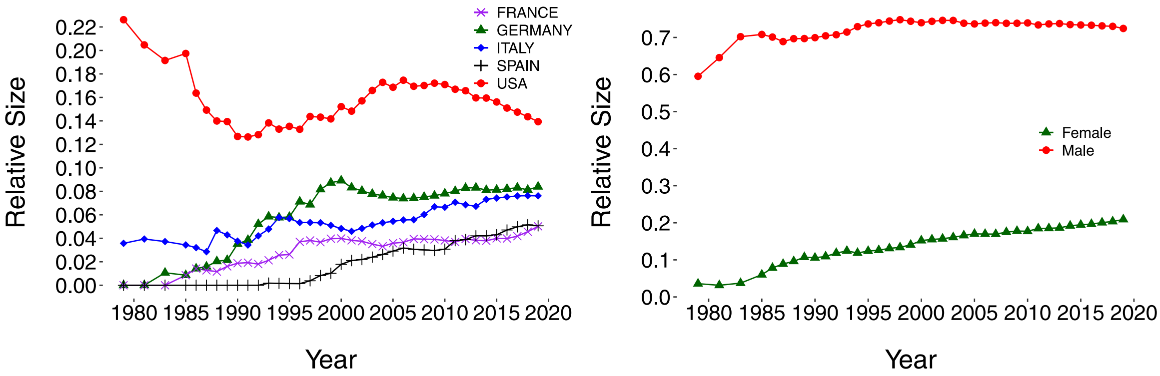


Fig. 5. Demographic attributes of authors over the years: country and gender.

single-paper authors. Specifically, from the initial 207 authors, 78 (37.68%) have node degree equal to or higher than 1. Moreover, many authors who wrote single-author papers also collaborate with other authors throughout the community history. For example, Jordi Cabot, Paul Johannesson and Tok Wang Ling show a high degree (24, 21 and 20, respectively) in the ER network (i.e., they are highly connected), but have also published single-author papers. We notice some peaks of single-author papers mainly in the first years of the conference; however, there is a decreasing trend of such papers in recent years. Such an evolution is a sign of a mature, engaging community. It may have started with single authors publishing their work; but now, most authors have their own research groups or developing teams (formed by coworkers and students, for example) such that writing a paper alone is not needed anymore.

4.1.1. Demography

We now investigate two demographic attributes by analyzing the country and gender (sex) of the authors. Before diving into this analysis, we manually evaluate the quality of our method for retrieving the authors' countries and genders.

Manual Validation. In order to manually verify whether gender and countries were correctly assigned, we selected a sample of 231 authors randomly (10% of the total of 2311 authors). The average number of papers for authors from the sample is 1.81 (mean value for all authors is 1.61). Regarding gender, after searching for the authors' Google Scholar page and Google Images, when it was not possible to verify gender with the first approach, we notice our method was able to correctly identify the gender of 94.8% authors.

For authors' countries, as most authors from the sample have only one paper (76.6%), we checked whether the country of the author according to our method was equal to the country of the author's first publication (obtained from ACM Digital Library[6] and SpringerLink[7] resources). Our method correctly associated the country of 88.13% authors. As authors may have multiple papers over years, we verified whether the country of authors' first publication has changed when compared with their last publication. Out of the 54 authors from the sample (23.4%) who have more than one ER paper, only six (11.1%) have changed their country, which might be an indication that our method for associating authors with their countries may still hold for those who have published more than one paper in the conference.

Despite the potential limitations our demographic methods might have, for instance, being limited by the accuracy of automatic gender predictors based on the name of authors, these results reinforce the quality of our methods to associate authors' countries and genders with minimal noise.

Data Evaluation. Regarding the authors' countries, we notice that ER, as an international conference, has attracted publications from 63 countries all over the world, being the United States the country with the highest number of authors as in 2019, 322 (13.93%). The other countries with more associated authors are Germany, Italy, Spain and France, with 194 (8.39%), 176 (7.61%), 117 (5.06%) and 116 (5.01%), respectively. These countries are followed by Australia, Brazil, China, Canada and UK, comprising together 18.69% of the authors. These results corroborate the internationalization of the ER community. Fig. 5 (left) shows the percentage of authors from the five most popular countries in terms of publications separated by year. We notice that the USA has reduced its relative presence over the years, which indicates that the conference has become more globalized over time.

Now considering the authors' gender, Fig. 5 (right) shows a large presence of male authors in the ER conference, at around 70%. Nonetheless, the participation of female authors is increasing over time, starting at 4% in the 1970's and reaching near 21% of the authors in 2019. Such value is way above the average in the field of Computer Science, which stays between 9% and 15% as pointed out by many studies [38]. The increasing percentage of women is an amazing accomplishment for the community. Still, there is much to be done to close the gender gap. Suggestions include having meetings specific for women during the conference to improve networking, establishing a mentor program for female graduate students and young researchers, and many other actions that would be synced with initiatives in the field [39]. In addition to authors who have gender-neutral names, many authors who

[6] https://dl.acm.org/.
[7] https://link.springer.com/.

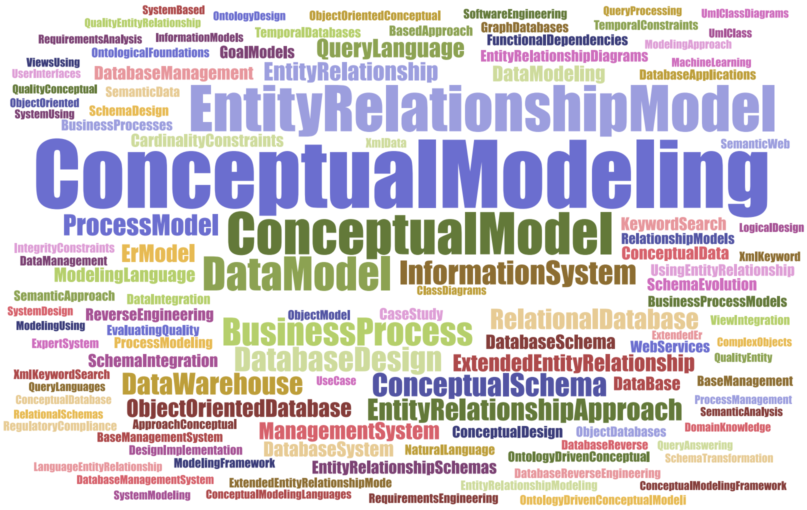


Fig. 6. Frequent terms from all titles of ER papers.

published in first-years have abbreviated names and, thus, we were not able to infer their gender. This explains the high number of unlabeled gender authors in the first years of the conference observed in Fig. 5 (right), the reason why the sum of the relative size of the genders does not sum 1.0.

4.2. Topic analyses

Now, changing the focus from the authors to the content of the papers, we analyze the main topics present at the conference four decades based on the title of its papers. Table 2 shows the top 20 most frequent bigrams (top) and trigrams (bottom), ranked by their frequency.[8] It also shows the first and the most recent year in which the bigrams/trigrams appeared at ER. For bigrams, *Conceptual Model*(*ing*) and *Entity-Relationship Model* are at the top, as expected. They are closely followed by *Data Model* and *Business Process*, and other terms that are part of the conceptual life cycle. Then, for trigrams, *Business Process Model* (BPM) leads the pack closely followed by *Entity-Relationship Model*. Note that *Entity-Relationship* appears again in the table with different flavors, such as *Extended Entity-Relationship Model*. Other trigrams are equally interesting such as those with the terms *XML*, *Ontology* and *UML*. All of them show the diversity and wide range of research interest within ER community. Another interesting analysis considers the time interval in which such terms appear in ER titles. Time frames emphasize that some terms are at the core of the whole ER history (e.g., *Conceptual Model/ing*), whereas others appear within closed time windows (e.g., *Object-Oriented Database System* from 1986 to 1995). New trends are also clear with more recent initial year, such as *Business Process* from 1994 on, *Data Warehouse* from 1998 on and *Ontology-Driven Conceptual Modeling* from 2016 on.

To make such analysis more visual, Fig. 6 illustrates the most frequent terms found in titles of all ER papers by means of a wordcloud.[9] The figure considers other words and combinations not shown in the previous table, such as *Web Services*. Moreover, many published papers focus on understanding the impact that conceptual modeling techniques might inflict on databases, business strategies and processes, thus making the appearance of such terms also frequent. Other common terms are *Database Schema*, *Cardinality Constraints*, *Schema Evolution* and *Modeling Language*, which are also topics closely related to the ER conference and well covered by it.

Figs. 2 and 3 coupled with Tables 1 and 2 clearly show that the ER conference has increased not only in number of authors and collaborations over the years, but also in diversity of topics. Putting all together, there are authors, usually with long careers, who have worked in distinct topics, which can be easily inferred from information extracted from their paper titles. Examples include John Mylopoulos whose ER titles include keywords such as *Requirements*, *Software*, *Reasoning* and *Business*, and David W. Embley with *Web Data*, *Ontology*, and *Languages*. Nonetheless, the increasing participation of new authors may be explained by new topics addressed at the conference over the years. A first example is *Ontology*, a subject introduced by Veda C. Storey at ER 1997 and later expanded by the participation of Giancarlo Guizzardi and his group from ER 2006 onwards. Another topic that has attracted the attention of the ER community in recent years is *Business Process Modeling*, whose variant *Business Process Model* is the top trigram in Table 2. Such a topic first appeared in paper titles at ER 1994, with three papers, and over 50 authors have addressed it since then. With so many people being part of the ever growing ER community, the next two sections go deeper at the community from the authors participation and collaboration perspectives.

---

[8] In Table 2, when identifying the most frequent bigrams and trigrams, we merged the counts of terms in singular and plural (e.g., *Data Model* and *Data Models*), as well as of words with the same meaning but distinct spelling (e.g., *Modeling* and *Modeling*).

[9] Word Cloud Generator: https://www.jasondavies.com/wordcloud/.

Table 2
Top 20 most frequent bigrams and trigrams extracted from all titles of the ER papers, with the first and most recent year in which the term appeared at ER.

BIGRAMS	#	First	Recent
CONCEPTUAL MODELING	84	1979	2019
CONCEPTUAL MODEL	50	1979	2019
ENTITY-RELATIONSHIP MODEL	49	1979	2012
DATA MODEL	40	1979	2015
BUSINESS PROCESS	37	1994	2019
DATABASE DESIGN	29	1979	2014
CONCEPTUAL SCHEMA	26	1981	2019
INFORMATION SYSTEM	24	1979	2016
DATA WAREHOUSE	22	1998	2019
ENTITY-RELATIONSHIP APPROACH	22	1979	1993
PROCESS MODEL	21	1998	2018
QUERY LANGUAGE	21	1979	2006
RELATIONAL DATABASE	20	1986	2014
ENTITY-RELATIONSHIP SCHEMA	17	1979	2015
ENTITY RELATIONSHIP	16	1985	2016
OBJECT-ORIENTED DATABASE	16	1986	2007
ER MODEL	15	1983	2014
EXTENDED ENTITY-RELATIONSHIP	15	1981	1995
MANAGEMENT SYSTEM	15	1979	2018
CONCEPTUAL DATA	14	1979	2015
TRIGRAMS	#	First	Recent
BUSINESS PROCESS MODEL	9	2000	2013
ENTITY RELATIONSHIP MODEL	8	1993	2016
EXTENDED ENTITY-RELATIONSHIP MODEL	7	1981	1993
DATA BASE MANAGEMENT	6	1979	1988
ENTITY-RELATIONSHIP MODELING	6	1979	2007
CONCEPTUAL DATA MODEL	5	1979	2013
DATABASE MANAGEMENT SYSTEM	5	1983	1991
CONCEPTUAL DATA MODELING	5	1985	2011
BUSINESS PROCESS MODELING	5	1994	2015
XML KEYWORD SEARCH	5	2011	2015
ONTOLOGY-DRIVEN CONCEPTUAL MODELING	5	2016	2019
SEMANTIC DATA MODEL	4	1979	1993
DATABASE REVERSE ENGINEERING	4	1993	1998
CONCEPTUAL MODELING FRAMEWORK	4	2000	2016
CONCEPTUAL MODELING LANGUAGES	4	2003	2016
UML CLASS DIAGRAM	4	2006	2013
OBJECT-ORIENTED DATABASE SYSTEM	3	1986	1995
ENTITY-RELATIONSHIP QUERY LANGUAGE	3	1987	1991
EXTENDED ENTITY-RELATIONSHIP SCHEMA	3	1989	1995
EXTENDED ER MODEL	3	1990	2007

5. Influential authors

Even though there is no single way of measuring influence in a coauthorship network, we address influential authors along two distinct axes. First, we overview the most prolific authors according to their number of publications. Then, we assess influence by considering the degree of nodes and correlating it with the authors' number of publications.

Table 3 shows the top 20 authors in four distinct time intervals sorted by their cumulative number of papers at each decade. John Mylopoulos, who is the author with the highest degree in the network (highest number of collaborations), is also the one with the highest number of publications by 2019. Interestingly, the top five authors in the most recent decade have different paths within ER history: John Mylopoulos appeared in the list in the previous decade; Tok Wang Ling and David W. Embley have been part of the top 20 list the whole time; Giancarlo Guizzardi is a rising star, being part of the top 20 only at the last decade; and Veda C. Storey is present for the last three decades.

Continuing with the whole history perspective in Table 3, four authors appeared along the four decades (Peter Chen, Tok Wang Ling, Shamkant B. Navathe and David W. Embley), four along the first three decades (Carlo Batini, Antonio L. Furtado, Marco A. Casanova and Ramez Elmasri) and two along the last three decades (Paul Johannesson and Veda Storey). In other words, ER has shown a good mix of prolific authors along its history, with new productive authors assuming more prominent positions, as the first ones give more space, mostly due to retirement or new endeavors in other areas.

Next, we analyze influence in a different perspective by investigating the degree centrality as an influence metric. Table 4 shows the top 20 authors with the highest degree centrality values as viewed in 2019, with the first year they published at ER conference and the most recent one. The top 10 authors with the highest degree centrality also appear in Table 3, with the exception of Arthur H. M. ter Hofstede. We notice there is a low positive correlation between both rankings (Kendall tau correlation equals 0.25, *p-value* < 0.001), which means that ER authors with a high number of publications do not necessarily have a high degree, i.e., they do not

Table 3
Most prolific authors according to their number of publications accumulated by decade.

(a) 1979–1989		(b) 1990–1999	
Author	#	Author	#
Peter P. Chen	11	Peter P. Chen	11
Adarsh K. Arora	5	Tok Wang Ling	11
Carlo Batini	5	Shamkant B. Navathe	10
Asuman Dogac	5	Marco A. Casanova	8
Maurizio Lenzerini	5	Ramez Elmasri	8
Antonio L. Furtado	4	Paul Johannesson	8
Tok Wang Ling	4	Carlo Batini	7
Victor M. Markowitz	4	Valeria De Antonellis	7
Shamkant B. Navathe	4	Antonio L. Furtado	7
Peter A. Ng	4	Daniel L. Moody	7
Hirotaka Sakai	4	Veda C. Storey	7
Paolo Atzeni	3	Silvana Castano	6
Marco A. Casanova	3	David W. Embley	6
Ramez Elmasri	3	Martin Gogolla	6
David W. Embley	3	Michael Schrefl	6
Sushil Jajodia	3	Adarsh K. Arora	5
James A. Larson	3	Asuman Dogac	5
Ewing L. Lusk	3	Jean-Luc Hainaut	5
Michael V. Mannino	3	Gerti Kappel	5
David S. Reiner	3	Mong-Li Lee	5
(c) 2000–2009		**(d) 2010–2019**	
Author	#	Author	#
Tok Wang Ling	19	John Mylopoulos	39
David W. Embley	15	Tok Wang Ling	24
John Mylopoulos	13	Giancarlo Guizzardi	23
Peter P. Chen	11	David W. Embley	19
Paul Johannesson	11	Veda C. Storey	17
Il-Yeol Song	11	Juan Trujillo	16
Bernhard Thalheim	11	Bernhard Thalheim	14
Marco A. Casanova	10	Jordi Cabot	13
Antonio L. Furtado	10	Xavier Franch	13
Shamkant B. Navathe	10	Il-Yeol Song	13
Veda C. Storey	10	Paolo Giorgini	12
Ramez Elmasri	9	Mong-Li Lee	12
Jean-Luc Hainaut	9	Stephen W. Liddle	12
Mong-Li Lee	9	Antoni Olivé	12
Daniel L. Moody	9	Oscar Pastor	12
Wilfred Ng	9	Peter P. Chen	11
Carlo Batini	8	Lois M. L. Delcambre	11
Juan Trujillo	8	Paul Johannesson	11
Klaus-Dieter Schewe	8	Shamkant B. Navathe	11
Antoni Olivé	7	Eric S. K. Yu	11

collaborate with many distinct authors. For instance, Tok Wang Ling, who has the second highest number of publications, appears in the 13th position in the degree table. One possible explanation is that these authors might collaborate consistently with others that held previous collaborations, which directly impacts on their node degrees. Also, those authors have published within a relative large range of years, as we can see by the last two columns of the table. Overall, the time intervals show there is no relation between having published at the conference the longest and having a high/low degree.

As an effort to determine whether influential ER authors are also influential in worldwide scholarly communities, Table 4 shows the h-index[10] of the top 20 degree centrality researchers, as collected from Google Scholar[11] in September 2020. For those we found a Google Scholar page, most have a high all time h-index (on average 45.83) and the average h-index since 2015 (27.28) shows that many of them have kept a relative high h-index throughout their careers. Notice that John Mylopoulos, who is the ER author with the highest degree centrality, is also the one with the highest value of the all time h-index.

6. Analysis of the collaboration network

In this section, we analyze the ER collaboration network from three different perspectives: connected components and some network metrics, network groups and homophily.

[10] The h-index [40] of an author is the highest number of her publications with at least that many citations, e.g., an author with five papers with at least five citations each has an h-index equal to 5.

[11] https://scholar.google.com/.

Table 4
Top 20 degree centrality ER authors with their first and most recent year of publication, and respective h-index for all time and just since 2015 as for September 2020.

Author	Degree	First	Recent	h-index (All Time)	h-index (Since 2015)
John Mylopoulos	64	1994	2018	91	41
Giancarlo Guizzardi	39	2002	2019	47	33
David W. Embley	27	1985	2018	n/a	n/a
Oscar Pastor	26	1998	2018	48	25
Xavier Franch	26	2005	2019	43	25
Jordi Cabot	24	2003	2019	40	36
Stephen W. Liddle	23	1998	2018	22	11
Arthur H. M. ter Hofstede	23	1996	2018	74	44
Veda C. Storey	21	1989	2018	n/a	n/a
Paul Johannesson	21	1989	2011	34	19
Juan Trujillo	20	2002	2019	42	25
Panos Vassiliadis	20	2003	2017	44	26
Tok Wang Ling	20	1985	2019	39	19
Dolors Costal	18	1997	2019	16	9
Paolo Giorgini	18	2000	2017	55	31
Ricardo de Almeida Falbo	18	2013	2019	32	20
Angelo Susi	17	2009	2015	33	21
Mathias Weske	17	2008	2019	58	36
Nicola Guarino	17	2000	2019	50	32
Shamkant B. Navathe	16	1983	2015	57	28

6.1. Connected components

A *connected component* of an undirected graph is a subgraph in which any two vertices are connected to each other by paths. Fig. 1 gave an overview of the ER connected components and basic statistics. Now, Fig. 7 shows the number of connected components over time (Fig. 7a), the number of newcomers connected or not to the largest connected component — LCC (Fig. 7b), and the relative size of the LCC and the second largest connected component — SLCC (Fig. 7c). The number of connected components has been increasing over the years. Also, the largest connected component has a higher relative size when compared to the other components, with a clear increasing trend for recent years. Nonetheless, newcomers (ER first time authors) tend to *not* join the LCC but smaller or new components instead. Such trend explains the rising number of connected components over the years and indicates that the ER conference has been attracting new authors with no previous collaborations with ER frequent authors, thus creating large disjoint groups. We notice that such behavior is important for the good health of the conference over the long term, since it helps to renew its community.

Fig. 7c shows that the LCC trend started to increasingly rise after 2005. Such trend can be explained by large groups of authors collaborating with each other and, therefore, becoming part of the largest component. Furthermore, as the giant component becomes larger, many newcomers start to join directly the LCC as observed in Fig. 7b, which also increases the relative size of the LCC when compared to the SLCC and other components.

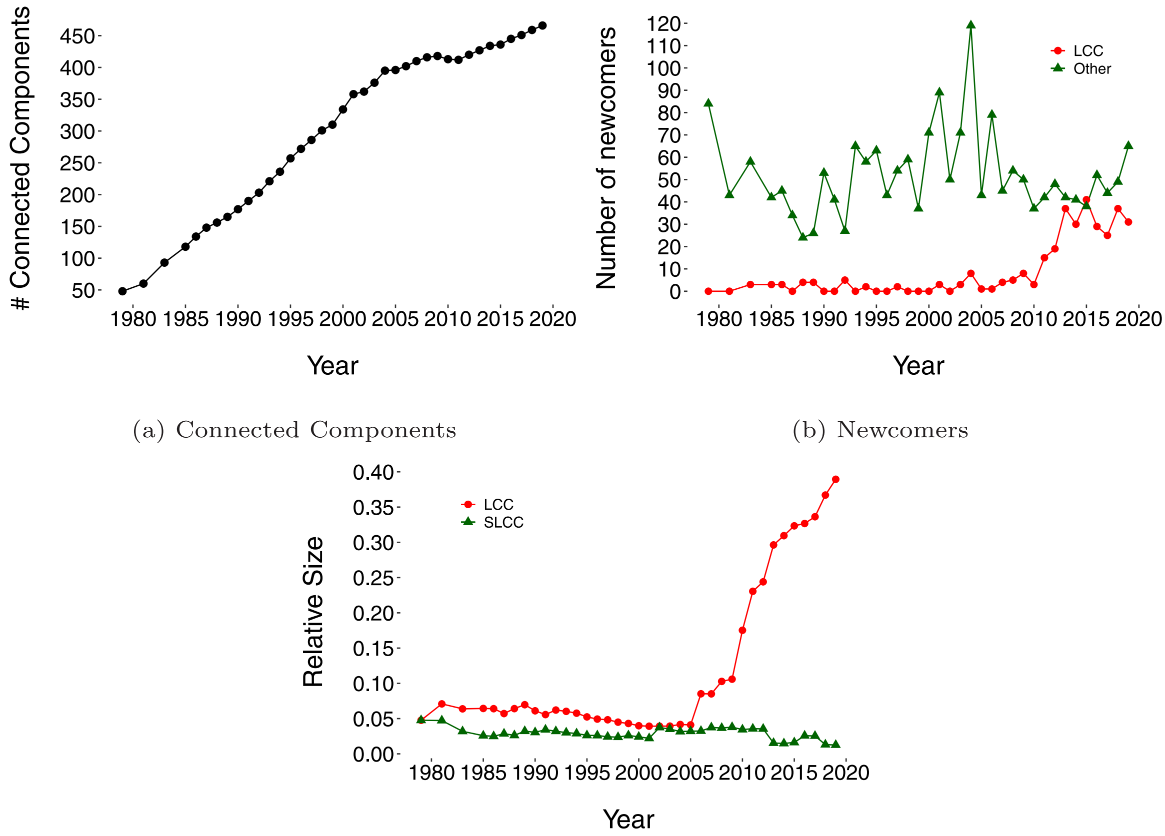
Fig. 8 shows the evolution of the *clustering coefficient* and of the *average shortest path length* for the ER network over the years considering the largest connected component. The clustering coefficient for a graph indicates whether its nodes tend to cluster together, whereas the average shortest path length indicates the number of steps along the shortest paths for all possible pairs of graph nodes. We compare the values of these metrics for the ER network (called *real*) with their equivalent *random* networks, i.e., the same number of nodes and edges randomly chosen.

In 2019, the LCC shows a high global clustering coefficient of 66.4% (Fig. 8 left), indicating dense groups of authors who publish together. A high clustering coefficient compared to its equivalent random network and a small average shortest path as low as its equivalent random network characterize the ER network as a *small-world network* [41] in the first three decades of the conference (Fig. 8 right). For recent years, the average shortest path of the ER network has become larger when compared with a random one. One possible explanation is that, as previously noted, after 2005 the LCC increased its relative size compared with the entire network, which created larger paths from one node to all the others in the LCC and, therefore, increased the average shortest path length for all nodes.

6.2. Network groups

Decomposing a complex network into groups (sets of highly connected nodes) is very important, as it may help to understand a-priori unknown features and properties of the network. In this section, we focus on discovering and analyzing collaboration groups inside the ER community.

In order to do so, we use a k-clique algorithm [42] that relaxes the notion of clique and has shown great success in detecting clusters on a large scale. A collaboration group is then defined as the maximal union of k-cliques that can be reached from each other by a series of adjacent k-cliques (they are adjacent if they share $k-1$ nodes). The k-clique algorithm returns a large number of groups inside ER, thus showing that collaboration has greatly increased over the years. Fig. 9 shows the number of *3*-cliques, *4*-cliques and *5*-cliques over the years.



(a) Connected Components (b) Newcomers

(c) Relative size of LCC and SLCC

Fig. 7. Number of connected components over time (a); number of newcomers connected or not to the largest connected component — LCC (b); and relative size of the largest connected component and second largest connected component — SLCC (c).

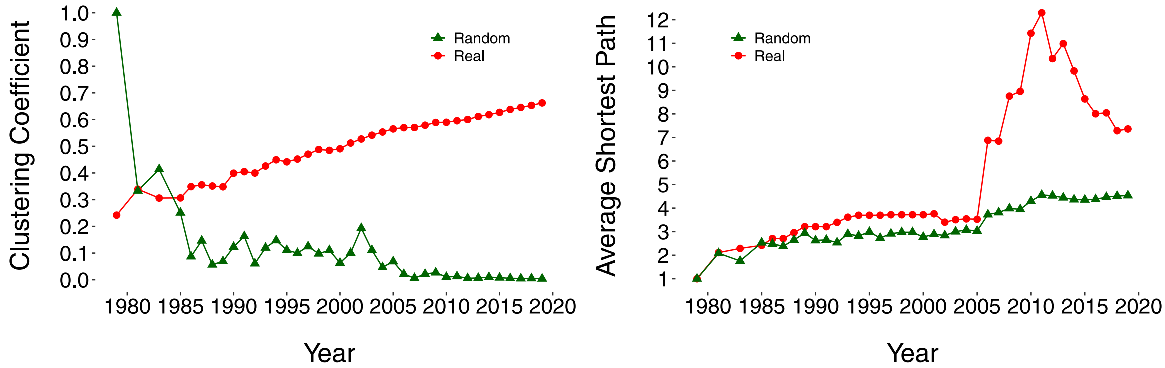


Fig. 8. Clustering coefficient (left) and average shortest path length (right) over time compared to their equivalent random networks.

For the complete network in 2019, there are 439 *3*-cliques, 256 *4*-cliques, and 103 *5*-cliques formed. The average number of authors per groups is 4.5 for *3*-cliques, 4.9 for *4*-cliques, and 5.9 for *5*-cliques, indicating that many groups are not formed by many authors. The largest *3*-clique has 98 authors, whereas the largest *4*-clique contains 26 authors. The largest *4*-clique group is composed of authors mainly from Brazil, with 10 of such authors being affiliated with the Universidade Federal do Espírito Santo. Also, most of these authors are connected with John Mylopoulos (Canada), who is one of the authors who creates a bridge to the second largest *4*-clique, composed of 14 authors. In the latter, six authors are affiliated with the Polytechnic University of Catalonia (Spain). Such results indicate that affiliation and, therefore, geographical location are important factors for creating groups within

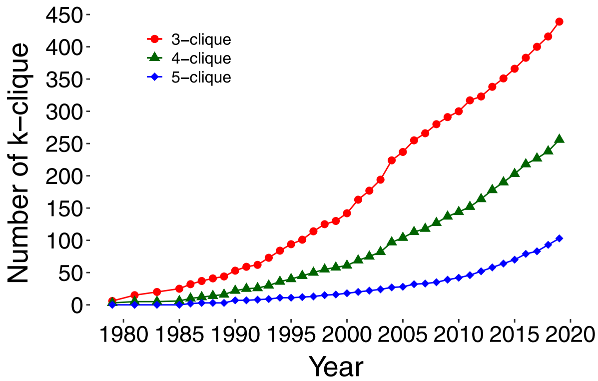


Fig. 9. Number of 3-cliques, 4-cliques and 5-cliques over the years.

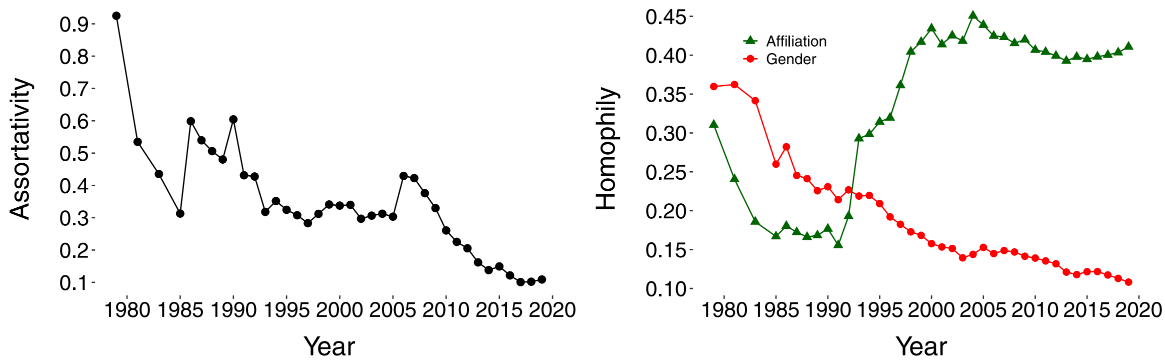


Fig. 10. Assortativity and homophily over time.

the ER community. To dive deeper into the impact such attributes might inflict on collaborations, we next investigate homophily on the ER network over time.

6.3. Homophily and further analyses

Homophily is the tendency of people to connect with similar ones. To better evaluate it, we first analyze the assortativity of the network, which measures the similarity of its connections regarding node degree. Then, we analyze homophily investigating the affiliation and gender of the authors to understand how collaborations are influenced by similar author characteristics.

The assortativity of the ER network (Fig. 10 left) has a decreasing trend, which indicates that nodes with high degree tend to connect to those with low degree. In other words, new authors with few collaborations tend to increasingly connect to authors with higher degree. This kind of behavior is usually observed between students and their supervisors, and between junior and senior researchers. This is potentially another indication that the ER network is always renewing, with young researchers being connected with senior ones, and so on.

Regarding affiliation and gender (Fig. 10 right), collaborations are becoming more diverse over time. We notice, however, there is higher affiliation homophily when compared to gender. Such a result is expected as there is a high number of distinct connected components (as observed in the previous section), and many of such components often include people from same affiliations. Overall, homophily is less than 0.5 for both metrics, indicating that dissimilar characteristics are observed more often than similar ones.

To complement the above analyses, Table 5 shows the top 12 institutions with the highest number of affiliated authors as viewed in 2019. Despite the highest relative number of authors being affiliated with institutions in the United States, we notice that the institutions with the highest number of affiliated authors are from distinct European countries. The top two institutions are Polytechnic University of Catalonia, with authors Xavier Franch, Dolors Costal and others, and University of Trento, with authors Paolo Giorgino, Fabio Casati and others. Also, institutions from authors who have appeared in other rankings are present here as well, such as: Queensland University of Technology (Arthur H. M. ter Hofstede), Brigham Young University (David W. Embley and Stephen W. Liddle), PUC-Rio (Antonio L. Furtando and Marco A. Casanova), National University of Singapore (Tok Wang Ling), and Polytechnic University of Valencia (Oscar Pastor).

Table 5
Top-12 institutions ranked by the number of distinct affiliated authors.

Institution	# Authors	Country
Polytechnic University of Catalonia	30	Spain
University of Trento	29	Italy
IBM T. J. Watson Research Center	20	USA
Queensland University of Technology	18	Australia
Politecnico di Milano	18	Italy
Brigham Young University	18	USA
RWTH Aachen University	16	Germany
University of Namur	15	Belgium
Pontifical Catholic University of Rio de Janeiro	15	Brazil
Eindhoven University of Technology	15	Netherlands
National University of Singapore	15	Singapore
Polytechnic University of Valencia	15	Spain

7. Conclusions

In this article, we went over the 40 years of history of the ER conference to carry out an in depth analysis of its collaboration network over the years. In particular, we investigate basic statistics, influential authors and metrics of the collaboration network, as well as collaboration groups and characteristics of authors in terms of demographic attributes.

Among our main findings, we show that the ER conference has an increasing participation of new institutions and authors, who publish in a wide range of topics that go from Conceptual Modeling to Business Process Modeling. On demographic attributes of authors, our in-depth analysis reveals the participation of female authors is increasing over time, although male authors still represent the largest fraction of the community. Authors are mainly affiliated to institutions in the United States, although there has been a large diversity of institutions and countries in recent years. Such a trend characterizes this international conference as a globalized one which welcomes research papers from all over the world.

Furthermore, even though the number of publications and the degree centrality metric are not definitive to pick influential authors, they help pointing highly prolific authors who have been active members of the community over the years. Likewise, our investigation on the collaboration network and its groups revealed the ER network is characterized as a small-world network for the first years of the conference. There is also an increasing number of connected components and many of such components are formed by newcomers who publish with other authors who are not directly connected to the giant component. The largest complete component corresponds to almost 40% of authors, and 45% of smaller components have three to 29 authors.

Finally, we investigate the tendency of authors to connect with similar ones by means of the assortativity and homophily of the authors given their gender and affiliation. Our study reveals the ER conference has been renewing its collaborations, where authors with high degree values collaborate with ones with low degree. Also, the analysis of homophily leads to presenting more dissimilar characteristics than similar ones, indicating that publications are becoming more diverse in terms of gender and affiliation.

Overall, we hope our analyses and findings help to provide the ER community with more information about its members and its evolution, leaving as future work a comparative study of the ER network with other known conferences in Computer Science.

CRediT authorship contribution statement

Lucas Henrique C. Lima: Performed data collection, Basic statistics and was in charge of all other metrics, Writing the paper. **Alberto H.F. Laender:** Designed the social study, Analyzing the ER evolution, Writing the paper. **Mirella M. Moro:** Analyzing the ER evolution, Writing the paper. **José Palazzo M. de Oliveira:** Analyzing the ER evolution, Writing the paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially funded by the authors' individual grants from CNPq, Brazil and FAPEMIG, Brazil, and also by institutional grants from the Ministry of Education's agency CAPES, Brazil - Finance Code 001. The authors would like to thank the reviewers for their constructive comments. All authors read and approved the final manuscript.

References

[1] P.P. Chen, The entity-relationship model - toward a unified view of data, ACM Trans. Database Syst. 1 (1) (1976) 9–36.
[2] P.P. Chen, The entity-relationship model: Toward a unified view of data, in: Proceedings of the International Conference on Very Large Data Bases, Framingham, Massachusetts, USA, 1975, pp. 173.

[3] P.P. Chen, 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., Springer Berlin Heidelberg, 2009, pp. 1–8.
[4] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. 74 (1) (2002) 47.
[5] A. Barabási, Network Science, Cambridge University Press, 2016.
[6] M.E.J. Newman, Coauthorship networks and patterns of scientific collaboration, Proc. Natl. Acad. Sci. 101 (Suppl 1) (2004) 5200–5205.
[7] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (7043) (2005) 814–818.
[8] S. Wasserman, K. Faust, Social Network Analysis - Methods and Applications, Structural Analysis in the Social Sciences, vol. 8, Cambridge University Press, 2007.
[9] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (6684) (1998) 440–442.
[10] Z. Bao, Y.C. Tay, J. Zhou, Sonschema: A conceptual schema for social networks, in: 32th International Conference on Conceptual Modeling, ER 2013, Springer, Hong-Kong, China, 2013, pp. 197–211, http://dx.doi.org/10.1007/978-3-642-41924-9_18.
[11] C. Chen, W.Z. Il-Yeol Song, Trends in citation analysis of the ER conference papers (1979-2005), in: Proceedings of the 11th International Conference on the International Society for Scientometrics and Informetrics, Society for Scientometrics and Informatrics, Leuven, Belgium, 2007, pp. 189–200.
[12] U. Dayal, C. Gupta, M. Castellanos, S. Wang, M. García-Solaco, Of cubes, DAGs and hierarchical correlations: A novel conceptual model for analyzing social media data, in: International Conference on Conceptual Modeling, ER 2012, Springer, Florence, Italy, 2012, pp. 30–49, http://dx.doi.org/10.1007/978-3-642-34002-4_3.
[13] Q. Wang, Network analytics ER model - towards a conceptual view of network analytics, in: 33rd International Conference on Conceptual Modeling, ER 2014, Springer, Atlanta, GA, USA, 2014, pp. 158–171, http://dx.doi.org/10.1007/978-3-319-12206-9_13.
[14] M.E.J. Newman, Who is the best connected scientist? A study of scientific coauthorship networks, in: Complex Networks, Springer, 2004, pp. 337–370.
[15] M.E. Newman, Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality, Phys. Rev. E 64 (1) (2001) 016132.
[16] C. Kozma, C. Calero-Medina, The role of South African researchers in intercontinental collaboration, Scientometrics 121 (3) (2019) 1293–1321, http://dx.doi.org/10.1007/s11192-019-03230-9.
[17] A.H.F. Laender, C.J.P. de Lucena, J.C. Maldonado, E. de Souza e Silva, N. Ziviani, Assessing the research and education quality of the top Brazilian computer science graduate programs, SIGCSE Bull. 40 (2) (2008) 135–145.
[18] Q. Luo, J.C. Xia, G. Haddow, M.A. Willson, J. Yang, Does distance hinder the collaboration between Australian universities in the humanities, arts and social sciences?, Scientometrics 115 (2) (2018) 695–715, http://dx.doi.org/10.1007/s11192-018-2686-x.
[19] J.P. Mena-Chalco, L.A. Digiampietri, F.M. Lopes, R.M. Cesar, Brazilian bibliometric coauthorship networks, J. Assoc. Inf. Sci. Technol. (ISSN: 2330-1643) 65 (7) (2014) 1424–1445.
[20] G.V. Menezes, N. Ziviani, A.H.F. Laender, V. Almeida, A geographical analysis of knowledge production in computer science, in: Procs Int'L World Wide Web Conference, Madrid, Spain, 2009, pp. 1041–1050.
[21] L.A. Digiampietri, et al., Brax-ray: An x-ray of the Brazilian computer science graduate programs, PLoS One 9 (4) (2014).
[22] G.R. Lopes, M.M. Moro, R. da Silva, E.M. Barbosa, J.P.M. de Oliveira, Ranking strategy for graduate programs evaluation, in: Proceedings of International Conference on Information Technology and Application, 2011, pp. 253–260.
[23] D. Arruda, F. de Lima Bezerra, V.A. Neris, P.R.D. Toro, J. Wainer, Brazilian computer science research: gender and regional distributions, Scientometrics 79 (3) (2009) 651–665.
[24] H. Lima, T.H.P. Silva, M.M. Moro, R.L.T. Santos, W.M. Jr., A.H.F. Laender, Assessing the profile of top Brazilian computer science researchers, Scientometrics 103 (3) (2015) 879–896.
[25] H. Lima, T.H.P. Silva, M.M. Moro, R.L.T. Santos, W. Meira Jr., A.H.F. Laender, Aggregating productivity indices for ranking researchers across multiple areas, in: Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries, 2013, pp. 97–106.
[26] S. Partelow, A.-K. Hornidge, P. Senff, M. Stäbler, A. Schlüter, Tropical marine sciences: Knowledge production in a web of path dependencies, PLoS One 15 (2) (2020) e0228613, http://dx.doi.org/10.1371/journal.pone.0228613.
[27] D. Purwitasari, et al., Identifying collaboration dynamics of bipartite author-topic networks with the influences of interest changes, Scientometrics 122 (3) (2020) 1407–1443, http://dx.doi.org/10.1007/s11192-019-03342-2.
[28] T.J. Ameloot, M. Marx, W. Martens, F. Neven, J. Van Wees, 30 years of PODS in facts and figures, ACM SIGMOD Rec. 40 (3) (2011) 54–60.
[29] A.L.C. Bazzan, V.F. Argenta, Network of collaboration among PC members of Brazilian computer science conferences, J. Braz. Comput. Soc. (ISSN: 1678-4804) 17 (2) (2011) 133–139.
[30] G. Maia, P.O.V. de Melo, D.L. Guidoni, F.S. Souza, T.H. Silva, J.M. Almeida, A.A. Loureiro, On the analysis of the collaboration network of the Brazilian symposium on computer networks and distributed systems, Journal of the Brazilian Computer Society 19 (3) (2013) 361–382.
[31] M.A. Nascimento, J. Sander, J. Pound, Analysis of sigmod's co-authorship graph, SIGMOD Rec. (ISSN: 0163-5808) 32 (3) (2003) 8–10.
[32] V.P. Freire, D.R. Figueiredo, Ranking in collaboration networks using a group based metric, J. Braz. Comput. Soc. 17 (4) (2011) 255–266.
[33] Y.-X. Liu, B. Lu, Q. Zhang, Empirical analysis of the coauthorship network based on DBLP, in: International Conference on Machine Learning and Cybernetics, Tianjin, China, 2013, pp. 1070–1076, http://dx.doi.org/10.1109/ICMLC.2013.6890752.
[34] L.H.C. Lima, G. Penha, L.M. de Alencar Rocha, M.M. Moro, A.P.C. da Silva, A.H.F. Laender, J.P.M. de Oliveira, The collaboration network of the Brazilian symposium on databases - 30 editions of history, J. Braz. Comput. Soc. 23 (1) (2017) 10:1–10:16.
[35] V. Cosentino, J.L.C. Izquierdo, J. Cabot, Metascience: An holistic approach for research modeling, in: International Conference on Conceptual Modeling, Springer, 2016, pp. 365–380.
[36] A.F. Smeaton, G. Keogh, C. Gurrin, K. McDonald, T. Sødring, Analysis of papers from twenty-five years of SIGIR conferences: What have we been doing for the last quarter of a century?, SIGIR Forum 37 (1) (2003) 49–53.
[37] X. Liu, J. Bollen, M.L. Nelson, H. Van de Sompel, Co-authorship networks in the digital library research community, Inf. Process. Manage. 41 (6) (2005) 1462–1480.
[38] S. Mattauch, K. Lohmann, F. Hannig, D. Lohmann, J. Teich, A bibliometric approach for detecting the gender gap in computer science, Commun. ACM 63 (5) (2020) 74–80, http://dx.doi.org/10.1145/3376901.
[39] D. Garcia, M. Charikar, E. Hearn, E. Lazowska, J. Reynolds, Institutions share successes, failures, and advice in moving the diversity needle, in: Proceedings of the 51st ACM Technical Symposium on Computer Science Education, 2020, pp. 331–332, http://dx.doi.org/10.1145/3328778.3366976.
[40] J.E. Hirsch, An index to quantify an individual's scientific research output, Proc. Natl. Acad. Sci. 102 (46) (2005) 16569–16572.
[41] J. Kleinberg, The small-world phenomenon: An algorithmic perspective, in: Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, Association for Computing Machinery, New York, NY, USA, 2000, pp. 163–170, http://dx.doi.org/10.1145/335305.335325.
[42] L. Cavique, A.B. Mendes, J.M.A. Santos, An algorithm to discover the k-clique cover in networks, in: Proceedings of the 14th Portuguese Conference on Progress in Artificial Intelligence, Springer, Aveiro, Portugal, 2009, pp. 363–373.

Lucas Henrique Costa de Lima holds a Master's degree in Computer Science from the Universidade Federal de Minas Gerais (UFMG, Brazil — 2019). He received his Bachelor's Degree in Information Systems at the same University in 2017. He was a fully funded scholarship recipient (CAPES) at the State University of New York — Stony Brook University (EUA, 2015–2016), and he held research intern positions at the Information Sciences Institute, University of Southern California — USC (EUA, 2016), and at the Max Planck Institute for Software Systems — MPI-SWS (Germany, 2018).

Alberto H. F. Laender holds a BSc degree in Electrical Engineering (1974) and an MSc degree in Computer Science (1979), both from the Universidade Federal de Minas Gerais, Brazil, and a PhD degree in Computing (1984) from the University of East Anglia, UK. He joined the Computer Science Department of the Universidade Federal de Minas Gerais in 1975, where he is a Full Professor and heads the Data Management Research Group. In 1997, he was a Visiting Scholar at HP Labs in Palo Alto, California. He has served on the advisory committee of several Brazilian research funding agencies, and was also a member of ACM SIGMOD's Advisory Board (2006-2010) and SIGMOD's Jim Gray PhD Dissertation Award Committee (2008-2011). Currently, he is a member of the ER Steering Committee. Prof. Laender has also served as a program committee member for several national and international conferences on databases, digital libraries and Web-related topics, among them SBBD, AMW, ER, VLDB, CIKM, SIGIR, JCDL, TPDL, WSDM, WWW and ICDE. He is a founder-member of the Brazilian Computer Society and one of the co-founders of Akwan Information Technologies, a Brazilian search technology company that was acquired by Google Inc. in 2005 to become its Research and Development Center for Latin America. Prof. Laender is a member of the Brazilian Academy of Sciences and of the Brazilian National Academy of Engineering, and in 2010 he was awarded the National Order of the Scientific Merit by the Brazilian President. He is the author of more than 200 refereed journal and conference papers. His current research interests include Data Management, Digital Libraries, Social Networks and Bibliometrics.

Mirella M. Moro is associate professor at the Computer Science department at UFMG (Belo Horizonte, Brazil). She holds a Ph.D. in Computer Science (University of California Riverside — UCR, 2007), and M.Sc. and B.Sc. in Computer Science as well (UFRGS, Brazil). She was a member of the ACM Education Council and the Education Director of SBC (Brazilian Computer Society, 2009–2015), the editor-in-chief of the electronic magazine SBC Horizontes (2008–2012), and associated editor of JIDM (2010–2012). Mirella has been working with research in Computer Science in the area of Databases since 1997. Her research interests include social networks analysis, query optimization, and hybrid database modeling. Her recent publications include papers on prestigious venues such as Scientometrics, Data & Knowledge Engineering, ACM Hypertext, IEEE/WIC/ACM Web Intelligence, TPDL, JCDL as well as JIDM and SBBD. She is also an advocate for increasing women participation in Computer Science, coordinating projects such as BitGirls.

José Palazzo Moreira de Oliveira is Full Professor at the Institute of Informatics at UFRGS (retired). He is a Visiting Professor and advisor for a master's and doctorate at PPGCC . Graduated in Electrical Engineering from the Federal University of Rio Grande do Sul (1968), Master in Computer Science from the Federal University of Rio Grande do Sul (1976) and Ph.D. in Informatics from the National Polytechnic Institute of Grenoble (1984). He has experience in the field of Computer Science, with an emphasis on Information Systems, acting mainly on the following topics: ontology, conceptual modeling, distance learning, database, information systems and systems on the Web. Guided the activities of 23 doctors and 60 masters.